National Community Managed Organisation (CMO) Outcome Measurement Project: Final Report

to the Mental Health Information Strategy Standing Committee

Australian Mental Health Outcomes and Classification Network and
Community Mental Health Australia

Version 1.1
September 2013

Document Information

Title: National Community Managed Organisation (CMO) Outcome Measurement Project: Final Report to the Mental Health Information Strategy Standing Committee

Suggested citation for this document

Australian Mental Health Outcomes and Classification Network and Community Mental Health Australia (2013). *National Community Managed Organisation (CMO) Outcome Measurement Project: Final Report to the Mental Health Information Strategy Standing Committee 1.1* . AMHOCN, Sydney, NSW.

Compiled by Tully Rosen, Tim Coombs and Rosemary Dickson

Document History

Version	Date	Details
1.0	August 2013	Provided to Australian Government Department of Health and Ageing
1.1	September 2013	Provided to the Mental Health Information Strategy Standing Committee

National Community Managed Organisation (CMO)
Outcome Measurement Project
Final Report

Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

There has been increasing interest in routine outcome measurement in the community managed sector. However there is a need for greater knowledge about the outcome measurement activities in the sector, and a desire for guidance on what measure or measures may be suitable for use in the sector. At the March 2012 Mental Health Information Strategy Standing committee a project was instigated with 5 objectives:

1. to describe the current status of consumer outcome measurement in the Australian mental health community managed organisation (CMO) sector;
2. to identify good practice examples where consumer outcome measurement has been introduced within day to day service delivery in the mental health CMO sector, with a view to highlighting possible directions for implementation of consumer outcome measurement in the sector;
3. to describe the information infrastructure in place across the mental health CMO sector that supports the use of routine consumer outcome measurement, and the extent to which suitable information infrastructure is comprehensively available;
4. to review the available measures of consumer outcomes that may be suitable for use in the mental health CMO sector, taking account of the range of service types delivered by the sector; and
5. to develop recommendations on:
    - a short list of consumer outcome measurement instruments that offer most potential for use in Australia across the various service types; and
    - the information infrastructure development that would be required to introduce reporting of consumer outcomes as a component of future national dataset requirements covering the mental health CMO sector.

To achieve the project objectives, three core pieces of work were outlined. These included:

1. a review of the literature identifying measures that may be suitable for use in the mental health CMO sector;
2. a survey of mental health CMO service providers and a set of funder interviews identifying measures that are currently being used, the infrastructure in place to support the collection and examples of good practice in the use of outcome measures in the sector; and
3. a workshop that brings together key sector and funder stakeholders to develop agreement on a short list of measures that would be suitable for introduction in the sector and the information infrastructure development needed to support the collection of the agreed measures.

During 2012-2013, these three pieces of work were completed. A national survey was undertaken that resulted in one hundred and thirty two (132) CMOs validated responses. Major jurisdictional funders of CMO mental health services were interviewed about their programmatic use of outcome measurement. A summary of the results of these consultations are provided below and the complete report is provided in appendix 1.

Concurrently a review of the literature was undertaken aimed at identifying the psychometric properties of the measures that were being used in the CMO sector in Australia and the identification of measures that may be suitable for use in the sector. These measures were organised across seven outcome domains: recovery, cognition and emotion, functioning (activities of daily living and interpersonal relationships), social inclusion, quality of life, experience of service provision and multidimensional measures. One hundred and thirty six (136) measures were identified and the psychometric properties of each of these measures including validity, reliability and sensitivity are provided in appendix 2. A set of criteria for measures selection was identified and this produced a short list of thirty one (31) measures suitable for recommendation to the sector and to funders for use in routine collections by CMOs. These measures are reported across the seven outcome domains and the relative suitability of these measures across the different CMO service types are also indicated.

The results of these two activities were presented to a workshop of key sector and funder stakeholders in May 2013. Workshop participants agreed that:

1. Routine outcome measurement should occur within the CMO sector;
2. Routine outcome measurement should include the collection of a universal measure of consumer or carer experience of service provision, and then be supplemented by specific measures depending on CMO service type and program characteristics;
3. There should be production of a “guidebook” that builds upon the results of the current project and which outlines measures, data collection protocols and the preconditions necessary for the implementation of routine outcome measurement in the sector; and
4. The guidebook would be used to structure discussion between CMO peak bodies, service providers, consumers, carers and funders to enable the implementation of routine outcome measurement to the sector.

The three key project activities – the online survey, the review of the literature and the stakeholder workshop – have informed the development of a set of recommendations regarding the use of outcome measurement in the mental health CMO sector.

1 BACKGROUND

There are a variety of terms used in the mental health literature that although they have specific meanings are often used interchangeably. The Oxford English Dictionary for example defines the following terms as follows:

- **Tool**: a thing (concrete or abstract) with which some operation is performed; a means of effecting something; an instrument.
- **Measure**: relating to the determination of magnitude or quantity.
- **Scale**: compare, estimate or weight.
- **Questionnaire**: a formulated series of questions by which information is sought from a selected group, usually for statistical analysis.
- **Screen**: systematic examination of a large number of subjects, esp. for the detection of unwanted attributes or objects.
- **Inventory**: a detailed list of articles.
- **Needs assessment:** to have need or be in need (of something).

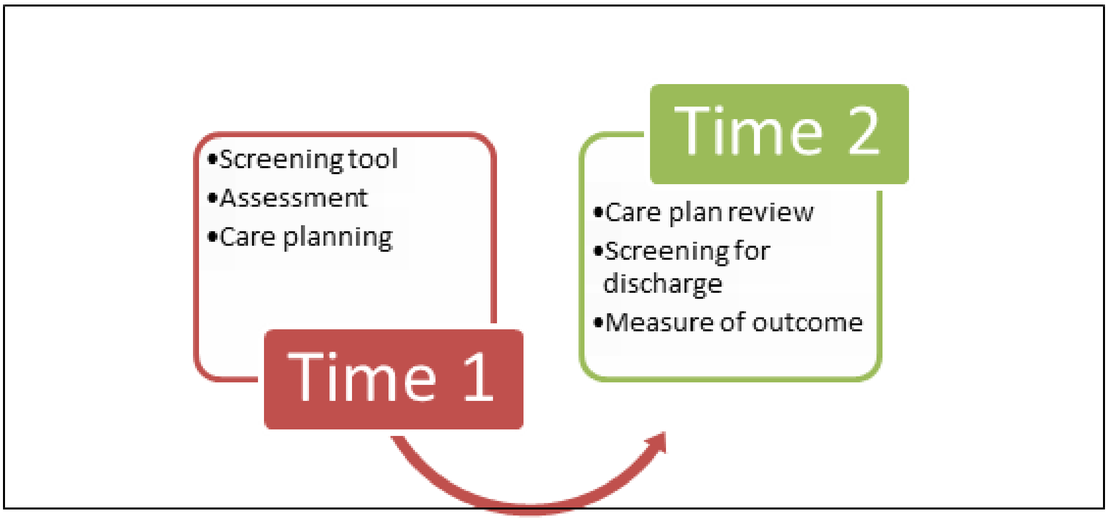
However, in the mental health literature these terms are often used interchangeably. For example, the General Health Questionnaire (Goldberg and Williams 1991) asks respondents to indicate the degree, on a 4-point response scale, they are currently experiencing of problems with strain, concentration, self-confidence, anhedonia, worry, decision making, hopelessness, worthlessness, mood (both depressed and happy), and functioning. Responses to these questions can be used as a screening tool (Ouimette, Wade et al. 2008).

The Camberwell Assessment of Need can be used to guide care, such as the provision of homework assignments by case managers (Kelly and Deane 2011) as well as a measure of the outcomes of care (Drukker, van Dillen et al. 2008). The Kessler -10, which was specifically designed as a screening tool, can also be used to measure the outcomes of care.

So, a tool, a measure, a needs assessment or a questionnaire can be used to screen a consumer for their eligibility for a particular program. A screening tool or inventory can be used to guide practice and support care planning and it is possible that a tool, questionnaire, inventory or scale can be collected on multiple occasions to demonstrate change and the outcomes of care.

Regardless of what they are called, how a standard measure is collected and how it is constructed will determine how it can be used. Figure 1 shows the potential multiple uses of standard measures. When collected just once at the start of care it can be used as a screening tool, an aid to assessment or the basis for a care plan. When collected again it can provide a tool to review care plans, screen for discharge or measure the outcomes of care.

Figure 1: Uses of standard measures



How these standard measures are constructed will also influence how they can be used. The way questions are structured has an important impact on the usefulness of a standard measure. For example, a standard measure such as the Dissociative Experiences Scale (Carlson, Putnam et al. 1993) asks respondents to indicate if they have ever experienced an episode of dissociation. While this is a perfectly legitimate question to ask if you are screening people into a program that caters for this type of disorder, it is of no use as a measure of outcome because there is no potential for change.

The breadth of information being collected also influences how a standard measure may be used. Using a standard measure like the Becks Depression Inventory (Beck, Ward et al. 1961) may be suitable for a program that focuses on the treatment of depression. In this case, the standard measure can be used to screen for the condition or monitor the success or failure of the program. This would be less suitable for a program that deals with consumers with multiple problems in a number of areas where broad spectrum standard measures like the Camberwell Assessment of Need (Phelan M, Slade M et al. 1995) or the Health of the Nation Outcome Scales (Wing, Beevor et al. 1998) may be more suitable.

Regardless of what they are called, standard measures can be tools to create theoretical insight, create a new or common language to assist in building bridges between workplaces and researchers and a way of improving practice (Kristensen 2010).

2 NATIONAL SURVEY RESULTS

A comprehensive mental health CMO consultation was performed through a nation-wide sector survey.

Four kinds of outcome measurement tool use were investigated:

- Established quantitative tools – tools that have been published and/or psychometrically validated[1], where responses are able to be counted, averaged and calculated. e.g. HoNOS, K10, CANSAS, etc.
- Modified quantitative tools – the modification or combination of established tools to suit a service's objectives, where some responses are able to be counted, averaged and calculated.
- Custom-made quantitative tools – service-developed forms or tools.
- Custom-made qualitative tool use – service-developed forms or tools.

The survey was also designed to question CMOs on the domains of outcome that they measure, and those they are interested in measuring, as well as establishing the categories of service that CMOs are using specific tools for. A set of training and information infrastructure questions was also included in the survey.

132 organisations completed valid responses to the survey. This is estimated to be almost half the total number of specialised mental health CMOs using outcome measurement in Australia. Of the CMOs responding to the survey, 23% indicated that they operated in more than one jurisdiction, and 9% operated nation-wide.

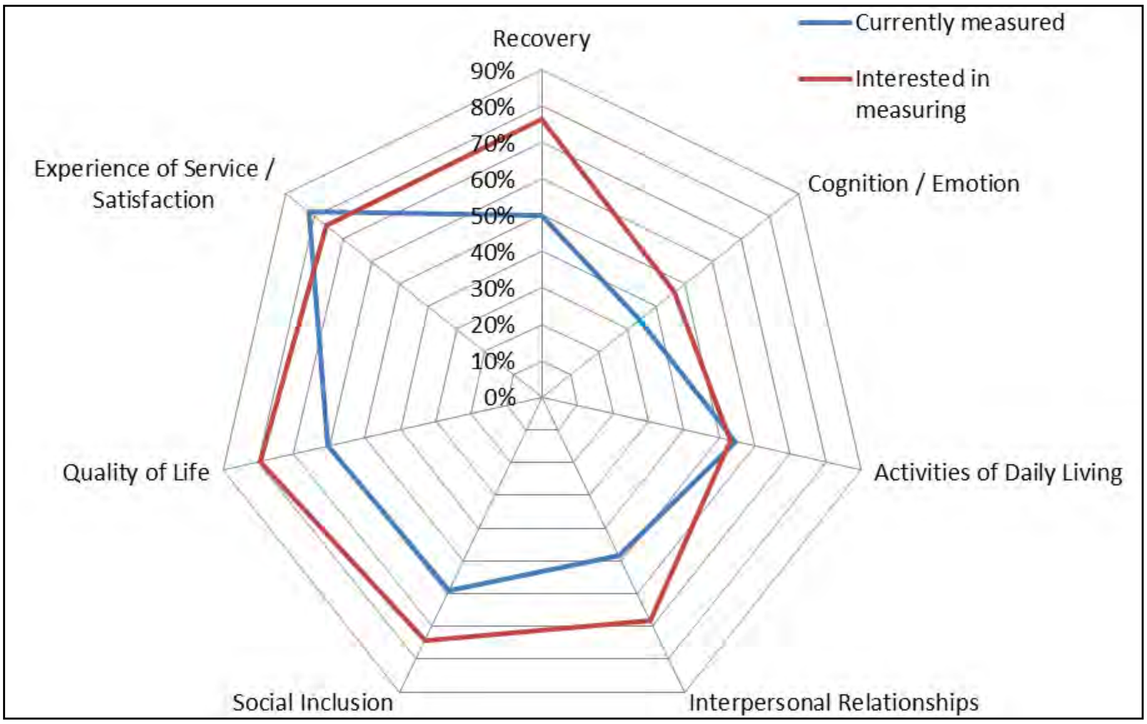
2.1 Domains of Outcome

CMOs were asked two sets of questions involving the set of outcome measurement domains identified in the project literature review. They were first asked which domains they *currently* measure in their organisation, and then they were asked which domains they *are interested* in measuring. The percentage of organisations responding to each domain are displayed below.

---

[1] The definition of “established tool” in this survey was intentionally defined as an instrument that has been published and/or psychometrically validated. This was to maximise the capacity of the survey to collect information on tools that are gaining in popularity but which have not had the opportunity to be scientifically scrutinised or validated.

Figure 2: Outcome domains – Percent of CMOs currently measuring and interested in measuring



2.2 Types of Outcome Measurement Tools

CMOs reported the use of a diverse range of established, standardised and non-standardised outcome measurement tools. They used an average of three different outcome measurement tools across their programs, and provided an average of six different categories of mental health service. CMO responses indicated that commonality of tool use may vary between services providing personalised supports, counselling and referral services, residential support and self-help.

Respondents were asked if their organisation utilised any Established, Modified, Custom-made (quantitative) or Qualitative tools.

Table 1: Types of outcome measurement tools in use

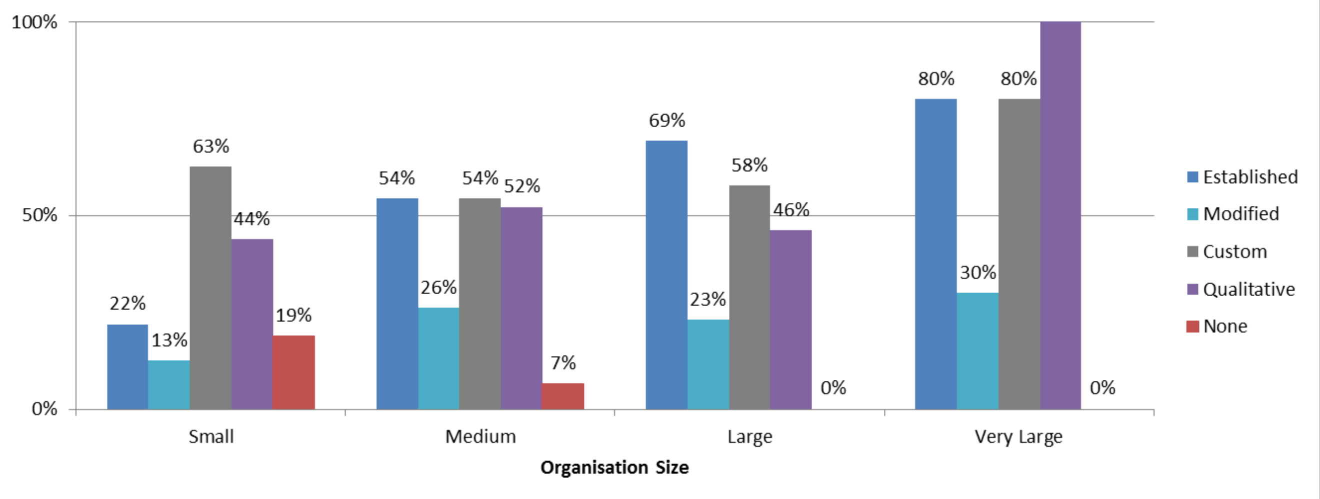
Tools Used	No. Orgs	% of Respondents
Established / formal (quantitative) tools	63	48%
Modified or a combination of established (quantitative or qualitative) tools	29	22%
Custom-made (quantitative) questionnaires or self-created tools	80	61%
Qualitative questionnaires or tools for outcomes	69	52%
None of the above*	11	8%

*The low response rate of CMOs not using outcome measurement tools should not be interpreted as indicative of the sector as a whole. This is due to the likelihood that CMOs not using outcome measurement tools were less likely to respond to the survey.

After specifying all established tools in use, some respondents indicated that up to 10 different established tools were in use in their organisation. These tools may have been combined into a single form or used in different ways for different programs, or under specific conditions.

Tool use counts were categorised by organisation size. A clear relationship appeared to exist between the size of a CMO and the likelihood of using outcome measurement tools. Larger CMOs are more likely to use established tools, and are also more likely to develop their own custom-made tools. Smaller CMOs are less likely to use outcome measurement tools overall.

Figure 3: Types of tools in use – percentage used by CMOs according to organisation size



2.2.1 Established Tools

A total of 62 organisations declared use of established individual outcome measurement instruments (published and/or are at some stage of validation). The eight most common established tools were the Kessler 10 (K-10), Behavior and Symptom Identification Scale (BASIS-32), Camberwell Assessment of Need Short Appraisal Schedule (CANSAS), Depression and Anxiety Stress Scales (DASS), Health of the Nation Outcome Scales (HoNOS), Life Skills Profile (LSP-16), 96. World Health Organisation Quality of Life (WHO-QoL) and the Mental Health Recovery STAR.

Table 2: Established tools in use by more than one CMO

Tool Name	No. Orgs
K10	22
BASIS-32	16
CANSAS	15
DASS	13
HoNOS	13
LSP-16	13
WHO-QoL	12
Recovery STAR	12
RAS	6
GAF	5
Personal Wellbeing Index	5
Homeless STAR	5
Recovery Interview	4
Mental Health Recovery Measure	3
ORS & SRS	2
APQ6	2
SF12	2
SIQ	2
Recovery Enhancing Environment Measure	2
PSI	2

A variety of other established tools were in use by an individual organisation only. Established tool usage varied across jurisdictions, however the top eight tools nationally also tended to high on the list for each jurisdiction.

Most established tools were utilised every three to six months.

Most CMOs indicated that their tools were rated by the consumer (90% of organisations), and three quarters of CMOs indicated that their tools were rated by the CMO worker. A minority of CMOs indicated that their tools were completed by family or carers, or by Local Health District (LHD) workers.

2.2.2 Modified Tools

There were 29 organisations that reported using modified or combined tools based on established tools. The most popular established tools that were being modified were the BASIS-32, CANSAS, HoNOS, LSP-16 and Personal Wellbeing Index.

Most modified tools were utilised every three to six months.

CMOs reported using their modified tools with 63% of clients and 20% of family or carers. Few were completed by workers. Only 2 organisations (7%) were required to use their modified tool as a condition of funding.

2.2.3 Custom-made (quantitative) Tools

There were 80 organisations that had developed their own custom tools containing some level of quantitative measurement. Most of these tools measured experience of service or service satisfaction.

Most custom-made tools were utilised every three to twelve months, and were primarily completed by the consumer or carer, although some were also completed by the CMO worker. One in five CMOs were required to use their custom tool as a condition of funding.

2.2.4 Qualitative Forms or Tools

There were 69 organisations that had used or developed qualitative (mostly open text) forms or tools intended to measure outcomes. Most of these tools measured experience of service or service satisfaction.

Most qualitative tools were utilised every three to twelve months, and were primarily completed by the consumer or carer, although some were also completed by the CMO worker.

2.2.5 Organisations that don't use outcome tools

There were 11 organisations out of 132 that did not use individual outcome tools of any kind. The main reasons provided were that they had not found any appropriate measures, as those currently available were too oriented toward medical or academic needs and language. Administrative burden and information infrastructure capacity limitations were also highlighted as significant issues.

2.2.6 Capacity to Collect Outcome Data

While a majority of CMOs use tools based on their written instructions, less than half of CMOs formally trained their staff to use their outcome tools or forms. CMOs mostly paid for their own training.

More than half (54%) of CMOs have implemented some kind of fully computerised data collection system. However this figure may be increased due to the same reporting bias that would limit responses from CMOs not currently collecting outcome data. Survey responses indicated that up to

47% of CMOs are currently capable of collecting client-level outcome data. Up to 77% would be capable with data system modification or expansion. A further 24% of CMOs indicated that they could only collect client-level outcome data if funding and resources were provided. Most organisations paid for their own data systems.

2.3 Funder and Stakeholder Consultation Results

The second component of project consultation was to investigate current activities, future plans and general attitudes toward outcome measurement in the CMO sector. Interviews were undertaken with funders of mental health CMO services, identified through the membership of the Mental Health Information Strategy Subcommittee (MHISSC), including consumer and carer representatives on current and planned CMO outcome measurement activities. Correspondence was also received from the Victorian Mental Health Carers Network and feedback was sought from the National Mental Health Consumer and Carer Forum. The final question (Q56) of the CMO sector survey was also designed to elicit general perspectives from CMOs on outcome measurement.

2.3.1 Consumer and Carer Perspectives

Feedback from consumer and carer representatives included that:

- Outcome measurement in the CMO sector is important as a feedback and communication mechanism for consumers and carers.
- Jurisdictions must work harder to develop standardisation for outcome measurement in the CMO sector.
- Further consultation with consumers and carers should occur for future activities resulting from this project.
- Many of the more established tools are getting “long in the tooth”. Some may need updating, and newer tools need to be further developed and supported.
- Consumer-rating and carer-rating of tools should be prioritised as good practice.
- The principle of striking a balance between tool brevity and psychometric validity should be established as good practice.
- The project should be focused on domains rather than service types to reduce the number
- De-identified data needs to be collected by CMOs and fed back to facilitate development of resources like the AMHOCN web Decision Support Tool.
- Trials of the carer outcome measurement tools identified in the AMHOCN report “Carer Outcome Measurement in Mental Health Services: Scoping the Field"[2] should be considered.

---

[2] Australian Mental Health Outcomes and Classification Network (2008). *Carer Outcome Measurement in Mental Health Services: Scoping the Field.* Sydney, New South Wales.

2.3.2 CMO Perspectives

CMOs were provided a final comment text box in the sector survey requesting their general views on outcome measurement. 77 organisations took the opportunity to provide comment on outcome measurement issues.

Major themes of the final comments included:

- Many CMOs indicated that they are planning or in the process of establishing routine outcome measurement in their organisations.
- There is a need for outcome measurement tools to be selected and developed appropriate to the context of CMO service provision.
- There is a need for more standardisation of tool use in the CMO sector.
- Resourcing, infrastructure and training capacity issues need to be considered.
- There is a desire for more information and guidance for CMOs regarding outcome measurement.
- Consideration also needs to be given to the administrative burden and issues regarding multiple reporting frameworks.

2.3.3 Government Funder Activities and Comments

Through initial liaison with MHISSC members, jurisdictional representatives with responsibility for funding and policy initiatives in relation to the mental health CMO sector in their State or Territory were contacted and interviewed. Detailed responses were received (either by telephone or e-mail correspondence) from all jurisdictions except for the Northern Territory Department of Health and Families.

The following pages contain a summary of funder activities, identified challenges to collecting outcome data in the CMO sector, potential processes for future outcome measurement activities, and other identified issues.

Few programs funded by jurisdictions are currently mandating the use of outcome measurement tools. Victoria and the Australian Capital Territory have previously recommended tool use to CMOs.

Many jurisdictions are making future plans to implement outcome measurement as a component of their CMO program evaluations. However, the two jurisdictions with firm policy developments are Western Australia and Victoria. Both jurisdictions are working with consumers, carers and CMOs to consider the use of standardised outcome measurement tools, and have indicated that this national project will help inform this work.

Table 3: Current activities and future plans by government funders of CMO services for outcome measurement tool use

	QLD	NSW	VIC	WA	SA	TAS	ACT	DoHA (D2DL)	DoHA (PIR)	FaHCSIA
Mandated Tools for CMOs	No	Yes	No	No	No	No	No	No	No	Yes
Details	N/A	APQ6 for RRSP only	N/A	N/A	N/A	Packages of Care (Cth) = CMOs must use an OM	N/A	N/A	N/A	"Evidence based tools" as per the National Standards. Reporting not mandatory.
Recommended Tools for CMOs	No	No	No	No	No	No	No	No	In development	Yes
Details	N/A	No, but there seems to be good CMO awareness of the LHD mandated suite of tools, NOCC	Policy in 2004 recommended the WHO-QoL, CANSAS and BASIS 32. Does not currently recommend tools, but is considering a new policy position.	N/A	Recommend the WHO-QoL on an informal basis. CANSAS & BASIS 32 also popular	However, CMOs commonly use WHO-QoL and Recovery Star	Offered tools like BASIS 32, & HoNOS, but CMOs didn't like them. K10 & LSP-16 gained more interest.	The program doesn't lend itself easily to OM – no "clinical" activity, no mandatory requirements on entry.	N/A	"Evidence based tools" as per the National Standards.

	NSW	VIC	QLD	WA	SA	TAS	ACT	DOHA (D2DL)	DOHA (PIR)	FaHCSIA
Future Plans for OM use	Maybe	Yes	No	Yes	Yes	Yes	Yes	Maybe	Yes	Yes
Details	Considering K10 (Boarding House HASI) and HoNOS (HASI Plus).	Will be developed in context of 2013-14 Victorian CMO reform. May entail mandating use of specific tools.	Not in scope and not a priority at this stage.	Underway - collaboration between the WAAMH outcomes measurement taskforce and Commission, CoMHWA and MHM2.	Will adopt OM tools from this project once there is consensus	Evaluation framework under review of which OM will be a part	Will incorporate OM into funding agreements on individual basis once there is consensus.	Have looked at OM tools & have done some work with Melbourne University. Waiting on this project.	National evaluation and monitoring project has commenced, including development of PIR MDS with client level OM tool(s).	Considering appropriate tools for new performance framework.
Evaluations of CMO OM use	Yes	Yes	Yes	No	Yes	No	No	No	No	Yes
Programs	Programs with public MH partners e.g. HASI	- Care Coordination - Intensive Home Outreach - Diversion & Substitution	Project 300	An evaluation was done by Deloitte in 2010-11 to assess the OM usage in WA CMO sector but otherwise no	A number of external evaluations on a range of programs	N/A	N/A	N/A	N/A	Targeted Community Care (Mental Health)
Tools Used	HONOS K10 SDQ	HONOS Basis 32 Life Skills Profile	HONOS		NOCC suite due to lack of consensus					Custom-made
Comments from Evaluation	Data collected by LHD workers	Data collected by LHD workers	Recommended use of social inclusion measure		NOCC tools for CMOs appeared to be a bad fit.					N/A

Further comments by government funders of CMO services on outcome measurement tool use

Challenges

- Not all CMO sector grant funding is managed centrally which increases complexity for mandating OM tools. (NSW)
- There are a huge array of CMO service models and types which makes it difficult to determine what 1-5 mandated tools might look like for such a wide and various sector. (NSW)
- Some states have not yet made the transition to outcome measurement, with outputs still the main activity being collected. (QLD)
- Lack of knowledge around outcome measurement in the CMO sector. (QLD)
- The need for funding to assist with the complicated and resource intensive nature of implementation. (QLD)

Issues for consideration

- It is important that an agreed substantive national rationale is reached regarding key objectives for the use of outcome measurement in the CMO sector, and that the preferred tool(s) for use by CMOs be determined on the basis of agreed parameters and/or selection criteria. (VIC)
- A number of interdependencies need to be considered, particularly the introduction of DisabilityCare Australia and the implications / opportunities this may present for the use of outcome measurement by CMOs. (VIC)
- It is very important to have a clear definition of outcome tools, for example as opposed to assessment tool, set out clearly as a part of this project. (NSW)
- You can’t mandate too many tools for use. (NSW)
- There are many synergies in terms of challenges between the OMP and the minimum dataset project, and the recommendations for a suite of OM tools must be attached to the minimum data set. (NSW)
- The need for appropriate infrastructure and the capacity to train and retrain people. (QLD)
- Looking at how practical the tools are to apply, ease of use, relevance, cost of implementation (eg license fees, cost to small and rural organisations, cost of ongoing data collection and reporting) capacity of tool to reflect individual differences, capacity for capturing valid / reliable data. (WA)
- Information about the tools and (face to face) education/training on how the tools can be used, to be made available to the sector (WA)
- WA is as yet not clear about which tool can be used by the sector to collect / report individual client level data on goals, which then has the potential to inform the sector’s performance on WA’s six mental health outcome areas (and identify areas that need improvement).

Future Directions

- Mandating tools is best left for the CMO sector to make recommendations on. (NSW)
- While Victoria is keen to align as much as possible with other jurisdictions to enable nation-wide outcome analysis and program and policy learning, Victorian CMO reform timelines may require Victoria to determine its preferred approach prior to any national decision. (VIC)
- Agreement should be reached with the sector about identified tools. (WA)
- All information gathered in this project should be made available to the CMO sector. (WA)
- Clear identification of an agency is required to continue working on CMO outcome measurement sector development. (ACT)
- Important to consider which selected tools are acceptable to CMOs and do not place additional reporting burden. (Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA))

3 REVIEW OF THE LITERATURE

A literature review was undertaken that aimed to identify the psychometric properties of the measures being used in the sector along with measures that may be suitable for use. The psychometric properties reviewed included validity, reliability and sensitivity to change.

Initially the literature was reviewed to determine what outcomes had been measured. This resulted in the identification of seven outcome domains. These included recovery, cognition and emotion, functioning (activities of daily living and interpersonal relationships), social inclusion, quality of life, and experience of service provision. These domains were used as key words to guide the search of the literature. While these proved a useful way of organising the search of the literature, it became clear early on that some measures, in fact, covered multiple domains. Therefore, for reporting purposes, a multidimensional domain was included. A description of each of these domains is in table 4.

One of the challenges of the literature review was the identification of measures that may be suitable for use in the CMO sector. Consequently, “Community Managed Organisations”, “Non-government organisations”, and “not for profit” were used as search terms along with the outcome domains identified above. However, the inclusion of these terms quickly limited the search results. Therefore the primary search terms were the outcome domains. However, when measures were identified, the literature often lacked a detailed description of the types of services or service settings within which measures were being tested. As a result, the literature review also relied on grey literature to identify measures that have been used or suggested for use in the sector. Finally, the results of the survey were also used to identify measures that were included in the literature review.

Given these challenges, 136 measures were identified through this process. It is important to note that not every CMO responded to the survey, so there may be other measures being used in the sector that are not included in this review. It is also important to note that there are a large number of mental health measures that have been created. The measures included in this review are only a subset of all possible measures.

3.1 Outcome Domains

Table 4: Outcome measurement domains

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
The personal process of individual recovery.	Individual consumer cognitive performance and emotional experience Individual carer cognitive performance and emotional experience	Simple and complex functional abilities are covered here including the ability to undertake activities of daily living consistent with developmental stage. The quantity and quality of interpersonal relationships consistent with developmental stage.	Education, employment, citizenship, stability of housing	General life satisfaction, physical health and wellbeing	Service satisfaction, consumer or carer experience of service provision Care or service co-ordination	Measures that capture information across multiple domains

3.2 Identified Measures

136 measures were identified as a result of the literature review process. These measures are listed in table 5.

Table 5: All identified measures

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
1. Recovery Assessment Scale (RAS) 2. Agreement with Recovery Attitudes Scale (ARAS) 3. Rochester Recovery Inquiry (RRI) 4. Consumer Recovery Outcomes System (CROS) 5. Crisis Hostel Healing Scale (CHHS) 6. Personal Vision of Recovery Questionnaire (PVRQ) 7. Recovery Interview (RI) 8. Recovery Attitudes Questionnaire (RAQ-16; RAQ-7) 9. Mental Health Recovery Measure (MHRM) 10. Reciprocal Support Scale (RSS) 11. Relationships and Activities that Facilitate Recovery Survey (RAFRS) 12. illness Management and Recovery (IMR) Scales 13. Ohio Mental Health Consumer Outcomes System (OMHCOS) 14. Peer Outcomes Protocol (POP) Recovery Measurement Tool (RMT) 15. Recovery Orientation (RO)a 16. Recovery Process Inventory (RPI) 17. Milestones of Recovery Scale (MORS) 18. Multi-Phase Recovery Scale (MPRM)a 19. Maryland Assessment of Recovery in People with Serious Mental Illness (MARS) 20. Mental Health Recovery Star (MHRS) 21. Questionnaire about the Process of Recovery (QPR) 22. Subjective Experiences of Psychosis Scale (SEPS) 23. Self-Identified Stage of Recovery (SISR) 24. Stages of Recovery Instrument (STORI) 25. Stages of Recovery Scale (SORS)	26. Kessler-10 (K-10) 27. Mental Health Inventory 38 (MHI-38) 28. Behaviour Symptom Identification Scales (BASIS-32®) 29. Depression, Anxiety, Stress Scale (DASS) 30. Beck Depression Inventory (BDI) 31. Beck Anxiety Inventory 32. Hamilton Depression Rating Scale 33. Hamilton Anxiety Rating Scale 34. State Trait Anxiety Inventory 35. Edinburgh Postnatal Depression Scale 36. Geriatric Depression Scale 37. Symptom Checklist 90 (SCL 90) 38. General Health Questionnaire-12 (GHQ-12) 39. Sphere-12 40. My Mood Monitor (M-3) 41. Suicidal Ideation Questionnaire (SIQ) 42. Columbia–Suicide Severity Rating Scale (C-SSRS) 43. Inventory of Complicated Grief-Revised 44. PTSD Checklist (PCL-C) 45. Dissociative Experiences Scale 46. Strengths and Difficulties Questionnaire (SDQ) 47. Child Behaviour Checklist (CBCL) 48. Devereux Early Childhood Assessment (DECA) 49. Anxiety Disorders Interview Schedules (Children) 50. Parenting Stress Index 51. Social Responsiveness Scale – SRS 52. Barriers to Adolescents Seeking Help Scale 53. Mini Mental State Examination 54. Cambridge Cognitive Examination CAMCOG 55. Rowland Universal Dementia Assessment Scale (RUDAS) 56. Warwick-Edinburgh Mental Well-being Scale (WEMWBS) 57. Working Alliance Inventory 58. Empowerment Scale 59. Brief COPE 60. Revised NEO Personality Inventory (NEO-PI-R) 61. General Health Seeking Questionnaire 62. Involvement Evaluation Questionnaire (IEQ) 63. Burden Assessment Scale (BAS) 64. Zarit Burden Scale 65. CarerQol-7D+VAS	66. Life Skills Profile] 67. Social Function Scale 68. Work and Social Adjustment Scale 69. Personal and Social Performance Scale (PSP) 70. Global Assessment of Functioning (GAF) 71. Social Functioning Questionnaire (SFQ) 72. The World Health Organisation Disability Assessment Schedule 2.0 (WHODAS II/2.0) 73. The Multnomah Community Ability Scale (MCAS) 74. Canadian Occupational Performance Measure (COPM) 75. Children's Global Assessment Scale (CGAS) 76. Parents' Evaluation of Developmental Status (PEDS) 77. Drug Use Disorders Identification Test (DUIT) 78. Alcohol Use Disorders Identification Test (AUDIT) 79. Specific Levels of Functioning Scale (SLOF) 80. Independent Living Skills Survey (ILSS) 81. Social Behaviour Schedule (SBS) 82. Child and Adolescent Functional Assessment Scale (CAFAS) 83. Columbia Impairment Scale	84. Social and Community Opportunities Profile (SCOPE) 85. Social Inclusion Questionnaire (SIQ) 86. Activity and Participation Questionnaire (APQ) 87. Staff Survey of Social Inclusion (SSSI) 88. EMILIA Project Questionnaire (EPQ) 89. Social Inclusion measure (SIM) 90. The Inclusion Web (IW) 91. Composite Measure of Social Inclusion (CMSI) 92. Australian Community Participation Questionnaire (ACPQ) 93. Evaluating Social Inclusion Questionnaire (ESIQ) 94. Living in the Community Questionnaire	95. Satisfaction with Life Scale (SWL) 96. World Health Organisation Quality of Life –Brief, Australian Version (Australian WHOQoL- BREF) 97. Purpose In Life (PIL) 98. Californian Quality of Life Survey (CA-QoL) 99. Quality of Life Questionnaire (QLQ) 100. Subjective Quality of Life Profile 101. SF-36 (or Medical Outcomes Study (MOS) 36-Item Short Form Health Survey) 102. Quality of Life Enjoyment and Satisfaction Questionnaire-short form (Q-LES-Q-SF) 103. Manchester Short Assessment of Quality of Life (MANSA) 104. Quality of Life Interview (QOLI) 105. Lancashire Quality of Life Profile (LQLP) 106. Quality of Life Index (QLI) 107. Satisfaction with Life Domains Scale (SLDS) 108. Quality of Life Scale (QLS) 109. Wisconsin Quality of Life Index – Canadian Version (CaW-QLI) 110. Brief Life Satisfaction Scale (BLSS) 111. Satisfaction with Life Scale (SWLS) 112. Personal Wellbeing Index 113. Personal Outcomes Measures	114. Mental Health – Consumer Perceptions and Experiences of Services (MH-CoPES) 115. National Research Corporation (NRC-Picker)/ National Health Service 116. Consumer and Carer Experience Questionnaires (C&CES) 117. Psychiatric Outpatient Experience Questionnaire (POPEQ) 118. The Mental Health Statistics Improvement Program Consumer Survey (MHSIP-Consumer) 119. Consumer Experience of Care Questionnaire (Australia) 120. Carer Experience of Service Provision Questionnaire (Australia) 121. Carer satisfaction interview 122. Questionnaire for carers of people who use social services (Wales) (QPSS-Wales) 123. Carer Participation Survey (CPS) 124. Consumer and Carer Experience Questionnaires (C&CES) 125. Carer Well-being and Support Questionnaire (CWS) 126. Disability and mental health service users and carers satisfaction survey 2009 127. The Personal Social Services Survey of Adult Carers in England 2009-2010 (PSSS-Adult Carers) 128. Mental Health Carers' Survey (MHCS) 129. Network analysis 130. Human services integration measure	131. Camberwell Assessment of Need (CAN) 132. Collaborative Goal Index/COMPASS 133. Health of the Nation Outcomes Scales (HoNOS) 134. Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA) 135. Health of the Nation Outcomes Scales 65+ (HoNOS 65+) 136. Outcome Rating Scales (ORS) and Session Rating Scales (SRS)

3.3 Selection Criteria

To support instrument selection, a number of criteria were established. These review criteria specified that the attributes for an outcome measure appropriate for mental health CMOs in Australia should:

- have been developed for use or used in the mental health sector;
- have been developed or used in Australia, with identified potential for further development;
- be able to be completed by either the consumer and/or CMO employee;
- be brief and easy to use (time and/or number items);
- yield quantitative data (does not exclude instruments that also yield qualitative data);
- have undergone scientific scrutiny and have demonstrated strong psychometric properties (e.g., of internal consistency, validity, reliability and sensitivity to change).

Using these selection criteria, a short list of measures was identified as suitable for use within the CMO sector (table 6). At this stage, these measures are only suggested. More detailed work is necessary to further establish the suitability of these measures.

3.4 Short List of Measures by Domain

Using the selection criteria above a short list of measures were identified across the seven outcome domains, see table 6.

Table 6: Sort list of measures by domain

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
1. Recovery Assessment Scale (RAS)† 2. Recovery Process Inventory (RPI)† 3. Illness Management and Recovery (IMR)† Scales 4. Stages of Recovery Instrument (STORI)† 5. Recovery Start†	6. Kessler-10 (K-10)† 7. Mental Health Inventory 38 (MHI-38)† 8. Behaviour Symptom Identification Scales (BASIS-32®)† 9. Strengths and Difficulties Questionnaire (SDQ)† 10. Involvement Evaluation Questionnaire (IEQ)† 11. Burden Assessment Scale (BAS)† 12. CarerQol-7D+VAS†	13. Life Skills Profile* 14. Work and Social Adjustment Scale* 15. The Multnomah Community Ability Scale (MCAS)* 16. Personal and Social Performance Scale (PSP)*	17. Social and Community Opportunities Profile (SCOPE)† 18. Activity and Participation Questionnaire (APQ6)† 19. Living in the Community Questionnaire†	20. Satisfaction with Life Scale (SWL)† 21. Manchester Short Assessment of Quality of Life (MANSA)† 22. World Health Organisation Quality of Life – Brief, Australian Version (Australian WHOQoL- BREF)†	23. Consumer and Carer Experience Questionnaires (C&CES)† 24. Psychiatric Outpatient Experience Questionnaire (POPEQ)† 25. Consumers Experience of Care† 26. Carers Experience of Service Provision†	27. Camberwell Assessment of Need – Short Appraisal Scale (CANSAS) †* 28. Collaborative Goal Index/COMPASS† 29. Health of the Nation Outcomes Scales (HoNOS)* 30. Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA)* 31. Health of the Nation Outcomes Scales 65+ (HoNOS 65+)*

† = client-rated

* = worker-rated

3.5 Short List of Measures by Service Type

The short listed measures were reviewed by the project team with the aim of identifying those measures that may be suitable for use across the different CMO service types. In table 7 the numbers correspond to the measure in table 6. Green indicates that the measure(s) and domain are suitable for the service type described in the row. Orange indicates that the measure(s) and domain may be suitable for this service type. The red indicates that the measure(s) or domains may not be suitable for this service type. The identification of the suitability of measure(s) and domains was a desk top exercise undertaken by the project team. Broader consultation and testing of these measures in the different CMO service settings will be required.

Table 7: Short list of measures by service type

Service Type	Recovery	Cognition and Emotion		Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
Counselling—face-to-face	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Counselling, support, information & referral—telephone	1-5?[3]	6-9		X	X	20-22	23-26	28
Counselling, support, information & referral—online	1-5?	6		X	X	20-22	23-26	28
Self-help—online	1-5?	6		X	X	20-22	23-26	28
Group support activities	1-5?	6-9		X	17-19?	20-22	23-26	28
Mutual support and self-help	1-5?	6-9?		X	17-19?	20-22	23-26	28
Staffed residential services	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Personalised support—linked to housing	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Personalised support—other	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Family & carer support	X	6-12		X	X	20-22	23-25	X
Individual advocacy	X	X		X	X	X	X	X
Care co-ordination	1-5?	6-9?	10-12?	13-16?	17-19?	20-22?	23-26	27-31
Service integration infrastructure	X	X		X	X	X	X	X
Education, employment & training	1-5?	6-9?	10-12?	13-16?	17-19	20-22?	23-26	X
Sector development and representation	X	X		X	X	X	X	X
Mental health promotion	X	X		X	X	X	X	X
Mental illness prevention	X	X		X	X	X	X	X

Factors that may influence measure selection include program design, age, Aboriginal and Torres Strait Islander status, Cultural and Linguistic Diversity, and intellectual disability.

---

[3] Question mark indicates that for some programs these measures may be suitable but this is yet to be determined.

4 RECOMMENDATIONS

1. That work continues to be undertaken to support the implementation of routine outcome measurement at the individual level within the community managed mental health sector in Australia.

2. That the implementation of routine outcome measurement in the community managed mental health sector entail the introduction of a single experience of care measure (one for consumers and one for carers) as the core of an agreed suite of measures for community managed mental health sector.

3. That a project be undertaken, guided by an expert advisory group consisting of technical and stakeholder experts, that develops a "guide book" to support the implementation of routine outcome measurement in the community managed mental health sector. This "guide book" builds on the work of the current project to:

    i. Outline the aims and principles of routine outcome measurement in the context of the community managed mental health sector;

    ii. Describe the processes of implementation for routine outcome measurement in community managed mental health services, and gives examples of good practice already occurring;

    iii. Describe the core consumer and carer experience of care measures, and a broader suite of measures across the outcome domains of: recovery, cognition and emotion, functioning, social inclusion and quality of life;

    iv. Identify measures from each domain that may be suitable to be used by community managed mental health services within each of their service types (as identified in the Australian Institute of Health and Welfare Mental Health Non Government Organisation Establishment National Minimum Data Set taxonomy); and

    v. Outline a suggested collection protocol for the identified measures.

4. The “guide book” provides the foundation upon which a set of discussions on measures and protocol, facilitated by CMHA peak bodies, occur at a jurisdictional level between funders and CMOs. These discussions are the first steps to enable a nationally coherent approach to the implementation of routine outcome measurement in the community managed mental health sector. These discussions should include the identification of the training, information infrastructure and capacity development necessary in each jurisdiction.

5 BIBLIOGRAPHY

Beck, A., C. Ward, et al. (1961). "An inventory for measuring depression." Archives of General Psychiatry **4**: 561-571.

Carlson, E. B., F. W. Putnam, et al. (1993). "Validity of the Dissociative Experiences Scale in screening for multiple personality disorder: a multicenter study." American Journal of Psychiatry **150**(7): 1030-1036.

Drukker, M., K. van Dillen, et al. (2008). "The use of the Camberwell Assessment of Need in treatment: what unmet needs can be met?" Social Psychiatry and Psychiatric Epidemiology **43**(5): 410-417.

Goldberg, D. and P. Williams (1991). A user’s guide to the General Health Questionnair. Windsor, NFER-Nelson.

Kelly, P. and F. Deane (2011). "Do Therapeutic Homework Assignments Address Areas of Need for Individuals with Severe Mental Illness?" Community Mental Health Journal **47**(2): 194-200.

Kristensen, T. S. (2010). "A questionnaire is more than a questionnaire." Scandinavian Journal of Public Health **38**(3 suppl): 149-155.

Ouimette, P., M. Wade, et al. (2008). "Identifying PTSD in primary care: Comparison of the Primary Care-PTSD Screen (PC-PTSD) and the General Health Questionnaire-12 (GHQ)." Journal of Anxiety Disorders **22**(2): 337-343.

Phelan M, Slade M, et al. (1995). "The Camberwell Assessment of Need: the validity and reliability of an instrument to assess the needs of people with severe mental illness." British Journal of Psychiatry **167**: 589-595.

Wing, J., A. Beevor, et al. (1998). "Health of the Nation Outcome Scales (HoNOS). Research and development." British Journal of Psychiatry **172**: 11-18.

6 APPENDICES

6.1 Appendix 1 - Report on the National Community Managed Outcome Measurement Project Survey and Consultation

Report on the National Community Managed Outcome Measurement Project Survey and Consultation

Final Report

June 2013

Authors

Tully Rosen, Policy Officer, Mental Health Coordinating Council (for CMHA)
Lucy Corrigan, Policy Assistant, Mental Health Coordinating Council (for CMHA)

Suggested reference

Community Mental Health Australia (2013). Report on the National Community Managed Outcome Measurement Project Survey and Consultation: Canberra.

Community Mental Health Australia is a coalition of the eight state and territory peak community mental health organisations established to provide leadership and direction in promoting the benefits of community mental health and recovery services across Australia. The peak bodies are the:

- Mental Health Community Coalition (ACT)
- Mental Health Coordinating Council (NSW)
- Mental Health Council of Tasmania
- Mental Health Coalition of South Australia
- Northern Territory Mental Health Coalition
- Psychiatric Disability Support Services of Victoria/VICSERV
- Queensland Alliance for Mental Health
- Western Australia Association for Mental Health.

CMHA provides a unified voice for over 800 community managed, non-government organisations who work with the millions of people affected by mental illness across the nation and who are members of, or affiliated with, the various coalition members.

CMHA contact details

E-mail: admin@cmha.org.au
Telephone: 02 6249 7756 (Secretariat – MHCC ACT)

Contents

Executive Summary

The routine use of standardised outcome measurement instruments to gauge consumer and carer wellbeing is an essential component of quality service provision in all mental health services. There is a clear need for community managed organisations (CMOs) to further develop their capacity to measure and report on service outcomes. However, little is known at a national level about the outcome measurement activities of CMOs delivering specialised mental health services in Australia.

The Department of Health and Ageing (DoHA) commissioned the alliance of state and territory mental health peak bodies, Community Mental Health Australia (CMHA), to work in partnership with the Australian Mental Health Outcomes and Classifications Network (AMHOCN) on a project investigating the use of individual consumer and carer outcome measurement tools in the mental health CMO sector. This report contains the results of the primary stakeholder consultations on current and future outcome measurement activities in Australia. The final project report will be tabled at the Mental Health Information Strategy Standing Committee (MHISSC) and will then be refined into a publication for broader release.

Two methods were employed to gain an understanding of the current status of outcome measurement in the sector: a detailed national survey was completed of CMOs specifically funded to provide mental health services in Australia, and a set of funder and stakeholder interviews were undertaken to gain a national view of funder activity and future plans.

The survey found that many CMOs operated across jurisdictional boundaries. CMOs reported the use of a diverse range of established, standardised and non-standardised outcome measurement tools. The eight most common established tools were the K10, BASIS-32, CANSAS, DASS, HoNOS, LSP-16, WHO-QoL and Recovery STAR. A smaller number of organisations reported that they modified or otherwise adapted established tools for use in their service. Most organisations also reported the use of custom-made tools or forms, many of which were used to determine whether clients were satisfied with service provision.

CMOs used an average of three different outcome measurement tools across their programs, and provided an average of six different categories of mental health service. CMO responses indicate that commonality of tool use may vary between services providing personalised supports, counselling & referral services, residential support and self-help.

CMOs reported the measurement of a variety of outcome domains, and many indicated a desire to move toward measuring outcomes relating to the personal recovery journey, quality of life, social inclusion and functional capacity for interpersonal relationships.

Levels of training and information infrastructure varied, indicating that some work would be involved in developing CMO capacity to measure and report on outcome data.

Funders of mental health CMO services provided summaries of their current outcome measurement policies and activities, and many indicated they are waiting for the final project report of this project to make future decisions. The Victorian Department of Health and the Mental Health Commission of Western Australian are in the process of implementing reforms that will involve the use of outcome measurement tools being used with or by CMOs.

This report will be incorporated into a final report, which will also contain a comprehensive literature review, a tool selection process and candidate set of tools, results of the stakeholder workshop, and future recommendations for outcome measurement in the CMO sector. The final report is scheduled to be tabled at MHISSC in October 2013. It will then be developed into a public document, and academic publication will be considered.

This document is a descriptive report of custom and practice for outcome measurement in community managed mental health organisations. It does not take a position on what the criteria should be for an established tool to be considered valid. An "established tool" is defined in this document as an instrument that has been published and/or psychometrically validated. Information from this document is an overview of usage and activity, and should not be used as the sole point of reference for selecting outcome measurement tools.

Strong consideration should be given to the fact that common usage of a tool does not necessarily indicate its appropriateness for the CMO service context, and that consumer, carer, CMO and funder perspectives should be carefully sought before selecting tools for a mental health program.

The literature review component of the National CMO Outcome Measurement Project will provide greater detail on the psychometric value of outcome measurement tools identified in this report.

Introduction

Little is known at a national level about the outcome measurement activities of community managed organisations (CMOs/NGOs)[1] delivering specialised mental health services in Australia. What is known is that there is no one standard outcome measurement tool, or suite of tools, used nationally by the majority of mental health CMOs.

In March 2009, the Productivity Commission's review of the not-for-profit sector proposed a measurement framework to guide the collection of data, evaluation approaches and reporting about that sector in Australia **(Productivity Commission 2010)**. The framework is a hierarchy of four contribution measures:

- Inputs (measures of resources used);
- Outputs (indicators of the level of activity undertaken);
- Outcomes (direct effects on activity participants); and
- Impacts (longer term effects on the participants and the community more broadly).

There is a clear need for CMOs to measure and report on their service outcomes, and to eventually develop a national CMO data collection process in partnership with government agencies. The routine use of standardised outcome instruments to gauge consumer and carer wellbeing is an essential component of quality service provision in all mental health services.

---

[1] Community Mental Health Australia uses the term "Community Managed Organisation" for organisations operating in the community managed mental health sector, as consultation has indicated that they prefer to be named according to what they are, rather than what they are not. These organisations have historically been known as "non-government organisations", or, in Victoria, Psychiatric Disability Rehabilitation and Support Services.

National documents set out the commitment by state and Commonwealth governments to work toward both the use of tools by CMOs and the need to collect outcome information in a nationally consistent manner. The Fourth National Mental Health Plan: *An agenda for collaborative government action in mental health 2009-2014* (Commonwealth of Australia, 2009) includes a priority area that outlines the importance of accountability across the mental health service system through better measurement and reporting of progress. The *Roadmap for National Mental Health Reform* (Council of Australian Governments, 2012) commits state and Commonwealth governments to increase the levels of mental health services (including CMOs) achieving accreditation against the *National Mental Health Standards*, which in turn stipulates the use by services of evidence-based outcome monitoring tools. Also, the *National Report Card on Mental Health and Suicide* Prevention (National Mental Health Commission, 2012) indicates that the National Mental Health Commission expects state and Commonwealth governments to be working towards a national system for measuring and reporting on client data, including outcomes. It specifically includes CMOs in this statement.

Routine use of outcome measurement tools can assist CMO clients in their recovery journey, as well as enabling national planning, policy, monitoring and quality improvement processes. So long as there is appropriate sector development and support, the results of this project will enable the Mental Health Information Strategy Standing Committee (MHISSC) to provide a national picture of the effectiveness of community managed mental health services. To maximise the usefulness of CMO outcome data, future national initiatives must ensure that selected tools are contextually appropriate to the CMO environment, are valid and reliable, are as consistent as possible between jurisdictions, and are incorporated into a national data collection.

Project Background

The Australian Institute of Health and Welfare (AIHW) *Preliminary Scoping Report* identified the issues, options and potential data development work required to obtain more detailed information about the specialised mental health services currently being delivered by CMOs in Australia (Australian Institute of Health and Welfare, 2010). The preliminary scoping report recognised that outcome measures are a vital element in enabling a better understanding of the role the CMO sector plays in the community.

In 2011 the AIHW, in partnership with CMHA, completed Phase 1 the NGO Establishments National Minimum Data Set (NMDS) Project. During stakeholder consultation it became clear that there was a strong desire in the CMO sector to achieve support in developing national consensus on outcome measurement. Eventually it was hoped that the NGO Establishments NMDS would collect client level and outcome data in order to enable national reporting of CMO service effectiveness, in line with current national reporting of public mental health services. While many MHISSC members acknowledged and appreciated the sentiment of this feedback, it was generally agreed during Phase 1 consultation that building consensus and information infrastructure capacity in the CMO sector may take a substantial amount of time, and that the initial NMDS be restricted to collecting establishment-level data until more standardised CMO reporting data became available.

The *Mental Health Non-Government Organisation Data Development Project: Phase 1 Final Report* (Australian Institute of Health and Welfare, 2011) contained a number of options and recommendations to achieve progress in nationally consistent CMO data collection, first of which was that a discrete project be funded to investigate the current levels of use and standardisation of outcome measurement tools in the mental health CMO sector.

In order to address this recommendation, the Department of Health and Ageing commissioned CMHA in 2012 to work in partnership with the Australian Mental Health Outcomes and Classification Network (AMHOCN) on the National Mental Health CMO Outcome Measurement Project. The key activities of this project were to:

- Investigate current national use of outcome measurement instruments in the mental health CMO sector, including those with a recovery focus, and identify good practice examples.
- Investigate the information infrastructure in place across the mental health CMO sector that supports the use of outcome measurement.
- Conduct a comprehensive literature review of outcome measurement instruments available for use by mental health CMOs.
- Conduct a national CMO and funder workshop to present and refine the findings.
- Identify a candidate set of outcome measurement instruments applicable for each of the National Minimum Data Set Taxonomy service categories.
- Deliver a final report to MHISSC with recommendations for activities that could be undertaken to encourage the use and standardisation of valid outcome measurement tools in the CMO sector.

This report contains the survey and stakeholder consultation components of this project.

Consultation Methodology

This project is designed to help guide CMOs toward a national vision of standardised outcome measurement. It also aims to facilitate future capacity building to enable client-level outcome data collection in the NGO Establishments NMDS, and to dovetail with AMHOCN's future outcome measurement activities

As well as more common client-level outcome measurement tools, there is an emerging interest in measuring CMO outcomes at the organisational level (NSW Consumer Advisory Group & Mental Health Coordinating Council, 2011) and to investigate the broader societal impacts of CMO service provision (Mental Health Commission Western Australia, 2012). However, as so little is currently known about CMO outcome measurement activities, this project remained focused on investigating individual consumer and carer outcome measurement instruments being used by organisations that are likely to contribute to the NGO Establishments NMDS[2].

In order to obtain a thorough view of the outcome measurement environment for mental health CMOs, the project steering group, comprised of DoHA, AMHOCN and CMHA representatives, identified the following set of stakeholders to consult:

- Community Managed (non-government) Organisations specifically funded to provide mental health services, with the exception of Medicare Locals.
- Jurisdictional funders of community managed mental health services in Australia (identified through membership of the Mental Health Information Strategy Standing Committee)
- Consumer Representation (MHCA designated MHISSC member)
- Carer Representation (MHCA designated MHISSC member)
- National Mental Health Commission's Contributing Life Project (Project officer)

A comprehensive CMO consultation was performed through a nation-wide sector survey. The CMHA Project Officer identified areas of investigation based on sector knowledge and subject matter expert correspondence. Survey drafts were revised by CMHA and AMHOCN staff, the Project Steering Group, and a small CMHA reference group. The remaining consultations were performed through one-on-one correspondence.

The final survey question list is provided in Appendix A.

---

[2] With the exception of for-profit community mental health organisations, which are estimated to currently comprise a very small percentage of the community managed mental health sector (Mental Health Coordinating Council, 2010).

Four kinds of outcome measurement tool use were investigated:

- Established quantitative tools – tools that have been published and/or psychometrically validated[3], where responses are able to be counted, averaged and calculated. e.g. HoNOS, K10, CANSAS, BASIS32, etc.
- Modified quantitative tools – the modification or combination of established tools to suit a service's objectives, where some responses are able to be counted, averaged and calculated.
- Custom-made quantitative tools – service-developed forms or tools.
- Custom-made qualitative tool use – service-developed forms or tools.

The survey was also designed to question CMOs on the domains of outcome that they measure, and those they are interested in measuring, as well as establishing the categories of service that CMOs are using specific tools for.

A national CMO outcome measurement initiative, in partnership with the sector, would require corresponding sector development to ensure that CMO skill levels and infrastructure were capable of collecting and reporting on client-level outcome information. For this reason a set of training and information infrastructure questions were included at the end of the survey.

Domains of Outcome

To provide a framework within which the various identifiable tools can be considered for the project, a set of domains of outcome measurement were identified based on a comprehensive review of outcome tool domains carried out by Mike Slade (2002) and preceding work. The domains of Social Inclusion and Recovery were also included to fit the future indicators and directions committed to by jurisdictional funders.

The developed outcome domains used for the CMO consultation were:

Recovery	**Emotional Distress & Cognition** (symptoms)	**Personal Functioning**		**Social Inclusion**	**Quality of Life**	**Experience of Service**
		Activities of Daily Living	**Interpersonal Relationships**			
Individual recovery as a personal process	Cognitive and emotional issues, and distress from anxiety and depression.	Simple and complex. Functional ability to undertake activities. Independence for developmental stage, self-care.	The quantity and quality of interpersonal relationships (incl. partner, family, friends).	Education, employment, citizenship, stability of housing.	General life satisfaction, wellbeing, physical health, personhood.	Service satisfaction, carer experience, how you were treated, met/unmet expectations of service provision.

---

[3] The definition of "established tool" in this survey was intentionally defined as an instrument that has been published and/or psychometrically validated. This was in order to maximise the capacity of the survey to collect information on tools that are gaining in popularity but which may not have had the opportunity to be scientifically scrutinised or validated.

Service Categories

The AIHW *NGO Establishments NMDS* Taxonomy was developed by a working group of the Mental Health Information Strategy Standing Committee (Australian Institute of Health and Welfare, 2012). It contains 17 categories of service activity designed to be a representation of the funded CMO activity in Australia. This taxonomy was used to establish the relationship between outcome measurement tool use and CMO service categories. The taxonomy consists of:

1. Counselling—face-to-face
2. Counselling, support, information & referral—telephone
3. Counselling, support, information & referral—online
4. Self-help—online
5. Group support activities
6. Mutual support and self-help
7. Staffed residential services
8. Personalised support—linked to housing
9. Personalised support—other
10. Family & carer support
11. Individual advocacy
12. Care coordination
13. Service coordination
14. Education, employment & training
15. Sector development & representation
16. Mental health promotion
17. Mental illness prevention

For descriptions of the NGO Establishments NMDS service categories see Appendix B

Survey Campaign

While the survey was developed centrally through the project steering committee, promoting the survey and encouraging CMOs to respond was a coordinated activity undertaken by each state and territory peak body. Promotional messages, articles and newsletters were developed by each peak (***for an example, see Appendix C***) and distributed to sector mailing lists beginning 2nd November 2012. The Mental Health Council of Australia provided their member mailing list and two promotional e-mails were distributed by the project officer. Promotional materials for the survey were also posted on the CMHA and AMHOCN websites.

At the end of the initial survey campaign deadline (14th December 2012) it was determined by the project team that a number of issues had created limitations with the survey response:

- The survey campaign was occurring during the Christmas / New Years holiday period. This caused communication delays with many potential respondents.
- The 81 completed responses fell short of the project target of 125, which was estimated to be the approximate number of CMOs required for the sample to be significantly representative.
- Responses were disproportionately skewed toward Queensland, with insufficient responses from New South Wales, South Australia and Western Australia.

The Steering Group agreed to extend the survey deadline to 25th Jan 2013. The project officer worked with the jurisdictional peak bodies to identify, personally contact and encourage non-responding organisations likely to be undertaking outcome measurement activities to complete the survey.

Funder and Stakeholder Consultations

The second component of project consultation was to investigate current activities, future plans and general attitudes toward outcome measurement in the CMO sector. A set of interviews were undertaken with funders of mental health CMO services, identified through membership of MHISSC, and other stakeholders on current and planned CMO outcome measurement activities. The final question (Q56) of the CMO sector survey was also designed to elicit general perspectives from CMOs on outcome measurement.

The results of these interviews (including CMO general comments) are provided after the survey report.

The National CMO Outcome Measurement Project also conducted two consultations after writing this report:

- Presentation and feedback session at National Consumer and Carer Forum / Register of Consumers and Carers Annual Forum (May 2013)
- Funder and stakeholder workshop with key MHISSC members to develop final recommendations (May 2013)

Survey Results – General Findings

Responses and Validation

Survey links went live on Friday 2 November 2012 and closed on Friday 25 January 2013. Peak bodies assessed the response list to ensure that most major CMOs in the sector provided a completed response, including the head offices of well-known national mental health organisations. Invalid responses were deleted at survey close.

Prior to inspection and data cleaning, the tally of survey responses was:

- **Completed: 147**
  A complete response was when the survey taker reached the last page of the survey.
- **Partial: 295**
  A partial response was when the survey taker left the survey part-way through, and might have intended to come back to finish later. Some partial responses were able to be converted to complete responses if enough pages had been completed.
- **Disqualified: 97**
  A disqualified response was when the survey taker selected options at the screening page that indicated they were a government entity, were a for-profit organisation, were a Medicare Local, did not have authority to respond for their organisation, or did not provide services funded specifically for mental health.
- **Abandoned: 259**
  An abandoned response was when a visitor did not progress past the first page.

After line-by-line inspection, 21 completed responses were removed either through disqualification or because they were superseded by a response from a more senior level of the organisation. Some organisations which share the same name but cover separate geographic boundaries (e.g. Centrecare, the fellowships, carers organisations) were left separated if they were found to be separately incorporated and governed. The partial responses were inspected and 6 responses, which had completed all tool use pages but not progressed beyond the information infrastructure page, were converted to Complete responses.

The final number of complete and valid responses was **132** organisations.

Response data underwent considerable inspection and cleaning, including:

- Removal of Complete/Partial responses with invalid data (e.g. fake responses)
- Removal of responses by divisions of organisations that were superseded by more senior respondents
- Conversion of Partial to Complete – as described above
- Technical cleaning of response data

Sample Representativeness

In order to establish the representativeness of the survey sample a firm estimate of the size of the Australian mental health CMO sector is required. However, due to low levels of research about the Community Managed (CMO) sector and the historically programmatic focus of CMO data collection, there is no consensus view on the number of mental health CMOs in Australia.

The *National Mental Health NGO Workforce Scoping Study* (National Health Workforce Planning and Research Council, 2011) estimates that approximately 800 organisations provide mental health services across Australia. This is an iterative estimate based on the combined number of unique member organisations within all state and territory mental health CMO peak bodies.

There is reason to believe that the target population for the National CMO Outcome Measurement Survey is substantially smaller than 800 organisations. Mental health peak body memberships include a proportion of “Type 3” CMOs, which are organisations that may provide mental health services but do not have specifically funded mental health programs. The *Mental Health Coordinating Council Sector Mapping Survey* (2010) of NSW mental health CMOs found that 14.2% of respondents were Type 1 (providing mental health programs only), 41.3% were Type 2 (providing mental health and other programs) and 44.5% were Type 3 (providing mental health services but no specific programs).

To maintain focus on mental health outcome measurement tools it was decided by the project management committee that the scope of responses should be limited to organisations that provide specific mental health programs (i.e. Types 1 and 2 CMOs only). Therefore the fourth survey screening question allowed respondents to proceed only if they were “funded specifically for people who are experiencing or have experienced mental illness, their family or carers.” This reduces the approximate target population size to approximately 450 organisations.

There is also a high likelihood that many CMOs that do not use outcome measurement tools did not complete the survey. Only 9% of organisations responding to the survey indicated that they did not use and kind of outcome measurement tool. The Mental Health Coordinating Council Sector Mapping Survey of NSW Mental Health CMOs (2010) found that approximately 60% of responding Type 1 CMOs used outcome measurement tools, and only 34% of responding Type 2 CMOs used outcome measurement tools. This suggests that the number of potential responding CMOs out of 450 who are likely to use outcome measurement tools would less than 300.

In this context a complete and validated response rate of 132 organisations is expected to be a highly representative sample of the outcome measurement activities of mental health CMOs in Australia.

Responding Organisation Characteristics

Of the 132 CMOs responding to the survey, 31 CMOs (23%) indicated that they operated in more than one jurisdiction, and 12 CMOs (9%) indicated that they operated nation-wide. The nation-wide CMOs included most well known nationally-funded organisations providing direct mental health services.

Table 1. Reported jurisdictions of CMO operation

Jurisdiction	No. Orgs
Queensland	41
New South Wales	46
Australian Capital Territory	30
Victoria	50
Tasmania	18
South Australia	21
Northern Territory	17
Western Australia	34
Multiple States (incl. National)	31
National (Covers all jurisdictions)	12

To gain a general sense of the size of CMOs in the sector, respondents were asked to approximate the Full Time Equivalent (FTE) number of employed staff in their organisation (or division). For the purpose of this report CMOs were then categorised as Small, Medium, Large or Very Large according to their approximate FTE staffing as listed below. These categories are used to cross-tabulate results where significant relationships were found between responses and CMO size.

Table 2. Approximate CMO Full Time Equivalent Staff Counts

CMO Size	FTE staff	No. Orgs	Percent of responses
Small	0-9	32	28%
Medium	10-99	46	40%
Large	100-499	26	23%
Very Large	500-3500	10	9%
(unsure/did not answer)	*N/A*	*18*	

For a full graph of the distribution of organisation size by FTE see Appendix D

CMOs were asked which categories of mental health service activity they undertook and were allowed to select multiple service types based on the AIHW *NGO Establishment National Minimum Data Set* Taxonomy (Australian Institute of Health and Welfare, 2012). These selections were then presented later in the survey to allow respondents to attribute specific service types to each outcome tool that is utilised.

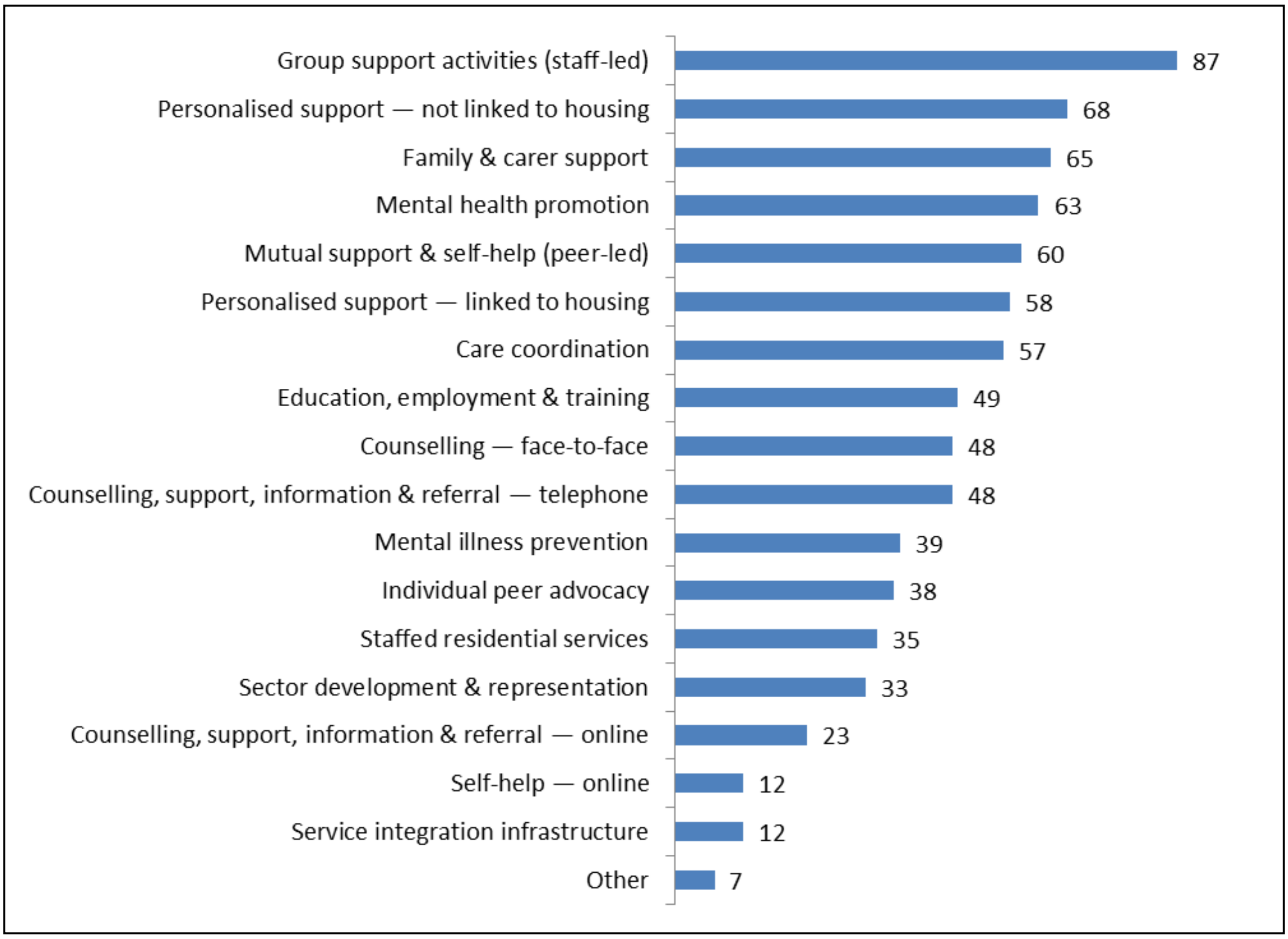
Table 3. Categories of mental health services provided by Responding CMOs

Service Category	No. Orgs
Care coordination	57
Counselling — face-to-face	48
Counselling, support, information & referral — online	23
Counselling, support, information & referral — telephone	48
Education, employment & training	49
Family & carer support	65
Group support activities (staff-led)	87
Individual peer advocacy	38
Mental health promotion	63
Mental illness prevention	39
Mutual support & self-help (peer-led)	60
Personalised support — linked to housing	58
Personalised support — not linked to housing	68
Sector development & representation	33
Self-help — online	12
Service integration infrastructure	12
Staffed residential services	35
Other	7

Other service types:

- Alcohol & Other Drug
- Clinical Services
- Community Development
- Research
- Community Housing and/or unstaffed accommodation

Figure 1. Categories of mental health services provided by Responding CMOs, sorted by no. of organisations



When the number of service categories provided by each organisation was totalled, it was found that the average organisation provided roughly 6 different categories of mental health service.

Table 4. Total number of service categories provided by each CMO

No. of service categories	No. Orgs
15 categories	1
14 categories	2
13 categories	3
12 categories	5
11 categories	7
10 categories	4
9 categories	13
8 categories	11
7 categories	12
6 categories	11
5 categories	14
4 categories	11
3 categories	11
2 categories	14
1 categories	13

Domains of Outcome

Organisation were asked two sets of questions involving the project team's developed set of CMO outcome measurement domains. They were first asked which domains they *currently* measure in their organisation, and then they were asked which domains they *are interested* in measuring. The results are displayed below. Many organisations indicated a desire to move away from measuring service satisfaction and activities of daily living. The domains of outcome CMOs were more interested in measuring were those involving the personal recovery journey, quality of life, social inclusion and interpersonal relationship functioning.

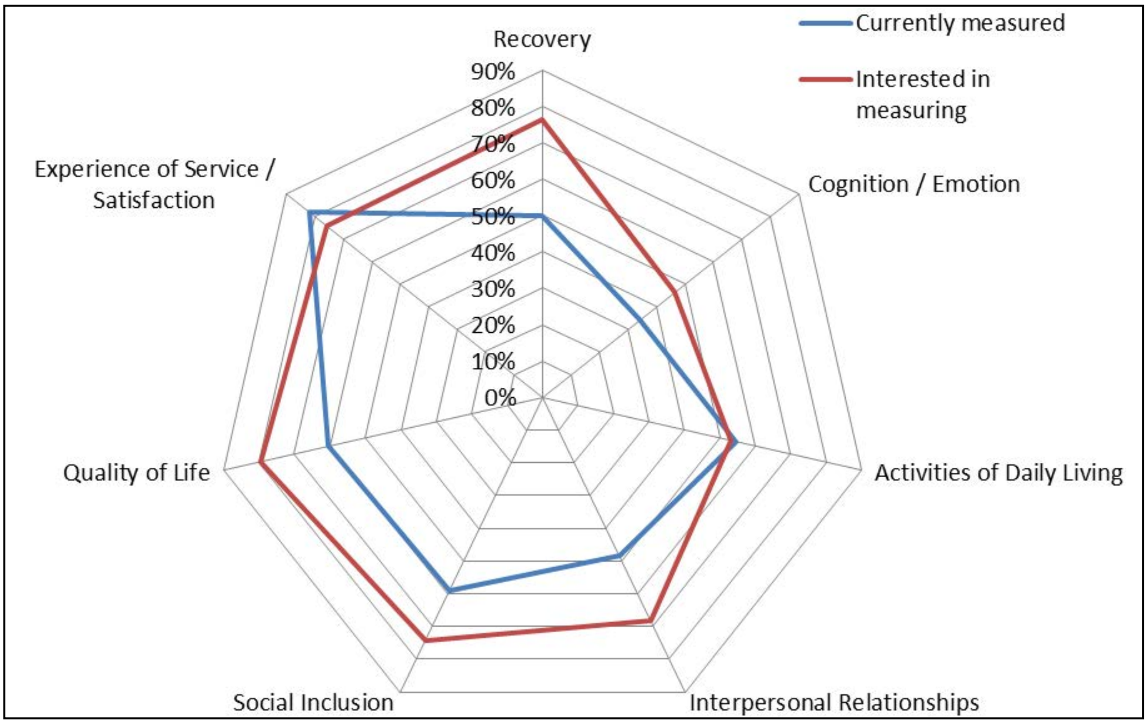
Table 5. Outcome domains – currently measured and interested in measuring

Outcome Domain	Currently measured (No. Orgs)	Interested in measuring (No. Orgs)
Recovery	66	101
Cognition / Emotion	45	61
Activities of Daily Living	72	70
Interpersonal Relationships	64	90
Social Inclusion	78	98
Quality of Life	80	105
Experience of Service / Satisfaction	108	100
Other	12	8

Other currently measured domains:

- Individual needs and/or goals
- Intention to access other forms of help or services
- Reduction or changed levels of hospital/ED presentations

Figure 2. Outcome domains – Percent of CMOs currently measuring and interested in measuring



Reasons for using outcome measurement tools

Responding CMOs were provided with a range of potential reasons for their use of outcome measurement tools. The results indicate that most organisations are using outcome tools for the purposes of assessing client needs and care planning, as well as for quality improvement (program evaluation and planning).

In spite of research indicating that routine use of outcome measurement tools can provide significant benefit to consumers and workers (Miller, Duncan, Brown, Sorrell & Chalk, 2006), less than half of respondents indicated that they used outcome tools as an aid to the therapeutic relationship.

Table 6. CMO reasons for using outcome measurement tools

Reason for using Outcome Tools	No. Orgs	% of Total
Program evaluation	104	79%
Identification of unmet client needs	99	75%
Program planning or development	97	73%
Development of individual care plans	83	63%
As evidence for use in funding applications	78	59%
Aiding the therapeutic relationship	53	40%
Mandatory - funding requirement	52	39%
Encouraged by funders (but not mandatory)	38	29%
Other	9	7%

Other reasons for using outcome measurement tools:

- Providing progress feedback to clients
- General quality improvement and innovation
- Providing evidence to tribunals or guardianship hearings
- Providing data for research

Survey Results – Specific Tool Use

Respondents were asked if their organisation utilised any Established, Modified, Custom-made (quantitative) or Qualitative tools. For descriptions of the difference between these types of tools see question 13 in Appendix A. CMOs indicating that they don't measure outcomes of any kind could select "none of the above" and were asked a small set of questions exploring their reasoning.

Table 7. Types of outcome measurement tools in use

Tools Used	No. Orgs
Established / formal (quantitative) tools	63
Modified or a combination of established (quantitative or qualitative) tools	29
Custom-made (quantitative) questionnaires or tools your organisation created itself	80
Qualitative questionnaires or tools for outcomes	69
None of the above*	11

*The low response rate of CMOs not using outcome measurement tools should not be interpreted as indicative of the sector as whole. This is due to the likelihood that CMOs not using outcome measurement tools were far less likely to respond to the survey.

After specifying all established tools in use, some respondents indicated that up to 10 different Established tools were in use in their organisation. These tools may have been combined into a single form or used in different ways for different programs, or under specific conditions.

When including Modified tools the highest number of tools in use remained at 10, however if including both Custom-made and Qualitative tools the highest number of tools in use rose to 12.

Table 8. Total number of tools in use across all service types within a CMO

Total number of Tools in Use	Established tools only (No. Orgs)	Established or modified tools (No. Orgs)	All tools/forms (No. Orgs)
12 tools	0	0	2
11 tools	0	0	2
10 tools	1	2	0
9 tools	2	2	6
8 tools	2	2	5
7 tools	2	2	9
6 tools	4	8	4
5 tools	10	11	9
4 tools	11	7	9
3 tools	8	7	13
2 tools	10	10	30
1 tools	12	21	32
	70 CMOs did not use established tools.	*60 CMOs did not use established or modified tools.*	*11 CMOs did not use any tools.*

Jurisdictional use of tools

Table 9. Tools in use by organisations operating in each jurisdiction

Jurisdiction	Total no. of Orgs	Established or modified tools	Custom-made tools	Qualitative surveys or tools	None
Queensland	41	22	31	26	2
New South Wales	46	28	31	28	2
Australian Capital Territory	30	23	21	20	0
Victoria	50	35	27	29	2
Tasmania	18	12	14	15	1
South Australia	21	12	14	15	1
Northern Territory	17	10	13	15	0
Western Australia	34	21	22	19	3
Multiple States (incl. National)	31	25	21	21	0
National	12	8	10	12	0

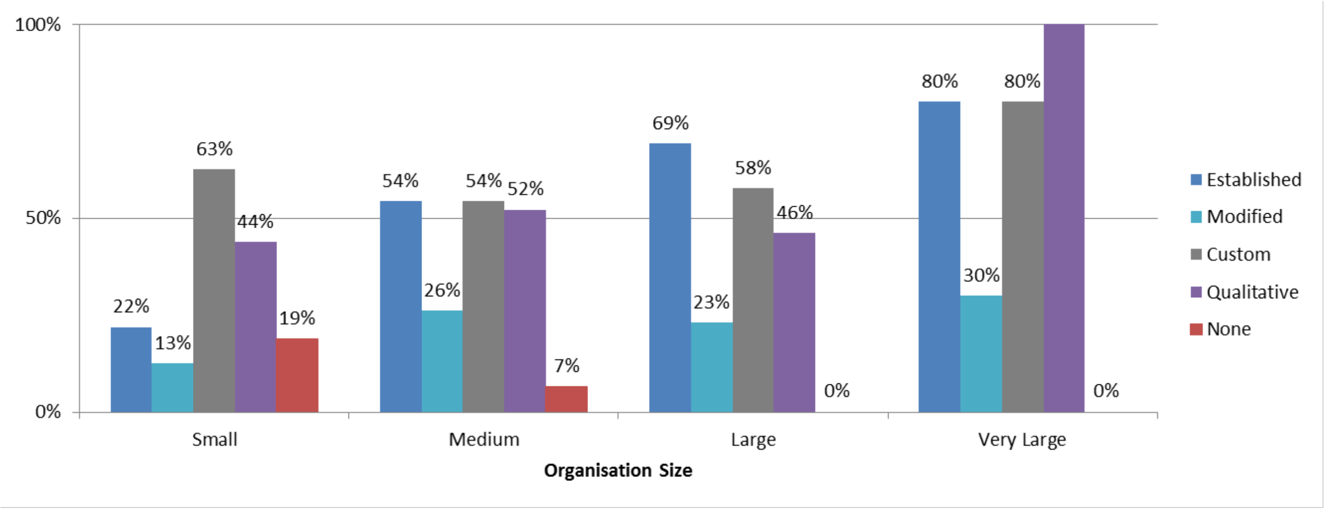
Relationship between organisation size and tool use

Tool use counts were divided into the categories of organisation size as described on page 15. A clear relationship appears to exist between the size of a CMO and the likelihood of using outcome measurement tools. Larger CMOs are more likely to use established tools, and are also more likely to develop their own custom-made tools. Smaller CMOs are less likely to use outcome measurement tools overall.

Table 10. Types of tools in use by organisation size

Org Size	Total No. Orgs	Established	Modified	Custom	Qualitative	None
Small	32	7	4	20	14	6
Medium	46	25	12	25	24	3
Large	26	18	6	15	12	0
Very Large	10	8	3	8	10	0

Figure 3. Types of tools in use – percentage used by organisation size



Established Tools

A total of 62 organisations declared use of established individual outcome measurement instruments (published and/or are at some stage of validation). Responses were verified by journal and grey literature searches to establish documented usage somewhere in the international mental health community. The most common established tools are listed below.

> ***The definition of "established tool" in this survey was intentionally defined as an instrument that has been published and/or psychometrically validated. This was in order to maximise the capacity of the survey to collect information on tools that are gaining in popularity but which may not have had the opportunity to be scientifically scrutinised or validated.***

Responding organisations indicated that they use a diverse range of established outcome tools. No one tool appeared to be the most outstandingly popular, however eight tools do appear to be in high usage across organisations. They are listed at the top of the following table.

Table 11. Established tools in use by more than one CMO

Tool Name	No. Orgs
K10	22
BASIS-32	16
CANSAS	15
DASS	13
HoNOS	13
LSP-16	13
WHO-QoL	12
Recovery STAR	12
RAS	6
GAF	5
Personal Wellbeing Index	5
Homeless STAR	5
Recovery Interview	4
Mental Health Recovery Measure	3
ORS & SRS	2
APQ6	2
SF12	2
SIQ	2
Recovery Enhancing Environment Measure	2
PSI	2

Many organisations indicated that they used tools that were not identified by any other organisations. These were the ADIS, AUDIT, BAS, BASH-B, BDI-II, Brief COPE, CAMCOG, CBCL, Collaborative Recovery Model Tools, Consumer Evaluation of Outcomes, CRAM, CSSRS, DECAC, DES II, DLC, DUDIT, EDI3, Edinburgh Postnatal Depression Scale, Empowerment Scale, EPDS, GAS, GHS-V, HADS, Home Inventory, HoNOSCA, ICG-R, IRIS, M3, MANSA, MDI, MHSIP, MHQ14, MMSE, NEO-P3, OL, PCL-C, PEDS, Personal Outcomes Measures (Council on Quality and Leadership), REE, RUDAS, SCL90, SISR, SOFAS, Sphere-12, SRM 1 & 2, STAI, STORI, WEMWEBS, Work and Social Adjustment Scale, Working Alliance Inventory, and one tool with an unknown name (provided by funder).

For a breakdown of established tool use by organisation size see Appendix E.

For usage characteristics of the 8 most common established tools see Appendix F.

For a breakdown of established tool use by service category see Appendix G.

For CMOs operating in each jurisdiction, the most common established tools were:

Table Set 12. Established tool use by jurisdiction

Queensland	No. Orgs
K10	11
DASS	7
HoNOS	5
CANSAS	4
WHO-QoL	4
GAF	4
LSP-16	3
RAS	2
Personal Wellbeing Index	2
APQ6	2
SF12	2

New South Wales	No. Orgs
K10	13
DASS	9
CANSAS	9
HoNOS	6
WHO-QoL	5
GAF	5
LSP-16	5
Personal Wellbeing Index	4
RAS	2
APQ6	2
SF12	2
Recovery STAR	2
SIQ	2

Australian Capital Territory	No. Orgs
DASS	8
LSP-16	8
K10	6
HoNOS	5
WHO-QoL	5
GAF	4
BASIS-32	4
CANSAS	3
Personal Wellbeing Index	3
Recovery Interview	3
SF12	2
Recovery STAR	2

Tasmania	No. Orgs
K10	6
WHO-QoL	5
DASS	4
GAF	3
HoNOS	2

Victoria	No. Orgs
BASIS-32	14
K10	10
WHO-QoL	10
HoNOS	8
CANSAS	7
Recovery STAR	7
DASS	6
LSP-16	4
GAF	4
RAS	4
Homeless STAR	3
Personal Wellbeing Index	2
SF12	2
Recovery Enhancing Environment Measure	2

South Australia	No. Orgs
K10	7
WHO-QoL	6
DASS	5
GAF	4
HoNOS	3
CANSAS	2
Personal Wellbeing Index	2
SF12	2

Northern Territory	No. Orgs
DASS	5
K10	4
WHO-QoL	3
GAF	3
HoNOS	2

Western Australia	No. Orgs
K10	11
DASS	6
WHO-QoL	4
GAF	4
HoNOS	4
Personal Wellbeing Index	3
SF12	2
Recovery Interview	2
CANSAS	2
RAS	2
Mental Health Recovery Measure	2

Table Set 13. Established tools used by organisations in more than one jurisdiction

Multiple (2+) Jurisdictions*	No. Orgs
K10	10
DASS	8
CANSAS	8
WHO-QoL	7
HoNOS	6
LSP-16	5
GAF	4
Personal Wellbeing Index	3
BASIS-32	3
SF12	2
RAS	2
Recovery STAR	2
APQ6	2

*includes national organisations

National Orgs (Covering all jurisdictions)	No. Orgs
K10	4
DASS	4
WHO-QoL	3
GAF	3
HoNOS	2

Tool completion

Respondents were asked who would complete an established tool that they used (CMOs answered for each tool). Most CMOs indicated that their tools were rated by the consumer (90% of organisations), and three quarters of CMOs indicated that their tools were rated by the CMO worker. Around one quarter of CMOs indicated that their tools were completed by family or carers, and around one in ten CMOs obtained tool data from Local Health District (LHD) workers.

Table 14. Person completing tool - all established tools

Tool Completed By	(No. Orgs)	% of Orgs using Est. Tools
Client*	56	90%
Family or Carer	15	24%
CMO Worker	46	74%
Other**	7	11%

*Respondents were instructed to count family or carers specifically using a family and carer service as a "client"

**The most common "other" person was a worker from the Local Health District. Individual respondents also identified research/evaluation consultants and staff from the Commonwealth Department of Education, Employment and Workplace Relations

The top eight most common established tools were further assessed for how many organisations using each tool had consumers, carers, CMO Workers and LHD Workers using the tool. In general, this enabled a delineation to be made between tools that are more likely to be completed by the consumer (K10, BASIS-32, CANSAS, DASS) and tools that are more likely to be completed by the workers (HoNOS, LSP-16, WHO-QoL, Recovery STAR)

Table 15. Person completing tool – most common established tools (% of CMOs using each tool)

Tool Completed By	K10	BASIS-32	CANSAS	DASS	HoNOS	LSP-16	WHO-QoL	Recovery STAR
Client	77%	94%	93%	85%	15%	62%	62%	62%
Family member or carer	5%	19%	7%	15%	0%	8%	8%	8%
CMO Worker	50%	63%	67%	38%	85%	77%	77%	77%
LHD Worker	5%	0%	0%	0%	15%	0%	0%	0%

Funding conditions

Respondents were asked whether each established tool was required to be used as a condition of funding. The only tool that a majority of CMOs were required to use as condition of funding was the LSP-16, though a number of other established tools were close to this amount.

Table 16. Requirement to use established tool as condition of funding

Tool	Required by Funder	Not Required	% Required
K10	7	15	32%
BASIS-32	5	11	31%
CANSAS	4	11	27%
DASS	2	11	15%
HoNOS	6	7	46%
LSP-16	7	6	54%
WHO-QoL	5	7	42%
Recovery STAR	1	10	9%
RAS	0	6	0%
GAF	2	3	40%
Personal Wellbeing Index	2	3	40%
Homeless STAR	0	4	0%
Recovery Interview	1	3	25%
Mental Health Recovery Measure	1	2	33%

Usage timeframes

Respondents were asked how frequently they used each established tool. Most CMOs indicated a usage of established tools between approximately quarterly or half-yearly increments.

Table 17. Established tools - usage frequency

Established Tool	At every contact	At least once a month	At least once every 3 months	At least once every 6 months	Less than once in 6 months
K10	1	0	4	3	3
BASIS-32	0	0	2	8	1
CANSAS	0	0	5	7	1
DASS	0	1	3	3	0
HoNOS	0	0	5	3	1
LSP-16	0	0	4	4	0
WHO-QoL	0	0	2	6	0
Recovery STAR	0	0	3	5	0
RAS	0	0	1	1	2
GAF	1	2	0	0	0
Personal Wellbeing Index	0	0	2	0	0
Homeless STAR	0	0	3	1	0
Recovery Interview	0	1	0	1	0
Mental Health Recovery Measure	0	0	1	1	0

For a breakdown of established tool use by organisation size see Appendix E.

For usage characteristics of the 8 most common established tools see Appendix F.

For a breakdown of established tool use by service category see Appendix G.

Modified Tools

29 organisations reported using modified or combined tools based on established tools. Tool modification can involve:

- Combination of multiple established tools into a new tool,
- Expansion, reduction or modification of question items in an established tool to suit the purpose of the mental health service,
- Language modification,
- Usage of tool in a method divergent to the tool's instructions (e.g. different rater, group usage, online or digital implementation, etc)

It is important to note that modification of an established tool can invalidate the psychometric properties of the instrument, which reduces the validity of any collected data. It is standard practice for a modified tool to undergo revalidation and peer-review before it is acknowledged as a viable alternative to the established tool's format.

The following established tools were utilised as part of a CMO's modified tool:

Table 18. Use of established tools within a modified tool

Tool used	No. Orgs
Multiple unspecified tools	4
BASIS-32	3
CANSAS	2
HoNOS	2
LSP-16	2
Personal Wellbeing Index	2
K10	1
WHO-QoL	1
Work and Social Adjustment Scale	1
SRS and ORS	1
Rethink Mental Illness UK satisfaction survey	1
Rosenberg's Self Esteem Scale (RSES)	1
Webqual	1
*No established tools were used**	5

*Most organisations indicating that they used a modified tool and that "no established tools were used" appeared to have misread the question and were re-coded as having developed a custom survey/tool. The remaining 5 responses indicated that their tools were based on unpublished work.

Of the 15 organisations using modified or combined tools:

- 5 organisations contextualised the tools to the service requirements,
- 4 organisations changed the language of the tools to be more person-oriented, to remove medical terminology and to minimise professional jargon,
- 3 organisations abbreviated or used subsets of existing tools,
- 2 organisations added the tools to larger surveys, and
- 1 organisation converted the tool to an exit survey

Tool completion

Table 19. Modified tools – person completing tool

Tool completed by	No. Orgs
Client	16
Family member or carer	4
Worker (from your service)	9
LHD Worker	1

Organisations reported using their modified tools with 63% of clients (with 17% variance) and 20% of family or carers (with 14% variance).

Funder requirement

Only 2 organisations (7%) were required to use their modified tool as a condition of funding.

Usage timeframes

Table 20. Modified tools - points of usage

Point of tool use	No. Orgs
Entry to service	9
Exit from service	9
Regular intervals	10
Under specific conditions*	3

*Specific conditions were post-program follow-up or by invitation for an online service.

Table 21. Modified tools – regularity of usage

Usage amount	No. Orgs
At every contact	1
At least once a month	1
At least once every 3 months	3
At least once every 6 months	4
Less than once in 6 months	1

Custom-made (quantitative) Tools

There were 80 organisations that had developed their own custom tools containing some level of quantitative measurement.

Descriptions of custom-made tools were individually inspected and coded to the domains of outcome they are intended to measure. Most tools were measures of service satisfaction.

Table 22. Custom-made tools – domains of outcome

Outcome Domain	No. Orgs	%
Experience of service/satisfaction	47	59%
Recovery	5	6%
Cognition/emotion	5	6%
Interpersonal relationships	4	5%
Activities of daily living	3	4%
Social Inclusion	3	4%
Quality of life	2	3%
*(Evaluation/other)**	23	29%

*When asked what their custom tool measures, a number of respondents specified a tool or form that intended to measure organisation-level outcomes or impacts, or involved some other form of broad evaluation of program logic (e.g. Results Based Accountability Framework).

A total of 16 custom tools measured a domain of individual outcome which was not “Experience of service / satisfaction.” These will be labelled “multiple domains.”

Table 23. Custom-made tools – person completing tool

Tool completed by	No. Orgs	No. Orgs (multiple domains)
Client	64	14
Family member or carer	24	4
Worker (from your service)	39	10
Other*	12	2

Organisations reported using their custom tools with 75% of clients (10% variance) and 18% of family or carers (8% variance).

Funding conditions

16 organisations were required to use their custom tool as a condition of funding, 3 of these were organisations using custom tools measuring multiple outcome domains.

Usage timeframes

Table 24. Custom-made tools – points of usage

Point of tool use	No. Orgs	No. Orgs (multiple domains)
Entry to service	33	8
Exit from service	29	6
Regular intervals	57	12
Under specific conditions*	18	3

*Specific conditions were funder audits, annual reviews, evaluation/quality research, post-program follow-up or by invitation for an online service.

Table 25. Custom-made tools – regularity of usage

Usage amount	No. Orgs	No. Orgs (multiple domains)
At every contact	8	1
At least once a month	2	4
At least once every 3 months	12	4
At least once every 6 months	19	3
Less than once in 6 months	14	4

Qualitative Forms or Tools

There were 69 organisations that had used or developed qualitative (mostly open text) forms or tools.

Respondents were asked if and how their qualitative tools were thematically coded. Most tools were either manually analysed or coded by staff members or they were not methodically analysed in any way.

Table 26. Thematic analysis of qualitative tools

Thematic analysis/coding	No. Orgs
Yes - by a staff member	41
No	23
Yes - by a computer program*	4
(no response)	1

*Respondents indicating that they used computer programs to perform thematic analysis specified software which does not appear to have capacity to analyse qualitative data (e.g. Survey Monkey, Carelink+, etc).

Descriptions of qualitative tools were individually inspected and coded to the domains of outcome they are intended to measure. Most tools were measures of service satisfaction.

Table 27. Qualitative tools – domains of outcome

Outcome Domain	No. Orgs	% of Orgs using qualitative tools
Experience of service/satisfaction	46	67%
Recovery	4	6%
Cognition/emotion	3	4%
Interpersonal relationships	1	1%
Activities of daily living	3	4%
Social Inclusion	3	4%
Quality of life	3	4%
(Unclear/unspecified)	11	16%

A total of nine qualitative tools measured a domain of individual outcome which was not “Experience of service / satisfaction.” These will be labelled “multiple domains.”

Table 28. Qualitative tools – person completing tool

Tool completed by	No. Orgs	No. Orgs (multiple domains)
Client	52	11
Family member or carer	20	3
Worker (from your service)	30	6

Organisations reported using their qualitative tools with 54% of clients (20% variance) and 15% of family or carers (11% variance).

Funding conditions

15 organisations were required to use their qualitative tool as a condition of funding, 1 of these was an organisation using the qualitative tool to measure multiple outcome domains.

Usage timeframes

Table 29. Qualitative tools – points of usage

Point of tool use	No. Orgs	No. Orgs (multiple domains)
Entry to service	17	2
Exit from service	28	4
Regular intervals	39	11
Under specific conditions*	16	3

*Specific conditions were evaluation/quality research and post-program follow-up.

Table 30. Qualitative tools – regularity of usage

Usage amount	No. Orgs	No. Orgs (multiple domains)
At every contact	4	1
At least once a month	2	2
At least once every 3 months	8	4
At least once every 6 months	14	4
Less than once in 6 months	10	2

Organisations that don’t use outcome tools

11 organisations out of132 indicated that they do not use individual outcome tools of any kind.

These CMOs were provided with a multiple-choice set of possible reasons, and also allowed to enter an “other” text response. The reasons provided were:

Table 31. Reasons for not using outcome measurement tools

Reason	No. Orgs
Have not found any appropriate measures - too clinical	7
Have not found any appropriate measures - too academic	7
Information system can't collect outcomes	6
Overly time consuming/costly to collect	5
Not a contractual requirement	4
Overly time consuming/costly to train	3
Gets in the way of our work	3
Have not found any appropriate measures - no tools for our service model	2
Not interested in measuring outcomes	2
Intend to measure outcomes in the future	1
"18 sources of funding - 18 reporting frameworks"	1

When asked how these organisations measure the effect that their services are having on their clients organisations indicated that they either relied on verbal feedback and/or regular reviews of the client’s care plan.

Respondents were asked there were any ways that outcome measurement tools could be designed or modified so that they would be more likely to use them. The responses are listed below.

Table 32. Respondent suggestions to increase likelihood of CMOs using outcome measurement tools

“Develop better survey/questionnaires.”
“Happy to use existing measurements as prompts in discussion, but not forms.”
“Provide relevant information (e.g. contact with staff, readmissions to hospital, medication compliance).”
“Would need to develop outcome tools for advocacy.”
“On-line tools.”
“We would need to work with an expert in this field to find out what would be applicable to measure outcomes.”
“We would need culturally specific measuring tools - i.e. not the western model measurement tools that are currently in use as they do not adequately or correctly capture the state our clients are in.”
“Using individual measures would undermine participation in our program.”

Survey Results – Capacity to Collect Outcome Data

Capacity to Collect Outcome Data - Training

In order to understand how CMOs currently train their staff to use outcome measurement tools, respondents were provided with a multiple-choice set of training activities. While a majority of CMOs use tools based on their written instructions, less than half of CMOs formally trained their staff to use their outcome tools or forms.

Table 33. Training undertaken by CMOs

Type of Training	No. Orgs	%
We read the manual or paper and informally trained on the job	78	64%
We ran formal training sessions at orientation	53	44%
We have regular retraining	40	33%
We had the developer of the measure come and provide training	22	18%
We do not undertake training	14	12%
We ran train the trainer workshops	11	9%
(did not answer)	11	9%

Table 34. Training source of funding

Who paid for the training?	No. Orgs	%
CMO	99	82%
Program funder	18	15%
LHD	4	3%

CMOs mostly paid for their own training.

Capacity to Collect Outcome Data – Information Infrastructure

In order to aid understanding of the mental health CMO sector's capacity to collect client-level outcome data, respondents were asked a set of questions on their current information infrastructure and their capacity to modify current systems to enable outcome data collection and reporting.

Table 35. Data collection – level of computerisation

Outcome and/or client data collection method	No. Orgs	%
Manual (paper-based)	59	46%
Computerised - MS Office	78	61%
Computerised - Dedicated system	57	45%
Computerised - Other*	3	2%

*Other systems were either statistical analysis software (SPSS/SAS) or online survey software (Surveymonkey).

These responses can be re-categorised as organisations that are fully computerised, those that have a combination of computerised and manual data collection systems, and those that have fully paper-based data collection methods.

Table 36. Data collection – types of data collection readiness

Outcome and/or client data collection method	No. Orgs	%*
Fully computerised (no manual collection)	68	54%
Manual + computerised (either)	43	34%
Manual (no computerised system)	16	13%

*Total equals 101% due to rounding error.

Table 37. Data collection – capacity to collect outcome data

Data system capable of collecting outcome data?	No. Orgs	Cumulative capacity
Yes	55	47% capable of collecting outcome data
No - but it could be modified/expanded	36	77% capable with system modification
No - don't have the resources to modify/expand our system	22	96% capable if resources were provided
No - it cannot be changed	5	
Not applicable	9	

Table 38. Data collection – Dedicated systems

System Name	No. Orgs
Carelink+	9
(purpose built system)	4
Microsoft CRM	3
SHIP and other systems	2
The Care Manager (DCA)	2
TRAK	2

Other dedicated systems used by individual organisations were:

> Access/DC, ADIS, Alchemy, Athena Penelope2, CHINTARO, Clubhouse Database, Community Services Statistics, EDI 3, Extranet, Framework-i, iCase, MACSIMS, Mavis, MedTech, meri, Navision, PSLOS/IBA, recordbase, Research Engine, SAGE Act pro, SMS, SRS, Thomnis, TOMS, and TRACCS.

Nine organisations did not specify the name of their dedicated data system.

Table 39. Data collection – funding for dedicated systems

52. Dedicated system provided by funder?	No. Orgs
No	42
Yes - Commonwealth funder	4
Yes - State Funder	7

19% of organisations had data systems provided by their funders. These systems were:

- Access/DC
- ADIS
- Alchemy
- Extranet
- MDS System - *unspecified*
- SHIP (2 organisations)
- SRS
- The Care Manager – DCA (2 organisations)
- TRAK

Perspectives - Consumer and Carer Representative Comments on Outcome Measurement in the CMO Sector

The MHCA-nominated consumer and carer representatives on MHISSC were invited to provide service user perspectives to the project's consultations. Below are the summarised transcripts of interviews carried out by telephone.

Interview Summary for Lei Ning – Consumer Representative

General comment

Outcome measurement is an important area and Australia has made substantial progress. There is increasing demand for reliable information on the consumer experience and the 10 year Roadmap also talks about measurement of consumer care. When talking to consumer representatives it can become controversial when specific tools are discussed.

Response to common tools identified in this report

- Recovery STAR is new and interesting, but not rigorous enough to rely on its data. It seems to have generally been welcomed by the mental health consumer community.
- K10 and DASS are very popular now, but mainly because of their brevity. They are too clinical.
- BASIS-32 is too old. It contains inappropriate language and is too focused on symptoms.
- HoNOS is clinician rated and hence we are less interested. I understand the desire for this information but it seems to run counter to the original intention of using outcome measurement tools.
- CANSAS is very good in theory, but in reality it doesn't seem to be used how it was intended. In CMOs the emphasis is too much on reporting to government rather than as a process to engage consumers.

Importance of tools being consumer-rated

Consumer completion of the tools is an important issue. The whole idea of outcome measurement to begin with was to use the tools collaboratively so that the consumer and worker can talk about their progress.

Comments on good practice

This principle should be at the forefront, but in reality this doesn't happen. Forms are sent out in the mail, feedback is rarely received, and clinicians don't seem to pay enough attention to consumer responses. The one exception to this is service satisfaction information – that doesn't necessarily require a collaborative situation.

Relevance of established tools to CMO sector

There should be a good balance between brevity and rigor. Tools must yield valid data. However, in order for that data to have real value the tools must be relevant and easy to use. Most clinical tools do not appear to be relevant to the CMO sector. There may need to be a special tool developed.

While there are a broad range of service types in the CMO sector, this does not mean that a large number of new tools are required. The whole sector should be focused on the same general domains of outcome.

Interview Summary for Jackie Crowe – Carer Representative

General comment

The jurisdictions need to work harder to ensure that we are working towards national standardisation for outcome measurement. That said, funders need to be careful when connecting outcome tool data to funding decisions. Most tools are not necessarily a good indicator of service effectiveness. You can't equate tool scores with money well spent.

Outcome tools are about having a conversation. They are for the consumers and carers. They are about doing a needs assessment at the beginning and end, and the ensuing conversations provide the therapeutic value.

Response to common tools identified in this report

- Recovery STAR is useful, but too focused solely on the consumer. It substantially excludes the family. Using such tools by themselves denies the opportunity to include families in support and decision-making.
- It is essential that tools are consumer-rated. However family inclusion is also essential, and many tools do not accommodate this.

Comments on good practice

Tools need to be used correctly, and by this I mean that workers shouldn't just use the tools because they've been told to. The tools should be a component of the care plan. They need to be used to help consumers and carers understand their progress.

National development

Consensus needs to be built on tool use. It should be nationally standardised and de-identified data needs to be collected by CMOs and fed back to resources like the AMHOCN tool data repository.

A National Outcome and Casemix Collection for the sector would be good. It is difficult hearing from organisations that tools may or may not be appropriate. If you are learning more about the client then that can't really be bad information.

The project team also received a written response from the Victorian Mental Health Carers Network advocating for the revisiting of AMHOCN's work on carer outcome measurement. See Appendix H.

Perspectives - CMO Comments on Outcome Measurement in the CMO Sector

CMOs were provided a final comment text box requesting their general views on outcome measurement. 77 organisations took the opportunity to provide comment on outcome measurement issues.

It is important to note that most CMOs made the assumption that this project is focused on developing the capacity for outcome measurement tool use to be implemented by CMO workers, not by LHD workers.

The comments were thematically analysed and are listed below.

Thematic Summary of Comment	No. of Organisations Making Comment
Plans for routine tool use are in development	17
Expressed appreciation for the value of outcome measurement	9
Need for contextually appropriate tools for CMOs	8
Need for standardisation of tool use in the sector	6
Different CMO service types/sizes/target populations need different types of tools	5
Lack of funding/infrastructure/training impedes capacity to measure outcomes	5
Positive feedback on the Recovery STAR	4
Desire for more information and support to be made available on outcome measurement for CMOs	4
Strong data system/infrastructure is enabling better outcome data collection	3
High administrative burden collecting multiple sets of data for multiple funders	3
Positive feedback on the national project	2
Organisation has lost funding / may be closing	2
Not interested in outcome measurement	1
Individual plans are a form of outcome measurement	1
Still implementing output data collection, implementing outcome measurement would be difficult	1
Importance of incorporating Qualitative information in service evaluations	1
Carer input needs to be considered during outcome measurement	1
Lack of flexibility from funders	1
Importance of incorporating qualitative information	1
Concept of outcome is different in intellectual disability sector	1
Outcome scores are complicated to interpret	1

For full responses by CMOs and their attributed thematic summaries see Appendix I.

Perspectives - Government Funder Activities and Comments Regarding Outcome Measurement in the CMO Sector

The following government agencies were approached with a request to conduct a telephone interview on their current and planned outcome measurement activities for the CMOs that they fund:

- Queensland Health
- NSW Health (InforMH)
- Victorian Department of Health
- Northern Territory Department of Health and Families
- Western Australian Mental Health Commission
- SA Health
- Tasmanian Department of Health and Human Services
- Mental Health ACT
- Commonwealth Department of Health and Ageing
- Commonwealth Department of Families, Housing, Community Services and Indigenous Affairs
- National Mental Health Commission - Contributing Life Project Officer

Initial contact was made with the MHISSC representatives of each member agency and responses were provided by the representative or by other staff as deemed appropriate. Further documentation was requested if applicable.

Detailed responses were received (either by telephone or e-mail correspondence) from all agencies except for the Northern Territory Department of Health and Families.

The following pages contain a summary of funder activities, identified challenges to collecting outcome data in the CMO sector, potential processes for future outcome measurement activities, and other identified issues.

Summary of Funder Activities

Few programs funded by MHISSC member agencies are currently mandating the use of outcome measurement tools. Victoria and Australian Capital Territory have previously recommended tool use to CMOs.

Many jurisdictions are making future plans to implement outcome measurement as a component of their CMO program evaluations. However the two jurisdictions with firm policy developments are Western Australia and Victoria. Both jurisdictions are working with consumers, carers and CMOs to consider the use of standardised outcome measurement tools, and have indicated that this national project will help inform this work.

The National Mental Health Commission has funded the National Contributing Life Project, which will be a direct communication and consultation mechanism for Australians to provide general system feedback and future guidance for the Commission. This project is still in development, and more information is available on the National Mental Health Commission website.

This table contains a summary of current activities and future plans by government funders of CMO services for outcome measurement tool use.

	QLD	NSW	VIC	WA	SA	TAS	ACT	DoHA (D2DL)	DoHA (PIR)	FaHCSIA
Mandated Tools for CMOs	No	Yes	No	No	No	No	No	No	No	Yes
Details	N/A	APQ6 for RRSP only	N/A	N/A	N/A	Packages of Care (Cth) = CMOs must use an OM	N/A	N/A	N/A	"Evidence based tools" as per the National Standards. Reporting not mandatory.
Recommended Tools for CMOs	No	No	No	No	No	No	No	No	In development	Yes
Details	N/A	No, but there seems to be good CMO awareness of the LHD mandated suite of tools, NOCC	Policy in 2004 recommended the WHO-QoL, CANSAS and BASIS 32. Does not currently recommend tools, but is considering a new policy position.	N/A	Recommend the WHO-QoL on an informal basis. CANSAS & BASIS 32 also popular	However, CMOs commonly use WHO-QoL and Recovery Star	Offered tools like BASIS 32, & HoNOS, but CMOs didn't like them. K10 & LSP-16 gained more interest.	The program doesn't lend itself easily to OM – no "clinical" activity, no mandatory requirements on entry.	N/A	"Evidence based tools" as per the National Standards.

(continued)

	NSW	VIC	QLD	WA	SA	TAS	ACT	DOHA (D2DL)	DOHA (PIR)	FaHCSIA
Future Plans for OM use	Maybe	Yes	No	Yes	Yes	Yes	Yes	Maybe	Yes	Yes
Details	Considering K10 (Boarding House HASI) and HoNOS (HASI Plus).	Will be developed in context of 2013-14 Victorian CMO reform. May entail mandating use of specific tools.	Not in scope and not a priority at this stage.	Underway - collaboration between the WAAMH outcomes measurement taskforce and Commission, CoMHWA and MHM2.	Will adopt OM tools from this project once there is consensus	Evaluation framework under review of which OM will be a part	Will incorporate OM into funding agreements on individual basis once there is consensus.	Have looked at OM tools & have done some work with Melbourne University. Waiting on this project.	National evaluation and monitoring project has commenced, including development of PIR MDS with client level OM tool(s).	Considering appropriate tools for new performance framework.
Evaluations of CMO OM use	Yes	Yes	Yes	No	Yes	No	No	No	No	Yes
Programs	Programs with public MH partners e.g. HASI	- Care Coordination - Intensive Home Outreach - Diversion & Substitution	Project 300	An evaluation was done by Deloitte in 2010-11 to assess the OM usage in WA CMO sector but otherwise no	A number of external evaluations on a range of programs	N/A	N/A	N/A	N/A	Targeted Community Care (Mental Health)
Tools Used	HONOS K10 SDQ	HONOS Basis 32 Life Skills Profile	HONOS		NOCC suite due to lack of consensus					Custom-made
Comments from Evaluation	Data collected by LHD workers	Data collected by LHD workers	Recommended use of social inclusion measure		NOCC tools for CMOs appeared to be a bad fit.					N/A

Challenges

- Not all CMO sector grant funding is managed centrally which increases complexity for mandating OM tools. (NSW)
- There are a huge array of CMO service models and types which makes it difficult to determine what 1-5 mandated tools might look like for such a wide and various sector. (NSW)
- Some states have not yet made the transition to outcome measurement, with outputs still the main activity being collected. (QLD)
- Lack of knowledge around outcome measurement in the CMO sector. (QLD)
- The need for funding to assist with the complicated and resource intensive nature of implementation. (QLD)

Issues for consideration

- It is important that an agreed substantive national rationale is reached regarding key objectives for the use of outcome measurement in the CMO sector, and that the preferred tool(s) for use by CMOs be determined on the basis of agreed parameters and/or selection criteria. (VIC)
- A number of interdependencies need to be considered, particularly the introduction of DisabilityCare Australia and the implications / opportunities this may present for the use of outcome measurement by CMOs. (VIC)
- It is very important to have a clear definition of outcome tools, for example as opposed to assessment tool, set out clearly as a part of this project. (NSW)
- You can't mandate too many tools for use. (NSW)
- There are many synergies in terms of challenges between the OMP and the minimum dataset project, and the recommendations for a suite of OM tools must be attached to the minimum data set. (NSW)
- The need for appropriate infrastructure and the capacity to train and retrain people. (QLD)
- Looking at how practical the tools are to apply, ease of use, relevance, cost of implementation (eg license fees, cost to small and rural organisations, cost of ongoing data collection and reporting) capacity of tool to reflect individual differences, capacity for capturing valid / reliable data. (WA)
- Information about the tools and (face to face) education/training on how the tools can be used, to be made available to the sector (WA)
- WA is as yet not clear about which tool can be used by the sector to collect / report individual client level data on goals, which then has the potential to inform the sector's performance on WA's six mental health outcome areas (and identify areas that need improvement).

Future Directions

- Mandating tools is best left for the CMO sector to make recommendations on. (NSW)
- While Victoria is keen to align as much as possible with other jurisdictions to enable nation-wide outcome analysis and program and policy learning, Victorian CMO reform timelines may require Victoria to determine its preferred approach prior to any national decision. (VIC)
- Agreement should be reached with the sector about identified tools. (WA)
- All information gathered in this project should be made available to the CMO sector. (WA)
- Clear identification of an agency is required to continue working on CMO outcome measurement sector development. (ACT)
- Important to consider which selected tools are acceptable to CMOs and do not place additional reporting burden. (FaHCSIA)

Next Steps

This report will be incorporated into a final project report, which will also contain a comprehensive literature review, a tool selection process and candidate set of tools, results of the stakeholder workshop, and future recommendations for outcome measurement in the CMO sector. The final report is scheduled to be tabled at MHISSC in October 2013. It will then be developed into a public document, and academic publication will be considered.

In order to facilitate tool identification in the project literature review, identified tools have been provided to AMHOCN for further investigation. Established tools identified in this report will be assessed for their psychometric properties, domains of outcome measurement, and other selection criteria. Custom and modified tools measuring domains *other than service satisfaction* that were uploaded through the survey mechanism will be reviewed as a final part of the literature review.

This report has uncovered a diverse range of outcome measurement tool usage that was not previously understood. The usage statistics provided in this report have enabled CMOs and funders to have a clearer perception of outcome measurement activities in the CMO sector. Identified tools in this report are also be valuable when developing data items on outcome measurement in future versions of government and sector minimum data sets, including the NGO Establishments National Minimum Data Set.

The training and information infrastructure responses in this report indicate that any realistic development of CMO outcome measurement capacity will require sector development activities to increase sector knowledge, skills and data system capacity for client-level outcome data.

References

Australian Institute of Health and Welfare. (2010). *Mental Health Non-Government Organisation Data Development Project: Preliminary Scoping Exercise.* Canberra, A.C.T.

Australian Institute of Health and Welfare. (2011). *Mental Health Non-Government Organisation Data Development Project: Phase 1 Final Report.* Canberra, A.C.T.

Australian Institute of Health and Welfare. (2012). *Mental Health Non-Government Organisation Establishments National Minimum Data Set: Specifications and collection manual. Version 2.3.* Canberra, A.C.T.

Commonwealth of Australia. (2009). *Fourth National Mental Health Plan: An agenda for collaborative government action in mental health 2009–2014.* Canberra, A.C.T.

Council of Australian Governments. (2012). *The Roadmap for National Mental Health Reform 2012-2022.* Canberra, A.C.T.

Mental Health Commission Western Australia. (2012). *Mental Health Outcome Statements.* Perth, W.A.

Mental Health Coordinating Council. (2010). *The NSW Community Managed Mental Health Sector Mapping Report 2010.* Sydney, N.S.W.

Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, M. B. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *Journal of Brief Therapy*, *5*(1), 5-22.

National Health Workforce Planning and Research Collaboration. (2011). *Mental Health Non-Government Organisation Workforce Project Final Report.* Adelaide, S.A.

National Mental Health Commission. (2012). *A Contributing Life: The 2012 National Report Card on Mental Health and Suicide Prevention.* Sydney, N.S.W.

NSW Consumer Advisory Group – Mental Health Inc. & Mental Health Coordinating Council. (2011). *Recovery Oriented Service Self-Assessment Toolkit (ROSSAT): A Recovery Oriented Service Provision Quality Improvement Resource for Mental Health Services.* Sydney, N.S.W.

Appendixes

Appendix A – Survey Question List

This is a transcript of a survey involving question logic, page piping and hidden items depending on earlier responses. For this reason the survey may appear to be longer than experienced by any one respondent.

Checks

1) My organisation is non-government / community-managed and not for profit.*

( ) Yes
( ) No

2) My organisation is *not* a Medicare Local.*

( ) Yes
( ) No

3) I have authority to answer this survey for my organisation (or branch), and have knowledge of our outcome measurement activities.

This survey should ideally be completed by the CEO or manager of your organisation, however if this is not possible, or your organisation is clearly delineated, please respond for your branch and encourage other branches of your organisation involved in mental health service provision to respond as well.*

( ) Yes
( ) No

4) My organisation (or branch) provides services that are funded specifically for people who are experiencing or have experienced mental illness, their family or carers. Alternately my organisation (or branch) is funded specifically for mental health promotion or mental illness prevention activities, including crisis counselling.*

( ) Yes
( ) No

Service location

5) In which states or territories does your organisation (or branch) operate?

(select all that apply)
[ ] Queensland
[ ] New South Wales
[ ] Australian Capital Territory
[ ] Victoria
[ ] Tasmania
[ ] South Australia
[ ] Northern Territory
[ ] Western Australia

Organisation details

These questions help us ensure the survey is a reliable cross-section of the sector.
Identifying information that CMHA collects in this survey will **not** ***be shared with any other agencies, including the Department of Health and Ageing. Only de-identified and aggregate data will be used in the final report unless express permission is provided.***

6) What is the name of your organisation or service?*

____

7) If your service is a branch of a larger organisation or owned by a parent entity please provide the name of that organisation.

____

8) What is your position?

( ) CEO / Director
( ) Manager
( ) Project Officer or Co-ordinator
( ) Research or Policy Officer
( ) Team leader
( ) Care Worker
( ) Other

Other - what is your job title?

____

9) Roughly how many Full Time Equivalent (FTE) staff are employed by your organisation?*

If you don't know, please answer "0"

____

Service types

10) Please indicate which mental health services are *primarily* provided by your organisation (or branch).*

Select the services that you are specifically funded to provide and/or you have staff where the activity is a major component of their role.

[ ] Counselling — face-to-face [?]
[ ] Counselling, support, information & referral — telephone [?]
[ ] Counselling, support, information & referral — online [?]
[ ] Self-help — online [?]
[ ] Group support activities (staff-led) [?]
[ ] Mutual support & self-help (peer-led) [?]
[ ] Staffed residential services [?]
[ ] Personalised support — linked to housing [?]
[ ] Personalised support — not linked to housing [?]
[ ] Family & carer support [?]
[ ] Individual peer advocacy [?]
[ ] Care coordination [?]
[ ] Service integration infrastructure [?]
[ ] Education, employment & training [?]
[ ] Sector development & representation [?]
[ ] Mental health promotion [?]
[ ] Mental illness prevention [?]
[ ] Other

Domains of outcome measurement

The Productivity Commission undertook a review of the not for profit sector in March 2009, and proposed a measurement framework to guide the collection of data, evaluation approaches and reporting about the not for profit sector in Australia. The framework consists of a hierarchy of four contribution measures:

- **inputs (measures of the resources used)**
- **outputs (indicators of the level of activity undertaken)**
- **outcomes (direct effects on activity participants)**
- **impacts (longer term effects on the participants and the community more broadly).**

The use of outcome measures can be individually and organisationally determined to meet the identified service or community need, or funding requirements. This survey is concerned with tools that aim to measure *outcomes* for *individual consumers and carers*.

11) What are the aspects of client outcomes that you *currently* measure?

[ ] Recovery [?]
[ ] Cognition / Emotion (incl. medical symptoms) [?]
[ ] Activities of Daily Living [?]
[ ] Interpersonal Relationships [?]
[ ] Social Inclusion [?]
[ ] Quality of Life [?]
[ ] Experience of Service / Satisfaction [?]
[ ] Other

12) What are the aspects of client outcomes that you are *interested in* measuring?

(including those you currently measure)
[ ] Recovery [?]
[ ] Cognition / Emotion (incl. medical symptoms) [?]
[ ] Activities of Daily Living [?]
[ ] Interpersonal Relationships [?]
[ ] Social Inclusion [?]
[ ] Quality of Life [?]
[ ] Experience of Service / Satisfaction [?]
[ ] Other

Outcome tool use

13) Thinking of the ways you measure individual client and carer outcomes, do you use any of the following?*

(Select all that apply)
[ ] Established / formal (quantitative) tools
Tools that have been published or psychometrically validated, where responses are able to be counted, averaged and calculated. e.g. *HoNOS, K10, CANSAS, BASIS32, etc*
[ ] Modified or a combination of established (quantitative or qualitative) tools
A modification or combination of established / formal tools to suit your service's objectives, where some responses are able to be counted, averaged and calculated.
[ ] Custom-made (quantitative) questionnaires or tools your organisation created itself
Forms or tools where most responses are able to be counted, averaged and calculated.
[ ] Qualitative questionnaires or tools for outcomes
Forms with mostly open text, qualitative responses.
[ ] None of the above

Reasons for using outcome measurement

14) What are the main reasons for your use of outcome measurement?

(Select all that apply)
[ ] Aiding the therapeutic relationship
[ ] Identification of unmet client needs
[ ] Development of individual care plans
[ ] Program planning or development
[ ] Program evaluation
[ ] As evidence for use in funding applications
[ ] Mandatory - funding requirement
[ ] Encouraged by funders (but not mandatory)
[ ] Other

15) What training has been provided to your workers on the outcome tools that you use?

(Select all that apply)
[ ] We do not undertake training
[ ] We read the manual or paper and informally trained on the job
[ ] We ran formal training sessions at orientation
[ ] We have regular retraining
[ ] We had the developer of the measure come and provide training
[ ] We ran train the trainer workshops

Who paid for the training?

[ ] [question("value"), id="26"]
[ ] Program funder
[ ] Other

Established tools

You indicated that you use established outcome measurement tools.

16) Which specific tools do you use?*

[ ] BAS [?]
[ ] BASIS-32 [?]
[ ] CANSAS [?]
[ ] CarerQol [?]
[ ] CGAS [?]
[ ] DASS [?]
[ ] GAF [?]
[ ] HoNOS [?]
[ ] HoNOS65+ [?]
[ ] HoNOSCA [?]
[ ] Homeless Star
[ ] IMR Scales [?]
[ ] IEQ [?]
[ ] Inclusion Web
[ ] K10 [?]
[ ] LSP-16 [?]
[ ] MHI [?]
[ ] Mental Health Recovery Measure
[ ] Personal Wellbeing Index
[ ] RUG-ADL [?]
[ ] RAS [?]
[ ] Recovery Interview
[ ] Recovery STAR

[ ] RPI [?]
[ ] SCOPE [?]
[ ] SDQ [?]
[ ] SF12 [?]
[ ] SF24 [?]
[ ] SIQ [?]
[ ] STORI [?]
[ ] WHO-QoL [?]
[ ] Work and Social Adjustment Scale
[ ] Other

Use of established tools
(Piping - Questions 17 to 22 are repeated for every tool specified in Question 16)

You indicated that you use the [specified tool – piped].

17) Who completes the tool?

(Select all that apply)
[ ] Client
[ ] Family member or carer
[ ] Worker (from your service)
[ ] Other

18) What percentage of your clients do you use this tool with?*

(including carers who are the primary clients of family and carer services)

____

19) What percentage of family or carers do you use this tool with?*

(excluding carers who are the primary clients of family and carer services)

____

20) Is use of [specified tool – piped] required as a condition of funding?*

( ) Yes
( ) No

21) At which points of service delivery do you use this tool with the client?

[ ] Entry to service
[ ] Exit from service
[ ] Regular intervals
[ ] Under specific conditions

Generally, how regular is the use of this tool?

( ) At every contact
( ) At least once a month
( ) At least once every 3 months
( ) At least once every 6 months
( ) Less than once in 6 months

Which specific conditions would trigger use of this tool?

____

22) Of the services you provide, which of these involve use of [specified tool – piped]?

(Piping ends)

Modified tool name

You indicated that you use modified or a combination of established outcome measurement tools.

23) What is your modified tool called?*

Modified tools

24) Have any established / formal tools been used to develop the [specified tool]?

[ ] BAS [?]
[ ] BASIS-32 [?]
[ ] CANSAS [?]
[ ] CarerQol [?]
[ ] CGAS [?]
[ ] DASS [?]
[ ] GAF [?]
[ ] HoNOS [?]
[ ] HoNOS65+ [?]
[ ] HoNOSCA [?]
[ ] Homeless Star
[ ] IMR Scales [?]
[ ] IEQ [?]
[ ] Inclusion Web
[ ] K10 [?]
[ ] LSP-16 [?]
[ ] MHI [?]
[ ] Mental Health Recovery Measure
[ ] Personal Wellbeing Index
[ ] RUG-ADL [?]
[ ] RAS [?]
[ ] Recovery Interview
[ ] Recovery STAR
[ ] RPI [?]
[ ] SCOPE [?]
[ ] SDQ [?]
[ ] SF12 [?]
[ ] SF24 [?]
[ ] SIQ [?]
[ ] STORI [?]
[ ] WHO-QoL [?]
[ ] Work and Social Adjustment Scale
[ ] Other
[ ] No established tools were used

25) Please briefly explain how these tools have been modified or combined.

Use of modified tool

You indicated that you use [specified tool].

26) Who completes the tool?

(Select all that apply)
[ ] Client
[ ] Family member or carer
[ ] Worker (from your service)
[ ] Other

27) What percentage of clients do you use this tool with?*
(including carers who are the primary clients of family and carer services)

____

28) What percentage of family or carers do you use this tool with?*
(excluding carers who are the primary clients of family and carer services)

____

29) Is use of [specified tool] required as a condition of funding?*

( ) Yes
( ) No

30) At which points of service delivery do you use this tool with the client?

[ ] Entry to service
[ ] Exit from service
[ ] Regular intervals
[ ] Under specific conditions

Generally, how regular is the use of this tool?

( ) At every contact
( ) At least once a month
( ) At least once every 3 months
( ) At least once every 6 months
( ) Less than once in 6 months

Which specific conditions would trigger use of this tool?

____

31) Of the services you provide, which of these involve use of [specified tool]?

Custom tool

You indicated that you use custom-made questionnaires / tools to measure outcomes.

32) What are the custom-made tools called?*

____

33) Please briefly explain the outcome information that you collect with these tools.

____

Use of custom tool

You indicated that you use [specified tool].

34) Who completes the tool?

(Select all that apply)
[ ] Client
[ ] Family member or carer
[ ] Worker (from your service)
[ ] Other

35) What percentage of clients do you use this tool with?*

(including carers who are the primary clients of family and carer services)

____

36) What percentage of family or carers do you use this tool with?*

(excluding carers who are the primary clients of family and carer services)

____

37) Is use of [specified tool] required as a condition of funding?*

( ) Yes
( ) No

38) At which points of service delivery do you use this tool with the client?

[ ] Entry to service
[ ] Exit from service
[ ] Regular intervals
[ ] Under specific conditions

Generally, how regular is the use of this tool?

( ) At every contact
( ) At least once a month
( ) At least once every 3 months
( ) At least once every 6 months
( ) Less than once in 6 months

Which specific conditions would trigger use of this tool?

____

39) Of the services you provide, which of these involve use of [specified tool]?

____

Qualitative tool

You indicated that you use qualitative (open-text) questionnaires or tools to measure outcomes.

40) What are these tools called?*

____

41) Please briefly explain the outcome information that you collect with these tools.

____

42) Are the responses thematically analysed or otherwise coded into quantitative data?*

( ) No
( ) Yes - by a staff member
( ) Yes - by a computer program

What is the name of this program?

____

Use of qualitative tool

You indicated that you use [specified tool].

43) Who completes the tool?

(Select all that apply)
[ ] Worker
[ ] Client
[ ] Family member or carer

44) What percentage of clients do you use this tool with?*

(excluding family and carers)

____

45) What percentage of family or carers do you use this tool with?*

(including those who are clients of family and carer services)

____

46) Is use of [specified tool] required as a condition of funding?*

( ) Yes
( ) No

47) At which points of service delivery do you use this tool with the client?

[ ] Entry to service
[ ] Exit from service
[ ] Regular intervals
[ ] Under specific conditions

Generally, how regular is the use of this tool?

( ) At every contact
( ) At least once a month
( ) At least once every 3 months
( ) At least once every 6 months
( ) Less than once in 6 months

Which specific conditions would trigger use of this tool?

____

48) Of the services you provide, which of these involve use of [specified tool]?

____

No tool use

49) What are the main reasons that you do not measure outcomes?*

(select all that apply)
[ ] Have not found any appropriate measures - too clinical
[ ] Have not found any appropriate measures - too academic
[ ] Overly time consuming/costly to collect
[ ] Overly time consuming/costly to train
[ ] Gets in the way of our work
[ ] Information system can't collect
[ ] Not a contractual requirement
[ ] Not interested
[ ] Other

50) How do you measure the effect that your services are having on your clients?

____

51) Are there any ways that outcome measurement tools could be designed or modified so that you would be more likely to use them?

____

Information Infrastructure

These last questions are designed to ascertain what information infrastructure is available for CMOs to collect outcome data.

52) How do you store and report on your client data?

(select all that apply)
[ ] Manual (paper-based)
[ ] Computerised - MS Office (e.g. Excel, Word, Access)
[ ] Computerised - Dedicated system (e.g. custom or off-the-shelf data system, client management system, CRM, etc)
[ ] Other

What is the name/brand of your dedicated system?

____

Was this dedicated system provided by a government funder?

( ) Yes - Commonwealth funder
( ) Yes - State Funder
( ) Yes - Local Health District
( ) Yes - Local Government
( ) No

53) Is your IT infrastructure, in its current state, capable of routinely tracking and generating detailed reports on individual clients?

( ) Yes
( ) No - but it could be modified / expanded
( ) No - we don't have the resources to modify or expand our system
( ) No - it cannot be changed
( ) Not applicable

Further details of unestablished measures

We would appreciate the opportunity to learn about modified, custom-made and qualitative outcome measurement tools. If you are able to provide documentation or examples please upload copies of your organisation's tools using the buttons below. You can save your progress while you find and/or scan your tools by clicking *"Save and continue survey later"* at the bottom of the page.

Alternately, leave contact details on the following page and a project officer will provide other options (e.g. fax, mail).

Follow-up details

55) Please provide the contact details of someone in your organisation who our project officer can talk to, should we wish to follow-up about your outcome measurement activities.

This person will be provided with a summary of the project findings.

Full Name: ____
Job Position: ____
Email Address: ____
Phone Number: ____

Do you have examples of your outcome tools that you would like the CMHA project officer to contact you about?

( ) Yes
( ) No

Other comments

56) Do you have any other comments about outcome measurement in the community managed sector?

Appendix B – AIHW NGO Establishments NMDS Service Category Descriptions

1. Counselling—face-to-face

Counselling services provide a structured process that is concerned with addressing and resolving specific problems, making decisions, working through feelings and inner conflicts, or improving relationships with others.

Counselling services operate through a range of mediums including face-to-face, telephone and online. This service type is intended only for services providing face-to-face counselling.

The counselling process will depend on the individual counsellor, the individual client and the specific issue.

2. Counselling, support, information & referral—telephone

Mental health support, information and referral services are those that provide support for people experiencing mental illness and which offer reliable referrals, information and self-help resources to empower people to take steps towards maintaining mental health and emotional wellbeing (Lifeline 2012).

3. Counselling, support, information & referral—online

Mental health support, information and referral services are those that provide support for people experiencing mental illness.

4. Self-help—online

Self-help—online services are interactive online programs which take people with a lived experience of mental illness through exercises to help them develop skills to handle life's challenges more effectively. Unlike *Counselling, support, information and referral–online*, services which fall under *Self-help–online* do not involve interaction with another person, only interaction with the online program’s content.

5. Group support activities

Group support activities are services that aim to improve the quality of life and psychosocial functioning of mental health consumers, through the provision of group-based social, recreational or prevocational activities. In contrast to services in the *Mutual support and self-help* service type, *Group support activities* are led by a member of the NGO.

6. Mutual support and self-help

Mutual support and self-help services provide information and peer support to people with a lived experience of mental illness and/or their carers. People meet to discuss shared experiences, coping strategies and to provide information and referrals. Self-help groups are usually formed by peers who have come together for mutual support and to accomplish a specific purpose.

7. Staffed residential services

Staffed residential services are those that provide overnight accommodation in a domestic-style environment, which is staffed for a minimum of 6 hours a day and at least 50 hours per week.

8. Personalised support—linked to housing

Personalised support services—linked to housing are services that provide personalised support that is coordinated with provision of social housing or privately negotiated housing at the point of entry into the program (but not tied to such indefinitely).

Personalised support services are flexible services tailored to a mental health consumer's individual and changing needs. They include a range of one-on-one activities provided by a support worker directly to mental health consumers in their homes or local communities.

9. Personalised support—other

Personalised support services—other are services that provide personalised support that is independent of housing arrangements (e.g. provision of social housing or privately negotiated housing) at the point of entry into the program.

10. Family & carer support

Family and carer support services are services that provide families and carers of people living with a mental illness support, information, education and skill development opportunities to fulfil their caring role, while maintaining their own health and wellbeing.

11. Individual advocacy

Individual advocacy services are those that seek to uphold the rights and interests of people with a mental illness, on a one-to-one basis, by addressing instances of discrimination, abuse and neglect.

Individual advocates work with people with mental illness on either a short-term or issue-specific basis.

12. Care coordination

Care coordination services provide a single point of contact (via a Care Facilitator) for people (and their families/carers) with lived experience of mental illness and complex care needs. Care Facilitators will be responsible for ensuring all of the patients' care needs, clinical and non-clinical, and as determined by a nationally consistent assessment tool, are being met.

13. Service coordination

Organisations providing service coordination establish a service platform which brings together appropriate mental health-related services, both existing and new, which have the objective of improving mental health. These services typically provide a range of support services aimed at improving the well-being and economic participation of people with a lived experience of mental illness.

14. Education, employment & training

Education, employment and training services are those which provide or support people with lived experience of mental illness, in gaining education, employment and/or training.

15. Sector development & representation

Mental health sector development and representation services engage with a wide variety of issues regarding the sustainability and development of the mental health sector. This includes information dissemination, advocacy, policy analysis, program development, and sector capacity building.

16. Mental health promotion

Mental health promotion services are those that work to maximize mental health and wellbeing among populations and individuals. It involves raising awareness about mental health/mental illness. This is achieved by providing information and/or education to the community to increase its capacity to support people with lived experience of mental illness.

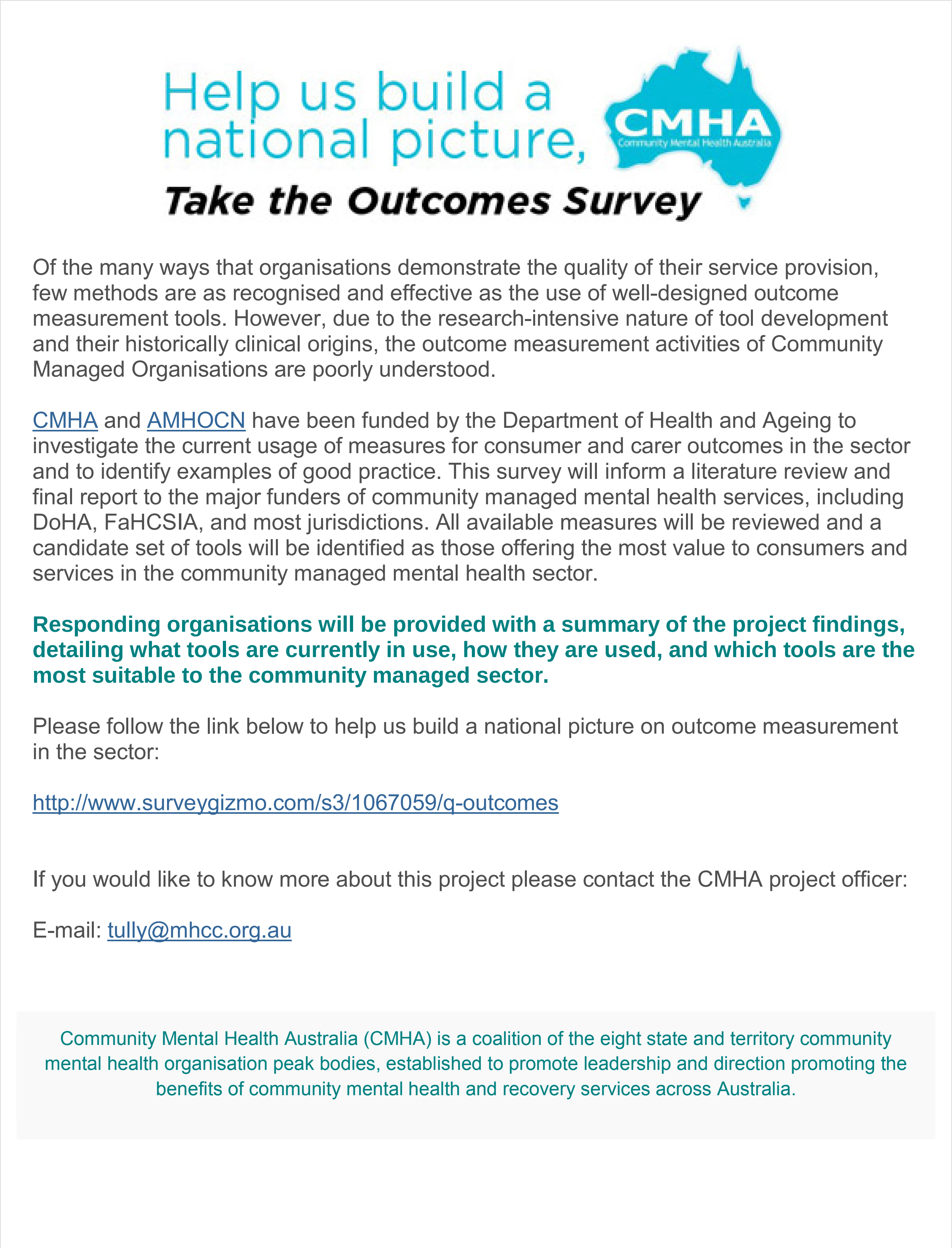
Discrimination reduction programs or campaigns aim to challenge prejudices towards people with a mental illness. Discrimination reduction involves the implementation of long-term strategies to educate the general public about mental health and encouraging a socially inclusive attitude toward people with a mental illness.

17. Mental illness prevention

Mental illness prevention services deliver activities that are designed to prevent the development of a mental disorder. Prevention endeavours to avoid the development of mental illness, generally through population based health activities, reduction of known risk factors and promoting factors that support resilience and coping.

Some service types do not lend themselves to the measurement of outcomes for individual consumers. So some of these service types will be excluded from the identification of potential outcome measures because they provide services to populations

Appendix C – Example Survey Campaign Promotion Material



Help us build a national picture,

Take the Outcomes Survey

Of the many ways that organisations demonstrate the quality of their service provision, few methods are as recognised and effective as the use of well-designed outcome measurement tools. However, due to the research-intensive nature of tool development and their historically clinical origins, the outcome measurement activities of Community Managed Organisations are poorly understood.

CMHA and AMHOCN have been funded by the Department of Health and Ageing to investigate the current usage of measures for consumer and carer outcomes in the sector and to identify examples of good practice. This survey will inform a literature review and final report to the major funders of community managed mental health services, including DoHA, FaHCSIA, and most jurisdictions. All available measures will be reviewed and a candidate set of tools will be identified as those offering the most value to consumers and services in the community managed mental health sector.

Responding organisations will be provided with a summary of the project findings, detailing what tools are currently in use, how they are used, and which tools are the most suitable to the community managed sector.

Please follow the link below to help us build a national picture on outcome measurement in the sector:

http://www.surveygizmo.com/s3/1067059/q-outcomes

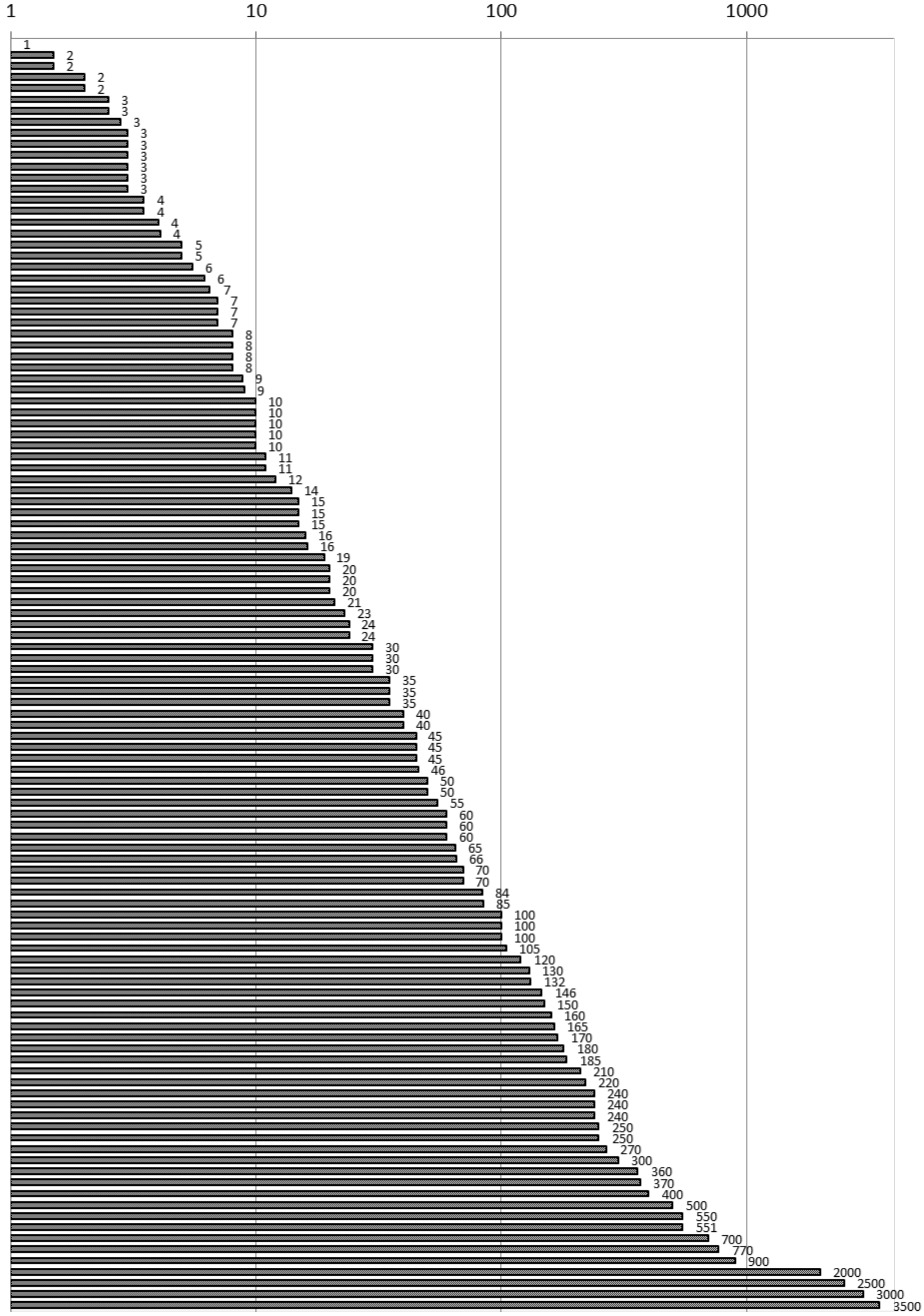
If you would like to know more about this project please contact the CMHA project officer:

E-mail: tully@mhcc.org.au

Community Mental Health Australia (CMHA) is a coalition of the eight state and territory community mental health organisation peak bodies, established to promote leadership and direction promoting the benefits of community mental health and recovery services across Australia.

Appendix D – Distribution of Organisation Sizes

Approximate number of full time equivilent (FTE) staff for each responding organisation – logorithmic scale



Appendix E – Use of Established Tools by Organisation Size

Established Tool	Small (No. Orgs)	Medium (No. Orgs)	Large (No. Orgs)	Very Large (No. Orgs)	*Unspecified size*	All Orgs
K10	2	6	6	5	*3*	22
BASIS-32	2	5	8		*1*	16
CANSAS	1	6	5	3		15
DASS	2	5	4	2		13
HoNOS	1	7	2	2	*1*	13
LSP-16	3	3	5	2		13
Recovery STAR	2	3	4	2	*1*	12
WHO-QoL		4	4	2	*2*	12
RAS		3	2	1		6
GAF	1	1	2		*1*	5
Homeless STAR		1	3	1		5
Personal Wellbeing Index		2	2	1		5
Recovery Interview	1	2			*1*	4
Mental Health Recovery Measure		3				3
APQ6			1	1		2
ORS & SRS		1		1		2
PSI	1			1		2
Recovery Enhancing Environment Measure			1	1		2
SF12		1	1			2
SIQ			2			2
ADIS	1					1
AUDIT				1		1
BAS			1			1
BASH-B		1				1
BDI-II			1			1
Brief COPE			1			1
CAMCOG		1				1
CBCL	1					1
Collaborative Recovery Model Tools		1				1
Consumer Evaluation of Outcomes			1			1
CRAM		1				1
CSSRS			1			1
DECAC	1					1
DES II		1				1
DLC	1					1
DUDIT				1		1
EDI3	1					1
Edinburgh Postnatal Depression Scale		1				1

Established Tool	Small (No. Orgs)	Medium (No. Orgs)	Large (No. Orgs)	Very Large (No. Orgs)	*Unspecified size*	All Orgs
Empowerment Scale	1					1
EPDS		1				1
GAS		1				1
GHS-V		1				1
HADS			1			1
Home Inventory	1					1
HoNOSCA			1			1
ICG-R			1			1
IRIS				1		1
M3				1		1
MANSA		1				1
MDI		1				1
MHQ14		1				1
MHSIP		1				1
MMSE		1				1
NEO-P3				1		1
OL			1			1
PCL-C		1				1
PEDS	1					1
Personal Outcomes Measures		1				1
REE				1		1
RUDAS		1				1
SCL90	1					1
SISR			1			1
SOFAS					1	1
Sphere-12		1				1
SRM 1 & 2		1				1
STAI		1				1
STORI			1			1
Unknown (provided by funder)		1				1
WEMWEBS			1			1
Work and Social Adjustment Scale			1			1
Working Alliance Inventory		1				1

Appendix F – Usage Characteristics of the Most Common Established Tools

K10

The K10 was used by 22 organisations.

Tool Completion

Tool completed by	No. Orgs	% of 22
Client	17	77%
Family member or carer	1	5%
CMO Worker	11	50%
LHD Worker	1	5%

Organisations reported using K10 with 49% of clients (15% variance) and 3% of family or carers (1% variance).

Funding conditions

7 organisations (32% of K10 users) were required to use K10 as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	18
Exit from service	7
Regular intervals	11
Under specific conditions	4

Regularity of Tool Use	No. Orgs
At every contact	1
At least once every 3 months	4
At least once every 6 months	3
Less than once in 6 months	3

Specific conditions	No. Orgs
Significant change in condition or circumstances	2
Online invitation	1
Research	1

BASIS-32

The BASIS-32 was used by 16 organisations.

Designated users

Tool completed by	No. Orgs	% of 16
Client	15	94%
Family member or carer	3	19%
CMO Worker	10	63%

Organisations reported using BASIS-32 with 71% of clients (11% variance) and 12% of family or carers (7% variance).

Funding conditions

5 organisations (31% of BASIS-32 users) were required to use BASIS-32 as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	11
Exit from service	6
Regular intervals	12

Usage amount	No. Orgs
At least once every 3 months	2
At least once every 6 months	8
Less than once in 6 months	1

CANSAS

The CANSAS was used by 15 organisations.

Designated users

Tool completed by	No. Orgs	% of 15
Client	14	93%
Family member or carer	1	7%
CMO Worker	10	67%

Organisations reported using CANSAS with 72% of clients (11% variance) and 3% of family or carers (0% variance).

Funding conditions

4 organisations (27% of CANSAS users) were required to use CANSAS as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	13
Exit from service	6
Regular intervals	13
Under specific conditions	2

Regularity of Tool Use	No. Orgs
At least once every 3 months	5
At least once every 6 months	7
Less than once in 6 months	1

Specific conditions	No. Orgs
Before or after carer education program	1
Significant change in condition or situation	1

DASS

The DASS was used by 13 organisations.

Tool Completion

Tool completed by	No. Orgs	% of 13
Client	11	85%
Family member or carer	2	15%
CMO Worker	5	38%

Organisations reported using DASS with 30% of clients (11% variance) and 0% of family or carers.

Funding conditions

2 organisations (15% of DASS users) were required to use DASS as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	9
Exit from service	8
Regular intervals	7
Under specific conditions	5

Regularity of Tool Use	No. Orgs
At least once a month	1
At least once every 3 months	3
At least once every 6 months	3

Specific conditions	No. Orgs
Pre, post and12-18 month follow up	1
Online invitation	1
Research	1
When attending carer retreats	1

HoNOS

The HoNOS was used by 13 organisations.

Tool Completion

Tool completed by	No. Orgs	% of 13
Client	2	15%
Family member or carer	0	0%
CMO Worker	11	85%
LHD Worker	2	15%

Organisations reported using HoNOS with 71% of clients (18% variance) and 2% of family or carers (0% variance).

Funding conditions

6 organisations (46% of HoNOS users) were required to use HoNOS as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	11
Exit from service	9
Regular intervals	9

Regularity of Tool Use	No. Orgs
At least once every 3 months	5
At least once every 6 months	3
Less than once in 6 months	1

LSP-16

The LSP-16 was used by 13 organisations.

Tool completion

Tool completed by	No. Orgs	% of 13
Client	8	62%
Family member or carer	1	8%
CMO Worker	10	77%

Organisations reported using LSP-16 with 72% of clients (14% variance) and 10% of family or carers (8% variance).

Funding conditions

7 organisations (54% of LSP-16 users) were required to use LSP-16 as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	13
Exit from service	8
Regular intervals	8

Regularity of Tool Use	No. Orgs
At least once every 3 months	4
At least once every 6 months	4

WHO-QoL

The WHO-QoL was used by 12 organisations.

Tool Completion

Tool completed by	No. Orgs	% of 12
Client	11	92%
Family member or carer	1	8%
CMO Worker	6	50%

Organisations reported using WHO-QoL with 59% of clients (11% variance) and 3% of family or carers (0% variance).

Funding conditions

5 organisations (42% of WHO-QoL users) were required to use WHO-QoL as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	8
Exit from service	4
Regular intervals	8
Under specific conditions	4

Regularity of Tool Use	No. Orgs
At least once every 3 months	2
At least once every 6 months	6

Specific conditions	No. Orgs
Before and after program component	1
Changes in service delivery or mental health need	1
Online invitation	1
Research	1

Recovery STAR

The Recovery STAR was used by 12 organisations.

Tool Completion

Tool completed by	No. Orgs	% of 12
Client	11	92%
Family member or carer	5	42%
CMO Worker	9	75%

Organisations reported using Recovery STAR with 49% of clients (19% variance) and 18% of family or carers (10% variance).

Funding conditions

Only 1 organisation (8% of Recovery STAR users) was required to use Recovery STAR as a condition of funding.

Usage timeframes

Point of tool use	No. Orgs
Entry to service	6
Exit from service	5
Regular intervals	9
Under specific conditions	1

Regularity of Tool Use	No. Orgs
At least once every 3 months	3
At least once every 6 months	5

Specific condition	No. Orgs
Client interest	1

Appendix G – Established Tool Use in Specific Service Categories

These cross-tabs report the number of CMOs reporting use of an established tool specifically for a category of service activity. It is important to note that there are different numbers of organisations providing each category of service. The first cross-tab lists CMO numbers, the second cross-tab lists CMO numbers as a proportion of all CMOs providing each category of service activity.

Number of organisations using established tools for each service category

Service Category	APQ6	BAS	BASIS-32	CANSAS	DASS	GAF	Homeless Star	HoNOS
Care coordination	1	0	4	3	0	0	2	4
Counselling — face-to-face	0	1	0	1	9	4	1	5
Counselling, support, information & referral — online	0	1	0	0	2	2	0	0
Counselling, support, information & referral — telephone	0	0	0	0	1	1	0	1
Education, employment & training	1	0	2	2	1	0	0	0
Family & carer support	0	0	2	2	3	0	0	1
Group support activities (staff-led)	0	0	6	6	4	0	0	2
Individual peer advocacy	0	0	0	0	0	0	0	0
Mental health promotion	1	0	1	1	2	0	0	1
Mental illness prevention	1	0	2	1	2	0	0	0
Mutual support & self-help (peer-led)	0	0	0	0	0	0	0	0
Personalised support — linked to housing	2	0	9	10	0	2	3	4
Personalised support — not linked to housing	2	0	10	13	2	1	0	6
Sector development & representation	0	0	0	0	0	0	0	0
Self-help — online	0	0	0	0	2	0	0	0
Service integration infrastructure	0	0	0	0	0	0	0	0
Staffed residential services	1	0	6	11	1	0	1	5

(continued) **Number of organisations using established tools for each service category**

Service Category	HoNOSCA	K10	LSP-16	Mental Health Recovery Measure	ORS & SRS	Personal Wellbeing Index	PSI	RAS
Care coordination	0	5	5	1	0	2	1	3
Counselling — face-to-face	0	10	2	1	1	4	1	1
Counselling, support, information & referral — online	0	4	0	0	1	1	0	0
Counselling, support, information & referral — telephone	0	3	1	1	2	1	0	0
Education, employment & training	0	1	0	0	1	2	1	1
Family & carer support	0	2	2	1	1	1	0	0
Group support activities (staff-led)	0	4	3	0	1	1	1	3
Individual peer advocacy	0	0	1	0	0	1	0	1
Mental health promotion	0	2	1	1	0	4	1	1
Mental illness prevention	0	2	0	1	0	2	1	1
Mutual support & self-help (peer-led)	0	0	0	0	0	0	0	0
Personalised support — linked to housing	0	5	6	0	1	1	1	3
Personalised support — not linked to housing	0	8	6	0	0	2	1	3
Sector development & representation	0	0	0	0	0	0	0	0
Self-help — online	0	2	0	0	1	1	0	0
Service integration infrastructure	0	0	0	0	0	0	0	0
Staffed residential services	1	4	8	1	0	1	1	4

Service Category	Recovery Enhancing Environment Measure	Recovery Interview	Recovery STAR	SF12	SIQ	STORI	WHO-QoL	Work and Social Adjustment Scale
Care coordination	1	2	4	0	0	0	2	0
Counselling — face-to-face	0	1	2	1	1	0	3	1
Counselling, support, information & referral — online	0	0	0	2	0	0	2	0
Counselling, support, information & referral — telephone	0	1	1	0	0	0	1	0
Education, employment & training	0	1	2	1	0	0	0	1
Family & carer support	0	2	3	0	0	0	2	0
Group support activities (staff-led)	2	0	6	1	0	0	5	0
Individual peer advocacy	1	1	1	0	0	0	1	0
Mental health promotion	0	0	1	1	0	0	2	0
Mental illness prevention	0	0	3	2	0	0	3	1
Mutual support & self-help (peer-led)	0	0	0	0	0	0	0	0
Personalised support — linked to housing	2	2	6	0	0	0	7	0
Personalised support — not linked to housing	2	2	6	0	0	0	7	0
Sector development & representation	0	0	0	0	0	0	0	0
Self-help — online	0	0	0	2	0	0	2	1
Service integration infrastructure	0	0	0	0	0	0	0	0
Staffed residential services	2	1	2	0	0	1	1	0

Number of organisations using established tools for each service category as a proportion of all organisations providing that service category.

Service Category	APQ6	BAS	BASIS-32	CANSAS	DASS	GAF	Homeless Star	HoNOS
Care coordination	1.8%	0.0%	7.0%	5.3%	0.0%	0.0%	3.5%	7.0%
Counselling — face-to-face	0.0%	2.1%	0.0%	2.1%	18.8%	8.3%	2.1%	10.4%
Counselling, support, information & referral — online	0.0%	4.3%	0.0%	0.0%	8.7%	8.7%	0.0%	0.0%
Counselling, support, information & referral — telephone	0.0%	0.0%	0.0%	0.0%	2.1%	2.1%	0.0%	2.1%
Education, employment & training	2.0%	0.0%	4.1%	4.1%	2.0%	0.0%	0.0%	0.0%
Family & carer support	0.0%	0.0%	3.1%	3.1%	4.6%	0.0%	0.0%	1.5%
Group support activities (staff-led)	0.0%	0.0%	6.9%	6.9%	4.6%	0.0%	0.0%	2.3%
Individual peer advocacy	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Mental health promotion	1.6%	0.0%	1.6%	1.6%	3.2%	0.0%	0.0%	1.6%
Mental illness prevention	2.6%	0.0%	5.1%	2.6%	5.1%	0.0%	0.0%	0.0%
Mutual support & self-help (peer-led)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Personalised support — linked to housing	3.4%	0.0%	15.5%	17.2%	0.0%	3.4%	5.2%	6.9%
Personalised support — not linked to housing	2.9%	0.0%	14.7%	19.1%	2.9%	1.5%	0.0%	8.8%
Sector development & representation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Self-help — online	0.0%	0.0%	0.0%	0.0%	16.7%	0.0%	0.0%	0.0%
Service integration infrastructure	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Staffed residential services	2.9%	0.0%	17.1%	31.4%	2.9%	0.0%	2.9%	14.3%

Service Category	HoNOSCA	K10	LSP-16	Mental Health Recovery Measure	ORS & SRS	Personal Wellbeing Index	PSI	RAS
Care coordination	0.0%	8.8%	8.8%	1.8%	0.0%	3.5%	1.8%	5.3%
Counselling — face-to-face	0.0%	20.8%	4.2%	2.1%	2.1%	8.3%	2.1%	2.1%
Counselling, support, information & referral — online	0.0%	17.4%	0.0%	0.0%	4.3%	4.3%	0.0%	0.0%
Counselling, support, information & referral — telephone	0.0%	6.3%	2.1%	2.1%	4.2%	2.1%	0.0%	0.0%
Education, employment & training	0.0%	2.0%	0.0%	0.0%	2.0%	4.1%	2.0%	2.0%
Family & carer support	0.0%	3.1%	3.1%	1.5%	1.5%	1.5%	0.0%	0.0%
Group support activities (staff-led)	0.0%	4.6%	3.4%	0.0%	1.1%	1.1%	1.1%	3.4%
Individual peer advocacy	0.0%	0.0%	2.6%	0.0%	0.0%	2.6%	0.0%	2.6%
Mental health promotion	0.0%	3.2%	1.6%	1.6%	0.0%	6.3%	1.6%	1.6%
Mental illness prevention	0.0%	5.1%	0.0%	2.6%	0.0%	5.1%	2.6%	2.6%
Mutual support & self-help (peer-led)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Personalised support — linked to housing	0.0%	8.6%	10.3%	0.0%	1.7%	1.7%	1.7%	5.2%
Personalised support — not linked to housing	0.0%	11.8%	8.8%	0.0%	0.0%	2.9%	1.5%	4.4%
Sector development & representation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Self-help — online	0.0%	16.7%	0.0%	0.0%	8.3%	8.3%	0.0%	0.0%
Service integration infrastructure	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Staffed residential services	2.9%	11.4%	22.9%	2.9%	0.0%	2.9%	2.9%	11.4%

(continued) **Number of organisations using established tools for each service category as a proportion of all organisations providing that service category.**

AIHW Category	Recovery Enhancing Environment Measure	Recovery Interview	Recovery STAR	SF12	SIQ	STORI	WHO-QoL	Work and Social Adjustment Scale
Care coordination	1.8%	3.5%	7.0%	0.0%	0.0%	0.0%	3.5%	0.0%
Counselling — face-to-face	0.0%	2.1%	4.2%	2.1%	2.1%	0.0%	6.3%	2.1%
Counselling, support, information & referral — online	0.0%	0.0%	0.0%	8.7%	0.0%	0.0%	8.7%	0.0%
Counselling, support, information & referral — telephone	0.0%	2.1%	2.1%	0.0%	0.0%	0.0%	2.1%	0.0%
Education, employment & training	0.0%	2.0%	4.1%	2.0%	0.0%	0.0%	0.0%	2.0%
Family & carer support	0.0%	3.1%	4.6%	0.0%	0.0%	0.0%	3.1%	0.0%
Group support activities (staff-led)	2.3%	0.0%	6.9%	1.1%	0.0%	0.0%	5.7%	0.0%
Individual peer advocacy	2.6%	2.6%	2.6%	0.0%	0.0%	0.0%	2.6%	0.0%
Mental health promotion	0.0%	0.0%	1.6%	1.6%	0.0%	0.0%	3.2%	0.0%
Mental illness prevention	0.0%	0.0%	7.7%	5.1%	0.0%	0.0%	7.7%	2.6%
Mutual support & self-help (peer-led)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Personalised support — linked to housing	3.4%	3.4%	10.3%	0.0%	0.0%	0.0%	12.1%	0.0%
Personalised support — not linked to housing	2.9%	2.9%	8.8%	0.0%	0.0%	0.0%	10.3%	0.0%
Sector development & representation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Self-help — online	0.0%	0.0%	0.0%	16.7%	0.0%	0.0%	16.7%	8.3%
Service integration infrastructure	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Staffed residential services	5.7%	2.9%	5.7%	0.0%	0.0%	2.9%	2.9%	0.0%

Appendix H – CMO General Comments on Outcome Measurement

The following letter was received from the Victorian Mental Health Carers Network on 21 January 2013.

Thank you for the opportunity to submit a response to the Outcome Measurement Project. Please find our response below.

Consumer Outcome Measures

It is important that any tool measuring outcomes for consumers also includes the carer's perceptions of the consumer's outcomes, as the carer perception may differ from that of the consumers.

Carer Outcome Measures

Over the last fifteen years, there has been an increased awareness of family and carer issues within mental health services. In particular, services are acknowledging that carers have their own support needs, which can include information, education, counselling, support groups and respite support. As a result, some organisations are now offering a range of services that meet these needs. At present, there are no requirements that outcomes are measured from these activities.

While there is currently a national project underway to develop a tool to assess the experience of carers, we believe that there also needs to be carer outcome measurements. Otherwise, services won't know whether or not the services that they are delivering to carers are effective.

In 2008, the Australian Mental Health Outcomes and Classification Network (AMHOCN) conducted a scoping study on carer outcome measurements. The study outlined several tools of evaluation and recommended two possible tools, with some modifications, for the use of carer outcome measures. These tools were Carers' Quality of Life-7D+ Visual Analogue Scale (CarerQol-7D+VAS) and Burden Assessment Scale (BAS).

While the study concluded that the time was not right to trial and adapt the measures, there have been significant changes to both state and federal mental health standards, policies and service systems. In particular we note that:

- The *National Carer Strategy* priority four, states: 'that carers are supported with appropriate, timely and accessible services'.
- The *National Mental Health Standard* 7.13, states: 'The Mental Health Service provides information about and facilitates access to services that maximise the wellbeing of carers'.

Furthermore, there has also been a significant increase in commonwealth funding of carer support services. As a result of the increased funding, the need for outcome measurement tools has increased.

It is essential that government and services know that policies and standards are implemented, and that money is well spent. Developing and implementing carer outcome measurements will make this possible.

Conclusion

We are happy to be of assistance in the adaptation or development of outcome measures for carers.

Appendix I – CMO General Comments on Outcome Measurement

The following is a list of general comments provided by community managed organisations when asked if they had any general comments about outcome measurement.

Minor corrections were made to spelling and grammar, and identifying information was removed. Six comments clarifying responses to earlier questions in the survey were also removed and addressed during the cleaning of the survey data.

CMO Comment (Q56)	Thematic Summary
Diagnostic tools are not very useful in a peer environment.	*Different CMO service types/sizes/target populations need different types of tools*
For intellectual disability outcome measurement is usually referring to outcomes of hours and numbers of clients for reporting to funding bodies. I have only provided details of qualitative and quantitative research but we also collect info on service delivery statistics.	*Concept of outcome is different in intellectual disability sector*
We are about to purchase an 'expensive' client database system which will allow us to use standardised tool such as the K10 HoNOS LSP etc and keep and track it electronically however have yet to decide on which are the best tool(s) to measure outcomes effectively and without adding burden to the clients	*In development; Strong data system/infrastructure is enabling better outcome data collection*
It is very difficult to get the focus to outcomes, while many of our staff are still struggling with Outputs, as in the past (prior to my time) they had limited accountability. Our database allows us to measure some aspects such as timeliness of writing case notes, dependency on the services, which clients have goals, whether these are overdue. We are working on creating an objective tool to audit file contents to identify if recovery practices are taking place	*Need for contextually appropriate tools for CMOs; In development; Still implementing output data collection, implementing outcome measurement would be difficult*
It would be great to see a universal tool that could be used for measuring outcomes.	*Need for contextually appropriate tools for CMOs; Need for standardisation*
I think that developing standard but flexible tools for assessing outcomes and impacts of the same service eg PHaMs or Day to Day Living, delivered by different providers in similar regions or cities is feasible and desirable. It is not feasible and would be counter-productive to use a single or a small number of tools to compare or assess different programs delivered in different regions by different providers to different subject groups of varying ethnicity, culture (around openness to the nature and possibility of recovery from mental illness), age, etc	*Need for contextually appropriate tools for CMOs; Need for standardisation; Different CMO service types/sizes/target populations need different types of tools*
Measuring outcomes in a small rural organisation is vastly different to a bigger city organisation These outcomes measures for a bigger organisation canot be simply imposed or transferred to a small community organisation	*Different CMO service types/sizes/target populations need different types of tools*

CMO Comment (Q56)	Thematic Summary
As we have just moved to outputs instead of inputs it is quite a leap to be talking outcomes - although i believe that we have been working with outcomes through the goals and strategies in the individual plans. We just have no way of collecting or reporting on actual outcomes at the moment. The lack of funding for software, hardware and training make it very difficult (LD gov only contribution fund our direct service provision). If we were funded properly we would have a much better chance of doing this. Thanking you for your time.	*In development; Lack of funding/infrastructure/training impedes capacity to measure outcomes; Individual plans are a form of outcome measurement*
It may be easier to measure outputs but it is much more valuable and relevant to focus on outcomes	*Appreciation of the value of outcome measurement*
Due to the nature of the community sector being severely underfunded and undervalued we find it extremely difficult to provide a Quality outcome measurement.	*Lack of funding/infrastructure/training impedes capacity to measure outcomes*
Great to see a little interest in client outcomes	*Positive feedback on the project*
I have completed this survey as requested but Advance Employment Inc. has just recently been informed that it was not successful in the recent DES Tender process	*Org has lost funding / may be closing*
Our organisation is likely to close very soon due to lack of funding	*Org has lost funding / may be closing*
We have only just commenced using Personal Outcomes Measures. Our first assessment was conducted in September. The data in this survey relates to that process. we intend to use this tool much more broadly to assess service Quality within our organisation. For more information on POMs see http://www.thecouncil.org/Personal_Outcome_Measures.aspx	*In development*
It is important that outcomes measure real change in the client and the client's circumstances, not just sessions attended, or visit made.	*Appreciation of the value of outcome measurement*
Our organisation is committed to outcome measurement, and has invested a considerable amount in the creation and (recently) the purchase of databases and systems that allow us to meaningfully analyse and use data from outcome measurement tools. We would like to expand this to allow people who use services to access their online records and tools, in keeping with recovery-centred practice.	*In development; Appreciation of the value of outcome measurement; Strong data system/infrastructure is enabling better outcome data collection*
It's a crucial component of ensuring Quality service	*Appreciation of the value of outcome measurement*
Evaluation needs to include personal recovery stories by the participants in the program. It is the changes and overcome challenges in their personal journey through the environment of support by the program that tells the real story.	*Importance of incorporating Qualitative information in service evaluations*
Need a dedicated database tool for mental health support work that is used across the sector.	*Lack of funding/infrastructure/training impedes capacity to measure outcomes*

CMO Comment (Q56)	Thematic Summary
Would certainly like information on available outcome measurement tools for mutual support self-help.	*Desire for more information and support to be made available on outcome measurement for CMOs*
I support the introduction of outcome measures however, measures need to be valid, reliable, easy to complete, not burdensome, no more than 3 separate measures and supported by IT that generates reports.	*Need for contextually appropriate tools for CMOs*
Mind Australia is currently trailing the MHEco project in collaboration with the Victorian Mental Health Carers Network. This will involve 300 carers and families. MHEco is focused on families and carers experience of the service.	*In development*
We are currently exploring and are likely to implement the Recovery Star into our practice. This might be something worth exploring for the sector because training I have done so far, indicate that there is a valuable reporting component to it which may be of use to agencies and across the sector.	*Positive feedback on the Recovery STAR; In development*
Regarding the question re IT data and reports. I have 4 teams but only one has a computerised data base / record.	*Lack of funding/infrastructure/training impedes capacity to measure outcomes*
No an area under development	*Not interested in outcome measurement*
I am sure we would have useful data but are unable to capture it.	*Lack of funding/infrastructure/training impedes capacity to measure outcomes*
Across the sector we need to do far more outcomes focused measuring. There needs to be a more comparable way of measuring so that we can compare apples to apples cross sectorally.	*Appreciation of the value of outcome measurement; Need for standardisation*
We are currently studying and plan on implementing the MSC technique	*In development*
WCS and ACT Mental Health have engaged the Centre for Mental Health Research to evaluate the program using peer interviewers and a survey tool for carers.	*In development*
Some outcomes that may seem small or difficult to measure are in fact huge steps for some clients living with a mental illness	*Outcome scores are complicated to interpret*
I hope we are able to get somewhere with this, it's important. Best wishes.	*Positive feedback on the study*
This is an area I feel passionately about and an area that I believe needs much attention.	*Appreciation of the value of outcome measurement; In development*
We are learning how to do it better. We could do with some help to do this.	*In development; Desire for more information and support to be made available on outcome measurement for CMOs*
We are a carer/families organisation and the measures developed are for consumers so don't reflect our needs	*Different CMO service types/sizes/target populations need different types of tools*

CMO Comment (Q56)	Thematic Summary
The Recovery Star has been well received by consumers and has allowed staff to work in a more holistic manner and identify areas that previously missed.	*Positive feedback on the Recovery STAR*
This is a really important project - glad to see it being conducted.	*Appreciation of the value of outcome measurement*
Each funding line requires different information to be measured making it difficult and time consuming for small agencies with limited resources to achieve. Qualitative outcome measurement tools would be more beneficial to identifying what is working for the target group within the sector and assist in service development and closing gaps.	*Importance of incorporating Qualitative information; High administrative burden collecting multiple sets of data for multiple funders*
Our tool is implemented cross our whole organisation, it is in the early stages, but all indications are that it is working well and is of great assistance to clients, their families/carers and support workers. As yet we don't have access to the web based evaluation tool, however once the remainder of programs have been trained a decision will be made as to how we evaluate, web or something we produce as an agency.	*In development*
We could do with a lot more and very interested in anyone with a Quality measurement tool. The tool we use in respite is very basic and would benefit from other ideas about how to improve, very necessary to have to receive true feedback about the impact we do or don't have to carers, clients, families	*Need for contextually appropriate tools for CMOs; Desire for more information and support to be made available on outcome measurement for CMOs*
Outcome measurement is an essential aspect of service provision. Client feedback has been essential for developing programs over 2 decades. Specific evaluations relevant to the client groups are most useful. Formal standard tools give only limited information and are often based on a medical model and pathology based. All programs/services rely on client outcomes to inform future service provision and preventative strategies.	*Appreciation of the value of outcome measurement; Need for contextually appropriate tools for CMOs*
There are too many options. Hence why we are trying to develop a tool that will potentially work across the NGO sector in the future.	*In development; Need for standardisation*
A must do topic to allow Boards and management to know that we are achieving mission (and that this quite different to what funders want).	*Appreciation of the value of outcome measurement*
Only to reiterate that currently the outcome measures available for use are western based models of assessment and there needs to be more culturally appropriate measuring tools available for use.	*Different CMO service types/sizes/target populations need different types of tools*
Looking forward to an outcome measurement tool to be recommended for implementation across mental health sector. Past Quality of Life Tool - no current measurement tool, waiting for sector outcome measurement tool.	*Need for standardisation*

CMO Comment (Q56)	Thematic Summary
I believe that some of the outcome measurement tools are not of great value to our clients and carers. Our clients and carers have indicated to us in the past that some measurement tools they find offensive and do not wish to complete for this reason. It seems that the wording with some makes carers feel as if they are not coping or do not care for their loved ones. We have tried to locate and develop tools that capture outcomes without relaying feelings of inadequacy for those completing them.	*Need for contextually appropriate tools for CMOs*
I am aware that many different tools are in use throughout the sector. Within our organisation along, there are different tools due to funding requirements of State and Federal Governments and also due to requirements of partnership ventures	*High administrative burden collecting multiple sets of data for multiple funders*
We are in the process of implementing the Basis 32 this year	*In development*
We have developed a Consumer led Client feedback process but have not yet implemented it.	*In development*
The Outcome STAR 'Homelessness' category is generally used for people who are homeless. However given the breadth and holistic nature of the rating areas, and the fact that the tool is more quantifiable in its measure, we as a service have adopted this outcome tool. The tool informs the assessment and case planning process, and is also used as a review mechanism of the distance travelled. It measures the following on a ten point scale; Self-motivation, Emotional and Mental Health, Meaningful use of time, Managing tenancy, offending behaviour, managing money and personal administration, social networks and relationships, drug and alcohol usage, and physical health.	*Positive feedback on the Recovery STAR*
We have only recently started developing outcomes measurement for the organisation and will be happy to share our knowledge. Caution needs to be exercised in regard to carer's input in respect to client's wishes.	*Carers' input needs to be considered during outcome measurement*
We have used the WHO-QoL and the Basis 32 for some years but are currently introducing the Recovery Star as we find the current tools are not sufficiently client focussed and do not measure change over time or client outcomes. This has been a considered approach involving training of all staff and new IT and at considerable cost to the organisation.	*In development; Strong data system/infrastructure is enabling better outcome data collection*
Although I have responded to the survey as accurately as possible, we are currently moving from Basis32, HONOS & LSP16 (generally considered by staff to be inadequate to capture outcome measurement effectively) to the Recovery STAR. The R.STAR is currently used by several PDRSS services within the Grampians Region with good results, and following recent external training is about to be adopted by the Balgartnie Centre. The R.STAR is expected to provide better tools for staff, members and carers to better construct recovery pathways, progress confidence and outcome measurement.	*Positive feedback on the Recovery STAR; Need for contextually appropriate tools for CMOs*
Measurement requests from funding bodies keep changing, and our database, like many others that other community sector providers have, is owned and managed externally with limited flexibility and change options.	*Lack of flexibility from funders; High administrative burden collecting multiple sets of data for multiple funders*

CMO Comment (Q56)	Thematic Summary
We have gone through an extremely long process (over a year) in developing and creating a tool that will effectively and efficiently measure outcomes for each individual. The backbone of this tool is that it will be meaningful and useful for clients and staff and is centred upon each individuals recovery needs. The tool is currently been evaluated.	*In development*
Very interested in all available assistance, and happy to contribute whatever we can moving forward. Thanks	*Desire for more information and support to be made available on outcome measurement for CMOs*
So many providers use different tools it makes it difficult to have any standard language across the sector. some standardisation would be useful.	*Need for standardisation*

Appendix J – List of Tables and Figures

Tables

Figures

4.2. Appendix 2 - Outcome measurement in the community managed mental health sector: A review of the literature

AMHOCN

Australian Mental Health Outcomes and Classification Network

‘Sharing Information to Improve Outcomes’

An Australian Government-funded initiative

Outcome measurement in the community managed mental health sector:

A review of the literature

Tim Coombs, Angela Nicholas, Sarah Anstey and Josh Onikul

Ver 1.1

August 2013

Document Information

Title: Outcome measurement in the community managed mental health sector: A review of the literature Version 1.0

Suggested citation for this document

Australian Mental Health Outcomes and Classification Network (2013). *Outcome measurement in the community managed mental health sector: A review of the literature Version 1.0* Sydney, NSW.

Document History

Version	Date	Details
1.0	20 May 2013	Provided for stakeholder consultation.
1.1	2 August 2013	Developed after stakeholder consultation

What is the Australian Mental Health Outcomes and Classification Network?

The Australian Mental Health Outcomes and Classification Network (AMHOCN) was established by the Australian Government in December 2003 to provide leadership to the mental health sector to support the sustainable implementation of the outcomes and casemix collection as part of routine clinical practice. It aims to support states and territories and to work collaboratively with the mental health sector to achieve the vision of the introduction of outcomes and casemix measures. AMHOCN consists of three components: a training and service development component supporting training in the measures and their use for clinical practice, service management and development purposes; a data bureau responsible for receiving and processing information; and an analysis and reporting component providing analysis and reports of submitted data. Currently, the Australian Government has contracted the following organisations to undertake these roles: The NSW Institute of Psychiatry (training and service development); Strategic Data Pty Ltd, (data bureau); and The University of Queensland (analysis and reporting). Further information regarding AMHOCN can be found at http://amhocn.org.

Acknowledgments

This review of the literature has been a collaborative effort by Tim Coombs, Josh Onikul, Sarah Anstey and Rosemary Dickson (AMHOCN) and Angela Nicholas. AMHOCN has also closely collaborated with Tully Rosen, representing Community Mental Health Australia, on key aspects of the work.

Feedback

Comments on the document are welcomed. Readers are encouraged to submit comments to Tim Coombs at timcoombs@live.com.au or Rosemary Dickson at rosemary.dickson@nswiop.nsw.edu.au

Contents

List of tables

Executive Summary

There has been increasing interest in routine outcome measurement in the community managed sector. However there is a need for greater knowledge about the outcome measurement activities in the sector, and a desire for guidance on what measure or measures may be suitable for use in the sector. At the March 2012 Mental Health Information Strategy Standing committee a project was instigated with 5 objectives:

1. to describe the current status of consumer outcome measurement in the Australian mental health community managed organisation (CMO) sector;
2. to identify good practice examples where consumer outcome measurement has been introduced within day to day service delivery in the mental health CMO sector, with a view to highlighting possible directions for implementation of consumer outcome measurement in the sector;
3. to describe the information infrastructure in place across the mental health CMO sector that supports the use of routine consumer outcome measurement, and the extent to which suitable information infrastructure is comprehensively available;
4. to review the available measures of consumer outcomes that may be suitable for use in the mental health CMO sector, taking account of the range of service types delivered by the sector; and
5. to develop recommendations on:
   - a short list of consumer outcome measurement instruments that offer most potential for use in Australia across the various service types; and
   - the information infrastructure development that would be required to introduce reporting of consumer outcomes as a component of future national dataset requirements covering the mental health CMO sector.

To achieve the project objectives, three core pieces of work were outlined. These included:

1. a review of the literature identifying measures that may be suitable for use in the mental health CMO sector;
2. a survey of mental health CMO service providers and a set of funder interviews identifying measures that are currently being used, the infrastructure in place to support the collection and examples of good practice in the use of outcome measures in the sector; and
3. a workshop that brings together key sector and funder stakeholders to develop agreement on a short list of measures that would be suitable for introduction in the sector and the information infrastructure development needed to support the collection of the agreed measures.

During 2012-2013, these three pieces of work were completed. A national survey was undertaken that resulted in one hundred and thirty two (132) CMOs validated responses. Major jurisdictional funders of CMO mental health services were interviewed about their programmatic use of outcome measurement. A summary of the results of these consultations are provided below and the complete report is provided in appendix 1.

Concurrently a review of the literature was undertaken aimed at identifying the psychometric properties of the measures that were being used in the CMO sector in Australia and the identification of measures that may be suitable for use in the sector. These measures were organised across seven outcome domains: recovery, cognition and emotion, functioning (activities of daily living and interpersonal relationships), social inclusion, quality of life, experience of service provision and multidimensional measures. One hundred and thirty six (136) measures were identified and the psychometric properties of each of these measures including validity, reliability and sensitivity are provided in appendix 2. A set of criteria for measures selection was identified and this produced a short list of thirty one (31) measures suitable for recommendation to the sector and to funders for use in routine collections by CMOs. These measures are reported across the seven outcome domains and the relative suitability of these measures across the different CMO service types are also indicated.

The results of these two activities were presented to a workshop of key sector and funder stakeholders in May 2013. Workshop participants agreed that:

1. Routine outcome measurement should occur within the CMO sector;
2. Routine outcome measurement should include the collection of a universal measure of consumer or carer experience of service provision, and then be supplemented by specific measures depending on CMO service type and program characteristics;
3. There should be production of a “guidebook” that builds upon the results of the current project and which outlines measures, data collection protocols and the preconditions necessary for the implementation of routine outcome measurement in the sector; and
4. The guidebook would be used to structure discussion between CMO peak bodies, service providers, consumers, carers and funders to enable the implementation of routine outcome measurement to the sector.

Chapter 1: Introduction

1.1 About the Project

The Australian Mental Health Outcomes and Classification Network (AMHOCN) and Community Mental Health Australia (CMHA) have undertaken a project focused on the use of outcome measures in mental health-related community managed organisations (CMOs) across Australia.

The project has included the following three components:

1. A literature review to identify instruments measuring mental health consumer outcomes:
    - relevant to the range of services provided by Australian community managed mental health services; and
    - that show the best potential for use across the sector.
2. A scoping survey to determine current use of outcome measures and existing supporting infrastructure in the Australian mental health CMO sector.
3. A working group that has provided advice on the findings of the literature review and the scoping exercise.

This report summarises the findings of the literature review outlined in point 1 above.

A final report will also be written that outlines recommendations regarding those measures that have the most potential for use in Australia across the various mental health CMO service types and that identifies information infrastructure development issues that would need to be considered to introduce reporting of consumer outcomes.

1.2 Report Structure

Chapter 1 of this report provides the rationale and background to the current review.

Chapter 2 outlines the methods used to conduct the review, including the review criteria and psychometric criteria examined for each measure that has been identified.

Chapter 3 comprises the largest component of this report and outlines the results of the review. The results are divided into seven sections reflecting seven outcome domains. The results of the review for each domain are described using two tables. The first table provides a brief description of each measure. The second table outlines the psychometric and usability properties identified throughout the review.

Chapter 4 is the discussion and recommendations section of this report. It summarises the findings of this review and makes recommendations about the possible use of those outcome measures in terms of meeting the most review criteria for each domain.

The measures identified in this review are included as Attachments 1 to 6. Each Attachment contains the measures for one recovery domain.

1.3 Background

The *Fourth National Mental Health Plan: An agenda for collaborative government action in mental health 2009-2014* includes a priority area that outlines the importance of accountability in the mental health system through the measurement and reporting of progress. [1]

CMOs form an essential component of the Australian mental health care system. However, these and other mental health organisations have an increasing need to provide evidence for the effectiveness of their services in assisting those with mental illness to better manage, and to recover from, their mental illness. This evidence can be used externally, to provide evidence of effectiveness to potential and existing funders, and internally, to assist quality improvement and needs-based planning.

In March 2009, the Productivity Commission's review of the not-for-profit sector proposed a measurement framework to guide the collection of data, evaluation approaches and reporting about that sector in Australia (Productivity Commission 2010).[2] The framework is a hierarchy of four contribution measures: inputs (measures of the resources used); outputs (indicators of the level of activity undertaken); outcomes (direct effects on activity participants); and impacts (longer-term effects on the participants and the community more broadly).

Generally, outcomes can be seen from a variety of perspectives [5]:

- individual service recipients/consumers;
- communities – community members and organisations; and
- services and service systems.

For the purposes of this project, the focus is on the outcomes for individuals who have contact with community managed/non-government organisations. The introduction of outcome measures attempts to measure whether a change has occurred for a consumer as a result of mental health care. Increasing focus is being given to the responsibility of service providers in the mental health sector to use outcome measures to contribute to the ongoing review and development of practice as well as to inform service planning, policy development and the broader community. Within the mental health field, the regular assessment of consumer outcomes has been seen as a priority.

In order to provide this evidence, CMOs need to establish standardised use of quality outcome measures that can demonstrate consumers' improvement in the variety of domains that have come to represent recovery. The Australian Institute of Health and Welfare (AIHW) report *Mental Health Non-Government Organisation Data Development Project* (2011) specifically recommended that a discrete outcome measurement project be funded to investigate the current use and potential standardisation of outcome measurement tools being used in the mental health CMO sector.[3]

The AIHW *Preliminary Scoping Report* (2010) identified the issues, options and potential data development work required to obtain more detailed information about the specialised mental health services currently being delivered by CMOs in Australia. The preliminary scoping report recognised that outcome measures are a vital element in enabling a better understanding of the role the CMO sector plays in the community.

Additionally, the AMHOCN Review of Recovery Measures (2010) found that there is growing use of outcome measurement instruments, among other tools, to assess recovery goals for consumers.[4] While this domain of activity is less developed than outcome measurement, it is a key agenda of the Commonwealth and state governments. Recovery measurement is an emerging construct that forms a key component of outcome measurement, and should be explored while reviewing outcome measurement tools, although it would be premature to specify a collection of recovery measurement "tools."

In 2011 the AIHW, in partnership with CMHA, completed Phase 1 of the NGO Establishments National Minimum Data Set (NMDS) Project. *The Mental Health Non-Government Organisation Data Development Project: Phase 1 Final Report* (2011) contained a number of options and recommendations to achieve progress in nationally consistent CMO data collection, first of which was that a discrete project be funded to investigate the current levels of use and standardisation of outcome measurement tools in the mental health CMO sector.

At its meeting in November 2011, the Mental Health Information Strategy SubCommittee (now known as the Mental Health Information Strategy Standing Committee (MHISSC)) agreed that a project should be funded to investigate the current use and potential standardisation of outcome measurement tools being used in the mental health CMO sector. A project plan was agreed at the next MHISSC meeting in March 2012.

The Australian Mental Health Outcomes and Classification Network, in collaboration with Community Mental Health Australia, were funded to undertake this project which has focussed on several core pieces of work:

- investigate the current national use of outcome measurement instruments in the mental health CMO sector, including those with a recovery focus, and identify good practice examples;
- investigate the information infrastructure in place across the mental health CMO sector that supports the use of outcome measurement;
- conduct a comprehensive literature review of outcome measurement instruments available for use by mental health CMOs;
- conduct a national CMO and funder workshop to present and refine the findings;
- identify a candidate set of outcome measurement instruments applicable for each of the National Minimum Data Set Taxonomy service categories.
- deliver a final report to MHISSC with recommendations for activities that could be undertaken to encourage the use and standardisation of valid outcome measurement tools in the CMO sector.

The *Report on the National Community Managed Outcome Measurement Project Survey and Consultation* (2013)(To add REF), which described the results of the survey of outcome measures currently being used across Australian CMOs, conducted as part of this project, has illustrated that there is no current standard outcome measure or suite of measures being used by all mental health

CMOs nationally. CMOs also often use more than one outcome measurement tool. A number of instruments are likely to be used to adequately reflect the range of consumer outcomes.

This review of the literature aims to identify measures that could be used by mental health CMOs, or are being used, and describe the domains that they measure and some of the psychometric properties.

1.4 Key Issues Underpinning the Project

1.4.1 Service type taxonomy

An important area of work that has strong linkages to this project has been the work undertaken by the AIHW on the development of a taxonomy which enables the categorisation of services delivered by mental health CMOs into like service types, facilitating the identification of the statistical counting units to be used for output reporting. The development of a service taxonomy, consisting of 17 service types, was central to the development of the Mental Health Non-Government Organisation Establishments National Minimum Data Set (MH NGOE NMDS).

It should be noted that some mental health NGOs are not required to report under the MH NGOE NMDS as they are outside the scope of the collection. Other NGOs will be funded to provide multiple services, not all of which will be in scope of the NMDS.

AMHOCN and CMHA have utilised the service type taxonomy to inform deliberations about the applicability of outcome measures. Some service types do not lend themselves to the measurement of outcomes for individual consumers, e.g. they may provide services at a population level and have thus been excluded in the scope of this project. Service types that are seen to be out of scope are those that focus on mental health promotion, mental illness prevention, sector development, individual advocacy and online self-help.

However, even with the exclusion of these service types, the mental health community managed organisation sector remains complex. The service types that are in scope for the purposes of the project provide a wide variety of services in many different ways and in different settings. There is also inherent variability within the different service types. For example, what actually occurs within counselling services “will depend on the individual counsellor, the individual client and the specific issue”. Any recommendation about which measure(s) may be suitable for demonstrating the outcomes of care for a particular service type will have to be capable of dealing with this variability, while being relevant to consumers, carers and workers in the community managed mental health sector.

The following table describes the 17 service types in the taxonomy and indicates those that are in scope or out of scope for the purposes of this project.

Table 1. Service types in /out of scope for this project

Service Type	In scope
Counselling—face-to-face	✓
Counselling, support, information & referral—telephone	✓
Counselling, support, information & referral—online	✓
Self-help—online	✗
Group support activities	✓
Mutual support and self-help	✓
Staffed residential services	✓
Personalised support—linked to housing	✓
Personalised support—other	✓
Family & carer support	✓
Individual advocacy	✗
Care co-ordination	✓
Service integration infrastructure	✓
Education, employment & training	✓
Sector development and representation	✗
Mental health promotion	✗
Mental illness prevention	✗

A description of each of the services types is provided below.

Counselling—face-to-face

Counselling services provide a structured process that is concerned with addressing and resolving specific problems, making decisions, working through feelings and inner conflicts, or improving relationships with others[5].

Counselling services operate through a range of mediums including face-to-face, telephone and online. This service type is intended only for services providing face-to-face counselling.

The counselling process will depend on the individual counsellor, the individual client and the specific issue.

Counselling, support, information & referral—telephone

Mental health support, information and referral services are those that provide support for people experiencing mental illness and which offer reliable referrals, information and self-help resources to empower people to take steps towards maintaining mental health and emotional wellbeing (Lifeline 2012).

Counselling, support, information & referral—online

Mental health support, information and referral services are those that provide support for people experiencing mental illness.

Self-help—online

Self-help—online services are interactive online programs which take people with a lived experience of mental illness through exercises to help them develop skills to handle life's challenges more effectively. Unlike *Counselling, support, information and referral–online*, services which fall under *Self-help–online* do not involve interaction with another person, only interaction with the online program's content.

Group support activities

Group support activities are services that aim to improve the quality of life and psychosocial functioning of mental health consumers, through the provision of group-based social, recreational or prevocational activities. In contrast to services in the *Mutual support and self-help* service type, *Group support activities* are led by a member of the NGO.

Mutual support and self-help

Mutual support and self-help services provide information and peer support to people with a lived experience of mental illness and/or their carers. People meet to discuss shared experiences, coping strategies and to provide information and referrals[6]. Self-help groups are usually formed by peers who have come together for mutual support and to accomplish a specific purpose[7].

Staffed residential services

Staffed residential services are those that provide overnight accommodation in a domestic-style environment, which is staffed for a minimum of 6 hours a day and at least 50 hours per week.

Personalised support—linked to housing

Personalised support services—linked to housing are services that provide personalised support that is coordinated with provision of social housing or privately negotiated housing at the point of entry into the program (but not tied to such indefinitely).

Personalised support services are flexible services tailored to a mental health consumer's individual and changing needs. They include a range of one-on-one activities provided by a support worker directly to mental health consumers in their homes or local communities [8].

Personalised support—other

Personalised support services—other are services that provide personalised support that is independent of housing arrangements (e.g. provision of social housing or privately negotiated housing) at the point of entry into the program.

Family & carer support

Family and carer support services are services that provide families and carers of people living with a mental illness support, information, education and skill development opportunities to fulfil their caring role, while maintaining their own health and wellbeing[9] .

Individual advocacy

Individual advocacy services are those that seek to uphold the rights and interests of people with a mental illness, on a one-to-one basis, by addressing instances of discrimination, abuse and neglect.

Individual advocates work with people with mental illness on either a short-term or issue-specific basis.

Care coordination

Care coordination services provide a single point of contact (via a Care Facilitator) for people (and their families/carers) with lived experience of mental illness and complex care needs. Care Facilitators will be responsible for ensuring all of the patients' care needs, clinical and non-clinical, and as determined by a nationally consistent assessment tool, are being met [10].

Service integration infrastructure

Service integration infrastructure includes services that provide infrastructure integration to establish a 'one stop shop' service platform that brings together an appropriate range of mental health-related services, both existing and new, which aim to improve the mental well-being and social participation of people with mental illness.

Education, employment & training

Education, employment and training services are those which provide or support people with lived experience of mental illness, in gaining education, employment and/or training.

Sector development & representation

Mental health sector development and representation services engage with a wide variety of issues regarding the sustainability and development of the mental health sector. This includes information dissemination, advocacy, policy analysis, program development, and sector capacity building[11].

Mental health promotion

Mental health promotion services are those that work to maximize mental health and wellbeing among populations and individuals. It involves raising awareness about mental health/mental illness. This is achieved by providing information and/or education to the community to increase its capacity to support people with lived experience of mental illness [12].

Discrimination reduction programs or campaigns aim to challenge prejudices towards people with a mental illness. Discrimination reduction involves the implementation of long-term strategies to educate the general public about mental health and encouraging a socially inclusive attitude toward people with a mental illness [13].

Mental illness prevention

Mental illness prevention services deliver activities that are designed to prevent the development of a mental disorder. Prevention endeavours to avoid the development of mental illness, generally through population based health activities, reduction of known risk factors and promoting factors that support resilience and coping [14].

1.4.2 Domains of outcome measurement

Prior to work on the review of the literature, it was necessary to identify a broad set of outcome domains, or elements, that may be suitable for demonstrating the outcomes of care for the different types of services offered by the mental health CMO sector. These domains were then used to guide the literature search and inform the analysis of the information gathered via the survey of the sector about tools currently being used. In order to articulate appropriate domains, it was first necessary to review seminal work undertaken by others in this area.

Andrews et al [15] identified five domains that could be measured routinely to determine the outcome of care at the individual level. These were described as:

1. Symptoms
2. Functioning
3. Quality of Life
4. Burden
5. Satisfaction with services

Bickman et al [16], in a review of outcome measurement for child and adolescent mental health services, identified nine domains, one of which is described as “multidimensional” because some measures actually capture information across multiple domains.

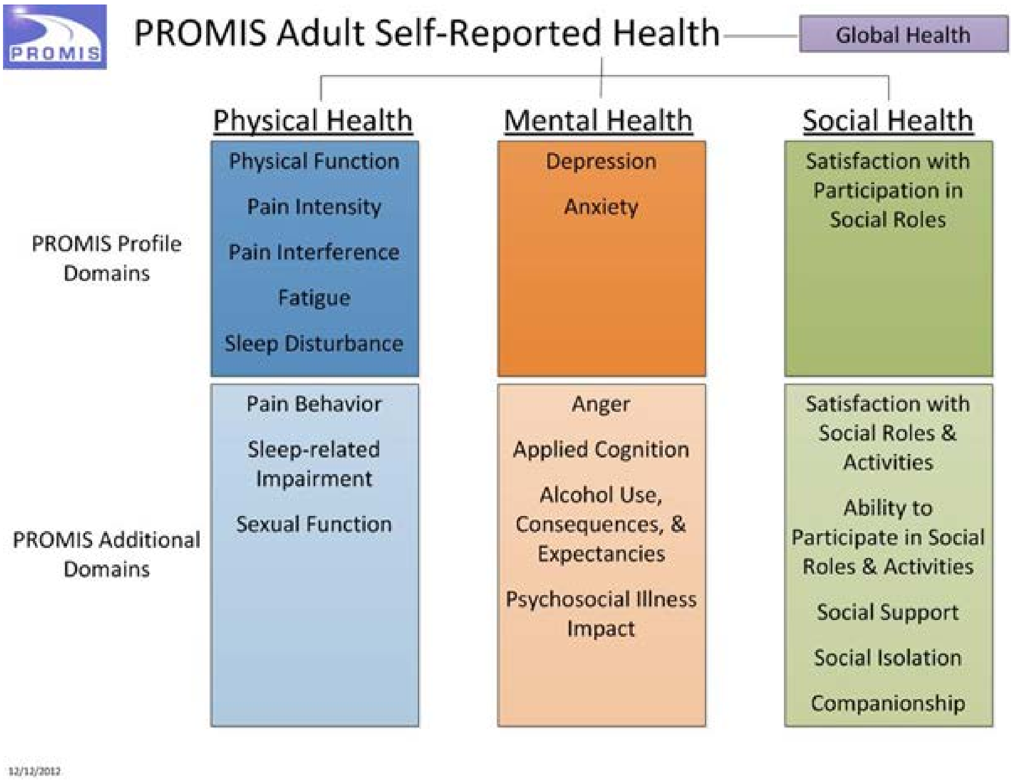
1. Symptoms
2. Functional impairment
3. Functional competence
4. Family functioning
5. Satisfaction
6. Self-esteem
7. Quality of life
8. Goal attainment
9. Multidimensional

Slade [17], following a systematic review of the literature of outcome measurement in mental health, identified seven domains. These were:

1. Wellbeing
2. Cognition emotion
3. Behaviour
4. Physical health
5. Interpersonal
6. Society
7. Services

In the United States, the Patient-Reported Outcomes Measurement Information System (PROMIS) Roadmap initiative (www.nihpromis.org), under the auspice of the National Institutes of Health (NIH), is a 5-year cooperative group program of research that is designed to develop, validate, and standardise item banks to measure patient-reported outcomes (PROs). The aim of this activity is to provide a framework that drives medical research into the 21st century and enable research activities to result in tangible benefits for people [18].

This framework identifies domains and subdomains that guide the research agenda and can be found at http://www.nihpromis.org/measures/domainframework1



Within the PROMIS framework, a specific mental health domain is populated by six subdomains:

1. Depression
2. Anxiety
3. Anger
4. Applied cognition
5. Alcohol use, consequences and expectancies
6. Psychosocial illness impact

However, for community managed organisations, the social domain may be more relevant for the purposes of measuring outcomes for the different service types within the sector.

The National Institute for Mental Health in the United Kingdom has produced a compendium of outcome measures that can be used in mental health [19]. Although not specifically organised into domains, they note that mental health is measured by instruments that gauge symptoms as well as quality of life, social functioning, social inclusion and self- reported perceptions of health status and recovery from illness. The identified measures can be generally described as falling in the following domains:

1. Symptoms
2. Social Inclusion
3. Health care and needs assessment
4. Patient perceptions of care
5. Quality of life and social functioning and wellbeing
6. Recovery
7. Social Functioning and functional disabilities

The Royal College of Psychiatrists [20] identified and recommended that, for adults, outcome measures might fall into six domains:

1. Effectiveness outcomes
    a. Patient identified goals
    b. Symptoms and Social functioning (as measured by the Health of the Nation Outcomes Scales)
    c. Condition specific measures
2. Quality of Life
3. Social Outcomes
4. Physical Health Measures
5. Patient safety outcomes
6. Patient and Carer Experience

In contrast, in another piece of work focussed on older persons, the Royal College of Psychiatrists [21] identified and recommended the use of outcome measures that fall into thirteen domains:

1. Global measures
2. Cognition
3. Delirium
4. Depression
5. Anxiety
6. Psychological therapies
7. Psychosis
8. Activities of Daily living
9. Quality of Life
10. Carers outcomes
11. Service Satisfaction
12. Recovery and wellbeing
13. Behaviour that challenges

In the mental health community managed sector in NSW, five domains of potential outcome measurement have been identified [22]:

1. Disease characteristics/perception of disease impact
2. Individual or environmental risk/ protective factors for the onset or relapse
3. General health status, psychological wellbeing and quality of life
4. Physical and psychological disability and Handicap
5. Needs, need satisfaction, satisfaction with services

This brief review indicates that there have been a variety of approaches to conceptualising and organising outcome measures for the purpose of reporting on those areas that are important to measure when demonstrating the outcomes of care. Each approach has its advantages and disadvantages. However, it is clear that there are some common domains for organising measures that demonstrate the outcomes of care.

It is also clear that, since some of the work described above has taken place, there have been revisions to both policy drivers and priorities which have resulted in changes to the focus on some aspects of care more than others. For example, when Slade [17] undertook his review of the literature in 2002, and subsequently identified outcome domains, recovery was identified as a subdomain of cognition/emotion and behaviour. While technically accurate, the Fourth National Mental Health Plan [14] has highlighted the importance of a recovery orientation to service provision and there has been both significant work on, and interest in, the measurement of recovery.

Community managed mental health services promote social inclusion and recovery by:

- providing consistent social, emotional and practical support to help people manage problematic areas of their lives better;
- reconnecting people and strengthening relationships with friends and family;
- supporting people to access and stay engaged with education and employment opportunities;
- supporting individuals to maintain stable housing options resulting in reduced needs for hospital admissions;
- facilitating access to physical healthcare services and providing support to achieve physical health;
- contributing to a sense of connections, belongingness and hope for a valued future for people living with mental illness; and
- promoting relationships between services, people affected by mental illness and the community. [23]

CMOs strive to promote recovery across a broad range of areas and it is therefore necessary to adopt a similarly broad approach to outcome measurement. This will ensure that the outcome measures used in the sector adequately reflect the range of domains in which people with a mental illness might experience improvement through their involvement with CMOs. Not identifying recovery and social inclusion as domains of interest would be a significant omission and would challenge the integrity of this project - including the literature review, survey and subsequent stakeholder consultation and engagement process.

Table 2 below maps the studies and reports identified above to seven domains that will be used to organise the presentation of measures that will be the focus of this literature review. No system is perfect and these domains represent a compromise for the purpose of simplifying what could be regarded as complex elements.

Table 2. Mapping of studies of domains

<table>
<tr><th>Study</th><th>Recovery</th><th>Cognition and Emotion</th><th>Functioning</th><th>Social Inclusion</th><th>Quality of Life</th><th>Experience of Service</th><th>Multidimensional</th></tr>
<tr><td>Andrews[15]</td><td></td><td>Symptoms
Burden</td><td>Functioning</td><td></td><td>Quality of Life</td><td>Satisfaction with services</td><td></td></tr>
<tr><td>Bickman[16]</td><td></td><td>Symptoms
Self esteem</td><td>Functional impairment
Functional competence
Family functioning</td><td></td><td>Quality of Life</td><td>Satisfaction</td><td>Goal Attainment
Multidimensional</td></tr>
<tr><td>Slade[24]</td><td></td><td>Cognition</td><td>Behaviour
Interpersonal</td><td>Society</td><td>Wellbeing
Physical health</td><td>Services</td><td></td></tr>
<tr><td rowspan="2">PROMIS</td><td rowspan="2"></td><td>Depression, Anger, Anxiety, Applied cognition</td><td colspan="2">Psychosocial illness impact</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td colspan="2">Alcohol use, consequences and expectancies</td><td></td></tr>
<tr><td rowspan="2">National Institute for Mental health</td><td rowspan="2">Recovery</td><td rowspan="2">Symptoms</td><td></td><td colspan="2">Quality of life and social functioning and wellbeing</td><td rowspan="2">Patient perceptions of care</td><td rowspan="2">Health care and needs assessment</td></tr>
<tr><td colspan="2">Social Functioning and functional disabilities</td><td></td></tr>
<tr><td rowspan="2">Royal College of Psychiatrists Adult</td><td rowspan="2"></td><td colspan="2">Symptoms and Social functioning</td><td rowspan="2">Social Outcomes</td><td rowspan="2">Quality of Life
Physical Health Measures</td><td rowspan="2">Patient and Carer Experience</td><td rowspan="2">Patient identified goals
Patient safety outcomes</td></tr>
<tr><td>Condition specific measures</td><td></td></tr>
<tr><td>Royal College of Psychiatrists Older Persons</td><td>Recovery and wellbeing</td><td>Cognition, Delirium, Depression, Anxiety, Psychological therapies, Psychosis, Carers outcomes
Behaviour that challenges</td><td>Activities of Daily living</td><td></td><td>Quality of Life</td><td>Carers outcomes
Service Satisfaction</td><td>Global measures</td></tr>
<tr><td rowspan="2">MHCC Outcomes Report</td><td rowspan="2"></td><td rowspan="2">Disease characteristics/perception of disease impact</td><td rowspan="2">Physical and Psychological disability and Handicap</td><td rowspan="2"></td><td rowspan="2">General health status, psychological wellbeing and quality of life</td><td></td><td>Individual or environmental risk/ protective factors for the onset or relapse</td></tr>
<tr><td colspan="2">Needs, need satisfaction, satisfaction with services</td></tr>
</table>

Table 3 describes the characteristics of each of the seven domains that will be used to organise the presentation of measures that will be the focus of this literature review.

Table 3. Outcome domain characteristics

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
The personal process of individual recovery.	Individual consumer cognitive performance and emotional experience Individual carer cognitive performance and emotional experience	Simple and complex functional abilities are covered here including the ability to undertake activities of daily living consistent with developmental stage. The quantity and quality of interpersonal relationships consistent with developmental stage.	Education, employment, citizenship, stability of housing	General life satisfaction, physical health and wellbeing	Service satisfaction, consumer or carer experience of service provision Care or service co-ordination	Measures that capture information across multiple domains

1.4.3 Criteria for selecting measures

Before deciding on the selection of a particular measure there are a number of methodological and practical decisions that need to be made [25]. These include identifying the:

- goals of implementing a measure;
- people who will participate in the collection, including the setting and timing of assessment;
- questionnaire(s) to be used;
- mode of administration and scoring of the questionnaire;
- process for reporting results;
- aids that will help with score interpretation;
- way people will respond to issues raised by the questionnaire(s); and
- evaluation of the impact of the collection of this information on practice.

In relation to the identification of which measures to use, Andrews et al identified six criteria upon which the selection of outcome measures should be based [15]:

- applicable;
- acceptable;
- practical;
- reliable;
- valid; and
- sensitive to change.

Services users' views are also seen as essential to the selection of measures [26].

Given these considerations, the following six criteria are proposed for the selection of measures:

1. have been developed for use or used in the mental health sector;

2. have been developed or used in Australia, with identified potential for further development;

3. be able to be completed by either the consumer and/or CMO employee;

4. be brief and easy to use (time and/or number items);

5. yield quantitative data (does not exclude instruments that also yield qualitative data);

6. have undergone scientific scrutiny and have demonstrated strong psychometric properties (e.g., of internal consistency, validity, reliability and sensitivity to change).

The selection of measures will be based on these criteria. This may result in the identification of a number of measures within each domain. Once this process has been undertaken then a measure or measures will be identified for the different service types.

Chapter 2: Method

This literature review involved two primary methods. First CINAHL, Medline, Psychinfo and Google Scholar were used to search the literature for published articles relating to the development or testing of outcome measures from the various domains. Searches were limited to literature published between 2002 and February 2013 and to those published in English. The literature review was also informed by the results of the survey of standard measures being used in the sector, so that all measures being used within the sector are included in this review.

Secondly, the review also used a "snowballing" technique where measures included in the published literature relating to another measure were also included.

2.1 Evaluation Criteria

To help focus the review of the literature and to support instrument selection, a number of criteria were agreed. These review criteria specified that an ideal outcome measure useable for mental health CMOs in Australia would:

1. have been developed for use or used in the mental health sector;
2. have been developed or used in Australia, with identified potential for further development;
3. be able to be completed by either the consumer and/or CMO employee;
4. be brief and easy to use (time and/or number items);
5. yield quantitative data (does not exclude instruments that also yield qualitative data);
6. have undergone scientific scrutiny and have demonstrated strong psychometric properties (e.g., of internal consistency, validity, reliability and sensitivity to change).

The psychometric properties of any measures should be considered before introduction into practice [27]. The psychometric evaluation criteria examined for each measure are described in Table 2.

Table 4. Psychometric evaluation criteria

Psychometric property	Definition
Validity	Refers to the extent to which the instrument measures what it intends to measure
Content validity	Refers to the instrument's comprehensiveness (i.e., how adequately the sampling of items reflects its aims)
Construct validity	Involves conceptually defining the construct to be measured by the instrument and assessing the internal structure of its components and the theoretical relationship of its item and subscale scores.
Criterion validity	Assesses the extent to which the instrument correlates with a 'gold standard' or more established measure of the same theme and can be split into content validity and predictive validity
Concurrent validity	Pits the instrument against a comparable measure or measures at the same point in time
Predictive validity	Assesses the instrument's ability to predict a future outcome, such as resource use or treatment response
Reliability	The extent to which a given instrument gives stable, consistent results, or can be considered as the inverse of the degree of error obtained from any measurement.
Test-retest reliability	Defined as the degree of agreement when the same instrument is completed by the same individual (or administered to the same individual by the same interviewer) at two different points in time.
Inter-rater reliability	Defined as the degree of agreement when the same instrument is administered to the same individual by different interviewers at the same point in time (In the case of test-retest reliability, kappas of ≥0.81 are regarded as very good, 0.61-0.80 as good, 0.41-0.60 as moderate, 0.21-0.40 as fair, and ≤0.20 as poor.[28, 29])
Internal consistency	Assesses the extent to which the items of a measure produce similar scores. For example, if a test is split in half and the score on one half is significantly worse than the score on the other half, then the measure does not have a high level of internal consistency. (In the case of internal consistency, a Chronbach's α of ≥ 0.90 is regarded as excellent, 0.80-0.89 as good, 0.70-0.79 as acceptable, 0.60-0.69 as questionable, 0.50-0.59 as poor, and < 0.50 as unacceptable.[30])
Sensitivity to change	Related to both validity and reliability – an instrument that is both valid and reliable, and which demonstrates change over time, can be regarded as being sensitive to change.
Feasibility and utility	Related to concepts such as ease of administration, and acceptability to stakeholders, including carers, clinicians and service types within the sector.

Chapter 3: Results

The results of the review of the literature are divided into seven sections reflecting the seven outcome domains. The results of the review for each domain are described using two tables. The first table provides a brief description of each measure. The second table outlines the psychometric and usability properties identified throughout the review.

3.1 Recovery

3.2 Cognition and Emotion

3.3 Functioning

3.4 Social Inclusion

3.5 Quality of Life

3.6 Experience of Service

3.7 Multidimensional

3.1 Recovery

Table 5. Profile of individual recovery measures

MEASURE	DATE	COUNTRY	DESCRIPTION
Recovery Assessment Scale (RAS)	1995	United States	The RAS was designed to assess various aspects of recovery from the perspective of the consumer, with a particular emphasis on hope and self-determination. The original instrument comprises 41 items, and a shorter version is also available, comprising 24 items. In both versions, each item is rated on a 5-point Likert scale. It covers five domains: *personal confidence and hope*, *willingness to ask for help, goal and success orientation*, *reliance on others*, and *no domination by symptoms*. [31] [32-35]. A 24-item Japanese version of the RAS has recently been developed. [36] [36] [36] [36] [36] [36] [36] [36] [36]
Agreement with Recovery Attitudes Scale (ARAS)	1996	United States	The ARAS was developed to assess consumers' changes in attitudes with respect to movement towards a recovery process. It comprises 22 items, each of which is rated on a 5-point Likert scale. [31, 37].
Rochester Recovery Inquiry (RRI)	1996	United States	The RRI is an open-ended, qualitative questionnaire that assesses consumers' views about *their psychiatric hospitalisations, their own illness, their relationships with other people,* and *the way in which they cope with stressful situations.* It comprises 32 questions. [31, 38]
Consumer Recovery Outcomes System (CROS)	1997	United States	The CROS was designed to measure elements of recovery beyond a reduction in symptoms. It has three versions, each of which measures the consumer's recovery from a different perspective: the consumer him/herself (consumer version); a family member or carer ('very important person' version); and a provider (staff version). Each of the three versions assesses the following domains of recovery: *hope for the future, daily function, coping with clinical symptoms,* and *quality of life*, as well as three items about *medication and substance use.* The consumer version also assesses *satisfaction with treatment,* and the staff version also assesses *service use.* The consumer and staff versions each contain 38 items, and the very important person version contains 33. All items are rated on a 4-point Likert scale. [33, 39]
Crisis Hostel Healing Scale (CHHS)	1998	United States	The CHHS was designed as an evaluation tool for the New York Crisis Hostel Project. It contains 40 items, each of which is rated on a 4-point Likert scale. These items measure 10 domains relevant to the concept of recovery from the perspective of the consumer, namely *self-esteem, confidence and internal self-control; feelings/hopefulness; altered states; self- and other-inflicted violence; spiritual awareness; physical well-being; medications; giving and getting care in relationships; perceptions/self-*

MEASURE	DATE	COUNTRY	DESCRIPTION
			acceptance; and *comfort and pleasure*. [19, 20]
Personal Vision of Recovery Questionnaire (PVRQ)	1998	United States	The PVRQ was designed to assess consumers' beliefs about their own recovery and assesses the following five factors: *support, personal challenges, professional assistance, action and help-seeking,* and *affirmation.* It comprises 24 items, each of which is rated on a 5-point Likert scale.[31, 32, 40, 41]
Recovery Interview (RI)	1998	United States	The RI is an open-ended, qualitative questionnaire designed to examine recovery from a personal perspective by eliciting rich information that can be analysed for themes. It comprises 31 questions.[31, 42]
Recovery Attitudes Questionnaire (RAQ-16; RAQ-7)	1998	United States	The RAQ was developed to compare attitudes about recovery across different groups, particularly consumers, providers, family members and carers, and members of the general community. The RAQ-16 comprises 16 items, and the RAQ-7 comprises seven. The items in both versions are rated on a 5-point Likert scale.[31, 32, 43]
Mental Health Recovery Measure (MHRM)	1999	United States	The MHRM is designed to assess the recovery process for people with psychiatric disabilities via seven domains: *overcoming stuckness, self-empowerment, learning and self-redefinition, basic functioning, overall well-being, new potentials,* and *advocacy/enrichment.* It consists of 30 items, each of which is rated on a 5-point Likert scale.[32, 44, 45]
Reciprocal Support Scale (RSS)	2002	United States	The RSS was designed to measure mutual support from the perspective of consumers taking part in a specific recovery-oriented mentoring and education program known as Leadership Class. It is made up of 14 items, each of which is rated on a 5-point Likert scale. [45, 46]
Relationships and Activities that Facilitate Recovery Survey (RAFRS)	2002	United States	The RAFRS was developed by researchers to identify the influences that consumers consider most significant in their recovery process. The RAFRS comprises 18 items, each of which is rated on a 4-point Likert scale. In addition, it contains two additional open-ended items. It assesses two domains related to recovery: *relationships* and *activities.*[45, 47]
Illness Management and Recovery (IMR) Scales	2004	United States	The development of the IMR Scales took place within the context of evaluating the IMR program, which is designed to promote illness management and advancement towards personal goals. The instrument does not purport to measure cohesive domains, but rather to assess a variety of aspects of illness management and recovery. It has two versions, allowing for an assessment of recovery from the perspective of the consumer him/herself (client version) and a provider (clinician version). Both versions contain 15 items, each of which is rated on a 5-point Likert scale.[32, 45, 48]
Ohio Mental Health Consumer Outcomes	2004	United States	The OMHCOS comprises three forms designed to capture consumer outcome information and includes a total of 138 items: Consumer Adult Form A (67 items), Consumer Adult Form B (39 items), and Provider Adult Form A (32 items). All draw heavily on existing instruments, and Adult Form A includes

MEASURE	DATE	COUNTRY	DESCRIPTION
System (OMHCOS)			recovery-related items from the Making Decisions Empowerment Scale and the Quality of Life Interview.[32, 45, 49]
Peer Outcomes Protocol (POP)	2004	United States	The POP was developed in the context of the POP Project, the remit of which was to provide mental health peer support programs with a validated evaluation protocol to measure domains of significance to people recovering from a mental illness. It is organised into seven modules covering *demographics, service use, employment, community life, quality of life, well-being* and *program satisfaction.* In total, it contains 241 items that take the form of closed- and open-ended questions and Likert scales.[45, 50-52]
Recovery Measurement Tool (RMT)	2004	United States	The RMT was developed to measure recovery from the perspective of individual consumers and is based on a model of recovery that incorporates elements such as stages and external influences. It comprises 91 items, each of which is rated on a 5-point Likert scale.[45, 53]
Recovery Orientation (RO)[a]	2005	United States	The RO was an attempt to empirically conceptualise the recovery orientation, which yielded four domains: *empowerment, hope and optimism, knowledge* and *life satisfaction.* It consists of 56 items, each of which is rated by the consumer on a 4-point, 5-point or 7-point Likert Scale.[32, 54]
Recovery Process Inventory (RPI)	2006	United States	The RPI measures the following domains of recovery from the consumer's perspective: *anguish, connectedness to others, confidence/purpose, others care/help, living situation* and *hopeful/cares for self.* It comprises 22 items, each of which is rated on a 5-point Likert scale.[32, 55]
Milestones of Recovery Scale (MORS)	n. d.	United States	The MORS is a provider-rated measure of a consumer's level of recovery. The stated rationale for the provider perspective is that recovery is highly subjective, and that observable behavioural correlates of recovery may be more objective. The MORS requires providers to indicate the point the given consumer has reached in his or her recovery based on an 8-point scale that considers levels of risk, engagement and skill.[32, 56]
Multi-Phase Recovery Scale (MPRM)[a]	2009	United States	This MPRM was developed specifically for a study and assesses four phases of recovery: *mourning and grief, awareness and recognition, redefinition and transformation,* and *enhanced well-being and quality of life.* It comprises 11 items, each of which is rated on a 4-point Likert scale.[32, 57]
Maryland Assessment of Recovery in People with Serious Mental Illness (MARS)	2012	United States	A 25-item self-report measure, the MARS was developed to measure recovery status in people with serious mental illness based on the Substance Abuse and Mental Health Service Administration (SAMSHA) definition of recovery. The MARS targets six domains: *self-direction/empowerment, holistic, non-linear, strength-based, responsibility* and *hope*.[58]
Mental Health Recovery	2008	United Kingdom	The MHRS assesses consumers' progress towards recovery from their own perspective. It does this across 10 dimensions: *managing mental health, self-care, living skills, social networks, work,*

MEASURE	DATE	COUNTRY	DESCRIPTION
Star (MHRS)			*relationships, addictive behaviour, responsibilities, identity and self-esteem,* and *trust and hope.* Each dimension is equivalent to a single item and is rated on a 10-point Likert scale.[59] Ratings on the scale require discussion between the service user and staff, lasting about 1 hour.
Questionnaire about the Process of Recovery (QPR)	2009	United Kingdom	The QPR was developed in conjunction with consumers specifically to measure recovery from psychosis. The 22-items were developed based on clinical interviews and are rated by the consumer on a five-point Likert scale from 0 (Disagree strongly) to 4 (Agree strongly). The QPR has two subscales: the intrapersonal scale is made up of 17 item scores assessing tasks for which the consumer is responsible and that are carried out to facilitate recovery, and the interpersonal scale of 5 items relates 'to individuals' ability to reflect on their value in the external world and on how recovery is facilitated by external processes and interpersonal relationships with others' (p. 148).[60]
Subjective Experiences of Psychosis Scale (SEPS)	2011	United Kingdom	The SEPS has 52 items in three categories: (1)(i) positive and (ii) negative impacts of psychotic experiences on mental health and wellbeing - 31 items; (2)(i) positive and (ii) negative impacts of support (such as treatment, friends and family, alcohol) - 8 items; and (3) (i) positive and (ii) negative dimensions of psychotic symptoms (frequency and pleasantness) - 4 items.[61]
Self-Identified Stage of Recovery (SISR)	2003	Australia	The SISR assesses the stage of recovery that a given consumer has reached from the consumer's own perspective. It consists of two parts: Part A requires respondents to choose one of five statements reflecting the five stages of the model that best reflects their current experience, and Part B consists of four statements reflecting processes of recovery, each of which is rated on a 6-point Likert scale.[62]
Stages of Recovery Instrument (STORI)	2006	Australia	The STORI is designed to capture the following stages of recovery from the consumer's perspective: *moratorium* (a time of withdrawal characterised by a profound sense of loss and hopelessness)*;* *awareness* (realisation that all is not lost, and that a fulfilling life is possible); *preparation* (taking stock of strengths and weaknesses regarding recovery, and starting to work on developing recovery skills); *rebuilding* (actively working towards a positive identity, setting meaningful goals and taking control of one's life); and *growth* (living a full and meaningful life, characterised by self-management of the illness, resilience and a positive sense of self). The STORI comprises 50 items, each of which is rated on a 6-point Likert scale. [32, 62]
Stages of Recovery Scale (SORS)	2011	Taiwan	The SORS is a 51-item self-report scale of recovery developed as an outcome measure for recovery-oriented services in Taiwan. It measures both the processes and outcomes of recovery. Processes measured are *regaining autonomy, disability management/taking responsibility* and *sense of hope*.

MEASURE	DATE	COUNTRY	DESCRIPTION
			Outcomes are *overall wellbeing, social functioning/role performance* and *helping others*. The scale was developed with a four-factor structure representing four stages of recovery: 1. Overwhelmed with the disability, 2. Struggling with disability, 3. Living with disability and 4. Living beyond disability.

Table 6. Psychometric properties of measures for individual recovery

MEASURE	PSYCHOMETRIC PROPERTIES	
Recovery Assessment Scale (RAS)	**Validity**	The RAS has been shown to have good concurrent validity. It has demonstrated significant correlations in the expected directions with the Rosenberg Self-Esteem Scale (RSES), the self-orientation domain of the Empowerment Scale (ES), the Social Support Questionnaire (SSQ), the Social Networks Scale (SNS), the Herth Hope Index (HHI), the Resilience Scale (RS), the Mental Health Recovery Measure (MHRM), the Self-Identified Stage of Recovery (SISR), the Camberwell Assessment of Need Short Appraisal Scale (CANSAS), the Community Integration Measure (CIM) and most stages of the STORI.[35, 63-70] It has demonstrated non-significant or negative correlations with the consumer-rated Kessler-10 (K-10) and the clinician-rated Brief Psychiatric Rating Scale (BPRS) and Health of the Nation Outcome Scales (HoNOS). RAS scores were also significantly higher for people in employment.[67, 71] While originally reported as having a two-factor solution, a Dutch study found the RAS items to load on just one common factor. The original five-factor structure of the RAS [72] was confirmed in a further Australian study using both exploratory and confirmatory factor analyses, with the five factors accounting for 51.66% of variance.[66]
	Reliability	The RAS has been shown to have good test-retest reliability (r = 0.88) over a period of 14 days.[35] The RAS has been shown to have good internal consistency (α = 0.93) [35] and (> 0.7 for all coefficients), although McNaught et al suggested a possible reduction from 24 to 21 items.[66]
	Sensitivity to change	The sensitivity to change of the RAS has not been tested.
	Usability/Acceptability	The RAS has been tested with an Australian sample of 168 participants with persistent psychiatric disability relating to a psychotic disorder from Queensland, NSW and Victoria area mental health services.[66] A study of the acceptability of six measures that measure personal recovery and that have undergone scientific scrutiny, two service-user consultants indicated the RAS as their preferred measure, noting that it was easy to complete, relevant to consumer-defined recovery, used positive language, and measured quality of life and a broad conception of recovery (p. 200).[73]
Agreement with Recovery Attitudes Scale (ARAS)	**Validity**	No psychometric research on validity was found.
	Reliability	The researchers reported an internal consistency of α = 0.87 (Ohio Demonstration Project, 1998 in Ralph et al, 2000)
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.

MEASURE	PSYCHOMETRIC PROPERTIES	
Rochester Recovery Inquiry (RRI)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.
Consumer Recovery Outcomes System (CROS)	**Validity**	The CROS-Consumer version subscales have shown moderate to high positive correlations (*r* = 0.30–0.78)* with the Behavior and Symptom Identification Scale (BASIS-32:) and the Wisconsin Quality of Life Index (WQLI) with the exception of the CROS Treatment Satisfaction subscale (*r* = 0.26 with the BASIS-32 Psychosis scale). [39] *Note that these data come from an unpublished manuscript.[74] [75]
	Reliability	Test-retest reliability of the CROS-Consumer version over eight days ranged from *r* *= 0.69 to 0.76 across the subscales (n = 102) indicating good test-retest reliability. Test-retest reliability of the five subscales on the Staff version (n = 106) over 11 days ranged from *r* = 0.80 to 0.89 indicating good to very good test-retest reliability. Using a sample of 585 staff and consumers, the five subscales have demonstrated internal consistency with Cronbach's α* ranging from 0.79 to 0.90 for the Consumer version and 0.84 to 0.89 for the Staff version (demonstrating good internal reliability. Inter-rater reliability for the Staff version was also tested and the subscale scores ranged from 0.47 to 0.65.[39] *Bloom and Miller (2004) [39]report on the psychometric of the CROS 3.0 but do not describe the statistical test used. We have made assumptions about the type of tests used, but these should be accepted with great caution in the absence of any other data. This data is from a non-peer reviewed source.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The CROS takes 5 to 15 minutes to complete. Questions for the CROS were developed and piloted in conjunction with mental health consumers in the US. The CROS is intended for use in inpatient, outpatient, community and residential settings and for use with consumers with severe mental illness, including schizophrenia, depression and bipolar disorder. The CROS does not require training to complete, but there are specific guidelines for scoring, and a fee for use.[74]
Crisis Hostel Healing Scale (CHHS)	**Validity**	*All available psychometric testing information for the CHHS comes from a non-peer reviewed source[76] and the original source was unable to be located. Factor analysis did not strongly support the constructs within the concept map developed as a basis for the scale.
	Reliability	The CHHS is reported to have an alpha of 0.89*, demonstrating good internal consistency. Test-retest

MEASURE	PSYCHOMETRIC PROPERTIES	
		reliability at six months with the non-treatment control group was 0.67*, also showing good test-retest reliability over a long period for consumers not receiving treatment.[76]
	Sensitivity to change	Although no specific statistics are provided, the authors report significant changes over time were shown for the treatment group.[76]
	Usability/Acceptability	The concept map for the tool was developed in conjunction with consumers and providers using and running the New York Crisis Hostel Project (1998). It is a face-to-face interview that takes about 15 to 20 minutes to complete. The scale was piloted with 110 users from outpatient centres and 'psychosocial clubs', then implemented with a further 225 people throughout the project.[76]
Personal Vision of Recovery Questionnaire (PVRQ)	**Validity**	PVRQ scores were correlated with scores on the Community Living Skills Scale (Smith & Ford, 1990), the Client Experiences Questionnaire: Life Satisfaction Measure (Greenley, Greenberg, & Brown, 1994), the Multidimensional Scale of Perceived Social Support (Zimet, Dahlem, Zimet, & Farley, 1988), and the Mastery Scale (Pearlin, Menaghan, Lieberman, & Mullan, 1981) in order to determine level of convergent construct validity. Resulting correlations were low to moderate (Enzfield, 1998), suggesting modest relationships between beliefs measured by the PVRQ and level of functioning, satisfaction with life domains, perceived social support and perceived control over one's life. (Enzfield, 1998) The five factors established through repeated principal-axis factor analyses accounted for 46% of variance. (Enzfield, 1998).
	Reliability	Internal consistency for the five sub-scales ranged from α= 0.57 to 0.70, suggesting questionable reliability. (Enzfield, 1998)
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The PVRQ was developed with a panel that included twelve mental health consumers in the US. The resulting scale was tested with a sample of 291 mental health consumers aged 18 to 74 with a diagnosis of schizophrenia, bipolar disorder or depression and receiving community- or consumer-operated mental health services. (Enzfield, 1998)
Recovery Interview (RI)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.
Recovery Attitudes Questionnaire (RAQ-16,	**Validity**	The RAQ has been tested with 844 participants with mental disorders, their family members, mental health professionals and members of the general public. Factor analysis, which reduced the scale to 7 items (RAQ-7)

MEASURE	PSYCHOMETRIC PROPERTIES	
RAQ-7)		comprised of two factors that explained 54% of the variance. [43]
	Reliability	The RAQ-21 showed good internal consistency ($\alpha = 0.84$), which increased to 0.86 when two items were removed. Cronbach's alpha for the two factors were 0.66 and 0.64, and for the full scale, 0.70. This suggests only questionable to acceptable internal consistency for this diverse sample. Test-retest reliability (n = 85) over 19 days for the RAQ-7 was 0.674 and 0.61 and 0.62 for the two factors, indicating substantial agreement in scores over the two time points. [43]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The scale was developed with a group including mental health consumers in the US. [43]
Mental Health Recovery Measure (MHRM)	**Validity**	*In 2005, the MHRM author [77] provided the following data, which have not been peer-reviewed, for the Compendium of Recovery Measures (Campbell-Orde, 2005). The MHRM total shows correlations of $r = 0.57$ with the Community Living Scale (N=180)(Smith & Ford, 1990, in Campbell-Orde), $r = 0.67$ with the Empowerment Scale (N=150)(Rogers, Chamberlain, Ellison, & Crean, 1997 – ref in Campbell-Orde), and 0.73 and 0.75 (both N=150)(with the Connor-Davison Resilience Scale (Connor & Davidson, 2003 – ref in Campbell-Orde) and the Resilience Scale (Wagnild & Young, 1993 – ref in Campbell-Orde). In an unpublished update of the psychometric properties of the MHRM using a normative sample of N = 71 from 17 community mental health centres and consumer-operated services across Ohio, Bullock claims that the MHRM again showed a significant moderate correlation ($r = 0.58$) with the Empowerment Scale.[78] The MHRM was also correlated with score from the Ohio Department of Mental Health outcome measures [79] and showed significant correlations with the Symptom Distress Scale ($r = 0.39$) and the Health and Safety Scale ($r = 0.39$), though not with the Quality of Life Scale ($r = 0.09$). The MHRM has alp shown significant correlations with the total RAS ($r = 0.89$, $P < .01$) and RAS subscales ($r = 0.24$-0.71, P < .05), the Self-Identified Stage of Recovery (SISR) subscales (A-stages or recovery, Spearman's Rho = 0.46, P < .01 and B–recovery processes, r = 0.80, P < 0.01), the Global Assessment of Functioning (GAF) (r = 0.24, P <.05), and the Kessler-10 (K10) (r = 0.5, P <.01). [80] Calveti et al (2012), in their review of recovery measures where psychometric properties are assigned a positive, indeterminate or negative rating, the MHRM is given a positive rating for content validity, but a negative rating for construct validity.[81]
	Reliability	Internal consistency (N=279 from community-based mental health centres) was $\alpha = 0.93$ for the MHRM total and ranged from $\alpha = 0.60$ (Overcoming stuckness) to 0.86 (Overall wellbeing) for the established seven sub-scales, suggesting questionable to good internal consistency across the subscales, but excellent internal consistency for the scale as a whole. Test-retest reliability (N = 18) was $r = .92$ at one week and 0.91 at two

MEASURE	PSYCHOMETRIC PROPERTIES	
		weeks, showing temporal stability.[77] In the 2009 update of the MHRM's psychometric properties, Bullock (2009) states that the MHRM total scale (N = 671) was $\alpha = 0.95$ (excellent).[78] In the Calveti review described above, internal consistency and test-test reliability of the MHRM both receive a positive rating (Calveti et al, 2012).
	Sensitivity to change	The MHRM discriminates between people at different levels of recovery, and shows significant change for consumers undertaking an evidence-based treatment program.[77] Bullock (2009) also states that the MHRM shows significant improvement (a change score of more than 10, reliable change index = 1.29) from pre- to post-treatment for 47% of consumers participating in the Wellness Management and Recovery program, while 12% show deterioration.[78]
	Usability/Acceptability	The MHRM is based on a model of recovery developed from 18 interviews with individuals with psychiatric disabilities.[82] It has subsequently been tested with a large group of consumers (N = 200) from inpatient, forensic and community health settings.[77] It is designed specifically for use with mental health consumers with severe mental illness. Calveti attribute the MHRM a positive rating for administrator-friendliness.
Reciprocal Support Scale (RSS)	**Validity**	One of the scale authors provided the following information for inclusion in the Compendium of Recovery Measures,[83] but this is not peer-reviewed. The RSS correlated with scores on the self-esteem scale of the Ohio Department of Mental Health outcome measures [79] ($r = 0.28$, $P < .05$). It is not stated whether other correlations were conducted but were not significant.
	Reliability	Internal consistency for the total RSS was reported as $\alpha = 0.95$ (excellent).
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The RSS was developed in conjunction with mental health consumers, their family and friends, service providers and advocates[83] and is intended for use with consumers with a severe mental illness or substance abuse and has been tested in an outpatient setting in the US. No training is needed for administration but there are guidelines for scoring.
Relationships and Activities that Facilitate Recovery Survey (RAFRS)	**Validity**	No psychometric research on validity was found.
	Reliability	Internal consistency was not computed owing to the large amount of missing data from 'not applicable' responses to the scale [84]; therefore, no psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The RAFRS was developed with input from consumers and has been used with consumers with a severe mental illness in the US.[84] Psychometric testing was conducted with the survey being read aloud to

MEASURE	PSYCHOMETRIC PROPERTIES	
		participants and not as self-report.
Illness Management and Recovery Scales (IMR)	**Validity**	The client version of the IMR has shown significant positive correlations with self-reported symptom distress on the Colorado Symptom Inventory (CSI), the Recovery Assessment Scale (RAS), the Coping Efficacy Scale (CES) and the Multidimensional Scale for Perceived Social Support (MSPSS). The clinician version of the IMR correlates with the clinician-rated functioning on the Multnomah Community Ability Scale (MCAS). The client and consumer versions of the IMR also correlate well with each other.[48, 85, 86] A study of 210 Hebrew-speaking, treatment-seeking consumers with severe mental illness suggest the scale has a three-factor structure rather than five factors that reflect the goals of the IMR treatment for which the scale was developed. The three factors: *coping with illness outcome, knowledge and goals* and *effective medication use/reduced alcohol and drug use*, accounted for 40.2% and 46.5% of variance on the consumer and clinician versions, respectively. The correlation between the three factors and the total IMR versions was low but significant ($r = 0.28$, $P < 0.01$). This provides limited evidence for the scale's construct validity.[85] Scores on the total IMR and for the three factors were correlated with measures of social support and efficacious coping: different patterns of correlation were shown between the client and clinicians versions, with the client total IMR and subscales showing positive, significant correlations with efficacious coping, total social support and most of the social support subscales (except for the medication factor, which showed no significant relationships). The clinician IMR total and subscales showed fewer significant correlations with social support and efficacious coping. Again, this demonstrates limited concurrent validity for the IMR.[85] In their 2012 review,[81] give the content validity of the IMR a positive rating, but the construct validity a negative rating owing to the low to moderate correlations reported with other measures and the lack of *a priori* hypotheses proposed. In the most recent published study relating to the psychometric properties of the IMR, the concurrent criterion validity of the IMR clinician version was supported by scores on the IMR showing significant correlational relationships with progress toward employment goals ($r = 0.37$, 14% shared variance), housing goals ($r = 0.22$, shared variance = 5%) and education goals ($r = 0.26$, shared variance = 7%). Evidence for construct convergence validity was established by significant ($P < .001$) relationships between the clinician IMR and clinician ratings on the Substance Abuse Treatment Scale –revised[48] ($r = 0.30$-0.67) and client ratings on the Recovery Markers Questionnaire[87] ($r = 0.07$-0.40).
	Reliability	The client and clinician versions of the IMR have both demonstrated good test-retest reliability ($r = 0.81$-0.82 and 0.78-0.81, respectively) over a period of two weeks.[48, 86] In a small American study, the client and clinician versions of the IMR both demonstrated acceptable internal consistency at two time-points ($\alpha = 0.68$ and 0.72 and $\alpha = 0.71$ and 0.80, respectively).[86] An Israel-based study (N = 210) showed lower internal consistencies for the complete IMR at $\alpha = 0.55$ for the client version and $\alpha = 0.73$ for the clinician version. When the scale items were reduced to three factors, the internal consistency for items measuring the three

MEASURE	PSYCHOMETRIC PROPERTIES	
		factors was client α = 0.50-0.74 and clinician α = 0.47-0.83. These data suggest limited reliability for the IMR scales.[85] Cavelti et al (2012) assigned the IMR a negative rating for internal consistency given these inconsistent findings. They did, however, assign a positive rating for test-retest reliability.[81] In a recent, very large scale American study (N = 9412), internal consistency for the IMR was reported as α = 0.82 (good).[88]
	Sensitivity to change	In a randomised controlled trial of a recovery program for people with schizophrenia, those receiving the test program compared with treatment as usual showed significantly greater improvements on their IMRS-client and –clinician scores, consistent with similar patterns of change in scores on the Psychosis Evaluation Tool for Common Use by Caregivers (PECC), which measures psychotic symptoms, depression and anxiety, suggesting sensitivity to change as a result of treatment.[89]
	Usability/Acceptability	The IMRS has been used in a pilot trial in both the US and Australia. In the pilot, 22 participants provided feedback on the IMRS and almost all perceived the scale as understandable, useful, respectful, useful, and as covering sufficient material. They also noted that it helped them better manage symptoms and recovery.[90] Cavelti et al (2012) assigned the IMR a positive rating for administrator-friendliness.[81]
Ohio Mental Health Consumer Outcomes System (OMHCOS)	**Validity**	Adult consumer forms A and B include the Symptom Distress Scale, which [91] reports as having adequate discriminant validity with the Beck Depression Inventory. Roth also reports significant correlations between the Making Decisions Empowerment Scale, which is a component of the OMCHOS, and symptom distress and quality of life indicators. No specific data are provided.
	Reliability	Roth[91] reports the following internal consistencies for subscales within the three forms of the OMHCOS: Adult consumer form A (N = 1376-1479) α = 0.77-0.93 (acceptable to excellent); Adult consumer form B (N = 887-888) α = 0.92-0.97 (excellent); and Provider adult Form A (N = 23 540) α = 0.72 (acceptable).
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	The OMHCOS has been used with people with severe mental illness or dual diagnosis and is being used as a standard outcome measure in Ohio's public mental health services.[91]
Peer Outcomes Protocol (POP)	**Validity**	Correlations between the POP scales and five established criterion scales were conducted (no n's provided but N = 100): POP Recovery scale and Social Acceptance Scale, $r = 0.55$ (moderate); POP Personshood scale and Rosenberg Self-Esteem scale, $r = 0.76$ (strong); Pop Recovery scale and Recovery Assessment Scale $r = 0.63$ (strong); Pop Empowerment scale and Empowerment Decision-making scale, $r = 0.46$ (moderate); and Pop Satisfaction scale and CSQ-8 Satisfaction scale, $r = 0.55$ (moderate).[92] These moderate to strong relationships lend support for the concurrent validity of the POP.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Internal consistency (N = 55-100) for the 12 scales ranges from $\alpha = 0.71$ (acceptable: Employment satisfaction) to 0.95 (excellent: Program satisfaction).[92] Of the 12 sub-scales, four have ‘acceptable’, six have ‘good’ and two have ‘excellent’ internal consistency, suggesting overall good internal reliability of the scales. Test-retest reliability over a two-week period (n = 41) ranged from $r = 0.47$ (moderate: Employment satisfaction) to $r = 0.88$ (very good: Health).[92] Of the 12 scales, two have ‘moderate’, eight ‘good’, and two ‘very good’ test-retest reliability, suggesting an overall good level of test-retest reliability for the POP.
	Sensitivity to change	No psychometric research on sensitivity to change was identified.
	Usability/Acceptability	Development of the POP involved consumer groups and their family and friends. The POP is designed for use in outpatient settings, peer-run programs and residential settings and has been used in the US with people with a serious mental illness. Face-to-face interviews are the only acceptable mode of administration and this takes approximately one hour to complete. Some training is necessary for administration and scoring.[92] The POP is free to download.
Recovery Measurement Tool (RMT)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	Developed with a group of consumers in the US. Completion time is about 20 minutes [93]
Recovery Orientation (RO)	**Validity**	NB: As the RO was not named by its developers, it is difficult to determine whether any further psychometric research has been conducted. The RO was developed from items from existing scales measuring various aspects of recovery. Its structure was tested using exploratory and then confirmatory factor analysis to arrive at its four-factor solution.[54]
	Reliability	RO has been shown to have acceptable to excellent internal consistency across its four dimensions ($\alpha = 0.75$-0.92).[54]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Useability/Acceptability	No psychometric research on usability was found.
Recovery Process Inventory (RPI)	**Validity**	Some evidence for the concurrent validity of the RPI has been established by the significant positive weak to moderate correlations ($r = 0.38$-0.55, $P < 0.05$) between the majority of its domains and various subscales of the Mental Health Statistics Improvement Program (MHSIP) Adult Consumer Survey, namely those related to service access, quality and appropriateness and perceived outcomes.[55]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	The RPI has been shown to have fair to good test-retest reliability across the six sub-scales (n = 185) (*r* = 0.36-0.63) over a period of 2 to 4 weeks.[55] The RPI has been shown to have acceptable to good internal consistency (N = 425) (α = 0.71-0.81) across five of its six sub-scales, with the remaining sub-scale having a poor level of internal consistency of α = 0.56 (Other's care/help scale).[55]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Useability/Acceptability	No psychometric research on usability was found.
Milestones of Recovery Scale (MORS)	**Validity**	MORS scores were correlated with scores on the Levels of Utilization System (LOCUS). A correlation of *r* = .49 was pre-set as an acceptable level of reliability. Correlations of *r* = 0.67 to .70 (95% CI, 0.66-0.90) were found for five of the seven LOCUS subscales.[94]
	Reliability	Inter-rater reliability of the MORS was reported as *r* = 0.85 (95% CI, 0.81-0.89), and *r* = 0.86 (95% CI, 0.80-0.90) at two sites, with two to five raters rating on the MORS each client whom they knew well; 49 raters in total rated 431 clients. Test-retest reliability over 10 to 20 days was *r* = .85 (95% CI, 0.81-0.87, N = 381).[94]
	Sensitivity to change	MORS scores collapsed into six stages of recovery were used in a study of participants with schizophrenia, bipolar disorder or depression and who had been homeless, incarcerated or at risk of being homeless or incarcerated, living in The Village in Long Beach, California, where there is a holistic focus on recovery. The MORS scores showed changes indicating increasing recovery over two years, with some cycling between stages (as would be expected).[95]
	Useability/Acceptability	The MORS is designed to measure recovery for administrative purposes, rather than to reflect an individual's process of recovery. Clinicians or other staff who know the consumer well rate recovery. Some concern by consumers that the staff-rated scale is disempowering [96]
Multi-Phase Recovery Scale (MPRM)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Useability/Acceptability	No research on usability was found.
Maryland Assessment of Recovery in People with Serious Mental Illness (MARS)	**Validity**	Principal components analysis and confirmatory factor analysis provide support for a single-factor model, for which all factor loadings were then 0.567 to 0.837.[58] In the initial development study, a single factor measuring self-efficacy and human agency accounted for 59% of variance in MARS scores. Other factors added modest, but non-significant amounts of variance. In the second study (N = 84), the four-component model (with a promax rotation), including the four subscales, accounted for 66.52% of the variance in the

MEASURE	PSYCHOMETRIC PROPERTIES	
		model.[97] In this study, canonical correlation analyses (CCA) showed significant relationships between the MARS subscales and measures of episodic stressors; hospitalisation; coping; community functioning; number of, and satisfaction with, social supports; internalised stigma and psychopathology. Using mediated multiple regression, MARS scores were also shown to fully mediate the relationships between psychopathology and community functioning, and stresses and community life.[97]
	Reliability	In the development study, one week test- retest reliability for the revised 25-item version (N = 25) was reported as excellent (*r* = .898).[58] In the initial development study involving 166 outpatients from two mental health clinics in two American states, the 25-item version of the MARS was reported as having demonstrated excellent internal consistency (α = 0.95).[58] In a second study involving 84 individuals with a diagnosed mental illness who acted as peer support providers to others with a mental illness, the internal consistency of the MARS subscales and full scale was reported as ranging from acceptable to excellent: α = 0.79 (Strengths) to α = 0.92 (Hope/holistic), full scale - excellent (α = 0.95).[97]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Useability/Acceptability	The MARS was developed with the assistance of a panel of mental health consumers and was tested with a sample of consumers with severe mental illness.[58]
The Mental Health Recovery Star (MHRS)	**Validity**	Convergent validity of staff-rated MHRS and staff-ratings on the Life Skills Profile measures of social functioning (n = 140) indicated that scores on the MHRS managing mental health, self-care and living skills had ICCs of 0.7-0.71 (95% CI, 0.61-0.78) (strong agreement) with the total LSP score; managing mental health and self-care MHRS subscales had ICC's of 0.7 and 0.71 (95% CI, 0.61-0.78) (strong agreement) with the LSP – self-care subscale; and social networks and the LSP - social contacts scale also had acceptable convergent validity with an ICC of 0.69 (95% CI, 0.59-0.77). Service users MHRS scores were correlated with scores on the Mental Health Recovery Measure (MHRM). No sub-scales showed an acceptable level of convergence (i.e. *r* >.70). The authors concluded that this pattern of convergence between the MHRS and the LSP and MHRM suggests that the MHRS is more likely to be assessing the social functioning aspect of recovery rather than the personal experience of recovery, as was the aim. (p. 69) An analysis of the convergent validity between the staff-only ratings on the MHRS and collaborative ratings with the service user showed that staff scored clients slightly more negatively alone than when collaborating with the client.[98] All MHRS items correlated at statistically significant levels with the total score (*r* = 0.50-0.77). Exploratory factor analysis showed evidence for a two-factor solution ('*internal management and personal relationships'* and '*external management and external relationships'*) explaining 48% of score variance with little item redundancy. [99]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Staff-only ratings of 138 service users showed two-week test-retest reliability, using Intraclass Correlation Coefficients (ICC) (%95 CI), of between 0.7 and 0.89 (0.61-0.90), showing strong agreement between the two time-points. Inter-rater reliability of the subscales of the MHRS (n = 85) ranged from ICC = 0.53 (95% CI, 0.38-0.69) for the Relationships subscale to ICC = 0.77 (0.67-0.85) for the Work subscale, indicating only a fair to moderate level of agreement between raters across the subscales. This level of inter-rater reliability lead the authors to recommend that the measure not be used as a clinical outcome measure at the current time.[98] The 12-day test-retest reliability of staff-client collaborated MHRS scores was good with ICC's on the MHRS sub-scores and total ranging from 0.71 – 0.82 (95% CI, 0.49-0.90).[98] Dickens et al showed good internal consistency for the MHRS (α = 0.85) in a UK study of 203 participants. The two factors each show either good (Factor 1 α = 0.85) or acceptable (Factor 2 α = 0.73) internal consistency.[99]
	Sensitivity to change	Dickens et al (2012) tested the responsiveness of the MHRS in a sample of 203 participants with moderate to severe mental health problems accessing a variety of UK community mental health services. Data was pulled from a database common to these mental health services. MHRS scores showed small but significant change from baseline to first follow-up (mean = 57 days, SD = 87.2, Cohen's *d* = 0.1- 0.26) and moderate change from baseline to second follow-up (Cohen's *d* = 0.2-0.4), with the exception of the 'responsibilities' domain, which showed no significant change.[99]
	Useability/Acceptability	The MHRS was developed with considerable input and feedback from UK mental health service users and is widely used in mental health services in the UK, where staff receive some training in its use.[99] In a study including service users and staff of UK community and inpatient mental health services, about 70% of staff (N = 125) took less than 30 minutes per consumer to complete the MHRS alone. The remainder took 30 to 60 minutes. The majority thought that it was easy to decide a score on the MHRS, it was easy to use and was useful in care planning. A smaller number thought the MHRS was useful as a clinical outcome measure. For 92 staff and service users who undertook collaborative ratings, relatively equal numbers took less than 30 minutes and between 30 and 60 minutes to complete the MHRS. About two-thirds said it was easy to score and most said it was easy to use. About half of staff said a collaborative score was easier to decide than deciding alone. 85% of consumers thought that the MHRS was useful for helping them and the staff to understand their progress and for helping plan the support they needed.[98] According to Onifade (2011)[100], the MHRS is fast establishing itself as the 'recovery tool of choice' in many UK services and a review of the use of the MHRS in the UK concluded that 27 organisations and 159 projects were using the MHRM at that time.[101]
The Questionnaire about the Process of	**Validity**	Principal components analysis identified two factors with items loading 0.4. Factor 1 accounted for 48% and the second for 6.8% of variance. Authors recommend validation with larger samples. Subscales of the QPR were show to correlate with the majority of subscales from measures of general health, empowerment, and

MEASURE	PSYCHOMETRIC PROPERTIES	
Recovery (QPR)		schizophrenia-quality of life.[60]
	Reliability	Test-rest reliability (n = 43) at about 20 days (SD = 6.65 days) was good to very good: Intrapersonal subscale, r = 0.874, p = .001; Interpersonal subscale, r = 0.768, p = 0.001. Internal consistency: Intrapersonal, α = 0.94 (excellent); Interpersonal, α = 0.77 (acceptable) (N = 111).[60]
	Sensitivity to change	The QPR was used in the UK as a secondary outcome measure in a small exploratory trial (N = 20) of cognitive therapy with people with schizophrenia-spectrum disorders not taking antipsychotic medication. Scores on the QPR increased from pre-treatment (Mean = 48.83, SD = 15.69) to post-treatment (Mean = 57.22, SD = 18.59) and follow-up (Mean = 60.96, SD = 18.8), consistent with increases on measures of psychotic symptoms (including a clinical interview) and functioning, suggesting sensitivity to change.[102]
	Usability/Acceptability	111 participants who had experienced psychosis were recruited via the UK's National Health Service and completed the QPR and other psychometric tests. Items were developed with service users and from in-depth interviews. The majority reported little distress at completing the scale, but those who did were those with lower recovery scores. Developed for a UK reading age of 14 to 16 years.[60]
Subjective Experience of Psychosis Scale (SEPS)	**Validity**	The negative impact of the psychotic experience sub-scale (1(ii)) and the Negative dimensions of the psychotic experiences sub-scale (3(ii) showed a number of significant correlations (n = 75) with PANSS total and subscales, the Psychotic Symptom Rating Scale (PSYRATS), the Process of Recovery Questionnaire (QPR), the Calgary Depression Scale, the Beck Hopelessness and Anxiety Inventories (BHI and BAI) and the Self-Esteem Rating Scale (SERS). Only scale 3(ii) correlated with the GAF. The only other significant correlations between SEPS sub-scales with all of these measures were between the BAI and negative support sub-scale (2(ii)) of the SEPS, and between the SEPS positive dimensions of the psychotic experience (3(i)) subscale and the PSYRATS Delusion scale (r = -.280), the QPR and the SERS.[61]
	Reliability	Intra-class correlations (n = 35) for all sub-scales were statistically significant, although the ICC coefficients ranged from 0.323 to 0.829. Positive and negative items of subscale 1 and subscale 3 did not correlate, although positive and negative items on subscale 2 showed a moderate significant correlation (ρ = 0.364, P < .001). Test-retest correlations at 1 to 2 weeks (n = 35) were moderate to strong (Spearman's ρ = .558 -.827) for all sub-scales except for the positive support component of subscale 2, which showed only a moderate to low correlation between time points (Spearman's ρ = .359).[61] Internal consistency of the scales (n = 100) ranged from questionable to excellent (α = 0.66 – 0.95): sub-scales 1 and 2 – acceptable to excellent, subscale 3 – questionable.[61]
	Sensitivity to change	Change scores for the SEPS were also correlated with change scores on the above measures to determine sensitivity to change. Relatively few correlations were shown, again mostly with subscale 1(ii) – the PANSS,

MEASURE	PSYCHOMETRIC PROPERTIES	
		QPR and SERS, Subscale 2(ii) also correlated with the QPR, and subscale 3(ii) again with PSYRATS Delusions, PANSS General and SERS positive.[61]
	Usability/Acceptability	The SEPS took participants between 5 and 30 minutes to complete. 75% (n = 100) reported no distress on completing the measure, 14% reported a little and the remainder reported moderate or 'quite a lot' distress.
Self-Identified Stages of Recovery (SISR)	**Validity**	The subscales of the SISR have been shown to correlate with the MHRM (subscale A-stages or recovery: Spearman's *Rho* = 0.46, *P* < .01; subscale B–recovery processes: *r* = 0.80, *P* < 0.01); the RAS total (Subscale a: Spearman's *rho* = 0.40, *P* < 0.01; Subscale B: *r* = 0.70, *P* < 0.01), and the majority of items of the SISR also correlate with the RAS subscales (*r* = 0.22-0.66, *P* < .05). The SISR subscales also correlate with the K10 (Subscale A *r* = 0.27, *P* < .05; Subscale B: *r* = 0.42, *P* < .01).[67]
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Useability/Acceptability	No research on usability was found.
Stages of Recovery Instrument (STORI)	**Validity**	The STORI has been shown to have moderate to good concurrent validity. At one extreme, the first subscale (representing the stage of *moratorium*) has been shown to have significant negative correlations with other recovery-related measures such as the Self Identified Stages of Recovery Instrument (SISR), the Recovery Assessment Scale (RAS), the Psychological Well-Being Scales (PWB), the Connor-Davidson Resilience Scale (CD-RISC) and the Adult State Hope Scale (HOPE), as well as with the Mental Health Inventory (MHI-5). At the other extreme, the fifth subscale (representing the stage of *growth*) has been shown to have significant positive correlations with these comparison instruments. The middle subscales (representing the stages of *awareness, preparation* and *rebuilding,* respectively) have generally been shown to have weaker, non-significant correlations with these comparison measures.[62] Most stages of the STORI (not stage 2) also show moderate to strong correlations with the RAS total score.[70] (Psychometric analysis of the STORI identified three clusters, rather than the expected five (representing the five stages of recovery) [62, 70] suggesting that the items do not discriminate sufficiently between stages of recovery.[62]
	Reliability	The internal consistency of the STORI sub-scales has been demonstrated as good to excellent in two separate studies. Andresen et al, in an Australian study (N = 94), reported α = 0.88-0.94, and Weeks et al (2011), in a UK study, reported α = 0.81-0.87 (N = 52).[62, 70] Weeks et al also reported test-retest reliability following a four-minute distractor task (N = 22) was above *r* = 0.9 (very good) for all stages.
	Sensitivity to change	No psychometric research on sensitivity to change was found.

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Useability/Acceptability	In a UK sample of both in- and out-patients from mental health services, most who experienced psychosis (n = 49), consumers mostly rated the STORI as easy to complete, not too long, the items were important to think about and was helpful for staff, although they were not sure that it helped them think about their own recovery. The STORI took an average of 11.4 minutes to complete (SD = 4.86).
Stages of Recovery Scale (SORS)	**Validity**	In a Taiwanese study aimed at developing the SORS,[103] (N = 470) found that overall well-being, regaining autonomy and overall recovery scores on the SORS were significantly lower in a psychiatric rehabilitation sample including consumers in various states of symptom control compared with and a sample living in the community who were socially engaged and had not been hospitalised for one year. Cluster analysis revealed four stages of recovery, with significant differences on all sub-scales between stages, and scores on all sub-scales increasing with progression through the stages of recovery. Cut-off scores for allocation to the stages were established. The overall SORS recovery score and the stages of recovery were correlated with author-developed measures of empowerment, social functioning and life satisfaction. Total recovery showed significant correlations of r = 0.49-0.61. A multiple regression analysis including the three measures, sex and diagnosis as predictors of total SORS recovery scores showed only the three scales to be significant predictors (Beta = 0.15-0.42, P = .000). MANOVA analyses also showed that scores on three scales could differentiate participants allocated to the different stages of recovery. These analyses provide initial evidence for the construct and discriminant validity of the SORS, though this has not yet been tested in an English-speaking sample.
	Reliability	Song and Hsu (2011) (n = 470) reported α = 0.97 for the complete scale and α = 0.80-0.95 for the sub-scales. Test-retest reliability and r = 0.72 (good) (P < 0.05, n = 55) over a 3- to 5-week test-retest period.[103]
	Sensitivity to change	No research on sensitivity to change was found.
	Useability/Acceptability	No research on usability was found.

3.2 Cognition and Emotion

Table 7. Profile of cognition and emotion measures

This is a broad domain that potentially can include thousands of measures. The measures identified here are a combination of those measures identified in the literature and in the survey. The measures currently reported in the survey range from the measurement of specific symptomatology to barriers to help seeking. The measures are arranged so that measures of similar constructs are together:

- Consumer measures:
    - Measures Distress, Anxiety and Depression
    - General Measures of Symptomatology
    - Specific Measures of Symptomatology
    - Measures for Children and Adolescents
    - Measures of cognitive capacity
    - Miscellaneous
- Carer measures.

MEASURE	DATE	COUNTRY	DESCRIPTION
Consumer measures			
Measures of Distress, Anxiety and Depression			
Kessler-10 (K-10)	2002	United States	The K-10 was developed by Kessler and colleagues for use as a measure of nonspecific psychological distress to be incorporated into the United States National Health Interview Survey (NHIS).[104] It is a self-report measure (but can also be administered via telephone or interview) that indicates the amount of time during the last four weeks the respondent has experienced the particular problem. The response scale ranges from 1 = *none of the time* to 5 = *all of the time*.
Mental Health Inventory 38 (MHI-38)	1983	United States	The MHI is a 38-item self-report measure designed to assess general psychological distress and wellbeing in a non-patient population.[105] Each item includes a description of a particular state of mind, which is generally scored on a six-point Likert scale. The respondent is required to indicate the frequency or intensity with which they have experienced this state in the past month.
Behaviour Symptom Identification Scales (BASIS-32®)	1986	United States	The BASIS-32® is a self-report measure developed by Eisen and colleagues at McLean Hospital in the United States as a consumer-oriented measure of symptoms and behavioural distress.[106] The instrument was originally developed and validated among inpatients, but subsequent studies have supported its use in outpatient and residential settings.[106-111] As a measure of mental health outcomes, the BASIS-32® has been widely used for research and quality-improvement purposes.
Depression, Anxiety, Stress Scale (DASS)	1995	Australia	The DASS is a self-report scale with 42 negative emotional symptoms rated on a 4-point Likert (0 to 3) for their severity over the last week. It captures depression, including dysphoria; hopelessness; self-deprecation and anhedonia; anxiety, including physical symptoms and subjective experience of anxiety; and stress, including relaxation difficulties, nervous arousal, agitation and irritability.[112]

MEASURE	DATE	COUNTRY	DESCRIPTION
Beck Depression Inventory (BDI)	1996	United States	The BDI is a self-report measure that was derived from clinical observation about the attitudes and symptoms frequently displayed by depressed psychiatric patients. The 21 symptoms and attitudes covered by the BDI are mood, pessimism, sense of failure, lack of satisfaction, guilt feelings, sense of punishment, self-dislike, self- accusation, suicidal wishes, crying, irritability, social withdrawal, indecisiveness, distortion of body image, work inhibition, sleep disturbance, fatigability, loss of appetite, weight loss, somatic preoccupation and loss of libido. Originally designed to be administered by a trained interviewer, the BDI is most often used as a self-report measure, which takes 5 to 10 minutes to complete. A total score is given by summing each of the 21 items rated on a 4-point Likert scale (0 to 3). In brief, the 21-item BDI-IA items were revised to correspond to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) mood disorder symptoms. Four BDI-IA items were replaced by four new items (agitation, worthlessness, loss of energy, and concentration difficulty), and the response options for 14 old items were revised. The original 1-week time frame for rating each item was revised to 2 weeks to be consistent with the DSM-IV. The total scale severity score for the BDI-II is obtained by summing the highest rating for each item. The total severity score ranges from 0 to 63. Cut-off scores are suggested, with scores of 0 to 13 being ‘minimal’ severity of depression, 14 to 19 being ‘mild’, 20 to 28 being ‘moderate’, and 29 to 63 being a ‘severe’ level of depression.
Beck Anxiety Inventory (BAI)	1988	United States	The BAI is a self-report measure that was created from three pre-existing scales: the Anxiety Checklist, the Physician's Desk Reference Checklist, and the Situational Anxiety Checklist.[113] The inventory consists of 21 items each describing a common symptom of anxiety. Respondents are asked to rate how frequently they have been affected by each symptom over the last week on a 4-point scale ranging from 0 = *not at all* to 3 = *Severely–I could barely stand it*. A total score is obtained by summing all items, and scores range from 0 to 63.[113]
Hamilton Depression Rating Scale (HAM-D)	1960	United Kingdom	The HAM-D is a clinician-completed set of scales developed for use with people diagnosed with a depressive disorder and to quantify the results of a psychiatric interview. The author indicates that the accuracy of the measure relies entirely on the “skills of the interviewer” [114 p, 56] The scale consists of seventeen variables: Depressed mood, Guilt, Suicide, Insomnia, initial, middle, delayed, Work and interests, Retardation, Agitation, Anxiety psychic, Anxiety somatic, Somatic symptoms gastrointestinal, Somatic symptoms general, Genital symptoms, Hypochondriasis, Loss of insight, Loss of weight. The interviewer rates these variables on a five point scale 0 Absent, I Mild or trivial, 2 or 3 Moderate and 4 Severe.

MEASURE	DATE	COUNTRY	DESCRIPTION
Hamilton Anxiety Rating Scale (HAM-A)	1959	United Kingdom	The HAM-A is a clinician-completed set of scales that was originally developed for people already suffering from 'neurotic' anxiety states. It is an interviewer-rated measure that covers 13 symptoms of anxiety including anxious mood; tension; fears; insomnia; cognitive changes; depression; and somatic symptoms including cardiovascular, respiratory, gastrointestinal, genito-urinary and general autonomic symptoms (headaches and sweating). The interviewers rate consumer symptoms on a 5-point scale: 0 = *none*, 1 = *mild*, 2 = *moderate*, 3 = *severe* and 4 = *very severe, grossly disabling*.[115]
State Trait Anxiety Inventory (STAI)	1983	United States	The STAI consists of two 20-item self-report measures. The STAI-State assesses how respondents *feel right now, at this moment* (e.g., *I feel at ease*, *I feel upset*), and the STAI-Trait targets how respondents *generally feel* (e.g., *I am a steady person*, *I lack self-confidence*). Respondents are asked to rate themselves on each item on the basis of a 4-point Likert scale, ranging from *not at all* to *very much* so for the STAI-State and from *almost never* to *almost always* for the STAI-Trait.[116]
Edinburgh Postnatal Depression Scale (EPDS)	1986	United Kingdom	The EPDS is a 10-item self-report questionnaire that has the mother describe how she has been feeling in the last 7 days. Responses are scored from 0 to 3 according to increased severity of symptoms, with a variety of anchors varying across questions but generally from 0 = *no, never* to 3 = *Yes, most of the time*. Seven of the ten items are reversed scored [117].
Geriatric Depression Scale (GDS)	1982	United States	The GDS was developed to identify depression in elderly people. It is a self-report that has elderly people agree or disagree (*yes/no*) originally with 30 statements expressing common manifestations of depression in later life.[118] These manifestations include lowered affect, inactivity, irritability, withdrawal, distressing thoughts, and negative judgments about past, present, and future. A short form of 15 items has been developed.[119] Aikman and Oehlert found the short form just as accurate in identifying depression as the long form of the measure.[120]
General Measures of Symptomatology			
Symptom Checklist 90 (SCL 90)	1976	United States	The Symptom Checklist [121] is a self-report measure originally designed for use in medical, clinical and non-clinical samples and based on the Hopkins Symptom Checklist[122]. The revised 90-Revised (SCL-90-R)[123] is a multidimensional self-report symptom inventory. The SCL-90-R consists of 90 items, each using five points from 0 = *no problem* to 4 = *severe/very severe*, yielding a total score from 0 to 360, called the Global Severity Index (GSI). The 90 items are intended to cover 10 areas of mental symptoms: somatization (12 items), obsessive–compulsive (10 items), interpersonal sensitivity (9 items), depression (13 items), anxiety (10 items), anger–hostility (6 items), phobic anxiety (7 items), paranoid ideation (6 items), psychoticism (10 items), and additional scales (7 items). The SCL-90-R is rated by the patient in relation to problems experienced during the last 7 days.[124]

MEASURE	DATE	COUNTRY	DESCRIPTION
General Health Questionnaire-12 (GHQ-12)	1979	United Kingdom	The GHQ is a self-administered screening questionnaire originally designed for use by general practitioners with the aim of detecting diagnosable psychiatric disorders.[125] There are several versions. The original comprehensive 60-item version (GHQ-60) has been reduced to a 30-item version (GHQ-30)[126], a 28-item version[125], a 20-item version and 12-item version[127]. Respondents indicate whether a symptom or feeling has increased, decreased or is unchanged using four response options: *Not at all*, *Same as usual, Rather more than usual* and *Much more than usual*. A total score is calculated by summing these ratings and total scores, depending on which version of the measure is being used. However, there are three approaches to scoring the GHQ known as Classic, the Corrected, and the Likert approaches.[128] The Classic approach (0–0–1–1) transforms the scale into a bimodal symptom presence scale by recoding scores of 0 and 1 as 0, and scores of 2 and 3 as 1.[129] The Corrected approach adopts the classic approach except for those items seen as reflecting chronic conditions where three of the four items are scored 1 (i.e. 0,1,1,1 for negatively worded items and 1,1,1,0 for positively worded items). The Likert approach involves attaching an ordinal scale to each item rating from 0 to 3.
Sphere-12	2001	Australia	The SPHERE-12 is a self-report measure that has two components: one scale has 6 items derived from the GHQ, which measure aspects of depression and anxiety; the other scale of 6 items measures fatigue and is made up of items from the Symptoms of Fatigue and Anergia (SOFA) scales.[130, 131] . The sphere is rated on a 3 point scale 0 never or some of the time, 1 a good part of the time, 3 most of the time.
My Mood Monitor (M-3)	2010	United States	The M-3 is a 27-item web-based tool that was developed to screen for psychiatric illness in primary care.[132, 133] It is rated on a 5-point Likert scale from 0 = *not at all* to 4 = *most of the time*. The first 19 questions ask about thoughts feelings and behaviours over the last week or more. The next four questions look at lifetime incidence of irritability, excitement and excess energy requiring little sleep. The final four questions look at the impact of these symptoms regarding role functioning or drug or alcohol use.
Specific Measures of Symptomatology			
Suicidal Ideation Questionnaire (SIQ)	1987	United States	The SIQ is a 30-item self-report inventory that assesses the frequency (on a 7-point scale ranging from *almost every day* to *never had the thought*) of a hierarchy of thoughts of suicide, ranging from possible thoughts of wishing one were not alive to specific and serious thoughts of actively killing oneself.[134] This original version was created for high school students, a 15 item “junior” version is also available for child and adolescents aged 11 to 15[135] and there is also a 25-item Adult version.

MEASURE	DATE	COUNTRY	DESCRIPTION
Columbia–Suicide Severity Rating Scale (C-SSRS)	2011	United States	The C-SSRS is a self-report measure designed to distinguish the domains of suicidal ideation and suicidal behavior. Four constructs are measured: (1) severity of ideation is rated on a 5-point ordinal scale in which 1 = *wish to be dead*, 2 = *nonspecific active suicidal thoughts*, 3 = *suicidal thoughts with methods*, 4 = *suicidal intent* and 5 = *suicidal intent with plan*; (2) the intensity of ideation scale comprises 5 items, each rated on a 5-point ordinal scale for frequency, duration, controllability, deterrents and reason for ideation; (3) the behaviour subscale is rated on a nominal scale that includes actual, aborted, and interrupted attempts, preparatory behaviour and non-suicidal self-injurious behaviour; (4) the lethality subscale assesses actual attempts and is rated on a 6-point ordinal scale - if actual lethality is zero then potential lethality of attempts is rated on a 3-point ordinal scale.[136]
Inventory of Complicated Grief-Revised (ICG-R)	1995	United States	The ICG originally consisted of 19 items. The ICG-R consists of 15 items focusing on separation distress and traumatic distress. It is rated on a 5-point Likert scale with respondents indicating how frequently, from 0 = *never* to 4 = *always*, they experienced the various emotional, cognitive and behaviours states described by the inventory (maximum range 15–75).[137, 138]
PTSD Checklist (PCL-C)	1991	United States	The PCL-C is a 17-item self-report measure of the 17 DSM-IV symptoms of PTSD. The PCL-C has a variety of purposes, including screening individuals for PTSD, diagnosing PTSD, and monitoring symptom change during and after treatment. It comes in three versions. Using a 5-point scale (1 = *not at all*, 5 = *extremely*), respondents indicate how much they have been bothered by each symptom in the past month.[139, 140]
Dissociative Experiences Scale (DES)	1986	United States	The DES is a 28-item self-report rating of dissociative experiences.[141] It quantifies the frequency and intensity of a wide range of experiences relevant to factors representing general dissociative tendencies [142]. Subjects respond on a scale from 0 to 100, with increments of 10. The DES-II score is an index of the average frequency of dissociative experiences (range 0–100)
Measures for Children and Adolescents			
Strengths and Difficulties Questionnaire (SDQ)	1997	United Kingdom	The SDQ was developed by Goodman in the UK as a brief screening tool that describes children's and adolescents' behaviours, emotions and relationships. It was designed to address some of the shortcomings of other similar instruments. That is, to be concise; to be applicable to both children and adolescents; to have versions that could be completed by both parents and teachers, as well as a self-report version (for older children/adolescents); to represent both negative and positive attributes; and to equally cover five dimensions (namely conduct problems, emotional symptoms, hyperactivity, peer relationships and prosocial behaviour).[143] In each version, the parent, teacher or child/adolescent is asked to consider the child's/adolescent's behaviour over the past six months and to consider whether the statement is *not true*, *somewhat true* or *certainly true*. A modification was made for the purposes of using the instrument at follow-up, such that any SDQ collected at baseline considers the previous six months and any follow-up SDQ

MEASURE	DATE	COUNTRY	DESCRIPTION
			considers the past month only. For the majority of items, a response of *not true* is scored 0, *somewhat true* as 1 and *certainly true* as 2. The exceptions are Items 7, 11, 14, 21 and 25, where the reverse scoring order applies. This scoring method yields a score on each scale of 0 to, and a total difficulties score of 0 to 40, generated by summing the scores from all of the scales except the Prosocial behaviour scale.[142, 144]
Child Behaviour Checklist (CBCL)	1983	United States	The first version of the CBCL was created in 1983[145] but has since gone through a number of developments. CBCL consists of 118 items and is a parent- and youth-report measure. It measures behavioural and emotional problems over the last 6 months. The CBCL contains eight narrow-band scales (Anxious/Depressed, Withdrawn, Somatic Complaints, Social Problems, Delinquent Behavior, Thought Problems, Attention Problems and Aggressive Behavior) and two broad-band scales (Internalizing Problems, Externalizing Problems). A Total Score can also be calculated. It can take 15 to 17 minutes to complete.[16]
Devereux Early Childhood Assessment (DECA)	1999	United States	The DECA evaluates within-child protective factors for children aged 2 to 5 and consists of 37 items that include two composite scales: Total Protective Factors and Behavioral Concerns. The Total Protective Factors Scale consists of 27 items that assess children's positive behaviors, and the Behavior Concerns Scale consists of 10 items that assess children's social and emotional problems.[146] Items all begin with the phrase: *During the past 4 weeks, how often did the child . . .;* that is then followed by a question about an observed behaviour. Items are then rated on a 5-point Likert-type scale ranging from *never* (coded as 0) to *very frequently* (coded as 4). Reports on administration time vary from 5 to 10 minutes[146], to 15 to 20 minutes.[147]
Anxiety Disorders Interview Schedules-Children (ADIS-C)	1988	United States=	The ADIS-C are semi-structured interviews originally developed around the DSM-111 and DSM-111-R [148] and updated for the DSM-IV[149]. They are completed by a clinician for child and parents. Detailed questions focus on the situational and cognitive cues for anxiety, intensity of anxiety, extent of avoidance, precipitating events, and history of the problem. For both the child and parent interviews, the total number of *yes* responses is calculated to obtain a total symptom scale score and to determine whether the total number of symptoms endorsed as *yes* is enough to meet the number of symptoms required to meet DSM-IV criteria. The schedule includes a Clinician Severity Rating (CSR) ranging from 0 to 8 indicating impairment: 4 or higher indicates a clinically significant level of impairment meeting DSM-IV criteria. The schedules require the use of clinical judgement to establish the presence or absence of criteria.

MEASURE	DATE	COUNTRY	DESCRIPTION
Parenting Stress Index (PSI)	1983	United States	The PSI was developed to facilitate early identification of children at risk for development of behavioural or emotional disturbance. It is a screening and diagnostic instrument that measures the relative magnitude of stress in the parent-child system[150, 151] Abidin [151] conceived of parental stress as a complex construct that represents a combination of parent, child, and family characteristics as they relate to a person's appraisal of his or her role as a parent. Parental distress, child difficulty, and parent–child dysfunctional interactions lead to more negative parenting, which in turn has a direct impact on a child's behavior. The original (full length) PSI consists of 101 items distributed across 13 subscales (plus an optional 19-item Life Stress Scale). Subsequently, a short form was created (the Parenting Stress Index-Short Form (PSI-SF)), which consists of 36 items and was intended as a brief assessment.
Social Responsiveness Scale (SRS)	1995	United States	The SRS is a 65-item rating scale that measures symptoms indicative of autism spectrum disorders.[152] It is a parent- and/or teacher-report measure of children's social impairments in naturalistic social settings that can be completed in 15 minutes. Questions are the same in both the parent and teacher form of the SRS. It generates scale scores for specific symptom domains (social awareness, social cognition, social communication, social motivation, and autistic mannerisms), as well as a total score that serves as an index of severity of social deficits in the autism spectrum. Higher total scores on the SRS indicate greater severity of social impairment. Total scores are converted to a T-score. As Constantino and Gruber [152] found strong differences between groups related to gender and whether teachers or parents made the ratings, T-scores are scaled in order to achieve equivalency for males and females and for parent and teacher ratings. T-scores less than 59 are considered to be in the normal range, T-scores of 60 to 75 are suggestive of clinically significant social impairment in the mild-moderate range, and T-scores of 76 or above indicate severe social impairment. Therefore, T-scores above 60 are suggestive of an autism spectrum disorder diagnosis.
Barriers to Adolescents Seeking Help Scale (BASH)	1997	United States	The BASH is a self-report measure specifically developed to create a psychometrically sound instrument that could be used with diverse populations and settings to formally assess barriers to adolescent help seeking.[153] Across fourteen categories that represent potential barriers to adolescent help seeking behaviour (affordability, alienation, confidentiality, family as sufficient to help, knowledge of resources, locus of control, peers as sufficient, perception of therapist, self-awareness, self-perception, self-sufficiency, stigma, time available and usefulness of therapy), 37 items ask respondents to self-report their agreement or disagreement with each item on a scale from 1 = *strongly agree* to 6 = *strongly disagree*. Higher scores indicate more resistance to seeking help. An abbreviated measure, the BASH-B, has been created which comprises 11 of the original 37 self-report items that specifically target belief-based barriers to seeking professional psychological help.[154]

MEASURE	DATE	COUNTRY	DESCRIPTION
Measures of cognitive capacity			
Mini Mental State Examination (MMSE)	1975	United States	The MMSE contains eleven questions, and the authors indicate that it requires only 5 to 10 minutes to administer and is therefore practical to use serially and routinely.[155] The examination is divided into two sections, the first of which requires vocal responses only and covers orientation, memory, and attention; the maximum score is 21. The second part tests ability to name, follow verbal and written commands, write a sentence spontaneously, and copy a complex polygon. Because of the reading and writing involved in Part II, some people with language or sight issues may have difficulty. The maximum total score is 30 and the examination is not timed.
Cambridge Cognitive Examination (CAMCOG)	1986	United Kingdom	The CAMCOG is the cognitive section of the Cambridge Examination for Mental Disorders of the Elderly (CAMDEX).[156] This cognitive section is the MMSE (above)[155] and the Abbreviated Mental Test Score (AMTS)[157]. The measure is derived from 60 items of the 68-item CAMDEX and a maximum score of 107 is possible.[158] Omitted from the CAMCOG score are naming two objects, registration and recall of three words, writing a sentence, and paper folding. The CAMCOG assess memory, orientation, praxis, perception, abstract thinking, language and calculation. It takes approximately 20 minutes to complete.
Rowland Universal Dementia Assessment Scale (RUDAS)	2004	Australia	The RUDAS was developed in a multicultural community in Australia. It takes only 10 minutes to administer. The items of the Memory (4-item grocery list) each correct response scored one point (maximum of four points). If a prompt was used, the maximum score was three, Gnosis (Body orientation, Fist/palm alternating task) commands should continue only until five were completed correctly (for a maximum score of five), Praxis Visuo-spatial drawing (Cube copying) scoring a maximum of 3 points for each of the cube characteristics, based on a square, all internal lines drawn, all external lines drawn, Judgement (Crossing the road) Each of the two necessary components (looking for traffic and safety) is scored out of two points. Any component which is prompted scores a maximum of one point. and Language (Animal generation) a maximum score of eight points for the number of new animals named in one minute [159]

MEASURE	DATE	COUNTRY	DESCRIPTION
Miscellaneous			
Warwick-Edinburgh Mental Well-being Scale (WEMWBS)	2007	United Kingdom	WEMWBS is a measure of mental well-being focusing entirely on positive aspects of mental health.[160] It was developed to monitor mental well-being at a population level and evaluate health promotion activities. A short version has been created.[161] It consists of 14 items as statements covering both hedonic wellbeing (positive affect, feelings of optimism, cheerfulness, relaxation) and eudaimonic, psychological aspects of wellbeing, including satisfying interpersonal relationships and positive functioning (energy, clear thinking, self-acceptance, personal development, competence and autonomy). Responders indicate what best describes their experience of each statement over the past two weeks using a 5-point Likert scale (*none of the time, rarely, some of the time, often, all of the time*). The Likert scale represents a score for each item from 1 to 5 respectively, giving a minimum score of 14 and maximum score of 70. All items are scored positively.
Working Alliance Inventory (WAI)	1979	United States	The WAI was developed to measure the therapeutic alliance between clinician and consumer and is based on Bordin's [162] pantheoretical tripartite model of bonds, goals and task[163]. It was originally a 36-item questionnaire but a short form of 12 items was developed:[164] four items for each of bonds, goals and task) are rated on a 5-point Likert scale ranging from 1 = 'never' to 5 = 'always'.
Empowerment Scale	1997	United States	A 28-item measure of the personal construct of empowerment as defined by consumers of mental health services. These 28 items are rated on a four-point Likert scale ranging from *strongly agree* to *strongly disagree*.[165] The authors believe the measure sets a framework for a clearer understanding of the imprecise and overused concept of empowerment
Brief Coping Orientation to Problems Experienced (Brief COPE)	1997	United States	Originally the COPE was developed to study the concept of coping with stressful life events. The Brief COPE was designed to be used in services because this is where psychologically impactful and highly meaningful stresses most often appear. It is a self-report measure that assesses both adaptive and dysfunctional coping strategies. It originally was 60 items long but has been reduced to 28 items.[166] The measure contains two items for each of 14 domains (active coping, planning, positive reframing, acceptance, humour, religion, using emotional support, using instrumental support, self-distraction, denial, venting, substance, behaviour disengagement, self-blame). Consumers rate their coping on a 4-point scale from 0 = *I haven't been doing this at all* to 3 = *I've been doing this a lot*.
Revised NEO Personality Inventory (NEO-PI-R)	1992	United States	The NEO-PI-R has 181 items and comes in two forms: "S" for self- reports and "R" for observer ratings. Items are answered along a 5-point Likert scale from *strongly disagree* to *strongly agree*. The inventory captures information on the five-factor model of personality that includes Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. [167] An abbreviated short-form comprising 60 items is also available (NEO-FFI).[167]

MEASURE	DATE	COUNTRY	DESCRIPTION
General Help Seeking Questionnaire (GHSQ)	2001	Australia	The GHSQ [168] is a self-report measure. It asks questions using a standard problem probe *If you were having [problem type], how likely is it that you would seek help from the following people?* Participants respond to each problem type by rating their help-seeking intentions on a 7-point scale ranging from 1 = *extremely unlikely* to 7 = *extremely likely* for each help source option.[154] The number of items will vary depending on the number of problem types and help source options that are being examined.
Carer Measures			
Involvement Evaluation Questionnaire (IEQ)	1987	Netherlands	The purpose of the IEQ is to measure levels of burden and distress among family carers of people with a severe mental illness. Studies have investigated the IEQ in the context of both schizophrenia and depression. The standard IEQ is a self-report 33-item instrument. The first 29 items aggregate into four distinct sub-scales, and the last four items form a 'generic' sub-scale, as observed in the previous four weeks. It is a self-report measure. Responses to all items are scored on a 5-point Likert scale, *Never, Sometimes, Regularly, Often and Always*, scored 0, 1, 2, 3 and 4 respectively.
Burden Assessment Scale (BAS)	1994	United States	The BAS was developed in the United States to measure both objective and subjective consequences of providing ongoing care to people with severe mental illness, predominantly schizophrenia. It is an interviewer-administered measure. Respondents are required to indicate whether they have experienced each of the types of burden *Not at all, A little, Some* or *A lot* in the past four weeks. These are scored 1, 2, 3, and 4 respectively.
Zarit Burden Scale (ZBS)	1980	United States	The ZBS is a 22-item self-report questionnaire of carer burden.[169] Areas assessed include the perceived health and psychological wellbeing of the caregiver, financial impact, social life, and relationship between the caregiver and the person with impairments. It is rated on a 5-point Likert scale where carers rate the frequency with which they experience stressful aspects of care, ranging from 0 = *never* to 4 = *nearly always*. Higher scores indicate greater impairment.
CarerQol- 7D+VAS	2005	Netherlands	The CarerQol 7D+VAS is a 7-item (plus visual analogue scale) self-report measure[170] with two components. The burden component assesses seven burden dimensions, and the valuation component uses a visual analogue scale to ascertain level of happiness. The burden component requires respondents to indicate whether they have *no, some* or *a lot* of problems or fulfilment/support regarding the given dimension. For problem-related items, *no, some* and *a lot* are scored 3, 2, and 1, respectively. For fulfilment/support items, reverse scoring applies, so a high score indicates a high level of burden. The valuation component allows for the calculation of a score from 0 = *completely unhappy* to 10 = *completely happy*.

Table 8. Psychometric properties of identified cognitions and emotion instruments

MEASURE	PSYCHOMETRIC PROPERTIES	
Consumer measures		
Measures of Distress, Anxiety and Depression		
Kessler-10 (K-10)	**Validity**	Strong content validity was a focus of the expert advisory panel set up to construct the measure.[104] K-10 scores have been shown to be significantly correlated with other instruments that measure symptomatology and/or disability, including the General Health Questionnaire (GHQ)[171], the Short Form 12 (SF-12)[171], the Comprehensive International Diagnostic Interview – Short Form (CIDI-SF)[172] and the World Health Organization Disability Assessment Schedule (WHO-DAS)[172]. Brookes et al [173] found a four factor structure, Negative Affect, Nervous, Agitation, and Fatigue. These factors are highly inter-correlated with the strongest correlation between negative affect and fatigue ($r = .83$), and the smallest (although still substantial) is between fatigue and agitation ($r = .62$).
	Reliability	The test-retest reliability of the K-10 was examined in pilot work undertaken prior to the conduct of computer-assisted telephone interviewing (CATI) health surveys in various Australian states. The individual items demonstrated only fair reliability, but the total score and the standardised total score demonstrated excellent reliability[174] Other studies have found similar results[104]. The concept of inter-rater reliability is not relevant for the K-10, since it is designed to be self-administered so there is only one rater.
	Sensitivity to change	The K-10 has demonstrated sensitivity to change in a study involving psychiatric inpatients. In this study the reliable change index for the K-10 was given as 7.58, indicating a 7-point change being needed to give 95% confidence that a reliable change occurred in the patient over the course of their inpatient stay. A reliable improvement was evident in 39% (n=27) of the patients and a reliable deterioration in 10% (n=7). A further 15 patients showed some improvements, 12 showed some deterioration and nine showed no change. [175]
	Acceptability/Usability	No published studies were found that have specifically examined the feasibility of the K-10 as a routine outcome measure or that have considered the acceptability and utility of the instrument from the perspective of consumers, carers, clinicians and/or managers. However, its brevity augurs well for ease of use and the K-10 is widely used in a range of settings, suggesting that many relevant stakeholders favour its use. In Australia, it has been incorporated into population surveys that have been conducted nationally [176, 177] and at a state/territory level [178-181] (although only the New South Wales Health Survey [182, 183] has incorporated the K10+). In addition, it is being widely used in primary care settings in Australia, where, as one of the key outcome measures of choice in the Better Outcomes in Mental Health Care initiative, it has received a positive response from GPs.[184, 185]

MEASURE	PSYCHOMETRIC PROPERTIES	
Mental Health Inventory 38 (MHI-38)	**Validity**	There is evidence of concurrent validity with studies showing the MHI correlates with other instruments in the predicated direction; for example, the Brief Pain Inventory (BPI)[186], the Medical Outcome Study Social Support Survey (MOS-SSS)[187], the Life Experience Survey (LES)[187], the Systems of Belief Inventory (SBI)[187], the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC)[188], the Positive and Negative Affect Schedule (PANAS)[189], the Dyadic Adjustment Scale (DAS)[189], the Purpose in Life Test (PIL)[190], the Life Regard Index (LRI)[190], the Sense of Coherence Scale (SOC)[190], the MOS 36-item Short-Form Health Survey (SF-36)[191] and the Behaviour and Symptom Identification Scale (BASIS-32®)[191]. Numerous studies have examined the internal consistency of the full-length MHI, as measured by Cronbach's alpha. In these studies, Cronbach's alpha has ranged from 0.63 to 0.93 for the subscales, 0.90-0.97 for the global scales and 0.93 to 0.97 for the total score[105, 191, 192]. Together, these studies indicate that the MHI has a high internal consistency. A number of studies have examined the structure of the MHI, considering how well the observed data fit the original model proposed by Veit and Ware [105]. Most have found support for the original structure.[193, 194]. However, some have only found support for the two- factor model (i.e., the global scores) [195, 196]. Others have found a poor fit across all models and/or suggested alternative solutions.[189] [197]
	Reliability	The internal consistency as measured by Cronbach's alpha has ranged from 0.63 to 0.93 (questionable to excellent) for the subscales, 0.90 to 0.97 (excellent) for the global scales and 0.93 to 0.97 (excellent) for the total score.[198] Generally, the test-retest reliability of the MHI has been reported as adequately high for subscale, global scale and total scores. Studies have reported correlations ranging from 0.56 to 0.97, indicating that the MHI is relatively stable over time.[198].
	Sensitivity to change	The sensitivity to change of the MHI has been tested by comparing consumer's global reports of mental health improvement, stability or deterioration with changes in MHI scores.[191]. Those who rated themselves as having improved showed corresponding improvements on the majority of MHI subscales, those who rated themselves as stable showed no change on the MHI, and those who rated themselves as having deteriorated showed a decline on the MHI.
	Acceptability/Usability	Stedman et al[191] asked consumers to rate the MHI's utility, using a quantitative scale that defined utility in terms of perceived relevance, effectiveness and usefulness. The MHI scored higher than either the BASIS-32 or the SF-36. Stedman and colleagues have also qualitatively explored the opinions of consumers and other stakeholders regarding the feasibility and utility of the MHI and have found them to report that it is comprehensive, easy to understand, user-friendly, acceptable and appropriate.

MEASURE	PSYCHOMETRIC PROPERTIES	
Behaviour Symptom Identification Scales (BASIS-32®)	**Validity**	Content validity has been assessed by soliciting comments from users on the instrument following its administration. Eisen et al[198] reported that the instrument was comprehensive. Other studies have had mixed responses. Concerns identified in the work by Cameron et al[198], Graham et al[199] and Stedman et al[191] related to ambiguous and complex language, an exclusive focus on difficulties, and issues with content areas. The BASIS-32® has been shown to correlate in the predicted direction with the following instruments: the Client Assessment of Strengths Interests and Goals – Self Report (CASIG-SR),[200] the Client Assessment of Strengths Interests and Goals (CASIG),[201] the Outcome Questionnaire (OQ-45),[202] the Short Form Health Status Profile (SF-36),[111, 191], [203] the Symptom Checklist (SCL-90),[107, 204] the Child and Adolescent Functional Assessment Scale (CAFAS),[107] the Hopkins Symptom Checklist (MSCL-43),[108] the Brief Psychiatric Rating Scale (BPRS),[108] the Mental Health Inventory (MHI),[191] the Camberwell Assessment of Need (CAN),[205] the Psychiatric Symptom Assessment Scale (PSAS),[109] the Outcome Assessment Program (OAP) Questionnaire,[206] the Health of the Nation Outcome Scales (HoNOS),[191] and Lehman's Quality of Life Interview (QOLI).[109] By contrast, performance on the BASIS-32® has been found to be unrelated to performance on the Social and Occupational Functioning Assessment Scale (SOFAS-Revised GAF),[109] the Mini Mental Status Examination (MMSE),[109] the Levels of Recovery from Psychotic Disorders Scale (LORS),[207] the Global Assessment of Relational Functioning Scale (GARF),[208] the Life Skills Profile (LSP)[191] and the Role Functioning Scale (RFS).[191] Studies conducted by Eisen and colleagues,[111, 203] and by Hoffman et al,[107] Russo et al,[109] and Chow et al [209] examined the subscale structure of the BASIS-32® and how well the observed data fit the five-factor model derived from the original sample of inpatients.212. However Eisen et al [75] suggested that the factor structure may not generalise well to ambulatory care settings, citing this as one of the reasons for the development of the BASIS-24®.
	Reliability	Only a few studies have examined the test-retest reliability of the BASIS-32®, but their findings are uniformly positive, suggesting that the instrument produces consistent results when rated by the same rater at different points in time.[191]
	Sensitivity to change	Several studies have assessed the sensitivity of the BASIS-32® to change by following consumers over time and comparing pre- and post-treatment ratings with independent pre- and post- measures of improvement, deterioration or stability. In general, these studies have found the BASIS-32® to be highly sensitive to change following treatment in both inpatient and outpatient samples. [107, 109, 111, 191, 203-205, 210, 211], The Impulsive and addictive behaviours subscale and the Psychosis subscale showed poorer sensitivity to change than the other three subscales.[109]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Acceptability/Usability	Empirical studies and published commentaries suggest that the BASIS-32® demonstrates adequate feasibility and utility. Stedman et al[191], for example, found that it was rated favourably by consumers in terms of its utility (defined in terms of its perceived relevance, effectiveness and usefulness), although it was not ranked as highly as the MHI. Others have found it to be applicable to a wide range of people receiving mental health treatment, not limited by diagnoses, symptom patterns or treatment setting, user-friendly, adaptable due to its alternative administration modes. It has also been found to place minimal burden on staff due to its brevity, the simplicity of its design and its absence of training requirements.214,
Depression, Anxiety, Stress Scale (DASS)	**Validity**	DASS scores have been shown to discriminate between anxious and depressed patients.[198] Concurrent validity has been demonstrated with high correlations between the DASS and other measures of anxiety and depression, such as the Beck Depression Inventory (BDI), the Beck Anxiety Inventory (BAI) and the State-Trait Anxiety Inventory STAI-T. [212] The DASS has demonstrated construct validity with Crawford et al finding [213]that the DASS consisted of three correlated factors corresponding to the depression, anxiety and stress scales.
	Reliability	Internal consistency is reported as good to excellent for the general population (α = 0.89-Anxiety, 0.94-Depression, 0.93- Stress and 0.96-total score) [214] and similarly good to excellent for a clinical population (0.89-Anxiety, 0.96-Depression, 0.93-Stress.[215]
	Sensitivity to change	The DASS has demonstrated sensitivity to change, in a study of acceptance and commitment therapy pre- post treatment scores for depression $t(53) = 6.74$; $p < .001$, Anxiety $t(53) = 4.99$; $p < .001$ and Stress $t(53) = 5.52$; $p < .001$.were all statistically significantly different [216]. Page et al [217]found the DASS sensitive to changes from admission to discharge from inpatient mental health units with statistically significant changes for Depression, $t(53)=6.74$; $p<.001$, Anxiety, $t(53)=4.99$; $p<.001$, and Stress, $t(53)=5.52$; $p<.001$.
	Acceptability/Usability	Valued characteristics of outcome measure instruments suitable for widespread use include brevity, low cost, self-report format, results that can be readily discussed with the patient and responsiveness to change in clinical status, the Depression Anxiety Stress Scales (DASS) seems to possess such characteristics[218].

MEASURE	PSYCHOMETRIC PROPERTIES	
Beck Depression Inventory (BDI)	**Validity**	The BDI-II has demonstrated excellent convergent and divergent validity with other measure such as the Hamilton Psychiatric Rating Scale for Depression, the Scale for Suicidal Ideation, and the Hamilton Rating Scale for Anxiety.[219] In people diagnosed with schizophrenia, the BDI-II has been found to correlate with other measures of depression including The Brief Psychiatric Rating Scale—Depression subscale (BPRS-D), Positive and Negative Syndrome Scale—Depression subscale (PANSS-D), Hamilton Rating Scale for Depression (HAMD), Montgomery Asberg Depression Rating Scale (MADRS) and Calgary Depression Scale for Schizophrenia (CDSS).[220] The BDI-II total scores correlate moderately and significantly with scores on self-report measures of hopelessness (r = 0.63), anxiety (r = 0.53) and suicide-related behaviours (r = 0.57).[221] In older persons, the BDI-II was significantly, positively, and strongly correlated with the depression subscales of the Collidge Axis II (CATI) (r = 0.66) and the Center for Epidemiologic Studies Depression Scales (CES-D) (r = .69).[222] Discriminant validity has been demonstrated through moderate negative correlations between the BDI-II and the A (Outgoingness), C (Ego Strength) and H (Boldness) scales of the 16PF questionnaire[223]. BDI-II scores were correlated to various scales of the Adolescent Psychopathology Scales (APS) in the expected direction with higher correlations with regards to Generalized Anxiety Disorder, Major Depression, Dysthymic Disorder, Social Phobia and lower correlations with Conduct Disorder, Oppositional Defiant Disorder, Substance Abuse, and Attention Deficit Disorder [224]. Factor analyses have identified three factors: a cognitive-affective dimension, the overt emotional upset, and the somatic–vegetative dimension of depressive symptomatology.[219, 225]
	Reliability	The measure has demonstrated high internal consistency, ranging from 0.83[226] and 0.86 [222] (good) and 0.92 (excellent).[221] The BDI has demonstrated good test-retest reliability with a Pearson product-moment correlation Coefficient = .79 for elderly depressed patients [227]. In a sample of 45 college students test retest reliability (average 3.2 days) was excellent at 0.96 [228]
	Sensitivity to change	In a study of 105 inpatients with major depressive disorder randomised to 5 weeks of either interpersonal psychotherapy or clinical management, an effect size of 0.71 was reported for the BDI used to measure symptoms before and after treatment[229]. Hiroe et al[230] in a study of forty hospital outpatients administered the BDI two weeks apart found the BDI sensitive to change in depression: a 5-point difference corresponded to a minimally important clinical difference, 10–19 points to a moderate difference, and 20 or more points to a large difference.
	Acceptability/Usability	The BDI has proven useful as a screening and diagnostic tool for depression in a variety of disorders and populations for example Parkinson's and Renal disease, diabetes and children and adolescents [231-234]

MEASURE	PSYCHOMETRIC PROPERTIES	
Beck Anxiety Inventory (BAI)	**Validity**	The BAI has demonstrated concurrent validity with positive correlations between the BAI and the HARS.[235] The results of this study support previous findings that the strongest quality of the BAI is its ability to assess panic symptomatology.[236] Several studies have found a correlation between the BAI and the BDI given the extensive overlap between these two types of symptoms. The BAI has been identified as an efficient screening tool for distinguishing between individuals with and without panic disorder.[236] Factor analysis has identified a factor structure to the measure that suggests it gathers information on the cognitive and somatic aspects of anxiety.[237]
	Reliability	The measure has demonstrated high levels of internal consistency.[237] De Ayala et al [238] reviewed 172 articles on the BAI 56.6% made no mention of reliability estimates, the remainder were analysed and the average coefficient alpha was .91 (minimum and maximum values of.83 and .95, respectively). The test-retest reliability estimates ranged from .35 to .83, with a mean reliability estimate of .66. Time intervals between administrations ranged between 7 and 112 days, with a mean of 32.1 days.
	Sensitivity to change	The measure has shown significant change pre and post internet-based cognitive behavioural therapy for severe anxiety.[239].
	Acceptability/Usability	The BAI has proven useful as a diagnostic tool for depression in a variety of populations including inpatient adolescents [240] and older outpatient adults[241] . It has also proven useful as a screening instrument for the general population [236]
Hamilton Depression Rating Scale (HDRS)	**Validity**	Heo et al [242] found that the HDRS-D correlates to other scales like the Montgomery-Asberg Depression Rating Scale (MADR). Bagby et al[243] in a review of the literature found that content validity was established because of the scales wide use and long clinical tradition the Hamilton depression scale seems to both define as well as measure depression, the measure has demonstrated convergent and discriminate validity. In terms of factor structure this review found studies that reported between two and eight factors and the authors conclude that the Hamilton Depression rating scale is multidimensional but the nature of that multidimensionality remains unclear.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Laenen et al [244], taking data from two longitudinal drug efficacy trials with HDRS measurements taken at baseline and after 1, 2, 4, 6, 8 and 10 weeks found R_t coefficient of around 0.50 and 0.60 for the two different trials. This reliability increased over the course of the trial and believe this could have been induced by a learning effect of the raters, stemming from gaining experience and/or enhanced familiarity with a patient during follow-up. Gonzalez-Pinto et al [245], in an inpatient observational study of people with bipolar disorder, found the internal consistency of HDRS -21 (standardised Cronbach α) was 0.67 at baseline and 0.80 at 4 weeks. The internal consistency of HDRS -5 was 0.71 at baseline and 0.71 at 4 weeks. For the HDRS -21, the Pearson correlation coefficient was 0.88, $p < 0.001$, and the ICC was 0.92, $p < 0.001$ (95% CI: 0.897–0.944). For the HDRS -5, the corresponding figures were 0.94 and 0.97 (95% CI: 0.955–0.975), respectively, both with $p < 0.001$.
	Sensitivity to change	Schneibel et al [229] administered the HDRS to 105 inpatients with major depressive disorder randomised to 5 weeks of either interpersonal psychotherapy or clinical management before and after treatment. They reported a large effect of 2.44.
	Acceptability/Usability	The measure has been in use for over 50 years.[246] To achieve adequate inter-rater reliability, training and retraining is required.[247]
Hamilton Anxiety Rating Scale (HARS)	**Validity**	The scale has demonstrated concurrent validity through correlation with other measures of anxiety, including the Covi Anxiety Scale (0.63).[248], State-Trait Anxiety Inventory (STAI), and the Hospital Anxiety and Depression Scale (HADS) [249] and the Beck Anxiety Inventory [235]. The HARS has demonstrated good construct validity [250]. The scale correlates considerably with the Hamilton Depression Rating Scale in samples of older adults, raising concerns about its discriminant validity[251].
	Reliability	The inter-rater reliability, using ICC of the measure, has been reported as HAM-A total score = 0.74; HAM-A total score of psychic anxiety = 0.73; HAM-A total score of somatic anxiety = 0.70.[252] Therrien and Hunsley,[253] in a review of measures, identify the internal consistency of the HAM-A in older adults diagnosed with generalised anxiety disorder as r = 0.77–0.80. Similarly, high inter-rater reliability has been shown with community samples and with older adults diagnosed with generalised anxiety disorder (r = 0.81–0.95).
	Sensitivity to change	The measure has demonstrated sensitivity to change; for example, in one study of 105 inpatients with major depressive disorder randomised to 5 weeks of either interpersonal psychotherapy or clinical management, an effect size of 2.44 was reported.[229]
	Acceptability/Usability	The measure has been in use for over 50 years.[246] To achieve adequate inter-rater reliability training and retraining is required.[247]

MEASURE	PSYCHOMETRIC PROPERTIES	
State Trait Anxiety Inventory (STAI)	**Validity**	While some have found only limited evidence of concurrent validity, as demonstrated by moderate correlations between the STAI-T and other measures of anxiety (r = 0.33–0.57). Slightly lower correlations were found between the STAI-S and other measures of anxiety. The inventory is substantially correlated with measures of depression (r = 0.41–0.70).[253]. In contrast, Mondolo et al [249] found high correlations between the State-Trait Anxiety Inventory (STAI), the Hospital Anxiety and Depression Scale (HADS) and the Hamilton Anxiety Rating Scale (HARS). Bieling et al [254] have questioned the factor structure of the STAI and conclude that the trait scale of the STAI does not assess 'pure' anxiety, but rather includes items that reflect depression and general negative affect. Content validity has been demonstrated by the convergence of DSM-IV diagnosis of generalised anxiety disorder and STAI items [255]. While, construct validity has been demonstrated by the consistent finding of increased state scores in exam stress situations and decreasing scores in relaxation conditions [116].
	Reliability	The internal consistency of the STAI has been reported as excellent: STAI State (∝= .95) and the STAI Trait (∝ =.93).[178] Test-retest correlations over intervals of 7 to 10 days are higher for the trait scale (∝= .84 and ∝= .83 respectively) than for the State scale (∝= .49 and ∝= .63 respectively) [256].
	Sensitivity to change	The intent of the T (trait)-anxiety scale is to characterize anxiety as a longstanding trait or characteristic of the person and as such is less responsive to change. In contrast the S (State)-Anxiety is more amenable to change [257]. Newham et al [258] in a review of the literature found that scores on the STAI appear amenable to change during pregnancy, both after a single session and multiple sessions of interventions designed to reduce maternal anxiety. The STAI has also been used to demonstrate change pre and post a cognitive behavioral intervention for anxiety in primary care [259]
	Acceptability/Usability	The measure has proven particularly popular, having been used in over 3,000 studies and translated into 30 languages.[260] Norms for older people have been identified.[261] However, concerns have been raised that the STAI is lengthy and easily misinterpreted by older adults.[262] . However others have concluded that the 'state' component is a brief and simple self-rating instrument covering a broad range of items that specifically describe different aspects of the intensity of anxiety. This feature qualifies it as a good measure for state changes in anxiety enabling a variety of experimental manipulations, both in clinical as well as subclinical populations[263].

MEASURE	PSYCHOMETRIC PROPERTIES	
Edinburgh Postnatal Depression Scale (EPDS)	**Validity**	The EPDS has been validated in a variety of populations.[264] Moran and O'Hara[265] found the correlation between the BDI and the EPDS at two weeks post-partum to be r = 0.70, p < 0.01. In a Danish population, Bergink et al [266] identified cutoff values for postnatal depression of 11 in the first trimester and 10 in the second and third trimesters as the scores that are the most adequate combination of sensitivity, specificity and positive predictive value. They note, however, that cut-off scores have varied across different populations. They also found across all three trimesters, significant high correlations between the EDS and the SCL-90 anxiety and somatization subscales (r > .50; p < .001). EPDS has also shown correlations with the GHQ-28 score: Spearman's correlation coefficient (SCC) = .77 and the Center Epidemiological Scale-Depression (CES-D) SCC = .84. Pallant et al [267], using a Rasch analysis, suggest that EPDS in its original 10-item form is not a viable scale for the unidimensional measurement of depression and an eight-item version (EPDS-8) would provide a more psychometrically robust scale.
	Reliability	The Cronbach's alpha for both the self-report EPDS and the partner version EPDS-P are $\propto = .85$ and $\propto = .80$, respectively.[198] Bergink et al[215] found test-retest reliability between 12 and 24 weeks' gestation to be r = .61 (p < .01); between 12 and 36 weeks, r = .55 (p < .01); and between 24 and 36 weeks, r = 0.63 (p < .01). Short test-retest reliability was tested with a mean time between administrations of 6.87 days (SD: 0.21; range: 4-10) and found to be Intraclass Correlation Coefficient (ICC) = .92.[268]
	Sensitivity to change	Matthey and Ross[170]-found that many women scoring at or above the cut-off score on the EPDS on a clinic visit for antenatal care when contacted via the telephone two weeks later scored below the cut-off score. The EPDS showed statistically significant change on nine antidepressant-free women with postpartum depression treated with repetitive transcranial magnetic stimulation, Friedman's test X2= 19.35 <.0005 [269]
	Acceptability/Usability	Guedeney and Fermanian[270] found that the EPDS was readily accepted by mothers and the measure was completed in approximately five minutes. The EPDS has been "translated" for use with Aboriginal and Torres Strait Islander women in North Queensland.[271] The measure has been used in a wide variety of populations for clinical trials of medication efficacy and population groups, adolescents, adults and older persons [246]
Geriatric Depression Scale (GDS)	**Validity**	Concurrent validity has been established with high correlations of .79 and .82 [118] between the GDS and the Zung (self-rated)[272] and Hamilton Depression Rating Scales (observer-rated)[114], respectively. Lucas-Carrasco [273] found significant associations between both the 15-item and 5-item GDS and the Barthel Index[274], the Cornell Scale for Depression in Dementia (CSDD)[275], and the WHOQOL-BREF[276]
	Reliability	Cronbach's alpha coefficients have been found to be .81 and .72 for GDS-15 and GDS-5, respectively, indicating good to acceptable internal consistency[273]. Pedraza et al [277] report test-retest reliability over a 15 month period of r = .68.

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Sensitivity to change	The 30-item version of the GDS has been shown as sensitive to change (effect size 1.63) over 15 months in individuals with rheumatoid arthritis.[278]
	Acceptability/Usability	In Australia the GDS is used extensively in practice. It has been identified as one of the most widely used self-rated mood questionnaires for older adults[277]. Aikman and Oehlert found the short form just as accurate in identifying depression as the long form of the measure.[120] Rapp et al[279] have used the GDS as part of a telephone-based cognitive assessment for elderly people.
General Measures of Symptomatology		
Symptom Checklist 90 (SCL 90)	**Validity**	The SCL-90 has demonstrated predictive validity, being able to distinguish between people with breast cancer and a physically healthy population[124] and between people who do or do not have depression after a stroke[226]. Discriminant validity has been demonstrated with small correlations between the SCL-90 and the HoNOS[280]. Using both a Mokken [281] and Rasch analysis[282] Olsen et al [283] found support for all subscales except the psychoticism subscale in a Danish community sample but conclude that the 63 non-psychotic items in the measure appear to reflect one broad dimension of distress. This conclusion has also been drawn from the results of a sample of involuntary consumers in acute psychiatric inpatient care [284]. Interestingly, Bjørkly [285] found that people with a history of violence under-reported symptomatology when compared to reference material.
	Reliability	The internal consistency of the SLC-90 has been reported as good .88 [226] to excellent .92 [286]. Test–retest (within 3 to 7 days) was established as r = .79 [286]. Inter-rater reliability of the checklist was tested by having nurses independently rate violent offenders Mean regression coefficient for the nine symptom dimensions and the three global indices was R =.87, (Range =.70–.97)[285].
	Sensitivity to change	In a prospective cohort study of five outpatient clinics in Norway, data was collected at admission and at discharge or, if still in care, 1 year after admission. Øiesvold et al [287] reported an effect size of 0.71. They also report on the proportion of consumers with reliable and clinically significant change (reliable change is based on the standard error or measurement, ensuring that any reported change is not due to measurement error and [280] clinically significant change in this study was defined as outcomes (post-treatment scores) falling beyond two standard deviations of the pre-treatment mean). Of 118 consumers 51% showed reliable change but only 9% showed clinically significant change[287].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	The SCL-90 has been recommended as a standard instrument for psychotherapy outcome research[288]. Subscales of the measure have been used to evaluate the effectiveness of internet-based interventions[289] Comparing pencil and paper with computer-based administration, Schmitz et al [290] found little difference between subscale scores and the global severity index except for subscales `Obsessive-Compulsive' and `Anger-Hostility', which scored slightly higher on the computer-administered versions.
General Health Questionnaire (GHQ)	**Validity**	Significant differences in GHQ scores have been demonstrated between groups with expected high scores (i.e., those who perceived themselves to be in ill health and those who were perceived as having helplessness, loneliness, anxiety, and depressive mood) and the groups with expected low scores on the scale (i.e., those who perceived themselves to be in good health and those who did not perceive themselves as having helplessness, loneliness, anxiety, and depressive mood) [291]. Child and Adolescents with family disruption, experiencing bullying and who seek help for a mental health problem are more likely to score highly on the GHQ.[292] There have been various analyses of the GHQ versions. In a large Australian study there was limited support for multiple factors in the GHQ-12, and a single score was felt more appropriate to report.[293]
	Reliability	Internal consistency has been reported as .79[294] to .90 [295], and test-retest reliability at one week for Pearson's r and ICC showed satisfactory results; the former ranged from .81 to .84 according to the three GHQ scoring methods.[247]
	Sensitivity to change	The measure has demonstrated change following treatment, with an effect size of .75 for people being treated for urological problems.[294]
	Acceptability/Usability	In Australia, concerns have been raised about the use of the measure with child and adolescent consumers.[296] There have also been concerns that given the generally low scores in reported studies like the Mental Health and Wellbeing Survey, the GHQ 12 is a less useful instrument for detecting mental illness than in other countries.[297]
SPHERE-12	**Validity**	In part, the concurrent validity of the PSYCH-6 component of the SPHERE-12 has been demonstrated through positive correlations [298] with the Hospital Anxiety and Depression Scales,[299] while the SOMA–6 component has been positively correlated [300] with the Functional Assessment of Cancer Therapy-Fatigue(FACT-F subscale-13 items).[301]
	Reliability	The internal consistency of the PSYCH-6 items has been demonstrated as .90 (excellent) and for the SOMA-6 as .80 (good). The test-retest over an "extended" period of time had an intra-class correlation coefficient of .81 for the PSYCH-6 and .80 for the SOMA-6.[131]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Sensitivity to change	While no specific analysis of the measure's sensitivity to change has occurred, it is has been used as a measure of the effectiveness of aromatherapy,[302] the impact of cancer treatment on fatigue and mood disorder,[303] and the relationship between viral illness, chronic fatigue syndrome and mood disturbance.[304]
	Acceptability/Usability	An Australian study raised concerns that the measure lacked specificity to be used as a screening tool in general practice.[130]
My Mood Monitor (M-3)	**Validity**	The M3 has been shown as useful as a screening tool, with the depression module having a sensitivity of .84 and a specificity of .80. The bipolar module has a sensitivity of .88, and a specificity of .70. The anxiety module has a sensitivity of .82 and a specificity of .78, and the PTSD module has a sensitivity of .88 and a specificity of .76. As a screen for any psychiatric disorder, sensitivity was .83 and specificity was .76.[132] No research on the structure of the M-3 was found.
	Reliability	No research was identified on the test-retest or inter-rater reliability of the M-3.
	Sensitivity to change	No research was identified on the sensitivity to change of the M-3.
	Acceptability/Usability	In a study of 647 adults (over 18) and older people who were seeking primary care at an academic family medical clinic, patients took less than 5 minutes to complete the M-3 in the waiting room, and less than 1% reported not having time to complete it. Eighty-three percent of fifty five (55) clinicians reviewed the checklist in 30 or fewer seconds, and 80% thought it was helpful in reviewing patients' emotional health.[132]
Specific Measures of Symptomatology		
Suicidal Ideation Questionnaire (SIQ)	**Validity**	Pinto et al [305], using ANOVA, found non-suicidal adolescents scored significantly lower on the SIQ than did hospitalised suicidal adolescents (both attempters and ideators) (F(2, 220) = 20.67, p < .001, n^2 = .16); attemptors and ideators did not significantly differ. Concurrent validity is supported, with the SIQ correlating with other measures of depression, hopelessness, anxiety, learned helplessness and self-esteem. All reported correlations are in the expected directions and are significant at the p < .001 level[135]. The Adult version of the SIQ correctly classified participants in one study as multiple attempters of suicide 80% of the time (Area Under the Curve (AUC) = .80, SE = 0.29 [95% CI, .74-.85]).[306] Kaminer et al [307] used the SIQ to assess the magnitude and course of suicidal ideation during outpatient treatment and aftercare for found that those you people who had attempted suicide ideation on the SIQ (M= 28.76, SD 22.34) than non-attempters (M=7.91, SD =8.73), t(17.04 = 3.79, p <.001)
	Reliability	The SIQ has demonstrated a high level of internal consistency, with Cronbach's alpha ranging from .91 to .97[305, 308]. Test-retest reliability over one to five weeks has been reported as .89.[308]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Sensitivity to change	The SIQ has been used to analyse the effectiveness of SSRI medications on treatment-resistant depression in adolescents, they found lower scores on the SIQ for responders (M=38, SD 20.6) to treatment than those who did not respond (M=45, SD 23.3) [309] Diamond et al [310]found a large effect size (d = 0.97) with a decrease in scores on the SIQ as a result attachment based family therapy.
	Acceptability/Usability	Pinto et al [305]found the measure internally consistent, it differentiates between hospitalized suicidal adolescents from hospitalized non suicidal adolescents and contains several critical items that can be used as a short screen for suicidal ideation.
Columbia–Suicide Severity Rating Scale (C-SSRS)	**Validity**	The concurrent validity of the (C-SSRS) has been demonstrated by correlations [311] between the C-SSRS and the Beck Scale for Suicide Ideation.[312] Posner et al [136] also found the C-SSRS severity subscale was moderately correlated with the worst-point score on the Scale for Suicide Ideation (r = .52, p < 0.001; effect size = 1.22, N = 472).[312] They also found a moderate correlation between the C-SSR severity subscale and the Suicidal Ideation Questionnaire- Junior [308] total score (r = 0.23, p < 0.01; effect size= 0.77). Posner et al [136] also found that baseline ratings on the C-SSRS (based on worst-point lifetime suicidal ideation and conducted before treatment), predicted suicide attempts during treatment (odds ratio (OR) = 1.45 [95% CI, 1.07-1.98], p = 0.2). In all, Posner et al [136] found that the data from three studies provide promising data on the convergent and divergent validity, predictive validity, sensitivity, specificity, sensitivity to change, and internal consistency of the C-SSRS.
	Reliability	The inter-rater reliability between interviewer-administered questions and computer-assisted administration has been tested. Mundt et al [311] found that the concordance (Kendall's tau-b) of the total ideation score for subjects with the most severe level of ideation was .85 (p < .01) between the human raters, .62 (p < .08) between Rater 1 and a self-report computer completed C-SSRS, and .81 (p < .01) between Rater 2 and a self-report computer completed C-SSRS. The internal consistency of the C-SSRS intensity subscale has been reported as excellent at $\alpha = .93$ for the period 'since last visit' and $\alpha = 0.94$ for 'past week'. Across different time periods (baseline, 4, 8, 16 and 24 weeks and on presentation to the emergency department) $\alpha = 0.73$.
	Sensitivity to change	Posner et al [136] assessed the behaviour subscales' sensitivity to change for three types of suicidal behaviours (aborted, interrupted and actual attempts). Changes in these behaviours as rated on the C-SSRS corresponded to behaviours identified by independent observers.
	Acceptability/Usability	Caution has been expressed about the use of the measure given that it is in the early stages of development [313] with most analysis focused on only aspects of the C-SSRS.

MEASURE	PSYCHOMETRIC PROPERTIES	
Inventory of Complicated Grief-Revised (ICG)	**Validity**	The ICG total score was correlated with the Beck Depression Inventory (BDI) total score ($r = 0.67, p < 0.001$), the Texas Revised Inventory of Grief [314] (TRIG) score ($r = 0.87, p < 0.001$), and the Grief Measurement Scale[315] (GMS) score ($r = 0.70$, $p < 0.001$)[316]. The ICG has also been correlated with measures of the intrusive aspects of post-traumatic stress disorder[317]. People with complicated grief experienced more life stressors, perceived less social support and achieved less clinically significant change in their mental health compared with those who did not experience complicated grief following the death of a spouse[318]. Using a modified version of the ICG, Boelen et al [319] found that lower self-concept clarity was associated with more complicated grief. Boelen et al [320], in a study of 1,321 bereaved individuals and using Item Response Theory, found that the ICG-R represents one dimension of complicated grief symptomatology. Their findings support the conclusion that complicated grief is better conceptualised as a disorder of disrupted cognitions or meaning-making processes rather than as a disorder of disrupted attachment. O'Connor et al[317] call into question the construct validity of the inventory having identified a substantial overlap between complicated grief and post-traumatic stress disorder and it is possible that the symptoms of complicated grief may be well accounted for by especially the intrusive component of post-traumatic stress disorder
	Reliability	The internal consistency of the 19-item ICG was shown as excellent at $\alpha = .94$, while the correlation between 6-month ICG repeat assessments was .80.[137]
	Sensitivity to change	Meert et al[321] found change in ICG scores for parents whose children had died in intensive care units. At 6 months ICG scores were 33.4 ± 13.6, while at 18 months, scores were 28.0 ± 13.5, representing an improvement in ICG score of 5.4 + 8.0 (95% CI, 4.1–6.8, $p < 0.001$). Shear et al[322] used the ICG to demonstrate the effectiveness of complicated grief therapy as opposed to interpersonal therapy. The response rate was greater for complicated grief treatment (51%) than for interpersonal psychotherapy (28%, $p = .02$) and time to response was faster for complicated grief treatment ($p = .02$).
	Acceptability/Usability	In Australia, Pini et al[323] found that Adult separation anxiety disorder occurs in a high proportion of adult psychiatric outpatients with complicated grief. No specific studies of the acceptability or utility of the measure were found.
PTSD Checklist (PCL-C)	**Validity**	Convergent validity has been demonstrated for the PCL-C, with high correlations[324] between the PCL-total scores and scores on other well established measure of PTSD, including the Impact of Event Scale (IES)[325], the Mississippi Scale for PTSD (MS-C)[326] and the Clinician Administered PTSD Scale (CAPS)[327].Factor analysis indicated a five-factor model – comprising re-experiencing, avoidance, numbing, dysphoric arousal, and anxious arousal in a group of elderly hurricane survivors[328]. Hem et al [329]found agreement between PCL-C and SCID-I using Cohen's kappa which equalled 0.71

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	The internal consistency of the checklist has been reported as α = .93.[330] Test-retest correlation coefficients for total scores on the PCL were .92 (p < .001) for immediate re-testers, .88 (p < .001) for a 1-week retest interval and .68 (p < .001) for a 2-week retest interval[324].
	Sensitivity to change	The checklist has been used to assess change in Vietnam veterans receiving treatment for PTSD and an effect size of .59 is reported. However, the authors caution that the checklist under-rated improvement compared with the 'gold standard' CAPS[331]. Although this may be the result of comparing a self-report with a clinician-rated measure.
	Acceptability/Usability	The PCL-C has been used as a screening tool for PTSD in relation to different disorders including breast cancer [139], as well as for non-clinical samples.[332] The uptake of the checklist is reported to be increasing [324].
Dissociative Experiences Scale (DES)	**Validity**	The DES demonstrates predictive validity, performing as a screening tool identifying people with and without personality disorders.[333] The scale has demonstrated criterion validity, in that DES depersonalization scores were able to predict depersonalization disorder diagnosis[334]. However, there are concerns regarding the content validity of the DES. During the measures development the validation process for the DES was not explicitly described, so there is insufficient information as to how depersonalization was defined and how items were matched to key aspects of the construct[335].
	Reliability	For the total scale the internal consistency has been reported as α = .93. For the Absorption subscale, α = .88; the Depersonalization and Derealization subscale, α = .87; and the Amnesia subscale, α = .70; and for a subset of eight DES items that measure pathological dissociation, α =.84.[336] Two-week test-retest reliability has been reported as .93 for the total score.[337].
	Sensitivity to change	No change has been observed between pre and post-treatment scores on patients admitted to a dissociative disorders unit. The authors speculate that this may be because the measure yields stable evaluations of the patient's own view of their history of dissociative experiences.[337] However, others have found change following treatment using cognitive processing therapy (effect size = .32).[338]
	Acceptability/Usability	The DES has been recommended as a useful as a screening tool for diagnostic purposes, but not appropriate for outcome research.[337]

<table>
<tr><th>MEASURE</th><th colspan="2">PSYCHOMETRIC PROPERTIES</th></tr>
<tr><td colspan="3">Measures for Children and Adolescents</td></tr>
<tr><td rowspan="3">Strengths and Difficulties Questionnaire (SDQ)</td><td>Validity</td><td>A number of studies have examined the construct validity of the SDQ by conducting confirmatory factor analyses of the parent-rated,[144, 339-342] teacher-rated [144] and self-report[144, 340, 343, 344] versions in a variety of languages. In general, these studies found five-factor solutions that corresponded with the original scales proposed by Goodman, and have resulted in only minor suggestions about alternative factor structures[345]. The exception was a study by Ronning et al[344], which found ‘a variable and questionable fit’ when data from the Norwegian self-report version of the SDQ were compared with the Goodman model[344].</td></tr>
<tr><td>Reliability</td><td>A body of studies has also considered the internal consistency of the SDQ, again considering the parent-rated [144, 340, 341, 345-348], teacher-rated[144, 345-347, 349] and self-report[144, 340, 343-346, 348, 350, 351] versions in a range of languages. Taken together, these studies suggest that the instrument has very good overall internal consistency, with Cronbach’s alphas of about .70 for total difficulties. The individual scales (including the Impact scale) also generally appear to have satisfactory to good internal consistency, although some – notably Conduct problems and Peer problems – have comparatively poorer internal consistency. The parent-rated and teacher-rated scales tend to have better internal consistency than the self-report scales. Several studies have purported to assess the test-retest reliability of the SDQ, although in all cases the period between the first and second administration of the instrument was arguably too long to expect the rating to remain the same (ranging from 3 to 4 weeks up to 12 months)[342, 349, 352]. Nonetheless, all these studies reported good stability across time, even for younger children. In terms of raters, teacher ratings were the most stable, and youth self-report ratings the least. In terms of scales, total difficulty scores and hyperactivity-inattention scores were the most stable; the impact score was the least reliable.</td></tr>
<tr><td>Sensitivity to change</td><td>Mathai, Anderson and Bourne[353] considered the sensitivity to change of the parent-rated, teacher-rated and self-report versions of the SDQ by following 130 consecutive new attendees at a Victorian Child and Adolescent Mental Health Service for six months. Changes in SDQ scores from baseline to follow-up were compared with changes in clinician-rated HoNOSCA scores over the same period. There was a significant improvement over time in the total difficulties score on all versions of the SDQ, which corresponded to improvement as measured by the HoNOSCA. There were also lowered levels of perceived difficulties and burden on the SDQ impact supplement, and a decrease in the overall impact. Having said this, it should be noted that the study suffered from substantial loss to follow-up, and it was unclear whether the reference period covered by the second SDQ was the previous six months or the previous month.</td></tr>
</table>

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	Given the relatively extensive psychometric testing of the SDQ, it is surprising to find that very few studies have asked respondents using the different versions of the instrument to comment on its feasibility and utility. The only study that provides any information in this regard is that of Goodman and Scott,[354] which found that mothers who had used the SDQ and the CBCL (below) were twice as likely to prefer the former.
Child Behaviour Checklist (CBCL)	**Validity**	The CBCL, like the SDQ, is able to detect inattention and hyperactivity as well as detecting internalising and externalising problems [354]. It has demonstrated an ability to discriminate between adolescents with and without bipolar disorder [355] and youth with and without anxiety disorders [356], and those referred for treatment and those not referred [357].The CBCL also correlates with other measures of depression, anxiety and arousal[358] such as the affect and arousal scale[359] and the Revised Child Anxiety and Depression Scale[360]
	Reliability	The internal consistency of the CBCL is excellent and has been reported for Total Problems, Internalizing, and Externalizing as .96, .89 and .91 [357]. Eight-week test-retest correlation coefficients of .87 and .76 [361], respectively, have been reported in Australia, along with an average 6-month test-retest reliability of .75.[362]
	Sensitivity to change	In a study of adopted Chinese girls, Tan [363] used the CBCL to identify behavioural disturbance and found that overall as the children got older, their behavioural adjustment worsened. He reported effect sizes for internalizing, externalizing and total problems of .36 (medium effect), .07 (small effect), and .06 (small effect) as the children moved through two-year age cohorts. Willemen et al [364] used the CBCL to assess the impact of stressful life events on psychopathology. The study showed that problem levels generally decline after referral. Stressful life events were associated with slower recovery of parent- and self-reported internalising, but not of externalising problems across time. In Australia, Starling et al [365] used the CBCL parent and youth report to monitor changes in the pattern of presentation of consumers to university-affiliated adolescent psychiatry service, the Rivendell Unit, between 1983 and 1998. They found that parents described increasing talk of self-harm over this period, but this was not reflected in adolescent self-reports. Lipman et al [366] used the CBCL to evaluate the effectiveness of a community-based intervention for boys at risk of antisocial behavior. They report a moderate effect size (.41) for children in the treatment group. The CBCL has also been used to monitor change during psychotherapy. [367]
	Acceptability/Usability	The CBCL is one of the most widely used measures of youth symptoms in the world [358]. There have been cut-points developed with an Australian sample[368]. Siddons and Lancaster [369] have undertaken an extensive review of the use of the CBCL in Australia and recommend its use with some caution around using norms from America. However, because of its length and concerns regarding its internal consistency, the CBCL was not recommended for routine use in Australian mental health services in a large routine outcome measurement scoping study.[16]

MEASURE	PSYCHOMETRIC PROPERTIES	
Devereux Early Childhood Assessment (DECA)	**Validity**	The DECA is able to predict membership in a problem-identified group compared with a community sample for both protective factors and behavioural concerns (69% and 71%, respectively)[370]. The factor structure proposed by the authors (Initiative, Self-Control and Attachment), has been replicated in a number of studies[371, 372]
	Reliability	The technical manual reports good to excellent internal consistency for all subscales initiative α = .84; self-control α = .86; attachment α = .76; total protective factors α = .91; behavior concerns α = .71. This good to excellent internal consistency has been replicated in a number of studies.[147, 372] As well as test–retest reliabilities over a 24-hour period for all subscales ranging from .55 to .80, inter-rater reliability between parents for all the subscales was fair to moderate, ranging from .21 to .44.[147] There was only fair agreement (.18-.30 across all scales) between parents and teachers when rating children using the DECA[373], although this is consistent with other cross-informant research.[374]
	Sensitivity to change	In Australia the DECA has been used to evaluate the effectiveness of a parenting program; however, no change was found.[375]
	Acceptability/Usability	The DECA is available in English and Spanish.[373] It has been identified as a useful screening tool, being used in a large community screening project, although no results were reported.[376]
Anxiety Disorders Interview Schedule-Children (ADIS-Children)	**Validity**	Concurrent validity of the ADIS-Children has been demonstrated, with Pearson correlations between the ADIS symptom counts[377] and the Diagnostic Interview Schedule for Children – Version IV[378] (r = .67), and with scores on the conduct component of the Behaviour Assessment System for Children 2nd Edition (BASC) (r = .33.)
	Reliability	Diagnoses derived from the ADIS have been shown to possess good to excellent inter-rater agreement. Agreement between raters for principle diagnosis has been reported as κ =. 0.92, and for individual anxiety disorders agreement has ranged from κ = .80 for generalised anxiety disorder to κ = 1.0 for specific phobia.[379]. It has been observed that there are poor levels of agreement between child and parent reporting using this instrument[380, 381], but agreement has been identified as better when anxiety is measured as a continuous rather than a dichotomous variable[382]. The ADIS has demonstrated acceptable test-retest reliability at 7 to 14 with intraclass correlation coefficients reported for the child (κ = .0.78 to 0.95) and parent interview (κ = .81 to 0.96)[383].
	Sensitivity to change	The clinician severity rating (CSR) component of the schedule has been used to demonstrate the effectiveness of treatment of children with attention-deficit/hyperactivity disorder (ADHD)[384] and the effectiveness of cognitive behaviour therapy (CBT) for children with anxiety disorders[385]. Symptom counts have been used to demonstrate the effectiveness of therapist-supported bibliotherapy.[386]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Acceptability/Usability	No specific studies of acceptability/ usability were found but the schedule has been used extensively as a research tool.
Parenting Stress Index (PSI)	**Validity**	McKelvey et al[387] in a study of low-income fathers found, using a confirmatory factor analysis, that a five-factor structure (general distress, parenting demands distress, parent self-rating and dyadic interaction) best fit the data. Using this five-factor structure, they demonstrated the measure performs so that scores on the PSI correlated in the expected direction; for example, fathers that reported better parenting undertook more play activities. They also describe construct validity of the measure through the inter-relationship of the five factors. However, others have found support for a two-factor model: parental distress and dysfunctional parent-child interactions.[388] Indeed, the parental distress subscale was significantly related to Global Severity Index scores on the SCL–90–R (r(185) = .54, p < .001); and the dysfunctional parent-child interactions subscale was related to parenting behavior as measured by the Conflict Tactic Scales (r(163) = .23 p < .01) and observed parent and child behavior (r(168) = -.22, p < .01). Costa et al [389] were able to demonstrate the complex interrelationship between parental stress and difficulties in the parent-child relationship. PSI difficult child (β = .36, p < .001) and PSI PC dysfunctional interactions (β = .19, p < .01) were significant predictors of CBCL internalizing symptoms, and PSI difficult child (β = .62, p < .001) was a significant predictor of child externalizing symptoms.
	Reliability	McKelvey et al[387] report that internal consistency is good (α = .76) for Parental distress but only acceptable for parent-child dysfunction (α = .65). Haskett et al[388] report acceptable to good internal consistency for the parental distress subscale (α = .78) and dysfunctional parent-child interactions subscale (α= .91) Test-retest correlations after one year between first and second assessments were r(21) = .61, p < .005 for the parental distress subscale; r = .75, p < .001 for the dysfunctional parent-child interactions subscale; and r = .75, p < .001 for the Total scale.
	Sensitivity to change	Gerdes et el[390] have used the PSI to assess change in parental stress following behavioural parenting training. They found that 10% to 55% of mothers achieved reliable and clinically significant change (Total Score RC Index = 8.74, CS (M+/- 2(SD))= 101.80). Hibel et al[391] used the PSI to identify the effects of parental stress on cortisol levels in working mothers. They found that parenting stress and workday interacted to predict maternal cortisol levels, so that mothers with parenting stress have higher cortisol levels on workdays than non-workdays. Interactions between job strain and parenting stress as measured by the PSI were significant for a.m. cortisol levels (β = .27, p < .001) and for cortisol awakening response (CAR) (β = .08, p = .04), but only on workdays. The interaction of job strain and parenting stress was not related to non-workday a.m. cortisol levels (β = .07, ns, or CAR, β= -.03, ns).

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Acceptability/Usability	The PSI-short form has become one of the most widely used instruments for measuring parenting stress across a culturally diverse families and children.[392] It has been used to compare services and its use is encouraged in practice.[393]
Social Responsiveness Scale (SRS)	**Validity**	Concurrent validity has been demonstrated, with correlations between the SRS and the Repetitive behaviour questionnaire (RBQ)[394] in the expected direction. For example, the autistic mannerisms subscale of the SRS correlated with the RBQ total (r = .419, p < .01) score. Concurrent validity has also been supported by a correlation between SRS total scores and psychological difficulties as measured by the total of four problem domains on the Strengths and Difficulties Questionnaire (SDQ) (rs = .704, p < .01).[395]. Wang et al [396]indicate that the scale is able to discriminate between child with autism spectrum disorders and controls as well as individuals with other psychiatric diagnoses. Given that these results are taken form a Taiwanese sample they indicate the scale has cross cultural validity.
	Reliability	The internal consistency (Cronbach's alpha) of the total raw SRS scores is excellent for females (α = .92, n = 253), and for males (α = .92, n = 247)[395]. Correlations between parent and teacher reports on the SRS are good (r = 0.79, n = 55, p = 0.0001).[397] Bolte et al [398] reported on 49 test and retests collected at 3 to 6 month intervals in a clinical sample. Test-retest reliability reached r = .97 and r = .95, respectively. Inter-rater reliability in this sample between mother and father ratings (n =172) reached r = .95 and r = .91, with all correlations significant at p < 0.001.
	Sensitivity to change	The SRS has been used in a number of outcome studies. In a study of the Multimodal Anxiety and Social Skill Intervention, White et al[399] reported an effect size of d = 1.18, and the Wilcoxon rank sum test (Z = -2.76, p < .01) indicated significant improvement from pre- to post-treatment. The SRS has also been used to test the efficacy of medication for the autism spectrum disorders, and Yui et al [400] report the number of individuals who achieved 50% improvements in the total SRS as treatment group = 14.3% vs. placebo group = 16.7% ([chi]^{2}= 0.01, p = 0.92).
	Acceptability/Usability	The brief nature of the SRS makes it user friendly and provides quick assessment, making it useful for large scale studies.[401] A strength of the SRS is that it is intended to detect grades of impairment, rather than a categorical yes/no presence of a particular symptom.[402]
Barriers to Adolescents Seeking Help Scale (BASH)	**Validity**	Higher BASH scores (indicating resistance to seeking help) were higher for those who had never been in treatment compared with those who were currently in, or had previously experienced, treatment, and scores were lower for those with more positive views on caregivers. There was no correlation found between distress and attitudes to help seeking.[153] A factor analysis has identified two factors “perceived need for autonomy” and “help seeking fears”.[403]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Reliability	During development, internal consistency using Cronbach's alpha was reported as α = .91. The shorter BASH-B version has good internal consistency of α =.84 [404]. The scale during development was reported as having good two-week test retest reliability 0.91.[153]
	Sensitivity to change	No research was identified that examined sensitivity to change of the BASH.
	Acceptability/Usability	The short form of the measure, the BASH-B, has been used in a number of studies to better understand help seeking.[403]
Measures of cognitive capacity		
Cambridge Cognitive Examination (CAMCOG)	**Validity**	The CAMCOG performs in a similar manner to other cognitive tests[405] including the MMSE[155], the cognitive component of the Functional Independence Measure (FIM)[406] and the Rotterdam short version of the CAMCOG (R-CAMCOG)[407]. It discriminates between those that develop dementia and those that do not.[408] There is no association between the severity of depression and scores on the CAMCOG.[409] The CAMCOG demonstrated diagnostic sensitivity and specificity being able to distinguish between normal persons from clinically diagnosed Alzheimer's Disease (AD) patients as well as distinguishing between individuals with an organic-dementing condition and normal adults. However, the CAMCOG was found to be less effective in distinguishing between AD and non-AD dementia patients and in distinguishing between patients suffering from organic dementia versus specified psychiatric disorders[410].
	Reliability	The internal consistency of the CAMCOG total score is high, even though it is made up of items assessing different aspects of cognitive functioning, with Cronbach's alpha reported as .82 and .89 for those screened as not having dementia and those assessed as have dementia, respectively.[411] The test-retest as a Pearson correlation Coefficient at 30 days was .86. The individual subscales performed less reliably than the total score.[411] Inter-rater reliability has been demonstrated, with intraclass correlation coefficients between raters for subscales ranging from .93 to 98.[412] In an Australian study, while the CAMCOG had high correlations with other cognitive screen instruments, it correlated less well with an informant measure of cognitive decline.[413]
	Sensitivity to change	The CAMCOG has been used to assess the rate of decline in cognitive function of non-demented elderly people, with a statistically significant rate of decline of -1.6 points per year (p <0.001).[414]
	Acceptability/Usability	The CAMCOG has been used extensively and version created in other languages, including Portuguese[412] and Hebrew[415]. It has been used to assess the impact of medication on cognitive functioning[416] and the effect of Vitamin B-12 on cognitive functioning.[417]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
Mini Mental State Examination (MMSE)	**Validity**	In a cross-sectional study, the MMSE distinguished 74 clients with early Alzheimer's disease from 74 healthy subjects matched for age and education levels[418], indicating predictive validity. However, The predictive validity of the MMSE as a screening test for dementia depends on educational level.[419] Indeed, it is possible to have a diagnosis of Alzheimer’s type dementia and receive a perfect score on the MMSE.[420] Criterion validity has been demonstrated with scores on the MMSE similar to scores on the Cambridge Cognitive Examination (CAMCOG) and Alzheimer’s Disease Assessment Scale (ADAS-cog)[421], which all have been shown to reflect a common dimension of global cognitive impairment.[422]
	Reliability	Internal reliability for the MMSE have been variously reported as .77 in a community sample [423], .96 in elderly hospitalised individuals[424], .62 for normal adults and .81 for those with Alzheimer’s [425]. Mitrusmna and Satz[426] found test-retest reliability ranged between good (.45) over a 1-year interval and moderate (.38) over a 2-year period. However, Lopex et al[427] calculated that if the true score was 22.65 the 95% confidence interval endpoints would be 17.81 (lower bound) and 27.50 (upper bound), which means for anyone who obtained a score of 23, an observer could only be 95% confident that the score actually lies between these two points. This range is so large that it covers the most commonly reported cut-off score for the measure of 24.[155]
	Sensitivity to change	A number of studies have assessed change in the MMSE over time. A change on the MMSE of more than 5 points over a 2-year period has been associated with the development of neurological disorder.[426] Hensel et al[428] indicate that small changes in MMSE can only be interpreted with great uncertainty and they have a reasonable probability of being caused by measurement error, regression to the mean or practise.
	Acceptability/Usability	Scores on the MMSE have been found to be influenced by the respondent’s level of education and should be used with caution as a measure of general cognitive function in subjects with less than five years of schooling.[429] Lopez et al argue that the MMSE has a number of flaws, including relatively low reliability, too many easy items, too many cut-off points, and a lack of standardised scores. However, it has been in use since 1975 and because of its ease of use has been taken up by many health care workers,[427] although some report difficulties with the MMSE and there is a cost for administration.[430] Although some cognitive self-report measure have been identified that are suitable for administration via the internet The MMSE is not one of them [431].

MEASURE	PSYCHOMETRIC PROPERTIES	
Rowland Universal Dementia Assessment Scale (RUDAS)	**Validity**	The face validity of the scale was established through the development process with a multidisciplinary advisory group reviewing potential domains and items[432] The predictive validity of the RUDAS has been demonstrated through identification of those consumers with and without dementia (κ = .58, p < .01, sensitivity = .88, specificity, .75).[433] Criterion validity has been demonstrated by the area under curve (AUC) of a Receiver-operating characteristic (ROC) analysis, finding no difference between the RUDAS and the "Gold Standard" MMSE.[155] MMSE (AUC = 0.82 [95% CI = .76-.87], p < 0.0001) and RUDAS (AUC = 0.83 [95% CI = .77-.88], p < 0.0001). Interestingly, this was slightly lower for The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), a self-report measure of cognitive decline (AUC = 0.77 [95% CI = .71- .83], p < 0.0001)[434], for which there was no significant difference between the areas under the curve (χ2 = 2.57, df = 2, p = 0.28).[435] While some have suggested the RUDAS does not appear to be influenced by language, education or gender[436], others have found the RUDAS to be influenced by level of education.[437]. The RUDAS accurately detects mild as well as more severe forms of dementia[438].
	Reliability	No research on the reliability of the RUDAS was identified.
	Sensitivity to change	No research on the sensitivity to change of the RUDAS was identified.
	Acceptability/Usability	The RUDAS seems little influenced by cultural factors.[435] Scores on the RUDAS have been compared across two modes of administration: face to face or via videoconference and mean scores, which differed by 0.04, lead the authors to conclude there was no significant difference between these two administrations types.[439]. The measure has proven
Miscellaneous		
Warwick-Edinburgh Mental Well-being Scale (WEMWBS)	**Validity**	Bartram et al[440] found scores on the WEMWBS to correlate negatively with anxiety (r = -0.69, p < 0.001) and depressive symptoms (r = -0.76, p < 0.001) indicating the measure has discriminant validity. Some support for the concurrent validity of the WEMWBS was shown by significant but low to moderate correlations between the WEMWBS and favourable psychosocial working conditions (low risk of work-related stress), including demands (r = 0.32), control (r = 0.45), managerial support (r = 0.48), peer support (r = 0.50), relationships (r = 0.37), role (r = 0.45) and change (r = 0.41) in a sample of 3200 Scottish veterinary surgeons. Stewart- Brown et al[441] examined the construct validity of the measure using a strict RASCH model approach and found that 7 of the 14 items performed to this standard. This produced a short form of the scale with a focus on psychological and eudemonic well-being, rather than hedonic well-being or affect.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Internal consistency has been reported as excellent for a group of students and a general population sample α = .89 (students) and α =.91 (general population)[160] and in a study of positive psychology interventions α =.92[442]. One-week test-retest reliability has been reported as good at .83.[160]
	Sensitivity to change	In Australia, the WEMWBS has been used to test the efficacy of two positive psychology interventions, and scores did change over 3 collection occasions.[442] Similarly, the measure has been used to evaluate the effectiveness of biofeedback therapy for women suffering postnatal depression.[443] Collins et al [444] calculated reliable and clinically significant change scores for the WEMWBS and reported a reliable change score of 6.93, meaning 69% of 134 participants reliably changed. Clinically significant change was assessed against the Jacobson and Truax [445] criteria that places the cut-off at the point halfway between the mean of the treatment group at the first collection and mean of a comparison group at the same collection occasion. By this criterion, 95 or 71% of the treatment group showed clinically significant change.
	Acceptability/Usability	An Italian version of the WEMWBS is available.[446] The WEMWBS is potentially valuable because it is a measure of mental well-being that focuses entirely on positive aspects of mental health. It has been used in national surveys of mental well-being in Scotland since 2006.[447]
Working Alliance Inventory (WAI)	**Validity**	Cattey et al[448] indicate that the WAI has good face validity. The WAISR (WAI short form) total score correlates with the Helping Alliance Questionnaire[449] total score (r = .71).[450] The short and the long forms of the WAI are highly correlated.[451] Bachelor[452] found that clients' and therapists' views on the alliance did not always agree. Ross et al [453], looking across therapists, clients and observers, found no difference in factor structure based on perspective but concluded that goals and tasks, given their high correlation (r = .90), may be a separate factor to bond (r = .69). Ultimately a three factor solution is generally agreed goals, tasks and bonds[454]
	Reliability	The WAI has demonstrated very good internal consistency, ranging across the subscales from .80 to .89[450]. Internal consistency for the total score of the short form has been reported as .92.[455] No studies examining test-retest reliability were identified.
	Sensitivity to change	Busseri and Tyler [451] found that changes in WAI scores were correlated with a composite improvement index which was made up of the number of target complaints (r = .42, p <.01) but these changes were less associated with changes in symptomatology (r = .19, p <.01).
	Acceptability/Usability	In a review of therapeutic alliance measures, Catty et al [448] identified seven measures of therapeutic alliance; all but the WAI had only been used once in research studies. It is therefore not surprising that it is described as the most commonly used measure of therapeutic alliance.[453]

MEASURE	PSYCHOMETRIC PROPERTIES	
Empowerment Scale	**Validity**	Rogers et al [456] found that the Empowerment Scale was moderately correlated with hope [457] (r = .67), a sense of recovery [35] (r = .67), the Segal [458] measure of personal empowerment (r = .45), and quality of life[459] (average of the subjective items, r = .44; item measuring satisfaction with life in general, r=.47). They found inverse correlations with measures of symptomatology, including the Colorado measure of psychotic symptoms[460] (r = -.39), and the Hopkins depression subscale [122](r = -.46). Corrigan et al[461] found a two-factor structure to the Empowerment Scale, with self and community orientations to empowerment. They found that self-empowerment was correlated with Lehman's Quality of Life Interview[459] (r = -.58) size of support network (Social Support Questionnaire[462], r = -.48) , self-esteem (Rosenberg Self Esteem Scale[463], r = .59) and psychiatric symptoms (BPRS[464], r = .51). Community-empowerment was associated with Needs and Resources Assessment[465] (NARA) total resources (r = -.46), Wechsler Adult Intelligence Scale—Revised (WAIS-R) vocabulary (r = .74). Chou et al[466] found that psychosocial rehabilitation increased empowerment, which increased quality of life.
	Reliability	Rogers et al[456] report an internal consistency of α = .82. Internal consistency for individual factors was factor 1-self-esteem, α = .82; factor 2-power and powerlessness, α = .59; factor 3-community activism and autonomy, α = .59; factor 4-optimism and control over the future, α = .45; and factor 5-righteous anger, α = .64. Corrigan et al [461] undertook a study of two-week test-retest reliability and found good reliability (r > 0.75) for six of the seven subscales: self-efficacy, powerlessness, self-esteem, effecting change, optimism/control over future, and group/community action. The reliability for the righteous anger scale was low (r = 0.38 p > .05)
	Sensitivity to change	No research was identified that explored the Empowerment Scale's sensitivity to change.
	Acceptability/Usability	The measurement of this construct is an important consideration with the increasing focus on recovery in mental health services.
Brief COPE	**Validity**	Cooper et al[467] divided the 28 items of the Brief COPE into 3 subscales: emotion-focused, problem-focused and dysfunctional. They found that dysfunctional coping was predicted by reporting more burden (β = 0.36; p < 0.001), more problem-focused coping (β = 0.31; p < 0.003) and less secure attachment (β = 0.25; p < 0.011). A factor analysis using a Greek sample found a slightly different factor structure than theoretically suggested, confirming that coping is complex and multidimensional.[468] In contrast, the factor structure of a Spanish sample [469] was almost exactly the same as that proposed during development[166].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	The internal consistency of the brief COPE has been reported as ranging from α = .67 to α = .88.[470] Cooper et al[467] reported the internal consistency for three subscales: emotion-focused, problem-focused and dysfunctional as ranging from acceptable to good at α = .72, α = .84 and α = .75, respectively. In a group of undergraduate students the measure has demonstrated acceptable internal consistency for individual items (range α =.45 to α.85), test-retest reliability at 6 to 8 weeks for individual items (range .48 to .89) along with convergent and divergent validity with correlations between the brief COPE and measures of personality[471]
	Sensitivity to change	Cooper et al[467] report that the measure is sensitive to change but this sensitivity varied across the subscales they identified.
	Acceptability/Usability	The measure has been used for a variety of research purposes ranging from understanding the coping strategies of those with mild brain injuries [472] to how medical students cope[473].
Revised NEO Personality Inventory (NEO-PI-R)	**Validity**	Factor analysis has generally supported the proposed five-factor structure of the NEO-PI-R[474], with only slight variation across cultures.[475] One of the creators of the inventory has viewed cross-observer agreement not as a form of reliability but as a form of validity in that self-reports and observer reports should correlate for personality traits, and found moderate agreement, with an intra-class correlation of .35 between pairs of samples (self-report/observer-report)(p < .001, α = 0.86).[476] Scores on the inventory correlate with the diagnosis of personality in the expected manner.[477, 478]
	Reliability	The internal consistency of the measure has been reported for the scales of the NEO, and ranged from good to excellent (α = 0.93 for neuroticism, .87 for extraversion, .89 for openness, .76 for agreeableness and .86 for conscientiousness).[479] In an Australian study, the test-retest reliabilities of all five scales were good and ranged from .80 (agreeableness) to .87 (openness) for the short form of the inventory, with internal consistencies of α = .87 (neuroticism), .80 (extraversion), .77 (openness), .75 (agreeableness) and 0.85 (conscientiousness) reported.[480] However, there is evidence that the reliability does vary across cultures.[481]
	Sensitivity to change	The inventory has been used to demonstrate change in personality with Alzheimer's disease[482] and changes in all domains of personality have been found pre- and post-drug rehabilitation treatment.[474]
	Acceptability/Usability	The measure has been used for a variety of purposes including use as part of employment selection processes.[483]
General Help	**Validity**	The GHSQ has found that the intention to seek help varies across types of problems and sources of help.[484]

MEASURE	PSYCHOMETRIC PROPERTIES	
Seeking Questionnaire (GHSQ)	**Reliability**	Using two problem-types: personal-emotional problem and suicidal thoughts, and three help-sources: doctor/GP, mental health professional (such as a counsellor, psychologist, or psychiatrist), telephone helpline (like Lifeline or Kids Help Line), the GHSQ has demonstrated excellent internal consistency, enabling the measure to be reported as an overall intentions scale.[403]
	Sensitivity to change	No studies were identified that examined the GHSQ's sensitivity to change.
	Acceptability/Usability	Primarily the measure has been used for a series of studies to better understand help seeking behaviour.[154, 168, 403, 484]
Carer Measures		
Involvement Evaluation Questionnaire (IEQ)	**Validity**	Testing of the psychometric properties of the IEQ is limited to only a few studies, but these have shown the measure to have satisfactory content, construct and concurrent validity. The IEQ, which can be regarded as a measure of objective burden, was pitted against a Dutch translation of the Maslach Burnout Inventory, which assesses subjective burden. The shortened version of the IEQ was also found to correlate well with the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) in samples of carers of consumers in inpatient and community settings[485]. The IEQ correlated well with emotional exhaustion and general subjective burden.[56, 486] In samples of carers of people with psychosis and carers of people with affective disorders, content validity was found to be satisfactory [487].
	Reliability	The relevance of test-retest reliability to the IEQ might be questioned, since it considers consequences for carers, which may change over time. Nonetheless, the EPSILON study found the test-retest reliability of the IEQ to be relatively high, at least .70.[488] The concept of inter-rater reliability is not relevant for the IEQ, since it is designed to be self-administered, so there is only one rater.
	Sensitivity to change	There are indications that IEQ may be sensitive to change, but further work in this regard is needed[485, 489].
	Acceptability/Usability	As part of work to identify a carer outcome measure, the Involvement Evaluation Questionnaire (IEQ) was seen as a potential candidate measure for testing in Australia.[486]

MEASURE	PSYCHOMETRIC PROPERTIES	
Burden Assessment Scale (BAS)	**Validity**	The content validity of the BAS has been tested in several ways. During its development in the United States, a carer advisory group of six family members of consumers with long-term mental illness reviewed the instrument for clarity and completeness. Following this, the BAS was piloted among two United States samples, one of carers who were nominated by a consumer and the other of carers who sought services for themselves. Both samples ranked items according to the nature of burden in similar ways, with 'worry about the future' receiving the greatest emphasis. These findings were interpreted as evidence of the BAS being a valid measure of burden.[490] The Swedish version of the BAS has also been examined for concurrent validity. Specifically, it was tested against the Quality of Life Index (QLI), the Global Assessment of Function scale (GAF) and the Clinical Global Impression scale (CGI) in two samples: parents of outpatients with schizophrenia and a reference group from the broader community. A correlation was observed between burden and both quality of life and severity of illness.[486, 491] Ivarsson et al [492]found a three factor structure to the BAS 1. activity limitation , 2, feelings of worry and guilt and 3, social strain
	Reliability	No research was identified on reliability of the BAS.
	Sensitivity to change	The original trial of the BAS in the United States found that the BAS has sufficient sensitivity to change to be used as an outcome measure for program evaluation. Specifically, it demonstrated that program participation reduced levels of burden over time.[490]
	Acceptability/Usability	As part of work to identify a carer outcome measure, the Burden Assessment Scale (BAS) was seen as a potential candidate measure for testing in Australia.[486]
Zarit Burden Scale (ZBS)	**Validity**	Concurrent validity has been demonstrated by correlating ZBI scores with a single global rating of burden (r = .71) and with responses to the Brief Symptom Inventory (r = 0.41).[493] Higher ratings of consumer dependence have also been shown to be associated with increased carer burden.[494]
	Reliability	Bachner and O'Rourke[493] could only identify 11 of 102 studies that reported test-retest reliability coefficients for the ZBS. The mean correlation coefficient was r = .59 over an average interval of 31.56 months (SD = 27.72); however, considerable variability was observed, ranging from r = .24 to .89. They argue that this variability could be the result of very short retest periods being compared to very long retest periods, or alternatively an indication that the measure is sensitive to change. Internal consistency is excellent at α = .92.[495]
	Sensitivity to change	The ZBS has been used to examine the effectiveness of case management services for the elderly but no significant changes were found at 4 and 12 months.[496]
	Acceptability/Usability	The ZBS has been used in a wide variety of studies of carer burden and is reported as one of the most commonly used measure of caregiver burden in the world.

MEASURE	PSYCHOMETRIC PROPERTIES	
Carer Qol-7D+VAS	**Validity**	The CarerQol-7D+VAS is a relatively new instrument, and only a small number of studies have been undertaken to assess its psychometric properties. Brouwer et al [497] in the initial testing of the measure indicated the instrument has good construct validity by comparing the burden component and the valuation component, and found that as scores rose on the burden component, they decreased on the valuation component and took this as evidence of goo internal consistency. They concluded that the measure had excellent concurrent validity by testing the burden component of the CarerQol-7D+VAS against the Carer Strain Index (CSI) and the Self-Rated Burden Scale (SRB), and tested the valuation component against the Process Utility (PU) instrument. Both the CSI and the SRB are concise instruments that assess carer burden, and the PU is a comparative measure of happiness between a current caregiving role and a hypothetical non-caregiving role. The concurrent validity of both the burden component and the valuation component of the CarerQol-7D+VAS were found to be excellent.
	Reliability	No psychometric research was identified that assessed the test retest reliability of the CarerQol-7D+VAS.
	Sensitivity to change	No research was identified that assessed the sensitivity to change of the CarerQol-7D+VAS.
	Acceptability/Usability	The CarerQol provides a good description of the impact of care giving on informal caregivers and can be seen as a useful instrument to include this effect in economic evaluations [498]. As part of work to identify a carer outcome measure, the CarerQol-7D+VAS was seen as a potential candidate measure for testing in Australia.[486]

Note that some measures of cognition and emotion were identified but excluded from the review because they did not fit the specific service types or the reviewers' perceived needs of the sector. For example, the Hospital Anxiety and Depression Scales[299] is an often-used measure of anxiety and depression, but is specifically designed for use in medical populations and so was excluded on this basis.

3.3 Functioning

Table 9. Profile of functioning measures

Functioning is a poorly defined concept [499, 500]; there is a lack of consensus concerning both its definition and how best to evaluate functioning in consumers with mental health disorders[501, 502]. However, it is now widely accepted that it is an important domain to measure. Whilst self-report measures provide consumers with the opportunity to be actively involved in care, the consumer's insight may impact on the results seen; however, ratings by others may be limited by poor knowledge of the person's day to day life.[503, 504]

MEASURE	YEAR	COUNTRY	DESCRIPTION
Life Skills Profile 39 (LSP-39)[505] **Life Skills Profile 16 (LSP-16)[506]** **Life Skills Profile 20 (LSP-20)[507]**	1989 1998 2001	Australia	The original LSP was produced by Rosen and colleagues in Sydney to measure constructs relevant to survival and adaptation in the community for individuals with schizophrenia and chronic mental illness.[505, 508] The LSP-16 was derived for use in the Australian Mental Health Classification and Service Costs (MH-CASC) Project [506, 509, 510] to reduce the rating burden on clinicians. The original developers were commissioned by the MH-CASC team to create the shortened version. The original LSP-39 comprised five subscales: communication, social contact, non-turbulence, self-care and responsibility[505]. An alternative set of subscales was proposed by Trauer et al[511]; Bizzare, Withdrawal, Self-Care, Compliance and Anti-social. Whilst the LSP-20 also comprises these 5 subscales[507], the LSP-16 comprises four subscales; withdrawal, self-care, compliance and anti-social behaviour[506]. The clinician rates the consumer's general functioning on each of the items in terms of his or her behaviour over the preceding three month period[512]. It is rated on a four-point scale rating from 0 (no problem), through 1 (slight problem) and 2 (moderate problem) to 3 (extreme problem). Individual item scores, subscale scores and total score can then be calculated. The LSP-16 can be regarded as a measure of impairment since high scores indicate poorer functioning. By contrast the LSP-20 can be regarded as a measure of strengths since higher scores indicate better functioning. The LSP-20 can be scored in either direction[507].

MEASURE	YEAR	COUNTRY	DESCRIPTION
Social Function Scale (SFS)	1990	United Kingdom	The SFS is a 79-item measure assessing social functioning across 7 domains: (1) withdrawal, (2) interpersonal behaviour, (3) prosocial activities, (4) recreation; (5) independence-performance; (6) independence-competence and (7) employment/occupation.[513, 514] This can be self-report or an informant interview, although it is generally administered to an informant. Items are scored on a 4-point scale, with higher scores indicating a higher level of functioning.[515] The reference period for the scale is unspecified.
Work and Social Adjustment Scale (WSAS)	1986	United Kingdom	The WSAS is a self-report scale of functional impairment attributable to an identified problem. It is a five-item scale with each item rated on a nine-point scale from 0 = *no impairment* to 8 = *very severe impairment*.[516]
Personal and Social Performance Scale (PSP) n.b. The PSP is the new version of DSM-IV Social and Occupational Functioning Scale (SOFAS)	2000	Italy	The PSP measures four main domains of person and social functioning: (1) socially useful activities, (2) personal and social relationships), (3) self-care and (4) disturbing and aggressive behaviours. It provides a single overall rating ranging from 1 to 100 with specific criteria for each 10-point interval. It is possible to rate more than adequate function from the 91 to 100, up to 70 ratings refer to only minor difficulties, from 70 to 31 manifest disabilities to various degrees, while under 30 indicates the persons functioning is so poor that intensive support and supervision is required[517]
Global Assessment of Functioning (GAF)	1987	United States	The DSM-III-R provided a rating scaling, the Global Assessment of Functioning (GAF), for measurement of overall psychiatric disturbance.[518] It is designed as an easy, brief measure that integrates within a single score three different dimensions of functioning: psychological, social and occupational. It can be either a single score (only the most severe of the symptom and functioning value recorded) or separate scores for symptoms (GAF-S) and functioning (GAF-F).[519] For both single scale and dual scale there are 100 scoring possibilities (1 – 100). The 100-point scales are divided into intervals each with 10-points (e.g. 31-40). The 10-point intervals have anchor points describing symptoms and functioning that are relevant for scoring.[520] A lower GAF score indicates greater dysfunction.
Social Functioning Questionnaire (SFQ)	1989	United Kingdom	The Social Functioning Questionnaire (SFQ) is an 8-item self-report instrument. Each item is rated on a 4-point scale (0-3), with a higher rating indicting greater problems with functioning. It that was adapted from the longer Social Functioning Schedule, a semi-structured interview, and measures an individual’s perception of functioning [521]. Items relate to the areas of work and home tasks, financial concerns, relationship and family, sexual activities, social contacts and spare time activities.

MEASURE	YEAR	COUNTRY	DESCRIPTION
The World Health Organisation Disability Assessment Schedule 2.0 (WHODAS II/2.0)	2009 (WHODAS 2.0)	International (involved piloting in 19 countries	The World Health Organisation Disability Assessment Schedule 2.0 (WHODAS 2.0) is an instrument developed by the World Health Organisation (WHO) to assess behavioural limitations and restrictions to participation experienced by an individual. It evaluates functioning in major life domains: (i) Cognition (understanding and communication); (2) Mobility (ability to move and get around); (3) Self-care (ability to attend to personal hygiene, dressing and eating, and to live alone); (4) Getting along (ability to interact with other people);(5) Life activities (ability to carry out responsibility at home, work and school); and (6) Participation in society (ability to engage in community and recreational activities)[522]. It is a 36-item questionnaire, which can be either self-administered or administered via interview, and assesses functioning over the past 30 days
The Multnomah Community Ability Scale (MCAS)	1994	United States	The Multnomah Community Ability Scale (MCAS) is a 17-item instrument measuring aspects of the client's community functioning in the past 3 months (past year for behavioural problems domain). There are 4 subscales: (1) interference with functioning (5 items); (2) adjustment to living (3 items); (3) social competence (5 items); and (4) behaviour problems (4 items). These items are rated on a 5-point scale according to difficulties in these areas, with a higher score indicating higher level of functioning.[523]
Canadian Occupational Performance Measure (COPM)	1994	Canada	The Canadian Occupational Performance Measure is a semi-structured interview designed to identify individual areas of difficulty in occupational performance across three areas; self-care, leisure and productivity. The first phase includes problem identification by the client. Each problem is rated on a scale from 1 (not important) 10 (very important). The client then selects the five most important identified problem areas. For each of these problems the client rates their current level of performance and their satisfaction with that performance on scales of 1 (unable to perform, not satisfied) to 10 (able to perform, extremely satisfied). Scores are totalled for both performance and satisfaction and average over the five areas. When reassessed the client reviews the identified areas of concern and rates their performance and satisfaction. The initial and reassessment are then compared and evaluated for change [524]
Children's Global Assessment Scale (CGAS)	1983	United States	The Children's Global Assessment Scale (CGAS) was adapted from the Global Assessment Scale (GAS) for adults[525] by Shaffer and colleagues of the Department of Psychiatry, Columbia University, United States[526] to provide an assessment of functioning for those aged 4 to 16. The CGAS is clinician-administered and provides a single global rating of a child or adolescent's lowest level of functioning over the previous two weeks. Ratings range from 1 - severe dysfunction to 100 - superior functioning, and the threshold of psychopathology is suggested to sit between 61 and 71.[527, 528]

MEASURE	YEAR	COUNTRY	DESCRIPTION
Parents' Evaluation of Developmental Status (PEDS)	1997	United States	The Parents Evaluations of Development Status (PEDS) is a 10-item questionnaire instruments used to identify parents' concerns about their child's development and behaviour[529]. The first question is open ended and prompts parents to describe any concerns about behaviour, learning and development. The following 8 questions cover concerns in different developmental domains with response options: no/yes/a little and the final question probes any additional concerns[530-533].
Drug Use Disorders Identification Test (DUIT)	2005	Sweden	The DUIT is an 11-item self-report questionnaire developed to screen individuals for drug problems.[534, 535] The 11 items in the DUDIT were chosen to give information on the level of drug intake and fulfilment of selected criteria for substance abuse/harmful use and dependence according to the ICD-10 and DSM-4 diagnostic systems. There is a long version, the (DUDIT-E), which identifies the frequency of illicit drug use (*D*) along with the positive (*P*) and negative (*N*) aspects of drug use and includes evaluation of treatment readiness (*T*) [536]. Nine questions are scored on 5-point scales, ranging from 0 – 4, and two questions are scored on 3-point scales, with values of 0,2 and 4. Total scores range from 0 – 44, with a higher score indicating a more severe drug problem.
Alcohol Use Disorders Identification Test (AUDIT)	1989	World Health Organization. Switzerland	The AUDIT is a 10-item self-report instrument designed to identify individuals whose use of alcohol places them at risk of alcohol problems or who are experiencing alcohol related problems.[537]It covers the domains of alcohol consumption, drinking behaviour and alcohol related problems[538]. Responses to each question are scored from 0 to 4, with a maximum total score of 40[538]
Specific Levels of Functioning Scale (SLOF)	1983	United States	The specific levels of functioning scale is a 43-item instrument covering 6 domains; (1) physical functioning, (2) personal care skills, (3)interpersonal relationships, (4)social acceptability, (5) activities and (6) work skills. Ratings are made on a 5-point Likert scale indicating the level of assistance the person needs to perform the task. Scores range from 43-215, a higher rating indicates better functioning. The scale also includes an open-ended question asking the informant if there are any other areas of functioning not covered by the instrument that may be important to measure[539]. The SLOF can be self-rated or rated by an informant, although most commonly it is rated by an informant. If it is not self-rated an additional item asks how well familiar the person is with the skills and behaviour of the person on a 5-point scale., "not well at all' to 'very well'[540].

MEASURE	YEAR	COUNTRY	DESCRIPTION
Independent Living Skills Survey (ILSS)	1985	United States	There are two versions of the ILSS ,a self report (ILSS-SR) and an informant report (ILSS-I). The ILSS-R is a 103-item instrument covering 12 areas of basic community living skills: personal hygiene (6 items), appearance and care of clothing (12 items), care of personal possessions and living space (9 items), food preparation (9 items), care of one's own health and safety (10 items), money management (10 items), transportation (7 items), leisure and recreational activities (13 items), job seeking (6 items), job maintenance (3 items), eating behaviours (9 items) and social interactions (9 items)[541]. Responses to each question are rated on a five-point scale (never, sometimes, often, usually, and always). The answers are scored from 0 (never) to 4 (always), then summed and averaged per functional area[541] . The ILSS-SR is a 70-item instrument covering 10 areas: personal hygiene (12 items), appearance and care of clothing (9 items), care of personal possessions and living space (6 items), food preparation (7 items), care of personal health and safety (7 items), money management (5 items), transportation (5 items), leisure and recreational activities (12 items), job seeking (4 items) and job maintenance (3 items)[541]. Responses to each question are a 'yes', 'no' or 'not apply'. Answers are summed (0=no, 1=yes) and then average per functional area[541].
Social Behaviour Schedule (SBS)	1986	United Kingdom	The social behaviour scale covers 21 areas of behavioural difficulties experienced or exhibited; spontaneous communication, incoherence of speech, odd or inappropriate conversation, inappropriate social mixing, hostility, demanding attention, suicide ideas or behaviour, panic attacks and phobias, overactivity and restlessness, laughing or talking to self, acting out bizarre ideas, posturing and mannerisms, socially unacceptable habits or manners, violence or threats, depression, inappropriate sexual behaviour, poor self care, slowness, underactivity, poor attention span, other behaviour . Most of the items are rated on a scale of 0 (no problem or acceptable behaviour) to 4 (serious problem) of the persons functioning over the previous month[542]. Two overall scores can be calculated: severe behaviour problems scale (BSS) which are those behaviours which have been rated 3 or 4; and the mild and severe problems score (BSM), includes those behaviours which have been rated 2,3 or 4[542].

MEASURE	YEAR	COUNTRY	DESCRIPTION
Child and Adolescent Functional Assessment Scale (CAFAS)	1989	USA	The CAFAS comprises 8 scales which assess the child or youth (6-17 years; School/Work, Home, Community, Behaviour Toward Self and Others, Moods/Emotions, Self-Harmful behaviour, Substance Use and Thinking. Two additional scales assess the environment; material needs and Family/Social support. The level of impairment is determined by the behaviour descriptor(s) which captures the most severe functioning in a specified time period[543]. For each scale the possible score ranges from 0 to 30 (by tens), and can range from 0-240 (if using the 8 scale version) or from 0-150 (if using the original 5 scale version) A higher score indicates a lower level of functioning[544]
Columbia Impairment Scale	1993	United States	The CIS is a 13-item scale that covers four domains; interpersonal relations, broad psychopathological domains (e.g. anxiety, depression or problem behaviour), functioning in job or school and use of leisure time, among youth aged 9-17 years [545]. There is a child and parent/carer version. Items are scored on a 5-point Likert scale ranging from 0(“no problem”) to 4 (“very big problem”), scores range from 0 – 52, with higher scores indicating greater impairment[546].

Table 10. Psychometric properties of functioning measures

MEASURE	PSYCHOMETRIC PROPERTIES	
Life Skills Profile 16, 20, 39 [506]	**Validity**	When developing the LSP-39, Rosen et al[505] conducted a principal components analysis that led to the retention of five components that together accounted for more than half (53.8%) of the total variance. These five components became defined as the original five subscales: Communication, Social contact, Non-turbulence, Self-care and Responsibility. Trauer et al[511] carried out further psychometric testing, replicating Rosen et al's method, and also identified five components that accounted for just over half (50.7%) of the overall variance. However, following a confirmatory factor analysis, they recommended the factor structure be modified with the alternative subscale structure: Bizarre, Withdrawal, Self-care, Compliance and Anti-Social behaviour, since this model better fitted the data. The correlations with the original subscales were .85, .95, .98, .97 and .90 respectively. Andrews et al[547] suggested the five subscales could be further divided into two dimensions, which they described as 'general impairment' and 'difficulty'. In testing the concurrent validity of the LSP-39, LSP-20 and LSP-16, the LSP-39 has been shown to perform well against the Health of the Nation Scales (HoNOS)[191, 508, 548, 549], the Katz Adjustment Scale[508], the Multnomah Communities Ability Scale[550], the Strauss-Carpenter Levels of Functioning Scale[550], the Global Assessment of Functioning Scale[550-552], the role functioning scale[191], the Quality of Life Scale[553], the interviewer-rated Quality of Life Scale[548], the Social Behaviour Schedule[552], the Research Associated Functional Level Scale[511] and the Global Assessment Scale[548]. However, it has demonstrated poor or mixed performance against the Basis-32[191], the Mental Health Inventory[191], the Short Form-36[191], the General Wellbeing Scale[553, 554], the Brief Psychiatric Rating Scale[508, 511], the Dysexecutive Questionnaire[548, 552], Cantril's Ladder[548]and the Affect Balance Scale[548]. It should be noted that several of the latter are self-report measures, so the poor levels of correspondence may reflect the expected gap between clinician and self-report ratings. In addition, some (e.g. the BPRS) are symptomatic scales, and it would be expected that symptoms might vary in intensity and not co-vary with the more robust and enduring, though often slower, changes in functioning or disability. The LSP-20 has been shown to produce comparable results to the LSP-39, and to the Positive and Negative Syndrome Scale (PANSS).[507] Rosen et al[505] found that low scores on the Responsibility and Non-turbulence scales of the original LSP-39 were characteristically scored by younger people. Other studies have found that total scores on the various versions of the LSP can distinguish between consumers based on their locus of accommodation and/or care, with higher levels of disability being at least moderately associated with those experiencing frequent changes in accommodation (as opposed to those in stable living environments) and/or living in

MEASURE	PSYCHOMETRIC PROPERTIES	
		long-term residential care settings or in the family home (relative to those living independently or semi-independently).[547, 554-558]. Still other studies have found the LSP to be able to discriminate between consumers on the basis of their levels of social functioning, as measured by factors like unstable employment, low-grade accommodation (e.g. hostel or refuge), welfare dependency, police contact and complaints by neighbours.[505, 547, 559] Several Studies have examined the predictive validity of the LSP-16 and LSP-39. In general, these studies have shown that the different versions of the instrument can predict outcomes relating to retention in the community[560], hospital readmission[554, 559], change in locus of care[555], length of inpatient stay[561, 562] and overall costs[554, 562]. An exception to this rule was a study by Parker et al[551], in which the other versions failed to support the predictive validity of the LSP-39.
	Reliability	A number of studies have examined the internal consistency of the LSP-39, as measured by Cronbach's alpha. In these studies, the internal consistency has been reported as moderately high, with subscale and total score correlations ranging from .64 to .88[505, 508, 511] and .93 to .94[191, 511, 550], respectively. The Communication subscale has been shown to have the poorest internal consistency.[505, 511] The few studies that have examined the test-retest reliability of the LSP have reported a high overall reliability score, albeit for the LSP-39 only.[15, 191, 508] Parker et al[508], for example, established high test-retest reliability for case workers, residential carers and parents, each of whom were asked to rate the same person with relatively stable chronic schizophrenia at two points in time (one month apart). A number of studies have assessed the inter-rater reliability of both the LSP-39 and the LSP-16. These studies found the overall agreement between pairs of raters on the LSP-39 to be fair to moderate [15, 191, 511] or moderate to good[505, 508, 511]. Some studies have found that raters of similar backgrounds are more likely to show high correlations between their respective ratings[508], whereas others have found the background of raters to have little bearing on levels of agreement[505].
	Sensitivity to change	A number of studies have examined the degree to which changes in LSP scores correlate with some external measure of change. Several of these studies have compared the LSP's ability to detect change against other more established measures of outcome. Stedman et al[191], for example, found significant associations between changes on the LSP and changes on the Global Change Ratings Scale, the Modified Clinical Global Impressions Scale (CGI), the Role Functioning Scale (RFS) and the HoNOS. The latter finding was reproduced by Parker et al[551] who also found a moderate association between the LSP and the Global Assessment of Functioning (GAF). Other studies have examined changes in LSP scores for different consumer groups that would be expected to show greater or lesser degrees of improvement depending on their treatment circumstances. Typically,

MEASURE	PSYCHOMETRIC PROPERTIES	
		these studies have found, as hypothesised, that the LSP demonstrates greater levels of improvement in those who participate in intensive case management than in those who undergo routine case management.[563-568] However, there have been some exceptions to this rule.[569, 570] Still other studies have used self-reported improvement or deterioration as the ‘gold standards’ against which to assess the sensitivity to change of the LSP. Stedman et al[191], for example, conducted an analysis of the LSP-39 scores over time for groups showing differing levels and directions of self-reported change. This study found that LSP scores worsened in the group who reported a decline in their levels of functioning, but there was no association between LSP change score and self-reported change for any other group.
	Acceptability/Usability	Published commentaries have reported that the overall feasibility and utility of the LSP-16 as moderately high.[551] However, only the work of Stedman et al[191] has really put this assertion to the test. These authors elicited the views of service providers about their experience with using the LSP via a purpose-designed utility questionnaire. The majority of service providers had little difficulty with the language and viewed the questions as relevant, useful and effective in measuring outcomes for consumers. Respondents in public sector psychiatric settings, in particular, rated the LSP more highly than other observer-rated measures.[191, 571]
Social Function Scale (SFS)	**Validity**	A number of studies have examined the construct validity of the SFS by conducting factor analyses. Single component solutions were found across all studies, accounting for 54.6% to 60% of variance.[513, 514, 572, 573] However, Hellevin et al[573] conducted further factor analysis in different cohorts of consumers. For those with Schizophrenia, a two-component solution was proposed with subscale 7 (employment) forming the second component. For those with bipolar disorder, the one-component solution was preferred. For healthy controls, a three-component solution was proposed with subscales 4 (recreation), 5 (independence-competence) and 6 (independence –performance) forming the 1st component, subscales 1 (social engagement/withdrawal) and 2 (interpersonal behaviour), forming the second component and subscale 7 (employment) the third component. Only a small number of studies have compared the SFS with more established measures. The SFS positively correlates with the Global Assessment of Functioning (GAF).[513, 573, 574]. Vazquez Morejon further found that some SFS scales showed a moderate correlation with the Social Behaviour Assessment Schedule (SBAS).[513] In the development of a new functioning measure, the SFS was used to assess its construct validity.[575] Studies have found that the SFS can discriminate between groups of consumers who were employed and unemployed[513, 573] and between consumers with schizophrenia and those with bipolar disorder. Those

MEASURE	PSYCHOMETRIC PROPERTIES	
		with schizophrenia tended to have lower functioning scores on the SFS[573, 576], although Dickerson et al[577] also found this it was not a significant difference. Consumers with positive and negative symptoms generally had lower functioning scores.[514] The SFS has discriminated between healthy controls who report better functioning than people with schizophrenia and bipolar disorder.[572]
	Reliability	A number of studies have examined the internal consistency of the Social Function Scale (SFS), as measured by Cronbach's alpha. In these studies, the internal consistency has been reported as moderately high, with subscale and total ratings ranging from .40 to .90[513, 514, 572, 573] and .80 to .89 [513, 514, 572, 573], respectively. The subscale score of interpersonal behaviour has been found to have low internal consistency at 0.40[513], although this was not found in the other studies. One study assessing the test-retest reliability of the SFS found that the instrument produces consistent results when rated by the same rater at different points in time, with correlations varying between .66 and .80 across all scales.[513] Considering the inter-rater reliability of the SFS, some studies have assessed self-report against another informant completing the SFS. These studies have found moderate to good overall inter-rater reliability.[513, 514] Vazquez et al (2000) found that there was a tendency towards a higher score in the self-report version; therefore, while there is a correlation, there is not complete concordance.[513]
	Sensitivity to change	Sensitivity to change has been evidenced by Barrowclough, who found that those consumers who improved on the social functioning scale following behavioural family intervention also had lower relapse rates, and their relatives showed significant change from high to low expressed emotion.[578]
	Acceptability/Usability	There are a number of authors who are positive regarding the usability and acceptability of the SFS. It has been suggested that it is useful as it effectively discriminates between groups of consumers[514, 573], and it is also appropriate for use with different consumer groups,[573] although it has also been suggested it is best used with stable disorders.[500] Although it has been proposed that the SFS requires little clinician time[513, 514], Brissos et al[501] proposed that it may be too long for routine use in practice. It's acceptability to consumers and carers has been proposed by Birchwood[514]; although, this has not been investigated with consumers and carers. It has been suggested that the SFS provides useful information about both abilities and activities[513, 514], is relevant to clinicians and researchers[514], and Vazquez et al [513] further suggests it can be used to assess change and progress over time. It has been reported as being used in a number of studies concerning the assessment of schizophrenia.[500] The SFS was selected as one of two scales to measure social functioning by a 48-member expert panel; it was rated highly by members in terms of sensitivity, practicality, usefulness to multiple raters and comprehensiveness of the scale.[515]

MEASURE	PSYCHOMETRIC PROPERTIES	
Work and Social Adjustment Scale (WSAS)	**Validity**	Three studies examined the factor structure of the WSAS as measured by a principal component analysis. Rogers et al[579], Mataix-Cols et al[580] and Jansson-Frojmark[581] all extracted a one-factor solution, accounting for 68%, 47 to 73% and 73.7% of the total variance, respectively. A number of studies have considered the concurrent validity of the WSAS by assessing the relationship between symptom severity and the WSAS, which is designed to measure functional impairment that might be attributable to an identified problem or disorder. The WSAS has shown to be moderately correlated with the Hamilton Rating Scale for depression[516], the Yale-Brown Obsessive Compulsive Checklist[516], the Fear Questionnaire Subscales[580], the one-item depression scale[580], the Eating Disorder Examination Questionnaire –version 4 subscales [582] and the Hospital and Anxiety Depression Subscales[582]. The WSAS has also been shown to have a positive correlation with the functioning item on Insomnia Severity Index[581], and the BSL version of the WSAS had a positive correlation with the functioning domains of the CORE-OM[579]. The WSAS has also been shown to be able to discriminate between groups of differing initial phobic severity (FQ total phobia).[580] Similarly, Tchnturia et al found that the WSAS could discriminate between those with anorexia nervosa and those of a healthy weight.[582] It can also discriminate those with a high number of physical symptoms, severe fatigue, depression, anxiety, poor sleep quality and poor physical fitness.[583]
	Reliability	A number of studies have examined the internal consistency of the WSAS. In these studies, Cronbach's alpha has ranged from .57 to .94, indicating a moderately high level of internal consistency.[516, 580, 581, 583, 584] The translated British Sign Language (BSL) version also showed good internal consistency at .88.[579] The internal consistency was also found to be marginally higher for self-report (.71-.90) than when used by an assessor (.57-.85).[580] Mundt et al found the test-retest reliability of the WSAS to be moderate, and similar reliability was found with each of the items, ranging from .70 to .75.[516] Jansson-Frojmark found high test-retest reliability at .90 to .99.[581] The study by Mataix-Cols et al examined the correlation between self-assessed and clinician ratings.[580] Agreement between self- and assessor-rated was high put imperfect (.55–.61). Consumers tended to rate themselves as more disabled on the WSAS total than did clinicians.
	Sensitivity to change	Two studies have examined change in WSAS over time in given settings, hypothesising that there should be a decrease in severity as the consumer nears the end of an episode. These studies both found a significant decrease in the WSAS from pre- to post-treatment.[580, 581, 583] One study also used consumer judgement as the standard against which to judge whether change has occurred and, if so, whether the WSAS is capable of detecting it. Those consumers who reported significant improvement

MEASURE	PSYCHOMETRIC PROPERTIES	
		scored significantly lower than for those indicating little or no improvement.[516]
	Acceptability/Usability	Many authors have proposed that the WSAS is a useful tool to measure disability with a range of consumer populations and additionally was able to discriminate between disorders[516, 580-583]. Both Jansson-Frojmark[581] and Mataix-Cols[580] et al further proposed a range of scores on the WSAS that would discriminate between levels of severity in insomnia and phobic disorders respectively. The adapted BSL version was also found acceptable in that population and has been adopted as part of the BSL health minds project, a specialised IAPT service for deaf people in some parts of the UK[579]. In regard to its feasibility or use as a routine outcome measure, the WSAS has been described as brief and simple[516, 580, 582, 583], and has the potential for use in care planning and monitoring of a person's progress.[582, 583] It has also been use in a controlled trial as a primary outcome measure to assess treatment effectiveness.[584]
Personal and Social Performance Scale (PSP)	**Validity**	Kawata et al[585] examined the subscale structure of the PSP using exploratory factor analysis. It was found that the four factor model defined by the original developers had a good fit, all factor loadings exceeded .40, ranging from .47 - .80. Kawata et al[585], suggested that different sets of items were all assessing different types of functioning. Examining the concurrent validity of the PSP, the PSP has shown to perform well against the GAF (0.91)[503, 586-589], SOFAS (0.91)[586, 589], Mini-ICF[586, 589], Strauss-Carpenter Level of Functioning[590] and Quality of Life Scale[585]. Although, Brissos et al[503] found that the correlation with the GAF was lower for the PSP disturbing and aggressive behaviour item. Several studies have also found that the PSP is associated with differing levels of symptoms; those with higher symptom severity, as measured by the PANNS, tended to have lower functioning scores, and those with lower CGI-S scores had higher PSP scores.[503, 585, 587, 590, 591] On further examination, it was found that the highest significant correlation was found between the negative symptom subscale of PANSS and total score of PSP[585-587, 589], although there were still significant correlations between the positive factor for PANSS and the general factor of PANSS[585-589]. Significant correlations were also found between neurocognition and PSP[503], although the category for disturbing and aggressive behaviour did not correlate significantly with 11 out 12 neurocognitive measures, suggesting a weaker association between neurocognition and function and such behaviours. The PSP has been shown to differentiate groups based on living situation: those living at home have higher functioning scores than those who were in institutional care[589, 591], and inpatients had lower scores than did outpatients[587]. Nicholl et al[592], have also found that a change in the PSP total score predicted future relapse.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	A number of studies have examined the internal consistency of the PSP as measured by Cronbach’s alpha. In these studies, α = .64 to .84, indicating that the PSP has a moderate to high level of internal consistency with low levels of item redundancy.[503, 585-589] Indeed, Kawata et al further found that the item correlations ranged from .43 to .65 and suggested that the reliability would be reduced if any items were deleted.[585] A number of studies examining the inter-rater reliability of the PSP have generally reported moderate overall reliability scores: 0.43–1.0.[503, 517, 586, 588, 589, 591] Interestingly, three studies have found that the highest level of agreement has been with ‘socially useful activities’ and the lowest with ‘aggressive behaviour’.[587-589] The test-retest reliability has been examined in a few studies which have reported high levels of reliability, ranging from .61 to .90.[586, 588, 590, 591]
	Sensitivity to Change	The ability of the PSP to detect genuine improvement or decline in functioning has been assessed in a number of studies. These studies have compared the PSP’s capacity to detect change against other more established measures of outcome. The PSP was found to perform commensurately with the CGIS-S[590, 591] and the PANSS[588, 590, 591]. Some of these studies have also examined clinically significant change on the PSP. Patrick et al[591] proposes a change of nine or greater may be clinically meaningful in patients with acute symptoms of schizophrenia, as determined by a change in PANSS scores, whilst Nasrallah proposed a seven-point improvement in the PSP may be clinically meaningful in a clinical setting.[590] Nicholl et al[592] suggests a 10-point decrement in PSP indicates a clinically meaningful decline in personal and social functioning as evidenced by a strong association with relapse status.
	Acceptability/ Usability	Most authors have been positive about the potential of the PSP. It is proposed that as it is brief and easy to use, it is appropriate for use in a clinical setting. [517, 585, 590, 592] Its multiple domains yielding a single score contribute to this ease of use[590], it requires minimal training[517], and has the ability to measure changes in social functioning over time[592]. In a review by Figueira et al[504] it is proposed that the PSP is one of the more widely used measures of social and person functioning; this strength lies in its ease of use with minimal training, as identified by other authors, and its reliability and sensitivity to change over time. However, its disadvantage is the degree to which its validity depends on the quality of the information available to guide clinical rating. Whilst Molodynski et al[593] proposed that the PSP had proved to be a popular and useful measure, a number of modifications have been proposed to improve its use. It is suggested that there needs to a be a normative population upon which to compare scores; rating of impairments associated only with schizophrenia should contribute to the score (although many studies have also used the PSP with other clinical groups), hostility and aggression should only be rated on the personal relationship domain, and a more rigid process for assigning the final score should be employed.

MEASURE	PSYCHOMETRIC PROPERTIES	
Global Assessment of Functioning (GAF)	**Validity**	There are limited studies assessing the content validity of the GAF, although there are many concerns that have been raised about its ability to measure the construct of function.[520] A study by Bacon et al[594], asked clinicians to list the three most important influences on their ratings on the GAF; the ratings were most influenced by symptom severity, followed by functional impairment and then a combination of the two. The authors propose that as the GAF ratings were strongly influenced by factors other than functional impairment, it is a limited measure of adaptive functioning/impairment. However, it could also be argued that the GAF is a measure of both symptom severity and functioning, though there should be caution if selected as purely a measure of functioning.[594] A number of studies have examined the construct validity of the GAF. [595, 596] Another study examined construct validity by determining elements that accounted for variance on scores; six significant predictors were found that accounted for 51.75% of variation in scores on the scale: access to and ability to use transportation, medication compliance, number of agency referrals, current living situation, current potential for violence and degree of social support. This suggests that the scale captures multidimensional information about social functioning.[597] Level of occupational functioning, however, was not a significant predictor.[597] A factor analysis by Hilsenroth et al.[596] confirmed the presence of two factors, but both related to the global score of symptoms and functioning. Pederson et al explored the validity of the separate GAF dimensions and found that both had concurrent validity by their associations with other measure of functioning and symptom distress, but equally they reflected different aspect of clinical impairment.[595] Although they are able to detect clinically significant differences between consumers, the results indicate that one global indicator of symptom distress and social dysfunction is appropriate, and it accounted for 90% of cases in the study.[595] The GAF has been shown to perform well against the Zung Depression test[598],Global Assessment of Relational Functioning Scale[596], self ratings of illness severity in outpatients[598], the Social and Occupational Functioning Assessment Scale[596], the self-reported SCL-90-R[596], SAPS[599], SANS[599], Social Behaviour Schedule[599], BDI[600], MADRS[600] and the SF-36[600]. However, the correlations with SAPS, SANS and Social Behaviour Schedule did not correlate significantly with the GAF on initial assessment.[599] GAF was found to not correlate with the Social Adjustment Scale global score[596] or the Inventory of Interpersonal Problems total score[596]. Similarly Roy-Byrne et al[601]found the GAF was unrelated to most measures of functional status on the Lehman Quality of Life Scale. In this study, however, it was more strongly correlated with measures of clinical status, PSAS[601]. Moos et al [602] also found that the GAF was more significantly associated with diagnoses and symptoms than with social or occupational functioning. Similarly, in a study comparing a range of measures for routine use in care, the GAF was found to perform the best in terms of an overall severity scale, but did not assess the needs of consumers

MEASURE	PSYCHOMETRIC PROPERTIES	
		for clinical and social needs as well as other measures.[603] While one study found that the GAF was independent of psychopathology[597], others have found it to differentiate between psychiatric diagnoses[602, 604, 605], previous inpatient care[602, 604], psychiatric symptoms[602, 604] and substance use[602, 604]. It was also found in one study that increased support and medication in the previous month was significantly associated with a higher GAF score[518], and that early increase in support levels showed improvement in scores[518]. The GAF, however, has been found not to discriminate between consumers with a dual and single diagnoses or between consumers who were homeless in comparison to those with residence.[601] Only a few studies have considered predictive validity of the GAF. In one study, those with greater impairment on the GAF were more likely to have received inpatient or residential care[602]; however, the ratings were not associated with the allocation or amount of service provision; indeed, those with more impairment were found to receive fewer services[602]. The authors suggest this finding casts doubt on the value of including GAF ratings as predictors of treatment outcome. Other studies have found that the GAF score has been shown to a have a significant negative correlation with length of inpatient stay.[606, 607]
	Reliability	Few studies were found that examined the internal consistency of the GAF. Greenberg[604], however, did examine internal consistency and found values ranging from α = .85-.86, indicating a high level of internal consistency. The inter-rater reliability of the GAF has been shown to differ between research and routine clinical settings. Moderate to high reliability has been found in a research setting, with ICC for single scale GAF scores ranging from .65 to .95[596, 598, 599, 608]and ICC for dual scale GAF scores ranging from GAF symptoms = 0.7[518] to GAF disability = 0.74[518]. The inter-rater reliability, however, has been shown to be much weaker when used in routine practice: .19 to .45.[600] In a study by Vatnaland[609], inter-rater reliability of researchers was compared with clinical staff; the ICC was much higher for researchers at .81 to .85 than routine care at .39 to .59. Similarly, Loevdahl et al found that reliability was unsatisfactory with untrained raters at .54[610].
	Sensitivity to Change	Very few studies have investigated the sensitivity to change of the GAF[520]. However, Schennach-Wolff et al found that change in function as measured by the GAF and the SOFAS correlated with a change in symptoms as measured by the PANSS.[611]
	Acceptability/ **Usability**	There are varied views regarding the use of the GAF in clinical practice. In an early field study, it was proposed that the GAF is a reliable, quick measure of disturbance in functioning which can be used readily used by multidisciplinary raters without extensive training.[597] In a later review by Burlingame et al[612],

MEASURE	**PSYCHOMETRIC PROPERTIES**	
		with the aim of selecting a measure for an inpatient service, the GAF performed well across a number of criteria, namely wide use, comparison with normative data, sensitivity to change, simplicity in administration and no cost. However, this was outweighed by concerns regarding accuracy as a result of combining a person's symptoms and functioning in a single rating. Some authors concur with Patterson et al[597], identifying that GAF can be used as a routine measure[597, 599] to facilitate monitoring over time in particular clinical services[597]. However, others raise concerns about its use for both monitoring consumer outcomes and service allocation[602, 604] and concerns about it measuring different constructs in a single rating are shared by others[520, 600, 613].Concerns with difficulties in rating the GAF have also been raised[519, 613], with a number of suggestions considered to improve its reliablity[519]. While some have proposed its ease of use and reliability with only minimal training[518, 599], others report low reliability and therefore acceptability in routine settings[600, 609].
Social Functioning Questionnaire (SFQ)	**Validity**	Validity Only a few studies have considered the concurrent validity of the SFQ in terms of its correlation with other established instruments that have been shown to validly measure related constructs. The SFQ has been shown to perform well against the Comprehensive Psychopathological Rating Scale[521], the California Quality of Life[614], the Mental Health Statistics Improvement Program Consumer Survey (MHSHIP)[614] and the Social Functioning Schedule[615]. A number of studies have examined the ability of the SFQ to discriminate between particular groups of consumers. Several studies have found high total scores to be associated with a diagnosis of personality disorder[521, 616-619], with a higher severity of symptoms[521, 617], and with a diagnosis of depression[616, 617]. In particular, the domains of close relationships, stress in completing tasks, use of spare time and family relationships have been found to have the greatest correlation with the symptom severity for people with personality disorder[617]. The SFQ has also been found to discriminate between groups of consumers differentiated on a range of treatment, service, and needs based indicators. The SFQ can discriminate between consumers with higher and lower education levels[614], between those in more or less supported accommodation[614], between people presenting at emergency in comparison to those in general practice[521] and is able to discriminate between those people discontinuing medication treatment and those who continued as well as those who discontinued with a different type of medication[620] Barrett et al[621] examined the costs associated with hospital, community and medication services for a cohort with anxiety disorder and found that higher scores on the SFQ were a significant predictor of

MEASURE	PSYCHOMETRIC PROPERTIES	
		higher costs.
	Reliability	Only one study has examined the internal consistency of the SFQ as measured by Cronbach's alpha. In this study Cronbach's alpha was .68 for the total score[622].
	Sensitivity to Change	The SFQ has been used to assess change in older persons with depression; however, in this particular study the SFQ showed no significant difference following CBT intervention[623]. In a separate study, however, the SFQ did show change over time in a group of consumers receiving intensive community support.[624] In a study using a problem solving approach in people with personality disorder, those in the intervention group had significantly better social functioning as measured by the SFQ compared with the control group.[625]
	Acceptability/ Usability	It is widely accepted that the SFQ's brevity as a measure of functioning lends to its use in routine clinical practice.[521, 618, 624] It has been reported that SFQ takes on average of four minutes to complete in comparison with the 20 to 30 minutes needed to complete the Social Functioning Schedule, from which it was derived.[521]
The World Health Organisation Disability Assessment Schedule 2.0 (WHODAS II/2.0)	**Validity**	A concern with the WHODAS II is fundamentally in its use as a self-report measure. Although Chopra (2004)[626] highlighted the importance of the consumer perspective, it has also been found that consumers with psychotic disorders often do not identify impairments in functioning owing to thought disorder or social difficulties. They generally reported little difficulty in self-care and only a small proportion highlighted problems with getting along with people.[627] This highlighted the disparity between clinicians' and consumers' viewpoints, and importantly raises concerns regarding the content validity of the WHODAS II in specific consumer groups. However, consumers have also reported that the assessments highlighted aspects of their disability, as well as identified issues that led to impairment in quality of life.[627] Usten et al[522] further reports that the instrument is valid across cultures and across a variety of diagnostic groups. A few studies have examined the factor structure of the WHODAS version 2.0 and II across different countries and language versions. Most studies found that the six-factor model was a good fit corresponding to the original scales proposed.[522, 628, 629] Garin et al[630], however, proposed an alternative with a seven-factor structure, suggesting life activities split into two factors (household and work or school). It did, however, support the global score. Factor analysis for the Chinese version differed[631], suggesting a different factor structure of 'self-care and household activities', 'getting along with people', 'getting around', 'understanding', 'communicating', 'participation in society' and 'family burden'; this factor analysis also excluded the following items: standing up, eating, sexual activities and

MEASURE	PSYCHOMETRIC PROPERTIES	
		living with dignity, and accounted for 72% of variance.[631] Numerous studies have considered the concurrent validity of the WHODAS 2.0/II by examining the correlation of individual, subscale and total scores to that of some other instrument that are viewed as an acceptable measure of similar constructs. The WHODAS II has been shown to perform well against the PANSS [628], HAM-D[628], SOFAS[628], CGI[628], the SF-36[632-634], the Lehman Quality of Life Interview[635] and the Quality of Wellbeing[636]. In one study, the subscales of 'getting around' and 'self-care' showed no relevant correlation coefficients with any of the other measures administered (PANSS-S, HAM-D, SOFAS, CGI).[628] The WHODAS 2.0 has been found to correlate well with the WHO Quality of life measure, London Handicap Scale and The Functional Independent Measure, but was less strongly correlated with the Short Form Health Survey[522]. It has also been found to perform well against the SF-36.[630] It has also been found that the WHODAS can discriminate between consumers with and without depression, as well as varying levels of severity.[635-639] However, the 12-item WHODAS II was not able to discriminate between the presence or absence of comorbidity amongst primary care consumers with depression.[637] Other studies have shown an ability of the WHODAS II to discriminate between consumers with and without schizophrenia[636] and between those who have higher levels of symptomatology, as measured by the PANNS[628]. Whilst McKibbon et al found differences between diagnostic groups, they did not find significant correlations between levels of positive and negative symptoms, cognitive performance of functional capacity.[636] The WHODAS-2 has been shown to differentiate between diagnostic groups[627, 632, 633, 636, 640] and disease severity[634]. For example, Chopra et al[627] found that those with multiple sclerosis had higher mean scores on the WHODAS II in comparison with those with mental health disorder, due to increased level of difficulty in activities dependent on physical limitation, whilst more people with a psychotic disorder rate problems in the areas of getting along. In a review by Noonan et al[641], it was found that while the effects of sociodemographic information such as age and education had demonstrated mixed effects in various populations, the patient reported variables, such as depression are consistently associated with lower WHODASII Scores. Similarly, WHODAS 2.0 has been shown to discriminate between severity of symptoms across a number of conditions, including mental health disorders.[522, 630]
	Reliability	Examining the internal consistency of the WHODAS II, Cronbach's alpha coefficients have been moderately high, with subscale and total scores ranging from .47 to .98.[628, 630-633, 636, 640, 642] The domains with the lowest internal consistency include self-care[632, 640] and getting along with others[633, 640, 642]. Internal consistency for the adolescent population was found to be .95 and for individual items from .79 to .86. It was noted that internal consistency improves in the adolescent population with removal of

MEASURE	PSYCHOMETRIC PROPERTIES	
		the sexuality question.[629]. A number of studies have examined the test-retest reliability of the WHODAS II and 2.0. They have found good stability across time, with correlations ranging from .75 to .98.[522, 628, 630, 636, 638, 640, 642] McKibbon et al[636] reported wide variability across questions, and Garin et al found that test-retest reliability unacceptable for ‘getting around’ at .19[630]. No studies were found that have examined the inter-rater reliability for the self-assessed version compared with the interview version.
	Sensitivity to Change	A number of studies have examined the degree to which changes in the WHODAS II/2.0 scores compared with another established measure of change. It has demonstrated sensitivity to change in comparison with the SF-36[522, 632, 633, 643] and the London Handicap Scale [522]. In the study by Perini et al[643], the WHODAS was most sensitive to change in the social phobia group when compared with the other measures. Pyne et al [644] found that while the WHODAS significantly correlated with changes in depression, it did not correlate with changes in schizophrenia-specific symptoms measured by the PANSS.
	Acceptability/ Usability	It has been suggested that WHODASII/2.0 is a useful and suitable tool for assessing disability and functionality, that it may have the potential to aid in a clinical settings[628], and that including the viewpoint of the consumer provides additional useful information[626]. Indeed, a 2009 review of the WHODAS II[645] reported that of the 51 studies reviewed there was agreement that it is a useful instrument for the assessment of disability, functioning and social participation across a broad range of areas, including mental health. The WHODAS 2.0 performs well across cultures, different subgroups of the general population and among those with mental health problems or addictions, and provides standard scores for the general population, allowing normative comparisons.[522]. Although there are shorter versions available for use it has been proposed that it is a lengthy tool [646, 647]. Uston et al [522] reported time for completion to be on average 20 minutes for the 36-item version and only 5 mins for the 12-item version. Kulnik et al[647] has raised concerns that the questionnaire in its current format tends to favour a medical construct of disability. With similar concerns, Chopra et al[626] and Uston et al[522] suggested that the consumers in their study had problems distinguishing difficulties attributable to their health condition from other difficulties. For example, some respondents identified issues other than health impacting on functioning, such as time or money, which is outside of the scope of WHODAS 2.0.[522] Chopra et al[627] suggests reframing the question to ask the consumer about their ability rather than difficulty may make the interview more acceptable. It has been reported that respondents found the questionnaire meaningful and relevant and that the 30-day time from was appropriate, highlighting that a longer time frame would be difficult to remember.[522] Additionally, it was proposed that a low proportion of missing values when using the WHODAS2.0 suggests easy

MEASURE	**PSYCHOMETRIC PROPERTIES**	
		completion for a wide range of consumers.[630]
Multnomah Community Ability Scale (MCAS)	**Validity**	A number of studies have examined the subscale structure of the MCAS. In the original development work a factor analysis was completed which confirmed the four-factor structure, and there were high correlations between the global rating and the items.[523] A later study by Hendryx et al[648], however, did not confirm this original factor structure, suggesting that there did not appear to be independent constructs. They proposed changes to item positioning and renaming of subscales. Corbiere et al[649] also found that the original factor structure did not fit and proposed a four-factor solution based on 12 items, rather than the original 17. Similarly, Bassani et al[650] found a poor fit with the proposed factor structure. All studies have found the most reliable and consistent subscale as social competence.[648-650] Bassanie et al suggest rather than removing or reorganising items, the subscales should be considered as indexes that measure aspects of disability.[650] A number of studies have considered the concurrent validity of the MCAS by examining the correlation between the total and subscales scores to that of another instrument that is viewed as an appropriate measure of a similar construct. The MCAS performed well against Client Satisfaction Questionnaire[648], Physical Health-SF-12[648], the Quality of social life from the Lehman Quality of Life Scale[648],the SOFAS[648], Brief Symptom Inventory[651], Brief Psychiatric Rating Scale[652], and the PANSS[653].The MCAS, however, was found not to significantly correlate with the Mental Health–SF-12[648], whilst the consumer self-report version did significantly correlate with the SF-12[651]. There was no significant correlation between the MCAS and participant-rated recovery measures (Recovery Assessment Scale)[654] and no association between MCAS and auditory hallucinations.[655] The MCACS has been found to discriminate between varying levels of service use, with those with higher service use having lower MCACS scores[523]; and between consumers based on hospitalisation and use of community resources[523, 656], severity of symptoms[656], and cognitive functioning [657]. Several studies have examined the predictive validity of MCACS. In the original pilot study it was found that consumers with higher MCAS scores were less likely to make use of the state hospital (during the two-year follow-up period) or to be admitted involuntarily to a local hospital (within the 18-month follow-up period) than those with lower scores.[523] This finding was confirmed in a validation study that found MCAS scores predicted subsequent hospitalisation.[658] In another study it was found that those with higher functioning had significantly lower total and community costs.[659]
	Reliability	A number of studies have examined the internal consistency of the MCAC, as measured by Cronbach's alpha. In these studies, moderate to high internal consistency has been reported for total score: .87-.90[523, 648] and sub-scale scores[648]. Internal consistency of the consumer-rated version has also

MEASURE	PSYCHOMETRIC PROPERTIES	
		found to be high at .82[651]. The initial pilot testing investigated the test-retest reliability over a 2-4week period, finding good reliability of .82[523] Lower reliability was found in the social network, social participation and medical compliance questions.[523] Test-Retest reliability, over a 2 week period, for the modified consumer self-report version was .91[651] for the total score and ranged from .82 to .90 across subscales. In the initial pilot testing of the instrument, good inter-rater reliability was found for subscale and total scores: .70-.78 and .85 respectively.[523] Follow-up studies have found similar reliability for subscale and total scores of .62-.99[653, 660] and .85-.96[653, 660], respectively. When comparing the self-report to that of the researchers/clinicians, inter-rater reliability was found to be fair: .2-.57.[651]
	Sensitivity to Change	The ability of the MCAS to detect improvement or deterioration has only been examined in a small number of studies. The simplest have examined change in MCAS over time in given settings, suggesting that an increase in functioning should occur over time. Some studies found improvements over time in a community setting[661, 662] and in long-term hospital care in the community[653]. It has also been found to be sensitive to change in comparison with changes on symptom ratings using the BPRS and self-report measures.[663]
	Acceptability/ Utility	Barker et al[656], suggest that the MCAS can be useful as a clinical assessment tool as it can highlight areas of client's functioning and can assess both symptoms and functioning problems, and may also be useful as an outcome measure for program evaluation and service payment allocation. The self-report measure was found to be acceptable from the consumer perspective and peer counsellors suggested it might help consumers track their progress[651]. It is suggested also to be easy to use, with 80% of consumers in one study completing it without assistance.[651]
Canadian Occupational Performance Measure (COPM)	**Validity**	The study by McColl et al[664] examined the construct validity of the COPM by testing the relationship of the theoretical constructs of satisfaction with performance, independent living and life satisfaction with the COPM through univariate and multivariant regression analysis. The COPM scores were significantly related to these theoretical constructs. Additionally, the subscale scores of performance and satisfaction were found to be significantly correlated at .68, but the correlation not so high for either scale to be redundant.[664] The COPM has shown to perform well against the Satisfaction with Performance Scaled Questionnaire[664], Reintegration to Normal Living Index[664], Life Satisfaction Scale[664] and the Perceived Problem Checklist[664], Occupational Self Assessment[665], Melville-Nelson Self-identified Goals Assessment[665], Global Severity Index[666], Modified PTSD Symptom Scale self-report[666] and

MEASURE	PSYCHOMETRIC PROPERTIES	
		Traumatic Institute Belief Scale[666]. However, it has shown to have poor or mixed performance against the GAF[667] and the Van du Toit Model of Creative Ability[667]. The COPM was able to discriminate between consumers with and without a clinically significant number of depressive symptoms; those with a clinically significant number of depressive symptoms identified more occupational performance concerns.[668]
	Reliability	Test-retest reliability has been examined in one study, where it was found to be moderately high at .84, where an initial and repeated measure were completed within 2 weeks[669]. In a review by Donnelley et al [524], it was reported that the original developers found test-retest reliability to be .63 -.84, although the time period for the second rating is not stated. This review also indicated that the inter-rater reliabilities for the COPM ranged between .63 and .89. The internal consistency was reported ranging between .41 - .56 for performance and .71 for satisfaction.
	Sensitivity to Change	Few studies have examined the sensitivity to change of the COPM in mental health services. However, two of these studies assessed the degree to which changes in the COPM correlated with another external measure of change. These found correlations with change on the COPM and changes in the GAF.[667, 670] Another study examined changes in COPM for a particular consumer group following an intervention; in this study, statistically significant changes in scores on the COPM were found following a posttraumatic stress program.[666]
	Acceptability/ Usability	Whilst the COPM is thought to be a client-centred approach to outcome measurement, there remain concerns regarding the time burden. Doige et al[670], for example, suggest that the use of the COPM is time consuming; however, this was weighed against the use of the measure promoting a client-centred approach to treatment in services. Similarly, Samsonraj et al[667] proposed that the use of the COPM was time consuming and not practical for routine use in practice. In a clinician survey it was found that only 50% thought the COPM was appropriate for routine use. Consumer perceptions of the COPM have been found to be positive; 75% reported that it had been helpful in identifying problems, and 100% reported have no problems understanding the questions.[664]
Children's Global Assessment Scale (CGAS) [513]	**Validity**	Several studies have considered the concurrent validity of the CGAS by examining the correlation of the CGAS to that of other established instruments which measure similar constructs. In general these studies have found that the CGAS scores demonstrate high correlations with independent measures of competence, intellectual and social functioning, and problem solving, and only moderate correlations with measures of symptomatology. Collectively these finding have been interpreted as evidence of the

MEASURE	PSYCHOMETRIC PROPERTIES	
		instrument’s concurrent validity, since the former constructs are what it purports to measure. [526, 528, 545, 671-673] Little has been reported on the predictive validity of CGAS. The exception is the work by Sourander and colleagues[674, 675]who found that CGAS ratings at admission were predictive of functioning and residential status at follow-up.
	Reliability	Schaeffer et al[676] examined the test-retest of the CGAS by presenting case vignettes to participating clinicians at an initial session and again six months later. Within-rater consistency was excellent for cases representing individuals with a range of diagnoses. Only the vignettes depicting individuals with isolated symptom disorders showed discrepancies between the two rating points[676] The design of the above study by Schaffer et al[526] also permitted an examination of the inter-rater reliability of the CGAS, since they could assess the agreement between raters at both time one and time two. They found excellent agreement between raters; a finding that has generally been supported by subsequent studies. Dyborg et al,[677] Bird et al [671] and Green et al[672], for example, also found good inter-rater agreement, particularly among experienced raters. Weissman[528], found good agreement between ratings made by mothers, children and psychiatrists, but Sourander et al[673] reported poor agreement between ratings provided by parents and teachers.
	Sensitivity to change	Weissman et al[528] observed patterns of CGAS scores for a group of children who had no current or previous psychiatric disorder at initial assessment, some of who progressed to a first onset within two years. The average difference between CGAS ratings at initial assessment and follow-up was significantly greater for this subgroup than for those who remained disorder-free. These results are indicative of the CGAS’s sensitivity to change in clinical status.
	Acceptability/Usability	The CGAS is generally regarded as a useful measure of child and adolescent functioning, providing more detailed information for guiding treatment decisions than diagnosis- or symptom-based measures alone.[527, 528, 677] Published commentaries do, however, put forward a number of criticisms about the measure. Firstly, concerns have been expressed about its vulnerability to rater manipulation, in that raters can assign scores below or above a particular cut-off point to suit their needs. Secondly, the global nature of the scoring has been criticised for failing to consider different domains of functioning in any organised manner. Finally, the instrument’s accuracy has been questioned, given its dependence upon the clinician’s observations of the consumer at the time of assessment and the availability of relevant background information[678].

MEASURE	PSYCHOMETRIC PROPERTIES	
Parents' Evaluation of Developmental Status (PEDS)	**Validity**	No studies were identified that examined the construct validity of the PEDS. Discriminant validity of the PEDS has been considered in a number of studies by examining its ability to discriminate between groups of consumers. For example, one study found that across all ages, concerns about behaviour and/or social skills as defined by the PEDS were associated with mental health problems.[679] PEDS was found to be sensitive to autism spectrum disorders and mental health disorders in children under 4 years and older.[680, 681] The measure has also been found to correlate with diagnostic tests of intelligence, language, academic achievement and adaptive behaviour skills.[681] In assessing concurrent validity by examining the correlation of the PEDS with other screeners, it has been found that it did not correlate with the Ages and Stages Questionnaire.[682] As a screening tool, the PEDS is designed to detect developmental problems and behaviour. The PEDS has been shown to have good sensitivity, of 74 to 79%, in identifying children with disabilities and good specificity, of 70 to 80%, in identifying children without disability.[530, 683] In agreement with these sensitivity ratings, Glascoe et al[679] found that the presence of concerns in either behaviour and/or social skills was highly sensitive to mental health problems, and 76% of children with issues in these areas as assessed by PEDS had significantly elevated scores on mental health problems.[679]
	Reliability	It has been reported in number of studies that the original developers Glascoe et al (1998) found high inter-rater reliability at .95[529, 533, 684]. However, in a later study in Australia, the inter-rater reliability between parents and carers was found to be low to moderate at best, with the highest reliability for gross motor being .40 and social-emotional, .37.[532] A number of studies have reported the test-retest reliability of the original developers Glascoe et al (1998) as moderately high .88, however the period between the first and second measure is not reported[529, 533, 684]. The internal consistency of the PEDS, as measured by Cronbach's alpha, has was found to be as .81 by the original developers, Glascoe et al (1998) as reported in a number of studies[529, 533, 684].
	Sensitivity to change	No psychometric research was identified that examined sensitivity to change of the PEDS.
	Acceptability/ Usability	The PEDS has been described as supporting decision making for service different types.[529] In a US national survey of paediatricians it was found that the use of PEDS has increased significantly between the years 2002 and 2009 (8% to 29%), although the percentage of use of this tool as well as other standardised outcome measures can still improve.[685] Schonwald et al[686] conducted a focus group to assess the acceptability of PEDS in routine practice;

MEASURE	PSYCHOMETRIC PROPERTIES	
		participants were positive about the tool and described it as easy to use, thought it saved time at client visits and provided an organised structure for discussing concerns with parents and caregivers; it also led to identification and discussion of other non-developmental concerns. In the same study they also found a significant increase in the identification of behavioural problems (in the 2-year old age group) and developmental concerns (3-year old age group) with the introduction of PEDS into routine practice.[686] Parents felt involved in care and considered their opinions were valued.[687] Similarly, in a parent survey, most found it easy to complete (98%) and thought it likely to be useful for health professionals (89%)[532]. When assessing the usability of the PEDS in Singapore, Kling et al[688] suggested modified scoring was needed to ensure its applicability in Singapore, as using the English version of ratings tended to increase the number of concerns when using the screening tool.
Drug Use Disorders Identification Test (DUDIT)	**Validity**	<u>Construct Validity</u> Only a few studies have conducted a principal component analysis to determine the factor structure and association between items of the DUDIT. In an early study, Berman et al (1995) found a three factor solution of dependence, drug-related problems and intensity of use[535]. In contrast a later study found a one factor solution with all item item-component loadings in the good to excellent range (.60 -.90) accounting for 64.91% of variance[534]. <u>Concurrent Validity</u> Other studies have investigated the concurrent validity of the DUDIT by assessing its correlation with other established instruments that have been shown to validly measure related constructs. The DUDIT has been shown to perform well against the DAST-10[534], and the drug problem severity domain and legal problem severity domain correlated well with the ASI-6[689]. One study found that it did not correlate with the AUDIT[689]. Another method to examine the concurrent validity is to consider the ability of the DUDIT to discriminate between groups of consumers. The DUDIT was able to distinguish drug from alcohol abusers[534]. ROC analysis has also been applied in a number of studies to explore the concurrent validity in relation to current diagnosis. In relation to dependency diagnosis an AUC of .93 was found[689] and in relation to current DSM-EV non-alcohol drug use disorder the AUC was at .83[690], which are both large and significant. <u>Predictive Validity</u>

MEASURE	PSYCHOMETRIC PROPERTIES	
		As a screening tool the DUDIT is designed to detect problematic drug use. The DUDIT has been shown to have good sensitivity (to identify people with problematic drug use) of .84 to.90 and good specificity (in identifying peoples without problematic drug use) of .75 to .85[534, 535, 689, 690]. There were however a number of differences in the cutoffs applied to identify problematic drug use, this ranged from 3 points (men) and 1 point (women) [690] , to 8 [534], to 12[689], to 25[535].
	Reliability	Internal Consistency A number of studies have examined the internal consistency of the DUDIT, as measured by the Cronbach's alpha. In these studies the, Cronbach's alpha ranged from .80 - .95 indicating that the DUDIT has a high level of internal consistency[534, 535, 689-692] Test-re-test Only one study has examined test-retest reliability , with the online version of the DUDIT. Participants were asked to rate the instrument at two points in time, one week apart, this resulted in a very low or non-existent correlation of 0.05[691]
	Sensitivity to change	There were no studies found examining sensitivity to change
	Acceptability/Usability	The DUDIT has been largely described as brief, easy to score and is feasible to introduce into a clinical setting[534, 535]. Indeed it is also thought appropriate to be used as an online tool, which offers flexibility to consumers and greater opportunity for collection to service providers[691].
Alcohol Use Disorders Identification Test (AUDIT)	**Validity**	Construct Validity A number of studies have examined the construct validity of the AUDIT by conducting a confirmatory factor analysis. Most studies have found that a 2-factor solution, alcohol consumption and alcohol-related consequences , provides the best data fit for AUDIT[693-696]. There have been however been a small number of studies proposing different factor solutions, including a one factor solution[697] and support for the original intent of a three factor solution for the AUDIT[698]. Concurrent Validity A number of studies have considered the concurrent validity of the AUDIT by examining the correlation between the AUDIT to that of another instrument that is viewed as an appropriate measure of a similar construct. It has been found to correlate with the AUS[699] the problem severity domain of the ASI-6[689] and the substance Abuse Index, which forms part of the Psychosocial Wellbeing Scale[699] and the SF-12[694]. One study found that it did not correlate with the DUDIT[689].

MEASURE	PSYCHOMETRIC PROPERTIES	
		Concurrent validity has been considered by a few studies, by examining the ability of AUDIT to discriminate between groups of consumers. It has been found to discriminate between those who were homeless compared to those with stable accommodation[693], those whose offences involved alcohol compared to those whose offences did not[693]between consumers with a history of psychiatric hospitalisation obtained compared to those with a penal/prison background (Hallinan et al 2011). It was also found to discriminate between those with an alcohol related diagnosis and those without[537] Concurrent validity has also been explored by conducting a ROC analysis in relation to current diagnosis. The AUC ranged from .83-.969[537, 689, 690, 699-702], indicating a high level of concurrent validity. Predictive Validity As a screening tool the AUDIT is designed to detect individuals whose use of alcohol places them at risk of alcohol problems or who are experiencing alcohol related problems. The AUDIT has been shown to have good sensitivity.67 - .97 and specificity .71-.98[537, 538, 689, 690, 699-703]. There was not however consensus over the most appropriate cut-off score when determining the sensitivity and specificity, ranging from 3 to 25 points.
	Reliability	Internal consistency A number of studies have examined the internal consistency of the AUDIT as measured by the Cronbach's alpha. In these studies a high level of internal consistency has been reported with values of .80 - .984[537, 689, 690, 693, 699, 700, 703]. The online version of the AUDIT tool was also found to have a high level of internal consistency .90-.93[691] Test-retest Only two studies were found to examine the test-retest reliability of the AUDIT. The online version reported a very low reliability, .20, however only a small number of participants chose to complete the AUDIT on a second occasion (within a week of the first collection)[691]. In contrast the study by Dybek et al[702]; found a high test-retest reliability for the total score, .95 and the correlation ranged from good to excellent on all items(.66 - .98) with exception of item 9 which had poor correlation .39. Participants in this study were asked to complete the measure within a month of the first completed test. It was proposed by the authors that the low ICC for item 9, " Have you or someone else been injured as a result of your drinking?", may be due to the high social undesirability of a positive response.
	Sensitivity to change	There are no studies examining the sensitivity to change of the AUDIT.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	The AUDIT has been described in the studies to be easy to use and are not a burden on staff time[537, 703] The AUDIT has been applied in a number of service setting, including forensic mental health services [689, 693], general practice[538, 700, 702], community mental health settings and outpatient clinics [534, 537, 690, 699-701], alcohol treatment units[703], as well as across ages groups from young people experiencing first episode psychosis[537, 690], to the older age population[701]
Specific Levels of Functioning Scale (SLOF)	**Validity**	Construct Validity Only the original developers of the instruments have examined the construct validity of the SLOF. Factor analysis found a six factor solution and accounted for 58-70% of the variance across a range of settings[540]. Concurrent Validity A number of studies have examined the concurrent validity of the SLOF be examining the correlations of subscale or total scores to that of some other instrument that is viewed as an acceptable measure of related constructs. The SLOF has been shown to perform well against the Beck Depression Inventory (BDI)[704, 705], Positive and Negative Symptoms Scale (PANSS)[704, 706], USCD Performance Based Skills Assessment (UPSA)[705-708], MATRICS Consensus Cognitive Battery (MCCB)[707, 708], Social Skills Performance Assessment (SSPA)[706], Burden Inventory for Relatives to persons with psychiatric disturbances (BIRP)[709], and the Neurocognitive Composite Score (NCS) from the Brief Assessment of Cognition in schizophrenia (BACS)[704, 706]. Interestingly, it was the clinician rated measure of SLOF that generally correlated well with other the more established instruments, particularly cognitive and performance based measures. The self-report of SLOF was found to not correlate significantly with the UPSA in a number of studies[705, 708, 710], SSPA[710], MCCB[705]. However, the study by Bowie et al found the self-reported SLOF was related to the self-report BDI and a self-rated Quality of Life measure[710]. A further study by Cardenas et al 2012, found that when self efficacy was low, as measured by revised self-efficacy scale (RSES), there was no significant correlation between SLOF and UPSA-B, but when self-efficacy was high, there was a significant relationship between the self-reported SLOF and the objective functioning measure[711]. Another method of examining concurrent validity of the SLOF has been to consider its ability to discriminate between consumer groups. Studies have found that the SLOF can discriminate between consumers who are employed or not employed[704, 707], those living independently versus those in supported accommodation[707], and

MEASURE	PSYCHOMETRIC PROPERTIES	
		those with financial responsibility versus those requiring assistance[707]. In contrast the study by Harvey et al, found that total SLOF scores were not associated with a range of functional milestones, including social (ever married or equivalent), vocational (ever employed, currently employed) and residential indicators(living independently, financially responsible)[707].
	Reliability	Internal consistency A number of studies have examined the internal consistency of the SLOF, as measured by Cronbach's alpha. In these studies the internal consistency has been reported as moderately high with correlations ranging from .59 -.95[540, 704, 709] Inter-rater reliability Schneider and Struening et al reported an inter-rater reliability of .42-.62, and noted that inter-rater reliability was greater when the client was more well known to the rater[540]. Several studies have examined the inter-rater reliablity of the SLOF, by comparing a clinician or informant rating with a self-report rating. These studies have found poor inter-rater reliability[708-710] In examining the differences further Sabbag et al 2012 found that greater discrepancy in scoring was related to greater symptom severity (PANSS) and poorer everyday functioning on the SLOF, a higher correlation was found when the consumer had a higher level of self-reported depression (BDI)[705, 708]. One study however reported good inter-rater reliability for all scales on the SLOF, apart from social acceptability and activities, this was however between informants and clinicians, rather than a self-report[709]
	Sensitivity to Change	The ability of the SLOF to detect genuine improvement or deterioration in functioning has been examined in one study, by assessing change over time in community setting and by comparing it to change on other established measures of outcomes. Dogan et al 1994 reported that the average item scores for interpersonal relationships, social acceptability, activities, work skills and total score improved significantly following community treatment. It was also found that the SLOF correlated with the WHOQOL-BREF, Family Assessment Device and Multidimensional Scale of Perceived Social Support, and the BPRS, which also reported significant change over time[712].
	Acceptability/Usability	There are no published studies that have specifically examined the feasibility of the SLOF as a routine outcome measure. A study by Leifker et al, however, reported that an expert panel considering the most appropriate functioning measures, determined the SLOF a candidate measure for routine use in mental health services based on its reliability, convergence, sensitivity, practicality, usefulness for multiple raters,

MEASURE	PSYCHOMETRIC PROPERTIES	
		relationship with symptoms and comprehensiveness, alongside 5 other measures[539]
Independent Living Skills Survey (ILSS)	**Validity**	Construct Validity Cyr et al (1994) examined the subscale structure of the ILSS by employing a principle component analysis. A two component solution was found, with the first factor (basic skills required to live independently), accounting for 36% variance and the second factor (higher level skills required to live and participate in the community) accounting for 21%[713]. This factor analysis did not however include the transportation scale due to identified numerical problems found as part of the factor analyses. Concurrent Validity Numerous studies have considered the concurrent validity of the ILSS, assessing its subscale scores and total scores in terms of their correlation with relevant scores on more established instruments that have been shown to validly measure related constructs. The ILSS in the main has not performed well against other established measures such as the Social Adjustment Scale-II[713], Perceive, Recall, Plan and Perform (PRPP) System of Task analysis[714], and Positive and Negative Symptoms Scale (PANSS)[715]. In contrast, the ILSS has performed well against the NOISE-30[713, 716], the Motility, Affect, Communication, Cooperation II (MACC-II)[716], and although low, a significant correlation with the Global Assessment Schedule (GAS)[541], and Brief Psychiatric Rating Scale (BPRS)[541] has been reported. Another method to examine the concurrent validity of the ILSS has been to consider its ability to discriminate between groups of consumers differentiated on treatment and service factors. Lower total scores on the ILSS have been associated with diagnosis of schizophrenia[713, 715], and with consumers who are in a longer stay units compared to those in shorter stay units[713]. Similarly some studies have found that the ILSS can discriminate between consumers based on a range of sociodemographic factors, lower scores have been related to gender, with men tending to have lower functioning than women[713, 715], age, with older individuals having lower functioning [715] and education level, lower level of educations based on schooling has been found to be associated with lower functioning[715]. Although Cyr et al (1994) and Perivoliotis et al (2004) found the ILSS could discriminate between consumers with different diagnosis, Bystritsky et al (2001) found the ILSS-SR did not discriminate between consumers with a diagnosis of OCD as opposed to schizophrenia. Similar levels of functioning was found in both groups of consumers[717]. Predictive Validity The ILSS-SR total was found to predict employability at a one-year follow up from the initial assessment,

MEASURE	PSYCHOMETRIC PROPERTIES	
		however there was not data to indicate the significance of this finding or the magnitude of the effect[541].
	Reliability	Internal consistency A number of studies have examined the internal consistency of the ILSS-SR and the ILSS-I, as measured by the Cronbach's alpha. In these studies the internal consistency of the ILSS – SR has been reported as marginal to high .453 - .93 [541, 713, 715]. The internal consistency of the ILSS-I has been reported as moderately high .67 – .944[541, 716] Wallace et al (2000) suggested a range of reasons for the internal consistency being higher for the informant version as opposed to the self-report, including the greater number of items in the informant report and the five point scale response may have a higher ceiling than the dichotomous scales, which would result in more normally distributed totals[541]. Inter-rater reliability No studies have been published on the inter-rater reliability of the ILSS-SR or the ILSS-IR. Wallace et al (2000) however has examined the correlation between the self report and the informant report version of the ILSS. A low to fair correlation between the ILSS-SR and the ILSS-I was found, with coefficients ranging from .276 (leisure) to .591 (food preparation), with an average of .444 (Wallace et al 2000). Test-Retest Wallace et al[541] and Cyr et al [713] have examined the test-retest reliability of the ILSS-SR and have reported a moderate to good reliability, with correlations ranging from .418 - .904 for subscales and .67 - .785 for total scores, with a test-retest period of 6 months and 2 months respectively. Wallace et al [541] also examined the test re-test reliability of the ILSS-IR, over a 6-month time frame, and reported moderate to good reliability with the exception of the job maintenance sub-scale (.340).
	Sensitivity to Change	A number of studies have examined changes in the ILSS scores for different consumer groups that would be expected to show greater or lesser degrees of improvement depending on their treatment circumstances. These studies found an improvement on the ILSS in consumers participating in a specific skills training program compared post intervention[541, 717, 718], and significant change on the ILSS was also found following Cognitive Behavioral Social Skills Training (CBSST)[719]. Whilst Liberman et al[718] noted associated improvements on the Social Activities Scale, GASS, BPRS, Granholm et al[719] found no associated improvements in symptoms, when comparing it to change on the ILSS.
	Acceptability/Usability	Although there has been no formal testing of the feasibility and utility of the ILSS, various authors have

MEASURE	PSYCHOMETRIC PROPERTIES	
		commented positively in this regard. Wallace et al [541], Perivoliotis et al[715], Menditto et al[720] and Mausbach et al[721] for example, have proposed that the ILSS is brief and easy to administer, with no specialised qualifications required and provides a structure to identify areas to target particularly for rehabilitation programs. Wallace et al [541] also describes that the development of the scale was specifically constructed to meet clinicians suggestions for useful information and each of the items content was reviewed for applicability and acceptability with diverse clinicians and consumers. In contrast whilst Perivoliotis et al[715] argues that it provides a good structure for discussion and direction of care, there is also an acknowledgement of the burden on staff, with the need to summarise the results of the ILSS across items and domains or scales. It is also proposed that some modification for use the older psychiatric population are required to improve is acceptability[715]
Social Behaviour Scale	**Validity**	Construct Validity A number of studies have examined the construct validity of the SBS by conducting principal component analysis[722-724]. These studies all extracted a four factor solution, accounting for 57.8% - 67% of the variance[722-724]. Whilst Curson et al and Harvey et al cited the factors as thought disturbance, social withdrawal, depressed behaviour, and anti-social behaviour, Lima et al described the factors as social withdrawal, embarrassing social behaviour, restless behaviour and hostility and violence[723]. Concurrent Validity Only a few studies have considered the concurrent validity of the SBS by comparing it against other more established instruments that measure similar constructs. The SBS has shown to perform well against the Manchester scale factors and items[722, 724], and the Social Role Performance Scale (SRP)[725]. Another method used to assess concurrent validity of the SBS, is to assess the ability of the SBS to discriminate between groups of consumers based on their clinical and/or treatment profile. Poyurovsky et al[726] found consumers with chronic schizophrenia with obsessive-compulsive disorder (OCD) were significantly more impaired on the SBS than those consumers with chronic schizophrenia without OCD, and Sturt and Wykes[725] found SBS could distinguish between consumers in long-term hospital and those engaged in active rehabilitation programs according with expectations. In another study Allen et al[727] found that the SBS was able to distinguish between those consumers that the treating team identified as having low expectation of discharge and those with a medium expectation of discharge in a long stay unit.
	Reliability	Internal consistency Two studies have examined the internal consistency of the SBS, as measured by Cronbach's alpha. In

MEASURE	PSYCHOMETRIC PROPERTIES	
		these studies, Cronbach's alpha has ranged from .766 - .88[723, 728] Test-Retest Reliability Only two studies have assessed the test-retest reliability of the SBS, with one of these being the original development work of Wykes et al In these studies the reliability was assessed as moderate, with a period between first and second administration of 30 days[728] and 9 months[542]. However in the study by Wykes and Sturt, 6 items: suicidal behaviour, panic attacks and phobias, acting out bizarre ideas, posturing and mannerisms, inappropriate sexual behaviour and underactivity failed to reach statistical significance[542]. Inter-rater Reliability Similarly few studies have examined the inter-rater reliability of the SBS. Those studies that have, reported a high inter-rater reliability[542, 723, 728], with the exception of some items, Wykes and Sturt found a low reliability for sexual behaviour problems[542] and Salvador-Carulla et al found low reliability for socially unacceptable habits, incoherence of speech, odd or inappropriate conversation and violence or threats[728]. However it was noted that most consumers did not score on these items, and when they did, there were discrepancies on the part of their raters. Inter-informant Reliability Some studies have also examined the reliability of different informants. Overall these studies have found adequate inter-informant reliability, though the reliability was found to be lower than inter-rater reliability[542, 723, 728].
	Sensitivity to Change	The ability of the SBS to detect improvement or deterioration has been explored in only a few studies. They have examined change in SBS over time in given settings. The study by Grinshpoon et al [729] found that there was significant change in some of the items of the SBS following 6 months of hostel residence, whilst there was no change in other items, proposing that this pattern may be expected for this population group. Significant change in the SBS has also been found in consumers following a CBT group[730] and after attending treatment in a rehabilitation service[731]
	Acceptability/ Usability	There have been no formal studies assessing the feasibility or usability of the SBS. The original developers propose that it is a measure that can detect changes in behaviour over time and can give an objective overview of a person's progress[542]. It has also been suggested that it provides information about a person's behaviour that is not easily measured through the use of other instruments and thus adds

MEASURE	PSYCHOMETRIC PROPERTIES	
		important information to the clinical picture[722]. It was noted in one study that during the training period, some raters had difficulty understanding the psychopathological items included in the SBS and this may limit who can reliably rate the instrument[728]
Child and Adolescent Functional Assessment Scale	**Validity**	**Construct Validity** To date, no studies have adequately examined the construct validity of the CAFAS. **Concurrent Validity** Numerous studies have considered the concurrent validity of the CAFAS, with many comparing the subscale scores and total scores with equivalent scores on other standardised measures. The following instruments have been shown to correlate well with the CAFAS in the predicted direction: Child Behaviour Checklist (CBCL)[732, 733], Child Assessment Schedule (CAS)[732], Parent Assessment Schedule (PCAS)[732], Burden of Care Questionnaire (BCQ)[732] and the Youth Self-report (YSR)[733]. It should be noted that Rosenblatt et al (2002) found that the CBCL and the CAFAS identified markedly different rates of clinically significant impairment within the same sample of youth who were diagnosed with serious emotional disturbance (SED)[734] . Additional evidence of the concurrent validity of the CAFAS comes from studies that have examined its ability to discriminate between groups of children and adolescents, on the basis of indicators of mental health problems, or treatment services accessed. These demonstrated that the CAFAS discriminated between children and adolescents with particular diagnoses [543, 733, 734], between children and adolescents who do and do not have problems with academic performance, school attendance and suspension[543, 732], and between those children and adolescents who have and have not been arrested, committed a crime or those who have had more probation violations [543, 732, 735]. CAFAS was also able to discriminate between children and adolescents who accessed different levels of care[543, 732, 736], who had and had not been hospitalised[543] and between children and adolescents with differences in educational outcomes[737]. **Predictive Validity:** Several studies have examined the predictive validity of the CAFAS. These studies have shown that the CAFAS can predict outcome relating to number of days in in out of family placements (i.e. residential unit or intermediate care)[736], likelihood of recidivism during the year after discharge from a juvenile justice residential placement[735], educational outcomes [737] and service utilisation and cost[732, 738]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Internal Consistency: Only the original developers have reported on the internal consistency of the CAFAS as measured by Cronbach's alpha. The CAFAS was reported as having good internal consistency ranging from .63 - .68[732]. Inter-rater reliability The inter-rater reliability was reported in the original development work with high inter-rater reliability of .84-.89 for the total score and moderate to high inter-rater reliability of .63-.95 for the subscales[732]. In a review by Bates et al (2001) it is reported that Ogles, Davis and Lunnen (1999) also tested the inter-rater reliability, with high reliability using vignettes .88-.94, but much lower reliability using case reviews .55 - .75[544]. Test Re-test reliability Bates (2001) in the review also notes only one unpublished study that examined the test-retest reliability. This study required the administration of the measure conducted a week apart, it was found to be adequate for both total and subscale scores (.82 - .95), although the substance use subscale was not reported on[544].
	Sensitivity to change	The ability of the CAFAS to detect improvement, or deterioration has been examined by only a few studies. Most of these have simply examined change in CAFAS over time in given settings, proposing that following particular interventions there should be an improvement in a child or adolescents functioning. CAFAS has shown to significantly change over time following inpatient programs [732], outpatient programs[732] and in community settings[739, 740]. The study by Hodges et al (2004) found that the pattern of outcome results also differed for subgroups based on the type and extent of impairment in the expected direction[739] One study examined the CAFAS ability to detect change against other established measures of outcome. Using this criteria Rosenblatt et al found that the level of agreement between CAFAS and the CBCL and the YSR in regards to the classification of change (i.e. positive change, no change and deterioration) was low[733].
	Acceptability/ Usability	In considering the feasibility and usability of the CAFAS one study explicitly sought the views of clinicians using the instrument in practice via survey[741]. This study reported that 85% of respondents were "satisfied" to "very satisfied" with the ease of establishing reliability with the instrument, although there were many who cited the time taken to achieve this was burdensome. Nearly two thirds (61%) stated they were 'satisfied' to 'very satisfied' with the clinical utility of the CAFAS. In the interviews clinicians

MEASURE	PSYCHOMETRIC PROPERTIES	
		were reported as stating it was useful to track change over time, that it was valuable in showing the family how the child has changed over time and it comprehensively covers a range of domains of functioning. In a review of the literature Bates (2001) found CAFAS being widely used, on a state wide level in the USA for performance outcome assessment and service eligibility determination, and in local services as an outcome measure in services and for service evaluation. He proposes that this widespread use may be due to ease of use; simple method, minimal cost, takes little time to complete and can be completed by non-professionals[544]. However, it was also noted that 18 children's mental health services of the Georgetown University Technical Assistance Centre are required to rate the burden of instruments on a 5-point Likert scale (1:low burden, 5:high burden. Of these 18 services, more than half gave a rating of 4 to 5 for the CAFAS[544]. The usability of the information from CAFAS has been reported in a number of studies, particularly from Michigan were it is a requirement for routine collection. These studies report that it is being used for a range of purposes - aggregated data is used to provide information on outcomes and continuous quality improvement to accrediting bodies and for program evaluation, and at an individual level to support case management, flow and decision making[739, 740].
Columbia Impairment Scale	**Validity**	Construct Validity Although the measure was originally developed to tap four functional domains, the original developers in their factor analysis suggested that it was indeed a unidimensional measure, with a single well-differentiated factor; impairment [742-744]. This single factor was also confirmed by Harris et al[745]. In contrast Singer et al, conducted exploratory factor analysis and found a 3 factor solution; school or work, socialising, and home or family, explaining a total of 59.8% of the variance [546]. Item 3, "problems with father figure" however, only loaded on the scale among mothers currently in a relationship with their child's father. Concurrent Validity Some studies have considered the concurrent validity of the CIS by examining the correlation of the CIS with another instrument that measures similar constructs. The CIS was found to correlate in the expected direction with Children's Global Assessment Scale (CGAS)[545, 743, 744], Child Behaviour Checklist (CBCL)[545, 745], Beck Depression Inventory (BDI) [746] and the Personal Adjustment and Role Skills Scale (PARS II)[745]. An alternative method of assessing concurrent validity is to consider the ability of the CSI to discriminate

MEASURE	PSYCHOMETRIC PROPERTIES	
		between different groups of consumers based on clinical, treatment or service factors. Several studies have found CIS to discriminate between children and adolescents with different diagnosis[545, 747, 748], between children and adolescents who have and have not been physically abused[749], who do and do not have difficulties at school[744] and between children and adolescents with differences in academic performance[744]. Studies have also found the CIS to discriminate between those receiving mental health services and those who were not[545, 744, 748], between those referred to mental health services and those not [744] and between those youth who sought help on the internet and those who did not[750]. Interestingly only one study noted significant gender effects; it found that girls had a greater impairment than boys on three of the five impairment scales, additionally girls who had contact with professional services or had a mental health diagnosis had higher impairment scores than their male counterparts[545]. A study by Harris et al assessed the ability of the CIS to correctly identify those children and adolescents with a mental disorder diagnosis and those without[745]. It was found that the CIS had a sensitivity of .24 - .64 and a specificity of .86 -1.0, with specificity increasing with a cut-off score of 12. This is in contrast with the original work which proposed a cut-off point of 15 to indicate impairment [743, 744]. Although sensitivity for the CIS was not ideal, it performed better than the PARS-II and CBCL[745].
	Reliability	Internal consistency Several studies have examined the internal consistency of the CIS, as measured by the Cronbach's alpha. In these studies, the internal consistency has been reported as moderately high with the child and parent version correlations ranging from .78[743, 744] and .78-.89[546, 743-745] respectively. Test Re-test reliability Only the original fieldwork trials has explored the test-retest reliability. Good reliability was found for the child version, .63 and very good reliability was found for the parent version .89[744].
	Sensitivity to change	There is a paucity of studies examining the ability of the CIS to detect sensitivity to change. A study by Bastiaens et al 2005 however, examined the ability of the CIS to detect change over time in comparison with the Global Assessment of Functioning (GAF) and the Health and Life Functioning Scale(HALFS)[751]. The CIS was found to detect significant change in consumers following community based treatment, and similar trends in change were found in the GAF and the HALFS. Similarly a study by Hamilton and Bridge et al found that changes in the CIS over time for adolescents in a community mental health service, correlated with changes in the BDI at 2 months and 4 months[746].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Feasibility/Usability	There are no published studies of the feasibility or usability of the CIS. It has been noted that it has been used in a variety of epidemiological research [747, 749], as well as in clinical settings[546] . Additionally, Singer et al proposes that is brevity and respondent base as well as it psychometric properties made it an ideal candidate for establishing eligibility for Supplemental Security Income (SSI) in the US[546] . Although the use of the measure has been described, there are no reports of how feasible it has been to use in practice, or if indeed the measure has provided useful information clinically. It has been noted that the psychometric properties of the parent version are better than that of the child version[742-744]

3.4 Social Inclusion

Table 11. Profile of social inclusion measures

MEASURE	DATE	COUNTRY	DESCRIPTION
Social and Community Opportunities Profile (SCOPE)	2012	United Kingdom	Designed for use either with the general population, for mental health service research, or as an outcome measure in mental health services, the SCOPE has both a short (48 item) version and a long (121 item) version, which take approximately 9 minutes and 37 minutes, respectively, to complete. The SCOPE can be administered as either self-report or an interview. The scale measures three domains: perceived opportunities, satisfaction with opportunities, and subjective wellbeing. Items measure participation and satisfaction with leisure and participation, housing and accommodation, safety, work, financial situation, self-reported health, education, and family and social relationships. Items are either marked against a 5- or 7-point Likert scales or responses are categorical (e.g. yes/no), using 'check-box'-style responses. Responses should be compared with national averages rather than aggregated to measure inclusion.[752]
Social Inclusion Questionnaire (SIQ)	2010	United Kingdom	The SIQ is a 23-item scale administered via self-report or interview. Responses are provided on a 5-point Likert scale and measure the three domains of social relationships, sense of community, and mental health services used. Items specifically measure whether the consumer feels accepted by neighbours and the community, feels accepted and involved in leisure activities, and is satisfied with friends and mental health workers. The questionnaire takes approximately 30 minutes to complete in self-report or up to an hour by interview. [753]
Activity and Participation Questionnaire (APQ)	2010	Australia	The APQ measures level of activity, satisfaction with activities, participation goals and desire to change level of activity, and assesses these across the areas of employment, seeking employment, unpaid work, education and training, and social and community participation. It also assesses readiness to change. The APQ has 14 possible items, although some items might be skipped depending on the response to initial questions. It can be administered by self-report or interview and takes less than 10 minutes to complete. Response forms vary across items: participation is measured using hours, employment is scored categorically, readiness to change is allocated to a stage of change based on the pattern of responses. This relative complexity in scoring requires some training. The APQ is designed primarily to enhance clinician-consumer discussions about social inclusion.[754]
Staff Survey of Social Inclusion (SSSI)	2009	United Kingdom	The SSSI requires either the consumer or a mental health service staff member who knows the consumer well to estimate their time spent in various activities over the last week. These activities are then allocated to the domains of employment, education, volunteering, arts, faith and culture activities,

MEASURE	DATE	COUNTRY	DESCRIPTION
			sport and exercise, local neighbourhoods, day centres and contact with family and friends. The staff member then ranks that level and type of activity to one of three levels of social inclusion. Administration requires training, but scores could be used at an individual or service level. [755]
EMILIA Project Questionnaire (EPQ)	2009	United Kingdom	Designed purely for clinical use with mental health service users, the EPQ comprises ten questions posed to the consumer to encourage them to reflect on their social inclusion over the past and coming year. Consumers provide responses verbally or in writing and the responses are analysed thematically. The questions are designed to encourage thought about participation in education, training, employment, meaningful unpaid activities and social networks. [756]
Social Inclusion measure (SIM)	2009	United Kingdom	The SIM comprises 19 items responded to on a 4-point Likert scale. Scores comprise the sum of items but has three domain sub-scores of social isolation, social relations, and social acceptance. Items refer to participation over the last three months and items assess building social capital, social acceptance, neighbourhood cohesion, security of housing tenure, leisure and cultural activities and citizenship. The SIM takes about 15 minutes to complete and is designed for use with mental health service users.[757]
The Inclusion Web (IW)	2008	United Kingdom	The IW measures two primary domains of social inclusion: people (personal relationships) and places (Institutions that matter to the individual). Types of participation considered include employment, education, volunteering, arts and culture, faith and meaning, family and neighbourhood, sport and exercise, and services. Information about these 16 areas are charted visually. A count of activities, total people and total places is used to calculate a 'clockspread' total. Software is available for scoring or administrators can be trained to convert the visual map into a summary score. The IW is designed to promote discussion about social inclusion between the clinician and consumer. [758]
Composite Measure of Social Inclusion (CMSI)	2008	Australia	The CMSI is administered via a face-to-face interview that takes approximately 42 minutes. For use with mental health consumers, the CMSI measures five domains: socially-valued role functioning, social support, absence of stigma experiences, integration in the rehabilitation community, and integration in the wider community. Activities measured include home duties and self-care, caring for others, engagement in rehabilitation, formal study or approved training, and competitive employment. The five domains have 15 levels, there are nine items related to stigma experiences that are rated on a 5-point Likert scale, and 20 items are related to community. A classification table is used to create a socially-valued role score using weekly hours of participation, performance standard, and support needed to perform role. [759]
Australian Community Participation	2007	Australia	The ACPQ measures informal social connectedness, civic engagement, and political participation. Comprised of 67 items requiring responses on a 7-point Likert scale, this self-report scale was developed for use with the general population and has not been tested with mental health service users. Social

MEASURE	DATE	COUNTRY	DESCRIPTION
Questionnaire (ACPQ)			inclusion factors measured include contact with immediate household, extended family, friends and neighbours; social contact with workmates; organised community activities; giving money to charity; voluntary sector activity; adult learning; religious observance; active interest in current affairs; expressing opinions publicly; community activism; and political protest. [754]
Evaluating Social Inclusion Questionnaire (ESIQ)	2006	United Kingdom	The ESIQ was originally developed for use as a self-report instrument, but in testing it was determined that it was better to be used as a semi-structured interview-guide for use with mental health service users. It is also suggested that the scale should be used qualitatively, rather than as a quantitative measure. The ESIQ comprises 18 items requiring scores on a 7-point Likert scale, which take approximately 20 minutes to complete. It measures the domains of the community, relationships and official services and the questions address the topics of community; leisure; education; work; housing; freedom to express beliefs; social life; stigmatisation; treatment by services; friends, family and neighbours; and fulfilment of potential. [760]
Living in the Community Questionnaire	2013	Australia	AMHOCN has been tasked with the development of this measure. A draft measure has been created and a proof of concept trial undertaken. Initial psychometric analysis is yet to be released.

Table 12. Psychometric properties of social inclusion measures identified

MEASURE	PSYCHOMETRIC PROPERTIES	
Activity and Participation Questionnaire (APQ-6)	**Validity**	Construct validity is reported to be 'good' on the basis of sound test-retest reliability and positive consumer feedback; but, it has not been evaluated independently of these properties [761]. Concurrent validity is not reported, but see discussion regarding the Composite Measure of Social Inclusion (CMSI)[759] below.
	Reliability	Stewart et al (2010),[761] following the development of the APQ-6, conducted two studies of the test-retest reliability of the questionnaire. These studies involved two separate samples from either NSW regional and metropolitan psychiatric rehabilitation and community mental health services or participants from a previous large scale study who were from the Brisbane region and who had previously been diagnosed with schizophrenia or schizoaffective disorder. In total, these studies resulted in 123 valid pairs (though valid pairs at the question level within each sample ranged from 32 to 62). Data was collected primarily by telephone using the same researcher on both occasions (84), but the remainder were self-completed by consumers, with the aid or a researcher as needed (39). Completion of the second administration occurred within 5 days of the first. Good to very good test-retest reliability was found for the majority of questions (κ = .62-.96; ICC = .69-.99), although some items questions yielded lower reliability: Q3a Total number hours in unpaid work, ICC (Qld) = 0.43; Participation in general community activities, κ (NSW) = .52; Readiness to change, κ (Qld, NSW) = .56, .55.[761] Notably, face-to-face administration did yield higher test-retest reliability for many, but not all, questions. The authors note that for some domains, a lower test-retest reliability could be expected, such as for social participation, which may well change substantially within a few days in comparison with the more stable domains of employment or study.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	With the exception of the Readiness to Change domain, the items used to assess the domains of the APQ-6 are taken directly or adapted from Australian Bureau of Statistics' surveys or the national Census. Therefore, the items have largely been widely used within Australia. In the studies of test-retest reliability described above, participants were also asked to provide feedback on their experience of completing the questionnaire; 39 NSW participants provided feedback and stated that the questionnaire took less than10 minutes to complete. All said they were either fairly or very confident in their answers. More than three-quarters said no questions were difficult; although, when asked to identify the most difficult question, one-third named the readiness to change question. Forty Queensland participants reported difficulties with the question regarding time spent on

MEASURE	PSYCHOMETRIC PROPERTIES	
		activities. Queensland researchers who administered the questionnaire stated that it had good clinical acceptability, was easy to administer verbally, but that, again, the readiness to change question was most difficult to administer due to the length of the response options. The authors state that minimal staff training and support was required for administration. Self-report administration resulted in greater amounts of missing data compared with telephone administration.
Australian Community Participation Questionnaire (ACPQ)	**Validity**	In terms of the construct validity of the ACPQ, a literature view initially informed the domains to be included, based on the concept of volitional community participation. Based on the literature, items aimed to tap into sixteen types of participation. Two rounds of pilot testing led to the development of the version of the questionnaire to be analysed for its psychometric properties.[754] Exploratory factor analysis, followed by one-factor congeneric modeling led to the identification of 14 factors relating to 14 types of community participation. Analysis of the association between participation and psychological distress as measured by the Kessler-10 indicated that 9 types of participation were significantly negatively related to distress (r = -.05 to -.20, N = 963), while Political protest was significantly, positively related to distress (r = .06). Entering the types of participation and socio-demographic information into a regression analysis predicting psychological distress, only 7 types of participation provided unique predictive power: contact with immediate household, extended family, friends and neighbours; organised community activities; religious observance; and active interest in current affairs. These participation types accounted for 8% of variance.
	Reliability	The internal reliability of the 14 scales ranged from .64 to .96 (questionable to excellent); however, only one scale fell below the acceptable level of α = .7[754]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on the usability and acceptability of the ACPQ was identified.
Composite Measure of Social Inclusion (CMSI)	**Validity**	Concurrent validity of the Socially Valued Role Classification Scale (SRCS), of which the CMSI is partly comprised, has been examined.[762] SRCS items show moderate to very good associations with some (though not all) relevant items on the APQ-6 [761] and the Work-related Self-efficacy Scale (WSS-37) [763] but poor correlations with relevant items on the Education-related Self-efficacy Scale (ESS-40) [762]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Internal consistency is acceptable to good (α= 0.74-0.85).[759] Socially valued role functioning and social support items suggest sound test-retest reliability (kappas not reported, *r* = 0.36-0.96 and *r* = 0.43-1.00, respectively) at 24 to 96 hours (n = 26 mental health service users). Majority of stigma experiences and community integration items do too (kappas not reported, r = 0.63-0.89 and r = 0.41-0.91, respectively, with removal of items with non-significant correlations).[759]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	Service users were involved in the design of the measures and training phase. Feedback indicated general satisfaction from consumers with the interview. No breaks were required during the administration and no negative effects were reported.
EMILIA Project Questionnaire (EPQ)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.
Evaluating Social Inclusion Questionnaire (ESIQ)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.
Inclusion Web (IW)	**Validity**	Construct validity was examined in context of assessing the coherence of the overall measure of clockspread. There were significant correlations for people and places in all domains except those of *arts and culture* and *faith and meaning*, suggesting that the notion of clockspread makes sense [758] Total scores for the Inclusion Web were shown to be normally distributed.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	Demonstrated to be sensitive to change when tested on consumers receiving standard and enhanced services. Consumers (N = 149) demonstrated small to medium improvements on almost all domains in terms of both people and places (total overall Z value for Wilcoxon matched pairs Signed rank test = -6.23, *P* = 0.00, effect size = 0.37) and in the overall clockspread score (Z = -5.94, *P* = 0.00, effect size = 0.57). [758]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Usability/Acceptability	No research on usability was found.
Social and Community Opportunities Profile (SCOPE)-Long Version	**Validity**	During the development of the SCOPE, the construct validity of the measure was addressed through the use of concept-mapping by a variety of groups, including mental health service users and mental health academics; scoping questions from previous UK-based surveys to match with the concepts identified during concept mapping and by consulting with an expert group. Preliminary testing involved a general population sample (n = 212), a group of participants with common mental disorders (N = 40), and two mental health service users group (n = 43 and n = 40). Construct validity: The SCOPE was shown to assess concepts that overlap with, but are not identical to, participation (as assessed by the ACPQ) [754] and social capital (as assessed by the Resource Generator-UK; RG-UK) [752, 764].Two subscales correlate with measures of social capital and community participation (at r = 0.33-0.48 and r = .42-.42 respectively (P < 0.01). Convergent and discriminant validity was assessed by comparing scores on the SCOPE scales between three groups of known mental health status: those with common mental disorders, mental health service users and those from the mentally healthy community. The mentally healthy group showed significantly different scores to the two other groups on Satisfaction with Opportunities (Effect size = .15, large), Perceived Opportunities (Effect size = .09, medium), and Overall Satisfaction with Inclusion (Effect size = .15, large). There were also considerable differences between groups on the objective items such as whether participants were earning an income from paid work and whether friends visit regularly.
	Reliability	Internal consistency varied greatly across the domains (α = .46-.84) and between user groups. The pattern of findings led the authors to conclude that the Changes to Opportunities items should be excluded from the short-form, but acceptable internal consistency achieved for the mental health service users group indicate they could be used within this population. Thy also concluded that two further scales (O14 and P13) should not be used as scales, although the items may be used individually.[752] Test-retest reliability was tested using a sample of university students (n = 102) who self-completed the SCOPE at baseline and again after two weeks (n = 26) and a second sample of students who completed an online version (n = 188, and n = 119 at follow-up).
	Sensitivity to change	Under investigation
	Usability/Acceptability	The mental health service users in the studies describe above were requested to complete an additional acceptability questionnaire. The SCOPE-long version took an average of 37 minutes to

MEASURE	PSYCHOMETRIC PROPERTIES	
		complete (range, 15-120 minutes, SD = 16 minutes) and most thought this too long (85%). More than three-quarters of participants thought the domains measured were important to their lives, Forty-five percent said that some questions were inappropriate, and half said there were some questions they would rather not have answered; in particular, questions about finances. Nineteen percent said some questions did not make sense, and 16% thought additional questions were needed (pets, sexual health etc.).
Social and Community Opportunities Profile (SCOPE)-Short form	**Validity**	Using a series of data reduction techniques, the number of items from the SCOPE-long form was reduced and the remaining items entered into principal components factor analysis. Items with the highest loadings on the domains (all above 0.7) were retained. Using this factor analysis and the results of Mokken scaling for polytomous items, 48 items were retained for psychometric testing within the short form. To test the discriminant validity of the short version, scoring from testing of the long version with mental health service users and people with common mental disorders and scores from the community sample were used to populate the short version. As with the long version, the short version Satisfaction with opportunities scale differed significantly between the mentally well and unwell groups (Effect size = .18, large). However, scores on the Perceived Opportunities scale did not differ significantly between the mentally healthy community group and the common mental disorders group. As with the long version, significant correlations were observed with measures of subjective wellbeing and community participation, lending support to the construct validity of the SCOPE-short form.
	Reliability	Internal consistency for the 8 items retained within the Satisfaction with Opportunities scale was α = .77 (acceptable), and for the five items within the Perceived Opportunities scale was α = .62 (questionable), although this relatively low level is not uncommon for scales with few items. For the latter scale, the inter-item correlation was .25, which is within optimal limits a scale with few items. The short and long versions of these scales were correlated and the correlations were .92 for the Satisfaction with Opportunities scale (84.3% shared variance) and .88 for the Perceived opportunities scale (77.6% shared variance). To analyse the test-retest reliability of the SCOPE-short form, 119 students completed the instrument at both baseline and 2-week follow-up. For continuous items, the correlation (r) ranged from .62 to 1.0 and for dichotomous items, k = .66-.97.[752]
	Sensitivity to change	A study of the sensitivity to change has begun, but only minimal data were available at the time of writing. These preliminary data have shown largely non-significant changes on SCOPE scores over time, but this may in part be due to some methodological challenges and to the small sample size. These

MEASURE	PSYCHOMETRIC PROPERTIES	
		studies are continuing.
	Usability/Acceptability	Students who completed the SCOPE-short form were also requested to complete an acceptability questionnaire. Twenty-seven students responded and the mean completion time of the SCOPE-short form was 8.7 minutes (range, 2-10 minutes), and 93% thought this was just the right amount of time. All but one respondent indicated that the domains measured were relevant to their lives. In the sensitivity to change studies, some managers and staff involved provided some feedback. While one service felt that the SCOPE was not relevant to their goal-oriented service, and thought that the 'census-style' questions were not useful to the clinical setting, the other services provided positive feedback about the ease of use and time taken, and the first service also stated that the domains seemed relevant to social inclusion.
Social Inclusion Measure (SIM)	**Validity**	A limited literature review and survey was undertaken to identify the concepts to be used within the SIM, lending to its content validity. Support comes from test of unidimensionality which found that each of the three scales (social isolation, social relations and social acceptance) correlate well with each other (r = .52-.70, P <.001), and very highly with the overarching model (r = .78-.91, P <.001) [757]. As evidence of the concurrent validity: Shown to correlate with the Clinical Outcomes in Routine Evaluation (CORE, a measure of mental health status)[765] (r =.58, P <.001) and an adapted empowerment measure[766] (r = -.62, P <.001).
	Reliability	Internal consistency of the three resulting subscales was α = .70 for social acceptance and .76 for both social isolation and social relations, all reaching the acceptable level of .70. The alpha coefficient for the whole scale suggests good internal consistency (α = .85). Some items were excluded as a result of this analysis.[757] Test-retest reliability has not yet been assessed.
	Sensitivity to change	No psychometric research on the sensitivity to change of the SIM was identified.
	Usability/Acceptability	The initial SIM was pilot-tested with 15 arts and mental health project participants and eight members of a service user research group. Users found the time taken to complete the SIM was acceptable but wording of some questions and the response format was altered based on the feedback.
Social Inclusion Questionnaire (SIQ)	**Validity**	Support for the construct validity of the SIQ comes from a principal component analysis that revealed seven factors that underpinned the concept of social inclusion and predicted 83.5% of total variance.

MEASURE	PSYCHOMETRIC PROPERTIES	
		Six of the seven factors showed stability, Three items cross-loaded on more than one factor and it was suggested that these be removed from the measure [753].
	Reliability	Internal consistency: Good (α= 0.80).[753] Test-retest reliability: Moderate or better. Over a two-week period (n = 51) 17 /23 items had kappas of >0.4 (acceptable) and 5/23 had kappa of >.6 (good).[753] The overall kappa range was .12 to .82 (poor to very good).
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	During development of the SIQ, the developers consulted with groups of mental health service users in a series of focus groups. Following feedback at each group, the questionnaire was modified until the final format was reached. An acceptability questionnaire was also administered during the pilot testing of the SI. Of the 69 participants, 84%found it easy enough to complete and 87% found in understandable. Of the 18 participants who provided additional comments, the most common themes were that an interview would be better than a questionnaire, the sexual health question was too intrusive, the questionnaire was too long and the order of the questions was not always right. On the positive side, that the SIQ was useful for respondents to voice their opinions was also a common theme.
Staff Survey of Social Inclusion (SSSI)	**Validity**	No psychometric research on validity was found.
	Reliability	No psychometric research on reliability was found.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Usability/Acceptability	No research on usability was found.
Living in the Community Questionnaire	**Validity**	Psychometrics yet to be released
	Reliability	Psychometrics yet to be released
	Sensitivity to change	Psychometrics yet to be released
	Usability/Acceptability	Psychometrics yet to be released

3.5 Quality of Life

Table 13. Profile of quality of life measures

MEASURE	DATE	COUNTRY	DESCRIPTION
Satisfaction with Life Scale (SWL)	2005	United States	The Satisfaction with Life Scale (SWL) is an 18-item self-report instrument developed as part of the evaluation of Mendota Mental Health Institute's Program of Assertive Community Treatment (PACT) to assess aspects of the subjective satisfaction with life of adults with serious mental illness. It contains 18 questions regarding current subjective satisfaction on four domains: living situation, work, social relationships, and self and present life. Each item is answered using a five-point Likert scale from 0 = *not at all* to 4 = *a great deal*.[767]
World Health Organisation Quality of Life – Brief, Australian Version (Australian WHOQOL- BREF)	2000	Australia	The Australian WHOQOL-BREF comprises 26 items that measure the broad domains of physical health, psychological health, social relationships and environment for the past two weeks. The WHOQOL-BREF is a shorter version of the original WHOQOL and contains one item from each of the 24 facets in the WHOQoL-100, plus two items from the overall quality of life and general health facet. All items are rated on a five-point Likert scale.[276]
Purpose In Life (PIL)	1964	China	The PIL is a 20-item self-report measure of the extent to which an individual perceives life to be meaningful. A higher PIL scale score indicates a higher level of perceived life meaning.[768] The measure covers two domains, despair (items which indicate a negative approach to life) and enthusiasm (items which indicate an excitement towards life). Responses are chosen on a 5 point Likert scale, and changes depending on the question being asked (e.g. 'In thinking of my life, I: 1)Often wonder why I exist – 5) always see reasons for being here)[769].
Californian Quality of Life Survey (Cal-QOL)	nd	United States	The Cal-QOL is a 40-item survey that was developed to assess patient-reported outcomes in the California Adult Performance Outcome System. The Cal-QOL was modelled after the Quality of Life Interview Short Form (Lehman 1988)[770], which is based upon a conceptual model that incorporates objective life conditions and subjective satisfaction with life conditions.[771] The constructs of the objective scales are family contacts, social contacts, finances, and arrests. Items on the objective scales ask about frequency of family and social contacts with response options including: at least once a day, at least once a week, at least once a month, less than once a month, not at all, or not applicable. Items asking about adequacy of finances are rated as yes/no. One item asks about frequency of arrests in the past month, with 5 response options ranging from 0 to 4 or more. The main constructs of the subjective scales are satisfaction with life, living situation, family relations, social relations, daily activities, leisure activities, safety, and health. Items on the subjective scales ask about level of feeling/satisfaction with various outcomes on a 7-point scale. Response options for these items are: delighted,

MEASURE	DATE	COUNTRY	DESCRIPTION
			pleased, mostly satisfied, mixed, mostly dissatisfied, unhappy, or terrible. [614]
Quality of Life Questionnaire (QLQ)	1997	United States	The QLQ is a self-report instrument comprised of 24 items that assess subjective quality of life in seven areas, each defined as a subscale of the instrument: living situation, finances, leisure, family, social life, safety and access to health care.[772] Consumers rate aspects of their lives on a scale from 1 = *terrible* to 7 = *delighted*.
Subjective Quality of Life Profile (SQLP)	1998	France	The SQLP is a self-report measure of subjective wellbeing divided into two parts. A core questionnaire consists of 27 items exploring four life domains: functional life, social life, material life and spiritual life, and an 'optional questionnaire' consists of a variable number of items chosen by the investigator from a bank of 54 items, depending on the particular points they wish to focus on (e.g., family, children, etc.).[773] For each item on the core section of the SQLP the participant indicates degree of satisfaction using a score ranging from -2 (very dissatisfied) to +2 (very satisfied), with a score of zero reflecting 'indifference'. On the optional section of the SQLP participants indicate performance or change, importance attributed, degree of anticipated change, and coping. Importance attached is scored on 3 levels from 0 (without importance) to 2 (great importance)[773]
SF-36 (or Medical Outcomes Study (MOS) 36-Item Short Form Health Survey)	1990	United States	The SF-36 is a multi-purpose, short-form health survey containing 36 self-report questions. It yields an eight-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index. The SF-36 has been used in surveys of general and specific populations, comparing the relative burden of diseases and in differentiating the health benefits produced by a wide range of different treatments.[774] The items are grouped under eight scales –physical functioning, role limitation due to physical functioning, bodily pain, general health, vitality, social functioning, role limitation due to emotional problems, and mental health – and one item on health transition. Higher scores indicated better health status. Responses range from yes/no, to 3, 5, and 6 choice responses across the questionnaire. [775]
Quality of Life Enjoyment and Satisfaction Questionnaire-short form (Q-LES-Q-SF)	1993	United States	The QLESQ-SF is a self-report measure of quality of life that consists of the first 14 items of the General Activities Scale of the full QLESQ, plus 2 additional items. It includes items covering such domains as physical health, economic status, relationships, living/housing situation, mood, work, medication (if applicable) and overall life satisfaction. Responses are recorded on a 5-point Likert scale. [776]
Manchester Short Assessment of Quality of Life (MANSA)	1999	United Kingdom	The MANSA is a quality of life scale written with the assistance of service users in order to ground the language at an appropriate level and style for all users. It has a mixture of subjective and objective components that address the client's view of satisfaction with life, work and education, leisure, safety, health, finance, social, living situation and family. Questions are answered either yes/no or using a seven-point Likert scale.[777]

MEASURE	DATE	COUNTRY	DESCRIPTION
Quality of Life Interview (QOLI)	1982	United States	The QOLI is an instrument that evaluates both subjective and objective components of quality of life. The subjective domains of the QOLI include global wellbeing (GWB) and eight sub-domains: work and education, leisure, family relations, social relations, finances, living situation, personal safety and health. A trained professional administers the interview, and items are rated on a seven-point scale ranging from 1 = *Terrible* to 7 = *Delighted*. These subjective experiences are related to objective indicators of external life circumstances, such as income, work status, housing and frequencies of social relations.[778]
Lancashire Quality of Life Profile (LQLP)	1991	United Kingdom	The LQLP combines objective and subjective measures in several areas of life. It is administered as a questionnaire led by a researcher. Subjective satisfaction with life is measured for perceived wellbeing and nine major life domains: work/education, leisure/participation, religion, finances, living situation, legal/safety, family relations, social relations and health. Subjective ratings for each domain and global wellbeing subscales are indicated on a seven-point Likert scale (1 = *my life couldn't be worse*, 7 = *my life couldn't be better*).[779]
Quality of Life Index (QLI)	1984	United States	The QLI[780] is a self-report questionnaire consisting of two sections of 32 items rated on a Likert-scale. One section assesses satisfaction in various domains of life and the other measures the importance of that domain. Subjects respond on a six-point scale from *very satisfied* to *very dissatisfied* and *very important* to *very unimportant*.[781] The QLI produces five scores: quality of life overall and in four domains (health and functioning, psychological/spiritual domain, social and economic domain, and family).
Satisfaction with Life Domains Scale (SLDS)	1981	United States	The SLDS consists of 15 questions assessing how a person feels (ranging from *delighted* to *terrible*) about a variety of life areas including housing, neighbourhood, food, clothing, health, roommate, friends, relationship with family, social interaction, job or day programming, meaningful activity, activity for fun, services in the area, economic situation and living place. The respondent is asked to indicate his or her feelings by choosing one of seven faces.[782]
Quality of Life Scale (QLS)	1984	United States	The QLS is a 21-item semi-structured interview administered by a trained clinician that provides information on symptoms and functioning during the preceding four weeks. Each item is composed of three parts: (1) a brief descriptive statement to focus the interviewer on the judgment to be made; (2) a set of suggested probes; (3) the seven-point scale with descriptive anchors for every other point. It evaluates QOL on 4 subscales: (1) Intrapsychic foundations (e.g. sense of purpose, motivation, empathy, and anhedonia), (2) Interpersonal relations, (3) Instrumental role, and (4) Common Objects and Activities.[783]
Wisconsin Quality of Life Index – Canadian Version (CaW-QLI)	2003	Canada	The CaW-QLI solicits information about the client from three perspectives: directly from the client, from a professional care provider, and from a family member (whenever possible). The client version of the W-QLI questionnaire provides a self-assessment of several separate domains, including satisfaction with different life domains, occupational activities, psychological wellbeing, symptoms, physical health, social relations/support,

MEASURE	DATE	COUNTRY	DESCRIPTION
			finances, and activities of daily living. Clients are also able to weight their perception of importance of each domain at the end of the scale. This is a unique property of the W-QLI that produces a global weighted score.[784]
Brief Life Satisfaction Scale (BLSS)	2003	United States	The BLSS is a 10-item self-report questionnaire that allows consumers to indicate their perception of the quality of their own lives.[785] The BLSS was developed by asking 11 clinicians to list in order of importance the areas of concern that most affect clients' satisfaction with their lives. The BLSS contains 10 target items that were most frequently listed by the clinicians, including family, friends and work or school. Items are rated on a five-point Likert scale from 1 = *Very dissatisfied* to 5 = *Very satisfied*.
Satisfaction with Life Scale (SWLS)	1985	United States	The SWLS is a narrowly focussed, five-question self-report measure that assesses global life satisfaction and does not tap related constructs such as positive affect or loneliness. The questionnaire is a self-report measure, answered using a five-point Likert scale. It is suited for use in different age groups across a variety of settings.[786]
Personal Wellbeing Index (PWI)	2006	International collaboration	The PWI contains eight items measuring satisfaction, each one corresponding to a quality of life domain: standard of living, health, achieving in life, relationships, safety, community-connectedness, future security, and spirituality/religion.[787] Respondents are asked to rate their satisfaction on a scale of 0 = *extremely dissatisfied* to 10 = *extremely satisfied*.
Personal Outcomes Measures®	1997	United States	The Personal Outcome Measures®[788] are a qualitative interview tool for evaluating personal quality of life and the degree to which organisations individualise supports to facilitate outcomes. People define outcomes for themselves. The outcomes are non-prescriptive, they have no norms and each person is a sample of one.[789]

Table 14. Psychometric properties of quality of life measures identified

MEASURE	PSYCHOMETRIC PROPERTIES	
Satisfaction with Life Scale (SWL)	**Validity**	The SWL was reduced to 18 items on the basis of the outcomes of a confirmatory factor analysis of the 21-item version and a four-factor structure was confirmed: living situation (items 1 to 4), social relationships (items 5 to 10), work (items 12 and 13), and self and present life (item 11 and items 14 to 18). Construct validity was tested further on the 18-item scale (between the instrument's four domains and clinically important life conditions of clients in the areas of symptoms, living and employment situations, and social relationships).[767] The measurement invariance demonstrated that the construct validity of life satisfaction was stable over a one-year period, which provides reasonable confidence that the factor structure is stable and that changes in the scores of the four SWL Subscales reflect actual changes, not measurement error.[790]
	Reliability	Internal reliability of the four subscales was examined using data from two samples at two different time-points. Acceptable reliability was evident for the living sub-scale (α = .74 and .76), while good reliability was found for both the social relationships self and present life (α = .80 and .81, and .83 and .82 respectively). The internal reliability of the work subscale was both questionable and acceptable (α = .61 and .74), believed to be because the subscale has only two items.[767]
	Sensitivity to change	No psychometric research on sensitivity to change was identified.
	Acceptability/Usability	The SWL is a brief, easily completed, freestanding scale that addresses several important domains related to subjective quality of life. The SWL scale was specifically created to measure life satisfaction among people with severe and persistent mental illness who receive community-based mental health services.[790]
World Health Organisation Quality of Life–Brief, Australian Version (Australian WHOQOL- BREF)	**Validity**	A test of each domain found them to be unidimensional, with all items loading on a single primary factor, providing support for the construct validity of the scale. The SF-36, the AQoL, the EQ5D, the HUI3, and the 15D were concurrently administered to participants and a number of significant correlations were identified, providing some support for the concurrent validity of the scale. The physical and psychological domains in particular demonstrate good construct validity. Individual items showed good discrimination between well, ill and very ill populations.[276]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	The majority of factors (physical health, psychological, and environment) yielded good to very good internal consistency (α = .77-.87) throughout a variety of samples (total, inpatient, outpatient, well). However, across these samples, the domain 'social relationships' (containing only 3 items) showed poor to acceptable internal consistency (α = .58–.74). [276] The test-retest Pearson correlations were above .8 for each of the domains over a two-week period, indicating very good test-retest reliability.[276] However, Kappas were low for all items, ranging from .23-.62, suggesting a relatively high degree of error (low reliability) at the item level. These findings suggest that the WHOQOL-BREF is not appropriate for individual assessment and is recommended for use at a population level.
	Sensitivity to change	Sensitivity to change was assessed through examination of differences between WHOQOL-BREF scores taken pre- and post-treatment for a subsample of depressed individuals from the LIDO study (n = 26). One of the eligibility criteria for the LIDO study was that patients be untreated at the beginning of the study, hence baseline scores were used as the pre-treatment measure. At three-month follow-up, patients were asked if in the previous three months they had "been counselled or given medications for the treatment of depression". For those who had received treatment, nine-month follow-up scores were used as the post-treatment measure. Paired t-tests (repeated measures design) were used to compare pre- and post-treatment scores for this group. For each domain, the post-treatment score was substantially higher, indicating improved quality of life following treatment for depression and providing evidence for the sensitivity to change of the WHOQOL-BREF.[276]
	Acceptability/Usability	The Australian WHOQOL-BREF is specifically modified and tested to Australian population norms. The WHOQOL instruments have been designed primarily for use at a population level.
Purpose In Life (PIL)	**Validity**	Two clear dimensions emerged in the factor structure of the PIL: despair and enthusiasm. Negligible overlap between the factors was evidenced by low Pearson correlations between the two factors. The first factor was found to be highly related to the Total test, whereas the second factor was observed to be only moderately related to the Total test.[769]
	Reliability	Cronbach's alpha estimates of internal consistency yielded the same pattern in both samples. Coefficients for the nine items constituting the first factor (Despair) were both good (.83 in sample one and .83 in sample two). For the five items in the second factor (Enthusiasm), the internal consistency was questionable (.65 and .63 in the two samples respectively).[769]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Acceptability/Usability	The PIL was tested on US high school students. No other psychometric research on acceptability/usability was found.

MEASURE	**PSYCHOMETRIC PROPERTIES**	
Californian Quality of Life Survey (Cal-QOL)	**Validity**	No psychometric research on validity was found.
	Reliability	Reliability of the Cal-QOL in the development pilot study was high (Cronbach’s alpha = .93) [771]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Acceptability/Usability	No psychometric research on acceptability / usability was found.
Quality of Life Questionnaire (QLQ)	**Validity**	Factor analysis using data from 971 consumers with serious mental illness support a seven-factor structure with a single second-order factor. Scores on the QLQ correlated significantly with the client’s level of functioning and satisfaction with services, providing evidence for concurrent validity of the QLQ.[772] Comparisons between employed and unemployed clients yielded significantly higher quality of life scores for those employed participants, providing support for the discriminant validity of the QLQ. Similarly, consumers who were attending the program against their will also showed significantly lower QOL scores than those attending at their will.[772] The QLQ has not been validated for different subpopulations and settings, including ethnicity, homeless people, clients receiving services in inner cities, and people in institutions. The study population (from the above study) was broad in several respects such as living situation, diagnosis, age and gender, but it was overwhelmingly white and substantially rural. The QLQ has not been tested for validity as a clinical or individual client assessment instrument.[772]
	Reliability	Evidence for the reliability of the QLQ is based on data gathered from 971 clients with serious mental illness who were receiving publicly funded mental health services at the time of the study. The Cronbach’s alpha reliabilities of the subscales based on the eight dimensions were all above .80, demonstrating good internal consistency. Internal consistency for the seven identified factors were good (α = .81-.89), except for the family and social life factors, which were excellent (α = .91 and .90, respectively) which were found to be equal to or better than the another measure of quality of life (The Quality of Life Interview).[772] The QLQ has not been tested for reliability as a clinical or individual client assessment instrument.[772]
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Acceptability/Usability	The QLQ has been used to facilitate the gathering of QOL information from clients in evaluation or program improvement. It has been tested in public mental health services. [772]

MEASURE	**PSYCHOMETRIC PROPERTIES**	
Subjective Quality of Life Profile	**Validity**	Psychotic patients' answers on their SQLP were compared with answers on another validated quality of life questionnaire (i.e., the Baker & Intagliata QOL Questionnaire, 1982). Strong correlations were found between responses on the two quality of life questionnaires, demonstrating concurrent validity. [773] Evidence for criterion validity include a statistically significant correlation found between depression scores on the MADRS, and both the global score of satisfaction and the sum of the satisfaction for each item on the SQLP. Answers to the SQLP and the Beck Depression Inventory for substance abuse patients were compared and significant correlations (from .30 to .5, $p = 0.002$) between depression scores and satisfaction items were also found.[773]
	Reliability	Limited evidence of reliability was found.[773]
	Sensitivity to change	For psychotic patients, changes were noted in quality of life at discharge. Three months after admission, there was a trend towards improved quality of life for psychotic patients using the SQLP; however, the changes were not statistically significant, with the exception of daily activities.[773] Limited evidence is available on whether or not the measure is sensitive to change over time.
	Acceptability/Usability	The SQLP has been used in many populations, including mental health consumers. [24]

MEASURE	PSYCHOMETRIC PROPERTIES	
SF-36 (also known as the Medical Outcomes Study (MOS) 36-Item Short Form Health Survey)	**Validity**	The construct validity for the two-factor structure was tested for schizophrenic patients and found to be consistent with the physical and mental health structures found for the general population. The Physical functioning, Role limitations resulting from physical problems, and Bodily pain scales all had a strong correlation (*r* > .70) with the hypothesised Physical Health dimension, and a weak correlation (*r* < .03) with the Mental Health dimension. Conversely, the Mental health, Role limitations resulting from emotional problems, and Social functioning scales loaded highly on the hypothesised Mental Health factor and had low factor loadings on the Physical Health factor. Outcomes of factor analyses provide stronger evidence that in this population, the Physical functioning, Role limitations resulting from physical problems, and Bodily pain subscales are unrelated to the underlying Mental Health factor and that the General mental health, Role limitations resulting from emotional problems, and Social functioning subscales are unrelated to the Physical Health factor.[26] 80 to 85% of the reliable variance in the eight SF-36 scales was accounted for by two factors based on physical and mental health.[27] To examine concurrent and discriminant validity of the SF-36 for this patient population, several correlation analyses were conducted using the Brief Psychiatric Rating Scale (BPRS), the Clinical Global Impressions (CGI) Scale, Severity Rating, and the Montgomery-Asberg Depression Rating Scale (MADRS). Mental Health, but not the Physical Health, component of the SF-36 showed linear relationships with widely used measures of psychiatric functioning. Correlations between the SF-36 Mental Health component and the BPRS, CGI and MADRS were -.31, -.15, and -.55, respectively. Conversely, the Physical component showed a modest linear relationship (*r* = -.19) to age, a variable that could be expected to be more highly correlated with physical, than with mental, status.[26]
	Reliability	Cronbach's Alpha Coefficients ranged from .76 for Vitality to .91 for Physical functioning and for Bodily pain. The Alpha Coefficient for the total number of SF-36 items was .90, demonstrating acceptable to excellent internal consistency.[26] Furthermore, a summary of 15 studies found most reliability statistics exceeded .80, indicating good internal consistency. [27]
	Sensitivity to change	Results from clinical studies comparing scores for patients before and after treatment have largely supported hypotheses about the validity of SF-36 scales based on psychometric studies.
	Acceptability/Usability	SF-36 has been used successfully in a variety of clinical settings because of its brevity, ease of administration and sound psychometric properties.[26]

MEASURE	PSYCHOMETRIC PROPERTIES	
Quality of Life Enjoyment and Satisfaction Questionnaire-short form (Q-LES-Q-SF)	**Validity**	In a study with 57 consumers who presented to a hospital for psychiatric treatment, almost all items of the Q-LES-Q-SF significantly correlated to the total score and other measures used in the study, with the correlations ranging from .41 to .81. These analyses provide some evidence for the construct and criterion validity of the scale.[791] In another study, concurrent validity of the Q-LES-Q as a measure of severity of illness was assessed by examining the correlation of the scale scores with other measures of severity of illness (such as the Clinical Global Impressions (CGI), Severity of Illness and Global Improvement Scales, and the Hamilton Rating Scale for Depression (HAM-D); correlations ranged from -.34 to -.68.[792] Based on a sample of 57 psychiatric patients who completed the questionnaire at intake, then again one week and four weeks later, the questionnaire scores could detect in 80% of cases those who had made significant change in their QOL according to the external criteria. In 100% of cases, scores from the Q-LES-Q-SF could exclude those who showed no real change, providing some evidence for the discriminant validity of the scale.[791] There was a high degree of correlation (ranging from .30 to .54) on a range of subscales between the CGI Global Improvement rating and the change scores of the Q-LES-Q.[792]
	Reliability	Stevanovic et al [791] found the internal consistency and test–retest coefficients were both excellent, at .9 and .93, respectively, when tested in a population of adults with a psychiatric disorder.[791] Test-retest reliability of the Q-LES-Q yielded correlations ranging from .63 to .89 across the range of subscales, with 54 subjects. Internal consistency was excellent, and yielded alpha coefficients of .90 to .96 across the various subscales. [792]
	Sensitivity to change	The Q-LES-Q – SF has been shown to possess appropriate measurement properties of an evaluative measure for assessing QOL changes in individual patients[791]. High correltations between change scores of the Q-LES-Q and HAM-D, and the Q-LES-Q and CGI Global Improvement scores, demonstrating they are sensitive to change [792]
	Acceptability/Usability	The Q-LES-Q – SF is a feasible measure, with good acceptability of the concept measured and is appropriate for evaluative purposes [791].
Manchester Short Assessment of Quality of Life (MANSA)	**Validity**	In a sample of 55 patients accessing public mental health services, correlations with the Lancashire Quality of Life Profile (LQLP) scale were shown to be around .82, indicating good concurrent validity. The MANSA subjective quality of life mean score also correlated with the BPRS total score at *r* = -.49, and with the BPRS sub score on anxiety and depression at *r* = -.42.[793]
	Reliability	Good internal consistency was demonstrated, yielding a Cronbach's alpha of .74 for satisfaction ratings. [793]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Sensitivity to change	In the construction of the MANSA, objective questions that in previous studies have neither discriminated between settings nor groups nor have been sensitive to change, were eliminated.[793]
	Acceptability/Usability	The measure has been used with people with severe mental illness (schizophrenia, bi-polar disorder) accessing community health services.[20]
Quality of Life Interview (QOLI)	**Validity**	The items comprising each of the subjective sub-scales had very high correlations with their own sub-scale totals—from .6 upwards (in four cases, the correlations exceeded .9). ‘Satisfaction with life in general’ had relatively few significant correlations with other areas of satisfaction, suggesting that the chosen sub-scales make important contributions to the overall score.[778] Construct validity was determined when the 23 items constituting the nine subjective scores for the QoLI were analysed, with the first eight dimensions accounting for 75.2% of the total variance. This analysis confirmed the eight dimensions previously defined by Lehman. Following varimax rotation, the structure remained very stable.[778] QOL indicators can differentiate among chronically mentally ill subpopulations and may offer sensitive discriminant outcome measures in this population.[770]
	Reliability	Cronbach’s alpha coefficients were calculated for each of the thirteen scores used in the factorial analysis. Consistency for all the subjective scales rated between questionable and good (.69 to .88).[778] In Lehman’s original study, internal consistency reliability coefficients (Cronbach’s alpha) were calculated for interview scales based separately upon inpatients and outpatients combined samples. Coefficients were found to vary from poor to good (.55-.88), with three subscales (privacy, autonomy, influence) found to be unacceptable (.35-.44).[770] Test-retest reliability correlations for a sample of 45 people tested one week apart revealed significant levels of stability for most interview items and scales.[770]
	Sensitivity to change	QOLI scores appear to be responsive to treatment interventions.[770]
	Acceptability/Usability	It is a structured, 45-minute Qualify of Life Interview for people with chronic mental illness.[21]

MEASURE	PSYCHOMETRIC PROPERTIES	
Lancashire Quality of Life Profile (LQLP)	**Validity**	The LQLP shows evidence of good construct validity and face validity.[794] Comparisons between the LQLP and the Social Functioning Scale (SFS) showed that there were substantial and significant correlations between comparable items and in other respects there was no association. Some of these correlations include shopping (LQOLP), associated with higher prosocial activity scores and higher independence scores (in both competence and performance, on the SFS). Similarly, going for a ride in a bus or car (other than to work) was also associated with the same three sub scores, as well as with better functioning in recreation. Having another family member in the home was associated with less social withdrawal and better functioning in recreation and with both independence subscales.[22] A study using a dataset from over 1500 cases concluded that mental health does not significantly alter the results derived from quality of life surveys, suggesting that QOL is a separate construct.[22]
	Reliability	The LQOLP is made up of a number of different sections, and each has been tested for aspects of reliability and consistency. Inter-rater reliability indicators were significant, yet modest. Internal consistency overall was good.[794] The alpha-coefficients for self-esteem (all greater than .6), affect balance (all greater than .6) and subjective well-being domains (all greater than .8) yield acceptable to good results. Internal consistency for each item varied across the five samples. All items achieved a minimum of acceptable internal consistency (never less than .75) in at least one sample. A number of other items (leisure and work), while achieving a high of .8 or more, can fall to nearer .5 in some samples. The least internally consistent items were found to be safety (.33-.80), religion (.45-.85) and living situation (.53-.76).[22] Only one test of inter-rater reliability was conducted, producing a fair, yet significant (r = 0.4) correlation between the two raters.[22]
	Sensitivity to change	The LQOLP has been used in service evaluations to demonstrate change.[794]
	Acceptability/Usability	The LQLP has had positive feedback from both respondents and interviewers.[22] It can be conducted with severely ill patients in a number of settings and was designed for use in operational contexts.[22]
Quality of Life Index (QLI)	**Validity**	Evidence for the content validity of the QLI was provided by a review of the literature and on the reports of patients undergoing haemodialysis.[795] Convergent validity of the QLI was supported by strong correlations between the overall (total) QLI score and a measure of life satisfaction (graduate students $\alpha = .75$; dialysis patient's $\alpha = .65$).[795] Factor analysis revealed four dimensions underlying the QLI: health and functioning, social and economic, psychological/spiritual, and family. These factors accounted for 91% of the total variance. Factor analysis of the four primary factors revealed one higher order factor, which represented quality of life [796]. Criterion validity was measured using groups based on self-reported levels of pain, depression, and success in coping with stress. Subjects who had less pain, less depression, or who were coping better with stress had significantly higher overall (total) QLI scores[797]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Test re-test reliability of the QLI has been reported for a two-week interval with correlations of .87 for graduate students and a correlation of .81 at one month for dialysis patients. The internal consistency has been shown as excellent in with this same sample of graduate students $\alpha = 0.91$ and dialysis patients $\alpha = .90$. [795].
	Sensitivity to change	The instrument has been used for people living with schizophrenia, and with non-white samples, as well as samples of low socio-economic status[781]
	Acceptability/Usability	No research was identified that examined the acceptability or usability of the QLI.
Satisfaction with Life Domains Scale (SLDS)	**Validity**	Carlson et al[798] conducted a validation study of the SLDS with 137 Spanish people with schizophrenia. They demonstrated divergent validity through non-significant correlations of the SLDS with unrelated constructs (including functioning, capacity and illness awareness), and construct validity through significant correlations of the SLDS with less psychopathology and better prognosis.
	Reliability	Suggesting concurrent validity, the SLDS total score correlates at r = .64 with the Bradburn Affect Balance Scale, and at r = .29 with the Global Assessment Scale.[782] In Carlson et al's study, Cronbach's alpha was .84, suggesting good internal consistency.[798] 93 pairs of data were used to test the test-retest reliability of the scale over 24 to 48 hours. ICC's ranged from .51 for satisfaction with neighbourhood services to .90 for total satisfaction; only three scales fell below 0.6.
	Sensitivity to change	No psychometric research on sensitivity to change was identified.
	Acceptability/Usability	No research on acceptability/usability was identified; however, the SLDS is short and easy to understand, taking just 10 minutes to complete.[798]
Quality of Life Scale (QLS)	**Validity**	Factor analysis yields evidence for good construct validity.[783] In order to empirically assess the plausibility of the four-factor model, a principal component factor analysis with varimax rotation was performed on 111 cases rated on the QLS by one of the authors. Results showed that these factors account for 73% of the variance of the QLS, with respective percentages of variance for the four factors being approximately 52%, 9%, 7% and 6%. [783] Similarities of the factor loadings in both magnitude and pattern were evident for both males and females.
	Reliability	The measure has demonstrated good test–retest reliability for nearly all items of the scale, categories and overall score. The internal consistency has been reported as ranging from 0.8–0.9 for the global score[799].
	Sensitivity to change	The measure has demonstrated sensitivity to change[800]. The QLS has been reported as substantially more sensitive to subtle change and treatment effects than the patient-reports[800]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	The measure has been widely used in the evaluation of psychopharmacologic treatments, predominantly in outpatients [801]
Wisconsin Quality of Life Index – Canadian Version (CaW-QLI)	**Validity**	Construct validity was assessed using Spearman correlations between domains, and between domains and the global score of the CaW-QLI. Each individual domain demonstrated significant correlations with the global score. For seven of the eight domains, these correlations were over .56 for the English, and over .50 for the French, clients. The correlations between domains were low, and ranged from .01 (occupational activities and activities of daily living) to .59 (psychological well- being and symptoms) in both the English and French sample. Convergent validity was tested using Spearman's correlations between the Spitzer's QL-Index and Uniscale, the SF-36 and the global CaW-QLI score. The correlations between the CaW-QLI global score and the Spitzer's QL-Index were .72 and .58 for the English and French clients, and .36 and .56 with the Uniscale. Discriminant validity with the SF-36 found correlations were higher with dimensions related to mental health (E: .53; F: .59) and role emotional (E: .51; F: .46), and lower with the dimensions related to physical functioning (E: .15; F: .27), bodily pain (E: .36; F: .25) and role physical (E: .40; F: .39).[784]
	Reliability	Test-retest reliability was determined by using Concordance Correlation Coefficients, which ranged from .47 to .76 among the QOL domains in the English population, whereas it ranged between .36 and .83 in the French. The test-retest reliability for the global score was .80 (E) and .85 (F) [784]. Internal consistency for the whole measure was calculated using Cronbach's alphas, yielding .78 for the English-speaking sample and .7 for the French-speaking sample. Alphas for the English speaking domains ranged from .33 to .86, and .08 to .88. for the French version.
	Sensitivity to change	No psychometric research on sensitivity to change was found.
	Acceptability/Usability	The average time to complete the CaW-QLI has been shown to be 25 minutes.[784]
Brief Life Satisfaction Scale (BLSS)	**Validity**	To determine dimensionality of the BLSS, factor analyses were performed on BLSS items in three samples: college students, non-referred adults and referred adults. Three factors: self-satisfaction, interpersonal satisfaction, and social role satisfaction, emerged in each sample.[785]. The measure was shown to correlate significantly with a range of measures, both on total score and subscales, including the Quality of Life Index, the Beck Depression Inventory, the Purpose in Life Scale, the PANAS Scales, the Hope, and the Satisfaction with Life Scale[785].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	The Cronbach alphas for the BLSS Total score ranged from .83 to .88 across the samples, indicating good internal consistency. Cronbach alphas for the subscales ranged from .54 to .83, with the lowest alphas found for the Interpersonal Satisfaction and Social Role Satisfaction subscales in the community mental health center sample. Overall, 90% of the alphas for the BLSS subscales were above .60, and 43.3% were above .70. Test-Retest reliability coefficients ranged from .56 to .87 in the non-referred samples and ranged from .46 to .68 for the referred sample.[785] BLSS scores correlate most highly and positively with measures of similar constructs and negatively with constructs associated with a low level of life satisfaction. Interpersonal Satisfaction and Social Role Satisfaction scores on the BLSS are most related with their corresponding subscales of the OQ-45, although, as expected, in a negative direction. Further, BLSS Self-Satisfaction scores are most highly related, although, as expected, in a negative direction, with the OQ-45 Symptom Distress Score, giving weight to divergent validity.[785]
	Sensitivity to change	Sensitivity to change was tested with paired t-tests in two groups (psychiatric outpatients, and drug and alcohol day program participants). Statistically significant differences were found between the two time-points in both samples (after one week for the drug and alcohol group, and between one and five weeks for the psychiatric outpatients).[785]
	Acceptability/Usability	The median time to complete the measure has been demonstrated as 57 seconds (range, 35 to 115 s).[785]
Satisfaction with Life Scale (SWLS)	**Validity**	Construct validity was established using principle factor analysis, yielding a single factor accounting for 66% of the variance.[29] Further evidence of construct validity was established by examining SWLS scores in a variety of samples and identifying that the lowest satisfaction comes from groups such as psychiatric patients, prisoners, students in poor and turbulent countries and abused women.[802] Convergent validity was tested on two samples, who completed SWLS and various other measures, including Cantril's (1965) Self Anchoring Ladder, Andrews and Withey's (1976) D-T scale, Fordyce's (1978) single item measure of happiness, Fordyce's (1978) percent time happy question, Campbell, Converse and Rodger's (1976) semantic differential-like scale, Bradburn's (1969) Affect Balance Scale, Tellegen's (1979) wellbeing subscale of his Differential Personality Questionnaire. Correlations were found across both samples from .50 to .75 (except for the negative affect scale, which correlated at -.32 to .37 for each sample). Scores correlate at .02 with the Marlowe-Crowne social desirability scale, demonstrating SWLS does not measure social desirability.[29] A group of therapy patients at the beginning of their therapy and a group 1 to 2 months into their treatment displayed a significant difference in life satisfaction scores, in the expected direction, providing some evidence for discriminant validity.[802]

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	A study of test-retest reliability was conducted with 76 students who completed the scale twice over a 2-month interval. Testing yielded a correlation coefficient of .82. Coefficient alpha for internal consistency was .87 (good).[29]
	Sensitivity to change	Within-group comparisons of some members of a pre-therapy group vs. post therapy group yielded a significant increase in SWLS scores.[802] Elderly carers of spouses with dementia who completed the SWLS at two separate interviews over 15 to 18 months as their care recipient showed objective declines in functioning reported significantly lower scores on the SWLS at time two.[802]
	Acceptability/Usability	Useful for clinical populations, or to measure change in life satisfaction due to an intervention, and is available in several languages.[802]
Personal Wellbeing Index (PWI)	**Validity**	The PWI is sensitive to differences in age, gender and geographical location.[803] The combination of both unique and shared variance by the eight domains typically explains about 30 to 60% of the variance when the domains are collectively regressed against 'Satisfaction with life as a whole'.[30] The eight domains also consistently form a single stable factor.[30] As a measure of convergent validity, a correlation of .78 with the Satisfaction with life scale (Diener, Emmons, Larsen, & Griffin, 1985) has been reported.[30] The Reports on the Australian Unity Wellbeing Index, that incorporate the Personal Wellbeing Index, indicate a level of sensitivity between different demographic groups that is consistent with the theory of subjective wellbeing homeostasis.[30]
	Reliability	Sixteen surveys of the Australian population have produced a maximum variation of 3.2 percentage points in subjective wellbeing.[30] Cronbach alpha lies within the good to acceptable range (.70 to .85) in Australia and overseas. Inter-domain correlations are often moderate at around .30 to .55 and item-total correlations are at least .50.[30] The index has demonstrated good test-retest reliability across a 1- to 2-week interval with an intra-class correlation coefficient of .84.[30]
	Sensitivity to change	No psychometric research on sensitivity to change was identified.
	Acceptability/Usability	Parallel forms of the PWI have been created to allow an appropriate version of the scale to be used with a number of population sub-groups. These four groups are the general adult population (aged at least 18 years), school-age children and adolescents, pre-school age children, and people who have an intellectual disability or other form of cognitive impairment.[30]

3.6 Experience of Service

Table 15. Profile of experience of service measures

MEASURE	DATE	COUNTRY	DESCRIPTION
Consumer Experiences of Services			
Mental Health – Consumer Perceptions and Experiences of Services (MH-CoPES)	2011	Australia	The MH-CoPES project aimed to develop a mechanism for consumer participation in service evaluation and quality improvement.[804] It produced a 27-item questionnaire for adult consumers. Items are rated on three colour-coded options: Needs lots of improvement (Red), Needs some improvement (Orange), Needs no improvement (Green).
National Research Corporation National Health Service (NRC-Picker)	1993 2001	United States United Kingdom Europe	The questionnaire comprises seven core dimensions: information and education, coordination of care, physical comfort, emotional support, respect for patient preferences, involvement of family/friends, and continuity of care. Response format varies depending on the question, but can be either dichotomous yes/no or Likert-type responses. The length of the questionnaire varies depending on the target population and the survey. For example, the NHS Community Mental Health Survey 2013 is 51 items long. A problem is defined as an aspect of health care that could, in the eyes of the patient, be improved upon.[805] A shorter 15-item version is available.[806]
Consumer and Carer Experience Questionnaire (C&CEQ)	2008	Australia	The C&CEQ were developed for the Department of Human Services, Victoria in order to seek consumer perspectives of service quality regarding Victorian public mental health services. Different versions of the questionnaires have been adapted for Child and Adolescent, Adult, and Aged Persons Mental Health Services and for Psychiatric Disability Rehabilitation and Support Services. Survey questions require Yes/No/Not Sure responses. The survey questions 'reflect consumer identified priorities, needs and expectations of services, and are linked to the National Standards for Mental Health 2010 to provide specific guidance for quality improvement and meaningful benchmarks for consumers and service providers'. Depending on the respondent's circumstances, up to 56 items may be completed. Questions cover topics such as time taken to access services, whether service providers listened to and respected consumers, support and information provided for consumers, and involvement in discharge planning.[807]
Psychiatric Outpatient Experience Questionnaire (POPEQ)	2004	Norway	The questionnaire comprises 11 items with five-point descriptive scales that are summed to give a single score, with 0 and 100 representing the worst and best possible experiences of care, respectively. Two example items are ''Did you have a say in choice of treatment package?'' and ''Do you feel that the health professional understands your situation?'' The response categories are *not at all*, *to a small extent*, *to a moderate extent*, *to a large extent* and *to a very large extent*.[808]

MEASURE	DATE	COUNTRY	DESCRIPTION
The Mental Health Statistics Improvement Program Consumer Survey (MHSIP-Consumer)	1996	United States	The MHSIP Adult Consumer Survey was developed to evaluate the performance of a mental health system on the domains of access, quality and appropriateness of services, and outcomes from the consumer's perspective.[809] Each domain in turn has a number of questions; for example, the access domain asks questions about convenience of the location of services, availability of a psychiatrist, financial barriers to care and frequency of access.
Consumers Experience of Care	2012	Australia	This national project aims to develop a psychometrically robust, recovery-orientated, consumer experience of care instrument, suitable for use in public adult mental health services building on existing tools and processes. A proof of concept trial has been completed but the report is yet to be formally released. • Questions 1 to 27 are referred to as experience questions. • Questions 28 to 31 are referred to as outcome questions. • Questions 34 to 35 are open-ended questions (free text). • Questions 36 to 42 are demographic questions.
Carer Experiences of Services			
Carer satisfaction interview	1998	Australia	To assess service compliance with the Australian National Standards for Mental Health Services (1996), a series of assessment tools were developed, titled the *Tools for Reviewing Australian Mental Health Services* (TRAMHS).[810] Among these tools was a Carer satisfaction interview, comprising a series of 31 possible questions addressing the areas of Rights, Consumer and carer participation (optional), Cultural awareness, Integration, Entry, Treatment and support, Supported accommodation (optional), Medication and other medical technologies, Inpatient care, Planning for exit, Exit and re-entry. The interview was designed to be administered in a group of up to ten carers with trained reviewers asking questions and taking notes. Carer responses are then rated with a score from 0 = *not met* to 3 = *exceeded*, indicating the level of satisfaction with each service component. Although originally administered in a group format, this may be adaptable to individual carers and interviewers. Time taken for the group format is estimated at 45 minutes, though it might take considerably less time for individual administration. It would appear that because the interview is designed to measure compliance with national standards, it is not designed to provide an overall 'score'.
Questionnaire for carers of people who use social services (Wales) (QPSS-Wales)	2008 (?)	Wales	The Social Services Inspectorate for Wales and the Audit Commission in Wales regularly conduct audits of social services, including mental health services. The carers' survey developed for this purpose comprises 32 self-report questions (including demographics and free text comments) assessing seven aspects of service provision: access to services, assessment, care management and review, range of services provided, quality of services provided, arrangements to protect vulnerable people, and success in promoting independence and social inclusion. Questions about services are answered on a four-point Likert scale from *strongly agree* to *strongly disagree*.

MEASURE	DATE	COUNTRY	DESCRIPTION
Carer Participation Survey (CPS)	2008	Australia	The Carer Participation Survey was developed as part of a study to determine if the introduction of practice standards into a district mental health service had an effect on the level and quality of carer participation. The survey consists of eight statements rated on a Likert scale from 1 = *strongly disagree* to 10 = *strongly disagree*. The eight items relate directly to aspects of the practice standards, including mental health workers' consultation with carers, provision of information to carers, and whether carer involvement enhanced care and treatment.
Consumer and Carer Experience Questionnaires (C&CEQ)	2008	Australia	The C&CEQ were developed for the Department of Human Services, Victoria, in order to seek consumer and carer perspectives of service quality regarding Victorian public mental health services. The C&CEQ are based on the earlier Carer survey of Victorian Public Mental Health Services 2003/2004[19.] (VPMHS 2003/2004), developed for the same purpose. Different versions of the questionnaires have been adapted for Child and Adolescent, Adult, and Aged Persons Mental Health Services and for Psychiatric Disability Rehabilitation and Support Services. Survey questions require Yes/No/Not Sure responses. The survey questions 'reflect consumer and carer identified priorities, needs and expectations of services, and are linked to the *National Standards for Mental Health 2010* to provide specific guidance for quality improvement and meaningful benchmarks for consumers, carers and service providers.'[20] Depending on their circumstances, carers may complete up to 56 items. Questions cover topics such as time taken to access services, whether service providers listened to and respected carers, support and information provided for carers, and involvement in discharge planning.
Carer Well-being and Support Questionnaire (CWS)	2009	United Kingdom	The CWS is a 49-item questionnaire (plus 14 items collecting background information) developed to assess outcomes for carers of people with a mental illness or with dementia. The CWS is the result of the redevelopment of the earlier Carers' and Users' Expectations of Services – Carer survey (CUES-C).[15] The CWS consists of two scales: Carer Support and Carer Well-being, the former scale being the most relevant to this review. Measured by the Support scale are five domains: information and advice for carers, involvement in treatment and care planning, support from medical and/or care staff, support from other carers, and taking a break (respite). The Support scale is made up of 17 items. The redevelopment of the CUES-C involved testing of the questionnaire items for acceptability with carers. The support scale of the CWS was used as the measure for the UK National Audit of Schizophrenia 2011[11].
Disability and mental health service users and carers satisfaction survey 2009	2009	Australia	The Disability and mental health service users and carers satisfaction survey (2009) was developed specifically for the Queensland Disability services, Department of Communities 2009 review of service user and carer satisfaction with government and non-government specialist disability services that are funded by the Queensland Government's Department of Communities. The survey is designed to measure satisfaction with accommodation support, community support, community access and respite services. The survey was primarily conducted using Computer-Assisted Telephone Interviewing (CATI), although email and hard copy

MEASURE	DATE	COUNTRY	DESCRIPTION
			survey forms were also available. The survey contains eleven possible mental health service-related questions that are rated on a five-point scale ranging from 1 = *Very dissatisfied* to 5 = *Very satisfied*. Responses to the survey do not seem to be reducible to a single total, rather, response options are collapsed into *Satisfied*, *In-between* and *Dissatisfied* and percentages of carers surveyed assigned to each category. This survey would need to be adapted for individual use.
The Personal Social Services Survey of Adult Carers in England 2009-2010 (PSSS-Adult Carers)	2009-2010	England	The PSSS-Adult Carers was developed to provide information to English Councils with Adult Social Services Responsibilities on their progress towards the goals set out in the Government's Carer Strategy. The survey consists of 58 self-report questions assessing the impact of caring on the carer's quality of life and their experiences of using social services. Domains assessed in this regard include satisfaction with social care support and services, views about aspects of service quality and views about quality of life. Response options vary per question but satisfaction questions are generally rated on a seven-point scale from *I am extremely satisfied* to *I am extremely dissatisfied*. Service quality questions are based on the response format *exceeded, met, almost met* or *not met.*
Mental Health Carers' Survey (MHCS)	2010	Australia	The MHCS was developed by the Mental Health Council of Australia as part of an ongoing commitment to assess whether the lives of carers of people with a mental illness 'are improving, the extent of that improvement, as well as the quality and availability of services available to assist them and the person they care for'[11] (p. 13). The domains measured relate to those identified in *Adversity to Advocacy*[5] and to issues raised by carers in workshops held to hear their concerns. The questionnaire consists of 50 questions, some of which comprise a number of items, such that carers may provide responses on up to 113 items. The 15 carer concerns addressed by the survey are the following: listen to and respect carers; integrated recovery-based care for the consumer; more and better trained staff at all levels; knowledge and information for carers; carer and consumer education for all professional groups and agencies; support systems, services and processes established for carers; acute care to be therapeutic and accessible; stigma, discrimination and isolation for carers and consumers; accommodation options for consumers at all levels of care; financial costs to carers; physical and mental health of carers; flexible respite options for carers; privacy and confidentiality issues; early intervention at each episode of care; and employment options for carers.
Carers Experience of Service Provision	2013	Australia	AMHOCN has been tasked with the creation of this measure. An initial draft has been created and a proof of concept trial is being undertaken.
Service Integration			
Network analysis	Various		Network analysis is a 'method for examining the issues of service system fragmentation and coordination'. (p.155)[811] Network analysts identify structural (administrative procedures that define roles) and process (flow of activities, direction and frequency of resource and information exchanges) dimensions and

MEASURE	DATE	COUNTRY	DESCRIPTION
			situational factors associated with forming and maintaining relationships. It is also imperative to measure the effectiveness of these systems as defined in terms of accessibility, adequacy, quality and attainment of goals.[811] Surveys designed for the purpose of measuring service integration tend to be developed or adapted for the particular purpose of that study.[811, 812]
Human services integration measure	2004	Canada	The Human services integration measure[813] quantifies the 'extent, scope and depth of effort' (p.3) of service integration as perceived by local service providers. It also identifies the 'sectors, services or agencies [that] are connected and are collaborating well with each other and which sectors and/or agencies could enhance their collaborative efforts' (p.3). The measure is based on a three-dimensional model of integration, also developed by the authors of the measure, to understand integration among human services. Extent of integration involves quantifying the services involved in the partnership. Scope of integration is the number of services that are aware of or link with others. Depth of integration is a rating from 0 = non-awareness to 4 = collaboration that rates the depth of links among services. Representatives from all services identified within a sector are asked to rate their depth of integration with other services on the list. Scope of integration and Depth of integration are further divided into 'perceived' and 'self-reported' scores, where 'perceived' reflects the perceptions of other services in relation to the focus service, and 'self-reported' is the focus service's rating of their relationship with other services. A total integration score measures the average depth of integration among all services within an area of focus.

Table 16. Psychometric properties of experience of service measures

MEASURE	PSYCHOMETRIC PROPERTIES	
Consumer Experiences of Services		
Mental Health – Consumer Perceptions and Experiences of Services (MH-CoPES)	**Validity**	Concurrent validity of the MH-CoPES has been demonstrated with a strong correlation between the total score of the 24 items of MH-CoPES and a question that asks “Overall, what do you think of the care you received at this community health centre?”(Spearman’s r = .68, p < 0.001, n = 106).[814]
	Reliability	Following state-wide roll out, Cronbach’s coefficient alpha = .96 (n = 1039).[814] Test-retest reliability was tested over two weeks but no significant correlation was found between the two time-points (Spearman’s r = -.31, p = .51, n = 7). There was, however, no significant difference between people’s responses on first and second completion of the Questionnaire (*t*(6) = 1.877, p = .11). From these results the authors conclude that it is not possible to draw conclusions about the test-retest reliability of the MH-CoPES Questionnaire for people using inpatient services.[814]
	Sensitivity to change	No research was identified that examined the MH-CoPES’ sensitivity to change.
	Acceptability/Usability	MH-CoPES only has three response options, and this prevents some forms of statistical analysis being undertaken. Consumers were involved in the development of the measure. Consumers who took part in the 2007 Trial and the 2008 Pilot were asked to provide feedback on the surveys via an evaluation questionnaire. 78% of consumers in the post-survey evaluations stated that they felt comfortable providing feedback via the questionnaire, and 61% thought that their feedback would be useful in improving services.[815]
National Research Corporation (NRC-Picker) National Health Service	**Validity**	No psychometric research was identified that examined the validity of the NRC-Picker.
	Reliability	No psychometric research was identified that examined the reliability of the NRC-Picker.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the NRC-Picker.
	Acceptability/Usability	No research was identified that examined the acceptability or usability of the NRC-Picker.
Consumer and Carer Experience Questionnaire (C&CEQ)	**Validity**	No psychometric research was identified that examined the validity of the C&CEQ.
	Reliability	No psychometric research was identified that examined the reliability of the C&CEQ.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the C&CEQ.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	No research was identified that examined the acceptability or usability of the C&CEQ.
Psychiatric Outpatient Experience Questionnaire (POPEQ)	**Validity**	In Norway, across 90 outpatient clinics the POPEQ was administered as part of a postal survey of 15,422 adults. 6677 (43.3%) patients responded to the questionnaire. Garratt et al[816] undertook a factor analysis of this sample and identified a one factor solution drawing the conclusion that the POPEQ was a unidimensional measure of outpatient experiences of psychiatric services. Bjørngaard et al[817] found variation across service types in experiences of care but attributed these to consumer factors rather than service provider factors.
	Reliability	Internal consistency for the total score was excellent: $\propto = .91$ and for theoretical subscale scores Outcomes $\propto = .81$, Clinician Interactions $\propto = .87$, Information $\propto = .81$.[816] Test-retest reliability is reported as excellent $\propto = .90$.
	Sensitivity to change	No psychometric research was identified on the sensitivity to change of the POPEQ.
	Acceptability/Usability	The measure has been used in a national survey so it demonstrates utility at that level [817]. However, its utility at an individual organizational level is less clear.
The Mental Health Statistics Improvement Program Consumer Survey (MHSIP-Consumer)	**Validity**	Factor analysis (n = 362) with oblique (Promax) rotation showed a three-factor solution accounted for 47% of variance.[818] The three factors were access to treatment (factor loadings of items = .57-.86), quality and appropriateness of care (factor loadings = .41-.85), and outcomes of treatment (.41-.75). Two items ('services I received were not helpful' and 'I experienced harmful medication side effects) did not load on any factor. Assessing criterion validity, Eisen et al correlated item scores with global scores for the MHSIP and these ranged from .01 to .52. The MHSIP item relating to recommendation of their health plan to others varied significantly between the six health plans valid for the participants, showing the discriminant validity of this item for this particular purpose. In a study of 459 consumers with severe mental illness in the public mental health system in the US, Jerrell[819], using a principal components analysis with varimax rotation. Jerrell also identified the three-factor solution identified by Eisen et al, with similar factor loadings for the three factors: 1) .57-.74; 2) .56-.75; and 3) .60-.80. Convergent validity was tested by correlating scores from the MHSIP with scores on the Consumer-to-Consumer Evaluation Team (CCET) questionnaire. Four of the CCET subscales showed good convergent validity with two of the MHSIP factors.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Cronbach's alpha for the three factors identified in the Eisen et al[818] study were .81, .89 and .89, respectively, demonstrating good internal consistency. Item-total correlations for these sub-scales ranged from .39 to .73. In the study of Jerrell[819], internal consistency was slightly lower: factor 1 α = .73, factor 2 α = .81 and factor 3 α = .79, though still acceptable to good. Subscale-total scale correlations were r = .77 to .88. Jerrell also assessed test-retest reliability (n = 186) over two weeks. Test-retest reliability varied across the three subscales: Access (r = .61), Quality and appropriateness (r = .60) and Perceived outcomes (r = .45) and was moderate.
	Sensitivity to change	No psychometric research was identified on the sensitivity to change of the MHSIP.
	Acceptability/Usability	An unpublished study by Ganju et al[820] states that consumers surveyed indicated that the MHSIP-Consumer assessed concerns that were important to them, and that it was easy to understand and complete. Participants in Eisen et al's study involving consumers of US behavioural health or substance abuse services took less than 20 minutes to complete the MHSIP survey (n = 362).[818] Just 11% of the respondents felt the survey was too long, and 21% found sections to be confusing or difficult, while 79% thought is covered important aspects of service. Percentages of missing data ranged from 2.7% to 6.4% for specific items.
Consumers Experience of Care	**Validity**	Results yet to be released
	Reliability	Results yet to be released
	Sensitivity to change	Results yet to be released
	Acceptability/Usability	Results yet to be released
Carer Experiences of Services		
Carer satisfaction interview	**Validity**	No psychometric research was identified on the validity of the CSI.
	Reliability	No psychometric research was identified on the reliability of the CSI.
	Sensitivity to change	No psychometric research was identified on the sensitivity to change of the CSI.
	Acceptability/Usability	No research was identified on the acceptability or usability of the CSI.
Questionnaire for carers of people who use social services (Wales) (QPSS-	**Validity**	No psychometric research was identified on the validity of the QPSS-Wales.
	Reliability	No psychometric research was identified on the reliability of the QPSS-Wales.

MEASURE	PSYCHOMETRIC PROPERTIES	
Wales)	**Sensitivity to change**	No psychometric research was identified on the sensitivity to change of the QPSS-Wales.
	Acceptability/Usability	No research was identified on the acceptability or usability of the QPSS-Wales.
Carer Participation Survey (CPS)	**Validity**	No psychometric research was identified on the validity of the CPS.
	Reliability	No psychometric research was identified on the reliability of the CPS.
	Sensitivity to change	No psychometric research was identified on the sensitivity to change of the CPS.
	Acceptability/Usability	Members of a study reference group, representing a number of mental health and carer organisations, took a draft of the Carer Participation Survey back to their respective organisations for comment before using the Survey in the study.[821] The outcome of this consultation was not reported.
Consumer and Carer Experience Questionnaires (C&CEQ)	**Validity**	Note: The Wallis Consulting Group Report (2008) from which these psychometric outcomes are taken does not show the exact results of the statistical tests conducted and only provides a summary of the results. Therefore, the veracity of these conclusions cannot be assessed. The surveys were developed in conjunction with carers, who assisted in determining central themes to assess. The content of the survey also reflects a number of the components of the Standard 7 criteria of the Australian *National Standards for Mental Health Services 2010,* both suggesting good content validity. The researchers excluded items that did not correlate with others and then conducted both principle components analysis and categorical principal components analysis on the resulting scales. They reported that the items loaded onto one single main factor (factor loading of .5 or greater), which they characterised as ‘the effectiveness of interpersonal care’.
	Reliability	The test-retest and inter-rater reliability of the scales were not tested.
	Sensitivity to change	Sensitivity to change was not determined; however, the scales were tested to determine the distribution of responses (i.e. that participants provide a range of responses and not just predominantly ‘yes’ to most items or ‘no/not sure’ to most items). Remaining items show an acceptable range of responses.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	The Consumer and Carer Experience Surveys (2008) represent a redevelopment of the earlier VPMHS (2003-2004) Surveys[19]. Consumers, carers and service providers were involved in the development and testing of the 2003-2004 VPMHS Survey. Following development of the surveys, the various forms were distributed across Victoria through Child and Adolescent, Adult, and Aged Persons Mental Health Services. During and following the Survey implementation period, information was gathered from service managers, staff, consumers and carers regarding the strengths and weaknesses of both the Survey itself and the implementation process [805] through a series of networks, focus groups, interviews and pilot studies. This feedback led to the recommendation that the language be simplified, the format improved, the length reduced and amendments be made to the rating scale before they are used again. The redevelopment of the 2003-2004 surveys aimed to develop both a survey tool and data collection methodology that overcame the concerns raised in regard to the earlier survey. The survey was administered primarily using Computer-Assisted Telephone Interviews (CATI), purely as a means of increasing the sample size for the study in which it was used.
Carer Well-being and Support Questionnaire (CWS)	**Validity**	Internal consistency was excellent for the Support scale (Cronbach's α = .97, n = 92) with item-total correlations ranging from .71 to .86 (Mean = .78). Factor analysis and item convergent/discriminant analyses showed support for the validity of the Support scale, with all items loading at > .4 (.71-.88, N = 351). The total two-factor model (Well-being + Support) accounted for 50.8% of variance. Participants completed a global question asking about their general satisfaction with the support received as a carer; those who showed lower satisfaction also showed significantly lower scores on the CWS-Support scale compared with those with higher levels of satisfaction (t(df) = -19.01 (337), P = < .01, n = 339). Carers who had received a carer training course showing higher scores on the Support scale (t(df) = -4.64(18), p < .01, n = 19). Support satisfaction was not correlated with age, although men did demonstrate higher satisfaction. (Note: this psychometric testing has now been published in a peer-reviewed journal: Quirk A, Smith S, Hamilton S, Lamping D, Lelliot P, Stahl D, Pinfold V, Andiappan M. (2012). Development of the carer well-being and support questionnaire (CWS), Mental Health Review Journal, 17 (3), 128-138. DOI 10.1108/13619321211287184)
	Reliability	Test-retest reliability was demonstrated using data from 92 pairs of carer data (ICC = .88) over a two-week period.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the CWS.

MEASURE	**PSYCHOMETRIC PROPERTIES**	
	Acceptability/Usability	The CWS represents a re-development of the CUES-C and the user testing and review process to develop a draft version of the CWS is outlined in detail in elsewhere (Quirk et al, 2009). Two CWS drafts were field-tested with carers of people with a mental health problem or dementia. CWS-v2, a 49-item version, was tested with 361 carers. Criteria used to measure acceptability: missing data (<5%), floor/ceiling effects and skew were all within the pre-determined acceptable range (Quirk et al, 2009). Qualitative feedback written on the surveys led to some further changes to the wording of some questions.
Disability and mental health service users and carers satisfaction survey 2009 - Queensland	**Validity**	No psychometric research was identified that examined the validity of the survey.
	Reliability	No psychometric research was identified that examined the reliability of the survey.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the survey.
	Acceptability/Usability	The Queensland Government Department of Communities conducts this satisfaction survey biannually. In 2009, the survey and data collection methodology was altered based on feedback from the 2006 survey to improve clarity and user friendliness[822].The survey development and implementation was overseen by a Survey Focus group from a range of representatives from departments related to Disability Services and from non-government organisations[822]. It is not clear whether consumers or carers were represented in this group.
The Personal Social Services Survey of Adult Carers in England 2009-2010 (PSSS-Adult Carers)	**Validity**	Development through a review of policy and research, combined with consultation with stakeholders including carers suggests good face validity of the items.
	Reliability	No psychometric research was identified that examined the reliability of the PSSS-Adult carers.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the PSSS-Adult carers.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Acceptability/Usability	The PSSS-Adult Carers survey was developed from the earlier Kent County Council Carers' Survey. Items from this survey and other sources was developed using a four-stage process to create the PSSS-Adult Carers survey. The initial reviewing of existing policy and research identified a list of topics (variables) that might be included in the survey. A list of these variables were sent to policymakers, carer representatives and Councils with Adult Social Service Responsibilities (CASSRs) who were asked to identify those variables most important to include in the survey. Carers' satisfaction with social care support and services and views on service quality were included. Cognitive testing in the form of interview with 30 carers, including carers of people with a mental illness, was undertaken to test the ease of understanding and answering the question, and to ensure that the questions were relevant to carers. The researchers provide the outcomes of these interviews to inform further development of this survey.[823] A version of this survey was subsequently used for the 2009-2010 survey of carers.[824]
Mental Health Carers' Survey (MHCA)	**Validity**	The items included in the MHCA reflect 15 major carer concerns outlined in the report *Adversity to Advocacy: the lives and hopes of mental health carers[825]).* This report was developed from the outcomes of 116 day long-workshops with mental health carers across Australia. This process suggests excellent face validity of the survey items.
	Reliability	No psychometric research was identified that examined the reliability of the MHCA.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the MHCA.
	Acceptability/Usability	While testing of individual survey items was not identified, the process used to develop the items, outlined under 'validity,' suggests that the item content would be acceptable to carers.
Carers Experience of Service Provision	**Validity**	Under development
	Reliability	Under development
	Sensitivity to change	Under development
	Acceptability/Usability	Under development
Service Integration		

MEASURE	**PSYCHOMETRIC PROPERTIES**	
Network analysis	**Validity**	In a US study of child and adolescent mental health system, Helfinger[811] carried out a network analysis at a demonstration site and two 'standard' comparison sites to identify whether a coordinated system of care was in place. Respondents were 131 representatives from 79 agencies. Participants completed the 'Fort Bragg Children and Youth Services Network Study' (FBNS) questionnaire that measures linkages between the respondent's agency and other providers and agencies identified in the network by asking questions about awareness of community agencies, frequency of interagency staff interaction, frequency of referrals among community agencies, amount of client-centred information exchange, extent of activity coordination among community agencies, mutual goal attainment, and satisfaction with relationships with other providers and agencies. This data was analysed using computer-based social network analysis programs to generate quantitative data on the structural dimensions of the inter-organisational network. They also completed the 'Assessing Local Service Systems for Military Dependent Children and Youth' (ALSS) instrument which assesses the problems experienced by military children and youth in the area; availability, adequacy, and quality of mental health and related services; service system performance (coordinated and responsive); and service system goal attainment. Descriptive data was generated from this instrument to allow comparisons between sites. Comparison of data between the demonstration site, which was developed specifically to reduce fragmentation of services and to provide a continuum of care, and two comparison sites showed that the demonstration site had achieved a more coordinated system and better outcomes for consumers accessing the services despite similarities in the structure of the system. This finding lends some support for the discriminant validity of the network analysis framework and of the two measures used in the study.
	Reliability	Measures tend to be developed or adapted for the specific purpose.
	Sensitivity to change	Measures tend to be developed or adapted for the specific purpose.
	Acceptability/Usability	Measures tend to be developed or adapted for the specific purpose.

MEASURE	PSYCHOMETRIC PROPERTIES	
Human services integration measure	**Validity**	Brown et al[813] pilot-tested the human services integration measure in two children's programs within one region to measure integration of services associated with each program. They report demonstrating face and content validity through ease of use by respondents and because respondents found the data valuable. The two different programs included in the study, which were known to be at different stages of development of their integration, showed differences in scores on the measure as predicted. This provides some evidence for the discriminant validity of the measure. Score on the human services integration measure also showed relatively weak correlations with a partnership synergy questionnaire, providing some evidence for concurrent validity of the measure.[826]
	Reliability	No psychometric research was identified on the reliability of the Human Services Integration measure.
	Sensitivity to change	No psychometric research was identified on the sensitivity to change of the Human Services Integration measure.
	Acceptability/Usability	Browne et al[813] reported that community experts found the measure helpful and that community leaders were able to use the measure and valued the information gained from it. Browne's Human Services Integration Model has since been used to evaluate the implementation of the PRISMA model of coordination-based integrated care in Quebec, Canada.[827]

3.7 Multidimensional

Table 17. Profile of multidimensional measures

MEASURE	DATE	COUNTRY	DESCRIPTION
Camberwell Assessment of Need (CAN)	1995	United Kingdom	The CAN consists of 22 items of need and has four sections for each item. Section 1 is scored by rating 0 = *no need*, 1 = *met need due to help given*, 2 = *unmet need* and 9 = *not known*. Sections 2 to 4 assess the amount and type of help needed and received. Because of its length, the CAN met with some resistance and the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) was developed, which consists of only section 1.[828]. Trauer et al have also developed a patient-completed version of the CANSAS – the CANSAS-P, with patients self-completing the scales with assistance from staff only as needed.[829]
Collaborative Goal Index/COMPASS	2013	Australia	Setting, working toward, and achieving, individually-set goals has been shown to improve well-being (see Clarke et al., 2009).[830] Indeed, the Collaborative Recovery Model (CRM) identifies the third component of recovery as 'collaborative life visioning and goal striving'.[831] The Collaborative Goal Index (CGI) or COMPASS, as re-named by Oades, Crowe and Deane[831] is a measure of goal attainment used in conjunction with the Collaborative Goal Technology (CGT) goal-setting intervention used during case management with people with a mental illness, who set a maximum of three, three-monthly goals aligned with their goals for recovery.[832] CGT was developed to operationalise component 3 of the CRM.[831] To calculate the CGI score, the mental health consumer and worker assign a score from 0 = lowest level of goal attainment to 2 = highest level of goal attainment to each goal set and multiply this by a level of importance score set by the consumer (10 points distributed across three goals). Scores for all goals are then added together, divided by 20 (the maximum possible score) and multiplied by 100 to gain a 'goal attainment percentage'. Scores above 50 suggest better than expected outcomes.
Health of the Nation Outcomes Scales (HoNOS)	1996	United Kingdom	The HoNOS was developed by Wing and colleagues from the College Research Unit of the Royal College of Psychiatrists in the United Kingdom as a means of assessing the extent to which the Government's Health of the Nation target 'to improve significantly the health and social functioning of mentally ill people' was being met. Specifically, it was designed as an instrument that could be used routinely by clinicians in the United Kingdom's National Health Service to measure outcomes for consumers with a mental illness.[833, 834]. Each item is rated on a five-point scale (0 = *no problem*; 1 = *minor problem*; 2 = *mild problem*; 3 = *moderately severe problem*; 4 = *very severe problem*), resulting in individual item scores, subscale scores and a total score. In assigning ratings, the clinician makes use of a glossary which details the meaning of each point on the item being

MEASURE	DATE	COUNTRY	DESCRIPTION
			rated.[835]
Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA)	1999	United Kingdom	Like the HoNOS, the HoNOSCA arose from the Health of the Nation strategy in the United Kingdom. The HoNOSCA was developed by the Department of Child and Adolescent Psychiatry at the University of Manchester, in conjunction with the College Research Unit from the Royal College of Psychiatrists, and with the assistance of a multi-disciplinary steering committee. It was welcomed as having the potential to fill a gap identified by several reviews [836, 837] – namely that existing instruments were too lengthy or too specific to be useful for measuring global outcomes for children and adolescents in routine clinical practice.[838, 839] The clinician is asked to draw on all available information and to rate each item on a scale of 0 = *no problem* to 4 = *severe problem* for the period under consideration. For Items 1 to 9, the clinician is asked to 'rate the most severe problem that occurred in the period rated'. Items 10 to 13 require 'a more general rating'.[840] A comprehensive glossary provides descriptions of the anchor points associated with each item.[841] Once completed, the HoNOSCA is scored in a manner that provides individual item scores, subscale scores and a total score. The total score comprises the items in Section A – Items 1 to 13 – only.[838, 839]
Health of the Nation Outcomes Scales 65+ (HoNOS 65+)	1999	United Kingdom	Two studies examining the psychometric properties of the HoNOS when used with older consumers concluded that although the HoNOS performed well, some modifications might make it more appropriate for use with this group. [842, 843] As a consequence, Burns and colleagues [844] from the College Research Unit of the Royal College of Psychiatrists began a process of modifying the general adult scale for older people. This process involved developing, piloting and evaluating the modified instrument, which became known as the HoNOS65+. Each item is rated on a five-point scale, with anchor points that more or less mirror the HoNOS (0 = *no problem*; 1 = *minor problem requiring no action*; 2 = *mild problem, but definitely present*; 3 = *moderately severe problem*; 4 = *severe to very severe problem*). Individual item scores, subscale scores and total scores can be calculated.[844] For each item, the clinician makes an assessment of the consumer's situation over the recent period. The overarching instruction is to 'rate the most severe problem that occurred in the period rated', but for Items 11 and 12 this is modified to 'rate the usual [situation]'.
Outcome Rating Scales (ORS) and Session Rating Scales (SRS)	2003/2000	United States	The ORS is a four-item visual analogue self-report outcome measure designed for tracking client progress in every session.[845-850] Each item requires the client to make a mark on a ten-centimetre line where marks to the left indicate more difficulties in the particular domain and marks to the right depict fewer difficulties. The SRS was designed as a 'working' alliance measure designed specifically for every session clinical

MEASURE	DATE	COUNTRY	DESCRIPTION
			use. The SRS is another four-item visual analogue scale that is based on encouraging patients to identify any alliance problems with their therapist so that the clinician may change to better fit patient expectations.

Table 18. Psychometric properties of multi-dimensional instruments identified

MEASURE	PSYCHOMETRIC PROPERTIES	
Camberwell Assessment of Need (CAN) and CANSAS	**Validity**	CANSAS-P: Agreement between scores on the CANSAS-P and the CANSAS patient rating ranged from ICC = .33 for unrated to .81 for unmet need [829] (poor to strong correlations), suggesting that there is relatively high consistency between the two forms, with the exception of the rating of 'unrated'. Correlation between a staff-rated CANSAS and the CANSAS-P (N = 89) varied across ratings, with the lowest agreement, ICC = .25, again for 'unrated' and highest, again, for unmet needs (ICC = .61). CANSAS: Reininghaus et al[851] tested the convergent, discriminant and predictive validity of the CANSAS in a sample of 98 patients with psychosis in the UK. Items on both the staff and patient versions loaded onto a common factor (factors loadings for CANSAS-clinician scale were .35 to .77; patient scale factor loadings were .20 to .79) and items also loaded onto domain factors for the domains measured. Items were also observed to load onto an 'unmet needs' factor (12 items significant). Loading of items from the staff and patient versions onto a common factor suggest convergent validity for the two forms. Findings also suggest that the Helping Alliance Scale (HAS)[852], measuring therapeutic alliance, and the CANSAS measure distinct concepts and therefore suggest discriminant validity. Indication of less common unmet needs on the unmet needs factor predicted psychiatric inpatient days and inpatient care, also suggesting predictive validity of the CANSAS. CAN: Phelen et al[853] correlated the item and summary scores of the CAN with scores on the GAF. While the former correlations were low, the latter was significant at r = -.51.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	Andresen et al[854] tested the inter-rater reliability of the CANSAS using three raters for a combination of 32 inpatients and day patients, most of whom had schizophrenia, at a NSW psychiatric rehabilitation unit. For items of need, agreement was highest with patient ratings, ranging from κ = .39 to 1.0 (fair to very good), with 64% of items above .70 (good to very good). Staff ratings showed less agreement: κ = .20 to 1.0 (poor to very good) with only 36% of items above κ = .70 (good to very good) and particularly low coefficient (.20 to .59) for items 10 and 11 (safety to self and to others). For ratings on level of need, correlations were generally high for both staff and patient ratings (r = .80-.91), with the exception of staff ratings for met needs (r = .53). The authors suggest that differences in staff ratings might result from differing understandings of 'need' and use of additional information by raters when attributing their ratings. They suggest better defining 'need' and adding more levels of need. Andresen et al's study[854] found lower levels of inter-rater reliability than studies by Phelan[853], and Hansson et al[855] that tested inter-rater reliability for the CAN. Agreement for Phelan's study was κ = .65 to 1.0 for patient ratings and κ = .74 to 1.0 for staff ratings (N = 60) and correlations for summary scores were r = 0.0 for patient ratings and r = 0.98 for staff ratings. Inter-rater correlation for the summary score was .98. Agreement in the Hansson et al study of the CAN, using fifteen pairs of raters, was also higher than for Andresen et al, who observed more than 80% agreement in 91% of ratings. Andresen suggest that these differences could result from raters from different backgrounds. Macpherson et al compared ratings on the clinician and patient versions of the CANSAS of 225 patients in the UK. Agreement between the forms ranged across items from κ = .61 to .84 (good to very good), with the exception of the safety to others scale, which, as observed by Andresen et al, had a particularly low level of agreement of κ = .33 (unacceptable). Test-retest reliability over a one week study in a study by Phelen et al[853] was κ = .19-1.0 for staff and .21 to .93 for patients (most were above .40). Test-retest reliability of the CANSAS-P was tested over about 12 days (n = 41), ICC = .36-.81 (poor to strong), with the lowest reliability observed for met need (ICC = .36) and unrated (ICC = .43).
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the CAN.
	Acceptability/Usability	Trauer et al[829] tested the acceptability to patients of the CANSAS-P (N = 90) using a standard set of questions and asking staff if patients had any difficulty completing the scale. Patients rated that they understood the purpose of the CANSAS-P and it was useful, but were least positive about the instructions and the difficulty of assigning ratings. Responses regarding other aspects such as the language used, the layout, the time taken, and guidelines were more moderate. More than 80% gave positive ratings overall. When asked about patient difficulties with completing the scale, staff stated that about one-third had difficulty with self-completion, while the remaining two-thirds had no difficulty.

MEASURE	PSYCHOMETRIC PROPERTIES	
Collaborative Goal Index (CGI)	**Validity**	In a study with 71 consumers with a psychotic disorder, Clark et al (2009)[830] demonstrated that those who achieved greater progress toward their goals also showed greater increases in self-confidence and confidence that they could achieve future goals, and reductions in aggression, self-harm and drug and alcohol use, lending some support for the construct validity for the CGI (although this was not the purpose of the study). Similarly, poorer goal attainment was related to greater symptom distress (K-10).
	Reliability	No psychometric research was identified that examined the reliability of the CGI.
	Sensitivity to change	No psychometric research was identified that examined the sensitivity to change of the CGI.
	Acceptability/Usability	No research was identified that examined the acceptability or usability of the CGI.
HoNOS	**Validity**	Several attempts have been made to explore the content validity of the HoNOS. Shergill et al,[842] Orrell et al[856] and McClelland et al[857] asked consumer/carer advocacy groups and mental health professionals to comment on whether the HoNOS items reflected areas of concern for them. In the main, respondents in these studies were positive, suggesting that the HoNOS was appropriate, well designed and thorough, and highlights consumers' problems quickly, indicating changes in their mental health status over time.[842, 856, 857] However, there were some reservations about specific items. Respondents were concerned about the restriction imposed by the rater being forced to indicate only one problem in Item 8 (Other mental and behavioural problems),[842, 857] and expressed concerns about the ability of Item 6 (Problems associated with hallucinations and delusions) to give an accurate picture of the symptoms and role performance of a person with schizophrenia[856]. They also felt that the social items (Items 10, 11 and 12) were problematic because the information needed to rate them is complex and/or not always available.[856, 857] In the main, the HoNOS has been shown to perform well against clinician-rated instruments such as the Role Functioning Scale (RFS)[858], the Brief Psychiatric Rating Scale (BPRS)[842, 856-858], the GAF scales for symptoms (GAF-s) and disability (GAF-d) from the Global Assessment Scale (GAS)[551, 842, 856, 857, 859, 860], the Life Skills Profile (LSP)[551], the Manchester Audit Tool (MAT)[861], the Behaviour Rating Scale from the Clifton Assessment Procedures for the Elderly (CAPE-BRS)[842], the Clinical Dementia Rating[842], the Mini-Mental State Examination (MMSE)[842], Schedules for Clinical Assessment in Neuropsychiatry (SCAN)[860, 862], the Broad Rating Schedule (BRS)[860], the Disability Assessment Schedule (DAS)[860], the Scale for Assessment of Negative Symptoms (SANS)[860], Location of Community Support Scale (LOCSS)[856], the Social Behaviour Scale (SBS)[856, 862], the Hamilton Rating Scale for Depression (Ham-D)[863], and the Positive and Negative Symptoms Scale (PANSS)[863].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	In studies of the internal consistency of the HoNOS, Cronbach's alpha has ranged from .59 to .76, indicating that the HoNOS has a moderately high level of internal consistency and low levels of item redundancy, supporting its use as a meaningful summary of severity of symptoms.[191, 842, 856-858, 864, 865] Having said this, Trauer has argued that the HoNOS should not be regarded as unidimensional, measuring a single, underlying construct of mental health status. Instead, it should be viewed in the context of its original intention, namely to provide a broad coverage of the problems typically experienced by consumers of mental health services.[864, 866] Comparatively few studies have examined the test-retest reliability of the HoNOS, but those that have generally report fair to moderate overall reliability scores.[842, 856, 867] Particularly low reliability scores have been reported for Item 1 (Overactive, aggressive, disruptive or agitated behaviour), Item 3 (Problem drinking or drug taking), Item 7 (Problems with depressed mood), and Item 10 (Problems with activities of daily living).
	Sensitivity to change	Most studies of the inter-rater reliability of the HoNOS have found that the overall agreement between pairs of raters is fair to moderate,[842, 862, 867] or even moderate to good,[856, 858, 860, 863] but that agreement is poor on particular items. Items identified as problematic include Item 4 (Cognitive problems)[862], Item 7 (Problems with depressed mood)[862], Item 8 (Other mental and behavioural problems)[858, 862], Item 9 (Problems with relationships)[856], Item 11 (Problems with living conditions)[856, 868] and Item 12 (Problems with occupation and activities).
	Acceptability/Usability	There has been considerable debate about the feasibility and utility of the HoNOS. The most positive authors have suggested although no instrument will fulfil all needs, the HoNOS is a comprehensive, user-friendly tool that is likely to have utility in routine outcome measurement (possibly with minor modifications).[857, 858, 863, 865, 869-871] Less enthusiastic published commentaries have argued that while it is acceptable to clinicians and feasible to administer during routine outcome measurement, it is of limited value in informing care planning, restricting its application to being a pure research instrument.[859, 872-875]

MEASURE	PSYCHOMETRIC PROPERTIES	
HoNOSCA	**Validity**	Several studies that have examined the correlation between the total HoNOSCA score and scores on other clinician-rated measures and have typically reported moderate correlations (r = .6 or above). This was the case in a study by Yates et al,[876] which compared the HoNOSCA with the Children's Global Assessment Scale (CGAS) and Paddington Complexity Scale (PCS), and in a study by Bilenberg where the 'gold standard' was the Global Assessment of Psychosocial Disability[877]. Studies that have evaluated the HoNOSCA against parent- and child/adolescent-rated instruments have typically produced lower correlations. Yates et al[876] found significant but at best only modest correlations between the HoNOSCA and the Behaviour Check List (BCL), the Strengths and Difficulties Questionnaire (SDQ), the Child Health Related Quality of Life Questionnaires (CHRQOL) and the Modified Harter Self-Esteem Questionnaire (Harter). Gowers et al found overall low levels of agreement between the HoNOSCA and the HoNOSCA-SR (a consumer-rated version of the instrument applicable to adolescents only) at an individual level, although some groups (e.g., outpatients with eating disorders) provided an exception to this rule.[878]
	Reliability	There are few published studies on the test-retest reliability of the HoNOSCA, and those which do exist are arguably studies of the sensitivity to change (or lack of change) of the instrument (see below), since they cover considerable time periods and consider stability in relation to other measures. Garralda et al [879] examined the test-retest reliability of the instrument over a six-month period, for consumers for whom clinicians indicated there had been no change on a global rating scale, and reported a figure of .69. Similarly, Brann [880] reported correlations of .80 over three months and .76 over five months when he examined the instrument's test-retest reliability, again in a group of consumers who were judged not to have changed over the given period. Several studies have considered the inter-rater reliability of the HoNOSCA, generally with reasonably positive results. These studies have consistently found that the majority of items in Section A demonstrate good or very good reliability, although there is less agreement about which items perform poorly. For example, Brann et al [881] reported a particularly low intra-class correlation (.06) for Item 10 (Problems with peer relationships), but Gowers et al [839, 882] found that this item, while not performing as well as some others, achieved an intra-class correlation of .77. There is also debate about the inter-rater reliability of Section B. In the original field work associated with the development of the HoNOSCA, Gowers et al [839, 882] found that the two Items comprising this section each had good inter-rater reliability: Item 14 (Problems with knowledge or understanding about the nature of the child or adolescent's difficulties) had an intra-class correlation of .73 and Item 15 (problems with lack of information about services or management of the child or adolescent's difficulties) one of .78. By contrast, the equivalent figures in a later study by Garralda et al [883] were .27 and .03, respectively.

MEASURE	PSYCHOMETRIC PROPERTIES	
	Sensitivity to change	Three different approaches have been taken to assessing the ability of the HoNOSCA to detect change, with individual studies often using a combination of these approaches. The first and methodologically weakest approach involves simply determining whether HoNOSCA scores change over time, with no reference to whether this reflects real change. In the original field work associated with the development of the HoNOS, for example, Gowers et al noted that 'the HoNOSCA demonstrated satisfactory sensitivity to change, with a mean overall reduction in total scores of 38% between rating points, on average nearly three months apart'[839, 882]. Manderson and McCune[884] made a similar observation in their study. These and other studies have considered factors that are predictive of particular patterns of change on the HoNOSCA, and have identified diagnosis and related symptomatology[877, 879, 884], initial case severity[879] and setting [878]. The second approach examines the correspondence between change as assessed by the HoNOSCA and change as defined by the difference between scores on other measures. Collectively, studies by Gowers et al[878], Garralda et al[879] and Bilenberg[877] have reported changes in HoNOSCA scores that are comparable in direction and magnitude with other clinician-rated measures, such as the CGAS and the Global Assessment of Psychological Disability (GAPD), and, to a lesser extent with parent- and/or consumer-rated measures such as the HoNOSCA-SR, the BCL and the SDQ.
	Acceptability/Usability	In considering the feasibility and utility of the HoNOSCA in routine outcome measurement, several studies have explicitly sought the views of clinicians.[839, 877, 882-884] Clinicians in these studies have been positive about the HoNOSCA's brevity and ease of use, its clinical utility, and its ability to be incorporated into routine practice (given adequate time and resources). The main concerns raised by clinicians about the instrument have been its applicability to children aged under five, its emphasis on child/adolescent symptoms and functioning, and its failure to take into account context.
HONOS 65+	**Validity**	Separate studies by Burns et al[844], Mozley et al[885], Spear et al[886], and Bagley et al[887] have examined the concurrent validity of the HoNOS65+. Most commonly, these studies have examined the correlations between the HoNOS65+ and more established clinician-rated measures that assess similar domains. Reasonable correlations have been observed between the HoNOS65+ total score and the Mini-Mental State Examination (MMSE)[844, 885, 886], the Crighton Royal Behaviour Rating Scale (CRBRS)[844] and the Barthel Activities of Daily Living Index.[844].

MEASURE	PSYCHOMETRIC PROPERTIES	
	Reliability	As yet, there are no published assessments of the test-retest reliability of the HoNOS65+. By contrast, several studies have examined the inter-rater reliability of the HoNOS65+. Burns et al[844] and Spear et al[886] both found inter-rater reliability to be good to very good for most items, whereas Allen et al were less positive[888]. In Burns et al's study, only Item 2 (Non-accidental self injury), Item 10 (Problems with activities of daily living), Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities) did not consistently perform well. In Spear et al's study, Item 4 (Cognitive problems), Item 5 (Physical illness or disability problems) and Item 9 (Problems with relationships) demonstrated only poor to moderate inter-rater reliability. In Allen et al's study, there were problems with a broader range of items, largely related to difficulties in interpretation of the glossary.
	Sensitivity to change	Only one study has considered the sensitivity to change of the HoNOS65+. Spear et al[886] found that consumers showed improvement on all HoNOS65+ subscales and on the HoNOS65+ total score between assessment and discharge from inpatient and community services. More particularly, they found that the discharge HoNOS65+ score and the change in HoNOS65+ scores showed moderate but significant correlations with the Clinician's Interview Based Impression of Change Scale (CIBIC+).
	Acceptability/Usability	Several studies have examined different aspects of the feasibility and utility of the HoNOS65+, including its acceptability to clinicians. In the original pilot of the HoNOS65+, Burns et al [844] asked raters whether or not they would find the instrument helpful in working with individual consumers; 39% indicated it would be very useful and 50% that it would be of some use. Spear et al[886] reported similar findings. In both studies, almost all respondents reported that it was easy to administer. Feasibility and utility have also been considered in terms of uptake, both at a national level and at a service level. Reilly et al[889] conducted a survey of old age psychiatrists across the United Kingdom, and found that 18% reported that the HoNOS65+ was being used in their service. Spear et al examined the proportion of episodes of care at which the HoNOS65+ was administered within a single service, and found completion rates of 96%.

MEASURE	**PSYCHOMETRIC PROPERTIES**	
Outcome Rating Scales (ORS) and Session Rating Scales (SRS)	**Validity**	Miller at al (2003)[890] conducted a US-based study of the psychometric properties of the ORS with a non-clinical group of 86 master's students, therapists and staff working at a community family agency and a clinical group of 435 clients attending the same agency. As well as the ORS, participants also completed the Outcome Questionnaire 45.2 (OQ-45.2), on which the ORS is based. The OQ-45.2 is a 45-item measure of client functioning with established internal consistency and test-retest-reliability and that has high validity coefficients with other established measures of anxiety and depression. As expected, the difference on ORS scores between the clinical and non-clinical group were large and significant (non-clinical mean = 28.0, SD = 6.8; clinical mean = 19.6, SD = 8.7, t-test P < .00001), providing some evidence for the discriminant validity of the ORS. Pearson product moment correlations between the ORS scales and the OQ-45.2 subscales ranged from .42 to .59, for the ORS total and OQ-45.2 subscales correlations ranged from .41 to .57, and was .59 for the total. These correlations give some moderate strength evidence of the concurrent validity of the ORS. Miller et al's study [890] was replicated by Bringhurst et al (2006)[891] with a non-clinical sample comprising 98 social work students. The mean ORS score of 29.9 (SD = 7.5) was similar to that reported in Miller et al's study. Correlations with the OQ were somewhat higher than in the earlier study, and once again suggest concurrent validity. An Australian study by Campbell et al (2009) involving a sample of 75 consumers referred for primary mental health care services again showed significant moderate to strong correlations between the ORS and the OQ-45 sub-scale and totals.[846] The ORS subscales and total were also significantly negatively correlated with the depression, anxiety and stress subscales of the DASS-21[892] (significant correlations ranged from -.34 to -.76), and positively significantly correlated with scores on the QOLS[893] (r = .49–.74) the General Perceived Self-efficacy Scale (r = .36–.53) and the Rosenberg self-esteem scale (*r* = .46-.67), also providing some evidence for concurrent validity of the ORS. Campbell et al's study also examined the validity of the SRS and found that the SRS had significant moderate correlations with the Working Alliance Inventory[894] (r = .37-.63)[846].
	Reliability	The non-clinical group in Miller et al's study repeated the ORS and OQ-45.2 four times over periods of one day to two weeks.[890] Internal consistency for this group was α = .87 at first administration and .96 at the last, and .93 (N = 336) for all administrations, suggesting an excellent internal consistency in a non-clinical sample.[890] According to the authors, this internal consistency is comparable to the internal consistency for the much longer OQ-45. Test-retest reliability across the three re-tests in the non-clinical group ranged from .49 (n = 77) to .66 (n = 86), which is moderate to good. In Bringhurst's study involving 98 social work students, the overall internal consistency was .93, again suggesting excellent internal consistency in a non-clinical sample.[891] Test-retest reliability for three weeks was .81 (n = 55, P < 0.001) (good). In the study of Campbell et al[846], internal consistency of the ORS was α = .90 (excellent). Inter-item correlations between the four ORS items ranged from r = .58 to .97, suggesting that the items are measuring a single underlying construct. Campbell et al's study also tested the reliability of the SRS and found the internal consistency to be excellent (α = .93) and the inter-item correlations to be consistently strong (r = .74–.86, P = .01).

MEASURE	PSYCHOMETRIC PROPERTIES	
	Sensitivity to change	A t-test for change between pre-and post- intervention scores on the ORS for the clinical group in Miller et al's study[890] showed a significant difference (pre-test Mean = 19.6, SD = 8.7; post-test mean = 25.7, SD = 8.7; P <.00001). A t-test for differences in scores for the non-clinical sample from first to final administration showed no difference (P > .1).[890] These outcomes support the ORS's sensitivity to change.
	Acceptability/Usability	The ORS completion rate over one year for a clinical sample was 89% at the 12-month time-point. By comparison, the rate of completion for the OQ-45.2 in a similar, but unrelated, sample was shown to be 25%.[890] The brevity and ease of completion of the ORS might lead to greater levels of completion among clinical samples. A Norwegian study by Sundet[895] examined feedback on the use of the SRS and ORS from four therapists working in a child and adolescent mental health service. Therapists' feedback supported the feasibility of the two scales (seen as 'useful', 'satisfactory') and the therapists expressed few complaints. Therapists saw use of the tools as a way to open conversation on process and outcome and of fostering collaboration with families. However, there were some concerns that use of a scale might disrupt therapeutic work, directing attention to non-therapeutic topics.

Chapter 4: Conclusion

A literature review was undertaken that aimed to identify the psychometric properties of the measures being used in the sector along with measures that may be suitable for use. The psychometric properties reviewed included validity, reliability and sensitivity to change.

Initially, the literature was reviewed to determine what outcomes had been measured. This resulted in the identification of seven outcome domains. These included recovery, cognition and emotion, functioning (activities of daily living and interpersonal relationships), social inclusion, quality of life, and experience of service provision. These domains were used as key words to guide the search of the literature. While these proved a useful way of organising the search of the literature, it became clear early on that some measures, in fact, covered multiple domains. Therefore, for reporting purposes, a multidimensional domain was included. A description of each of these domains is in table 19.

One of the challenges of the literature review was the identification of measures that may be suitable for use in the CMO sector. Consequently, “Community Managed Organisations”, “Non-government organisations”, and “not for profit” were used as search terms along with the outcome domains identified above. However, the inclusion of these terms quickly limited the search results. Therefore the primary search terms were the outcome domains. However, when measures were identified, the literature often lacked a detailed description of the types of services or service settings within which measures were being tested. As a result, the literature review also relied on grey literature to identify measures that have been used or suggested for use in the sector. Finally, the results of the survey were also used to identify measures that were included in the literature review.

Given these challenges, 136 measures were identified through this process. It is important to note that not every CMO responded to the survey, so there may be other measures being used in the sector that are not included in this review. It is also important to note that there are a large number of mental health measures that have been created. The measures included in this review are only a subset of all possible measures.

4.1 Outcome Domains

Table 19. Outcome measurement domains

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
The personal process of individual recovery.	Individual consumer cognitive performance and emotional experience Individual carer cognitive performance and emotional experience	Simple and complex functional abilities are covered here including the ability to undertake activities of daily living consistent with developmental stage. The quantity and quality of interpersonal relationships consistent with developmental stage.	Education, employment, citizenship, stability of housing	General life satisfaction, physical health and wellbeing	Service satisfaction, consumer or carer experience of service provision Care or service co-ordination	Measures that capture information across multiple domains

4.2 Identified Measures

136 measures were identified as a result of the literature review process. These measures are listed in table 20 which follows.

Table 20. All identified measures

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
1. Recovery Assessment Scale (RAS) 2. Agreement with Recovery Attitudes Scale (ARAS) 3. Rochester Recovery Inquiry (RRI) 4. Consumer Recovery Outcomes System (CROS) 5. Crisis Hostel Healing Scale (CHHS) 6. Personal Vision of Recovery Questionnaire (PVRQ) 7. Recovery Interview (RI) 8. Recovery Attitudes Questionnaire (RAQ-16; RAQ-7) 9. Mental Health Recovery Measure (MHRM) 10. Reciprocal Support Scale (RSS) 11. Relationships and Activities that Facilitate Recovery Survey (RAFRS) 12. illness Management and Recovery (IMR) Scales 13. Ohio Mental Health Consumer Outcomes System (OMHCOS) 14. Peer Outcomes Protocol (POP) Recovery Measurement Tool (RMT) 15. Recovery Orientation (RO)a 16. Recovery Process Inventory (RPI) 17. Milestones of Recovery Scale (MORS) 18. Multi-Phase Recovery Scale (MPRM)a 19. Maryland Assessment of Recovery in People with Serious Mental Illness (MARS) 20. Mental Health Recovery Star (MHRS) 21. Questionnaire about the Process of Recovery (QPR) 22. Subjective Experiences of Psychosis Scale (SEPS) 23. Self-Identified Stage of Recovery (SISR) 24. Stages of Recovery Instrument (STORI) 25. Stages of Recovery Scale (SORS)	26. Kessler-10 (K-10) 27. Mental Health Inventory 38 (MHI-38) 28. Behaviour Symptom Identification Scales (BASIS-32®) 29. Depression, Anxiety, Stress Scale (DASS) 30. Beck Depression Inventory (BDI) 31. Beck Anxiety Inventory 32. Hamilton Depression Rating Scale 33. Hamilton Anxiety Rating Scale 34. State Trait Anxiety Inventory 35. Edinburgh Postnatal Depression Scale 36. Geriatric Depression Scale 37. Symptom Checklist 90 (SCL 90) 38. General Health Questionnaire-12 (GHQ-12) 39. Sphere-12 40. My Mood Monitor (M-3) 41. Suicidal Ideation Questionnaire (SIQ) 42. Columbia–Suicide Severity Rating Scale (C-SSRS) 43. Inventory of Complicated Grief-Revised 44. PTSD Checklist (PCL-C) 45. Dissociative Experiences Scale 46. Strengths and Difficulties Questionnaire (SDQ) 47. Child Behaviour Checklist (CBCL) 48. Devereux Early Childhood Assessment (DECA) 49. Anxiety Disorders Interview Schedules (Children) 50. Parenting Stress Index 51. Social Responsiveness Scale – SRS 52. Barriers to Adolescents Seeking Help Scale 53. Mini Mental State Examination 54. Cambridge Cognitive Examination CAMCOG 55. Rowland Universal Dementia Assessment Scale (RUDAS) 56. Warwick-Edinburgh Mental Well-being Scale (WEMWBS) 57. Working Alliance Inventory 58. Empowerment Scale 59. Brief COPE 60. Revised NEO Personality Inventory (NEO-PI-R) 61. General Health Seeking Questionnaire 62. Involvement Evaluation Questionnaire (IEQ) 63. Burden Assessment Scale (BAS) 64. Zarit Burden Scale 65. CarerQol-7D+VAS	66. Life Skills Profile] 67. Social Function Scale 68. Work and Social Adjustment Scale 69. Personal and Social Performance Scale (PSP) 70. Global Assessment of Functioning (GAF) 71. Social Functioning Questionnaire (SFQ) 72. The World Health Organisation Disability Assessment Schedule 2.0 (WHODAS II/2.0) 73. The Multnomah Community Ability Scale (MCAS) 74. Canadian Occupational Performance Measure (COPM) 75. Children's Global Assessment Scale (CGAS) 76. Parents' Evaluation of Developmental Status (PEDS) 77. Drug Use Disorders Identification Test (DUIT) 78. Alcohol Use Disorders Identification Test (AUDIT) 79. Specific Levels of Functioning Scale (SLOF) 80. Independent Living Skills Survey (ILSS) 81. Social Behaviour Schedule (SBS) 82. Child and Adolescent Functional Assessment Scale (CAFAS) 83. Columbia Impairment Scale	84. Social and Community Opportunities Profile (SCOPE) 85. Social Inclusion Questionnaire (SIQ) 86. Activity and Participation Questionnaire (APQ) 87. Staff Survey of Social Inclusion (SSSI) 88. EMILIA Project Questionnaire (EPQ) 89. Social Inclusion measure (SIM) 90. The Inclusion Web (IW) 91. Composite Measure of Social Inclusion (CMSI) 92. Australian Community Participation Questionnaire (ACPQ) 93. Evaluating Social Inclusion Questionnaire (ESIQ) 94. Living in the Community Questionnaire	95. Satisfaction with Life Scale (SWL) 96. World Health Organisation Quality of Life –Brief, Australian Version (Australian WHOQOL-BREF) 97. Purpose In Life (PIL) 98. Californian Quality of Life Survey (Cal-QOL) 99. Quality of Life Questionnaire (QLQ) 100. Subjective Quality of Life Profile 101. SF-36 (or Medical Outcomes Study (MOS) 36-Item Short Form Health Survey) 102. Quality of Life Enjoyment and Satisfaction Questionnaire-short form (Q-LES-Q-SF) 103. Manchester Short Assessment of Quality of Life (MANSA) 104. Quality of Life Interview (QOLI) 105. Lancashire Quality of Life Profile (LQLP) 106. Quality of Life Index (QLI) 107. Satisfaction with Life Domains Scale (SLDS) 108. Quality of Life Scale (QLS) 109. Wisconsin Quality of Life Index – Canadian Version (CaW-QLI) 110. Brief Life Satisfaction Scale (BLSS) 111. Satisfaction with Life Scale (SWLS) 112. Personal Wellbeing Index 113. Personal Outcomes Measures	114. Mental Health – Consumer Perceptions and Experiences of Services (MH-CoPES) 115. National Research Corporation (NRC-Picker)/ National Health Service 116. Consumer and Carer Experience Questionnaires (C&CES) 117. Psychiatric Outpatient Experience Questionnaire (POPEQ) 118. The Mental Health Statistics Improvement Program Consumer Survey (MHSIP-Consumer) 119. Consumer Experience of Care Questionnaire (Australia) 120. Carer Experience of Service Provision Questionnaire (Australia) 121. Carer satisfaction interview 122. Questionnaire for carers of people who use social services (Wales) (QPSS-Wales) 123. Carer Participation Survey (CPS) 124. Consumer and Carer Experience Questionnaires (C&CES) 125. Carer Well-being and Support Questionnaire (CWS) 126. Disability and mental health service users and carers satisfaction survey 2009 127. The Personal Social Services Survey of Adult Carers in England 2009-2010 (PSSS-Adult Carers) 128. Mental Health Carers' Survey (MHCS) 129. Network analysis 130. Human services integration measure	131. Camberwell Assessment of Need (CAN) 132. Collaborative Goal Index/COMPASS 133. Health of the Nation Outcomes Scales (HoNOS) 134. Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA) 135. Health of the Nation Outcomes Scales 65+ (HoNOS 65+) 136. Outcome Rating Scales (ORS) and Session Rating Scales (SRS)

4.3 Selection Criteria

To support instrument selection, a number of criteria were established. These review criteria specified that the attributes for an outcome measure appropriate for mental health CMOs in Australia should:

- have been developed for use or used in the mental health sector;
- have been developed or used in Australia, with identified potential for further development;
- be able to be completed by either the consumer and/or CMO employee;
- be brief and easy to use (time and/or number items);
- yield quantitative data (does not exclude instruments that also yield qualitative data);
- have undergone scientific scrutiny and have demonstrated strong psychometric properties (e.g., of internal consistency, validity, reliability and sensitivity to change).

Using these selection criteria, a short list of measures was identified as suitable for use within the CMO sector (table 21). At this stage, these measures are only suggested. More detailed work is necessary to further establish the suitability of these measures.

4.4 Short List of Measures by Domain

Using the selection criteria above a short list of measures were identified across the seven outcome domains, see table 21.

Table 21. Short list of measures by domain

Recovery	Cognition and Emotion	Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
1. Recovery Assessment Scale (RAS)† 2. Recovery Process Inventory (RPI)† 3. Illness Management and Recovery (IMR)† Scales 4. Stages of Recovery Instrument (STORI)† 5. Recovery Star†	6. Kessler-10 (K-10)† 7. Mental Health Inventory 38 (MHI-38)† 8. Behaviour Symptom Identification Scales (BASIS-32®)† 9. Strengths and Difficulties Questionnaire (SDQ)† 10. Involvement Evaluation Questionnaire (IEQ)† 11. Burden Assessment Scale (BAS)† 12. CarerQol-7D+VAS†	13. Life Skills Profile* 14. Work and Social Adjustment Scale* 15. The Multnomah Community Ability Scale (MCAS)* 16. Personal and Social Performance Scale (PSP)*	17. Social and Community Opportunities Profile (SCOPE)† 18. Activity and Participation Questionnaire (APQ6)† 19. Living in the Community Questionnaire†	20. Satisfaction with Life Scale (SWL)† 21. Manchester Short Assessment of Quality of Life (MANSA)† 22. World Health Organisation Quality of Life – Brief, Australian Version (Australian WHOQOL- BREF)†	23. Consumer and Carer Experience Questionnaires (C&CES)† 24. Psychiatric Outpatient Experience Questionnaire (POPEQ)† 25. Consumers Experience of Care† 26. Carers Experience of Service Provision†	27. Camberwell Assessment of Need – Short Appraisal Scale (CANSAS) †* 28. Collaborative Goal Index/COMPASS† 29. Health of the Nation Outcomes Scales (HoNOS)* 30. Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA)* 31. Health of the Nation Outcomes Scales 65+ (HoNOS 65+)*

† = client-rated

* = worker-rated

4.5 Short List of Measures by Service Type

The short listed measures were reviewed by the project team with the aim of identifying those measures that may be suitable for use across the different CMO service types. In table 22, the numbers correspond to the measure in table 21. Green indicates that the measure(s) and domain are suitable for the service type described in the row. Orange indicates that the measure(s) and domain may be suitable for this service type. The red indicates that the measure(s) or domains may not be suitable for this service type. The identification of the suitability of measure(s) and domains was a desk top exercise undertaken by the project team. Broader consultation and testing of these measures in the different CMO service settings will be required.

Table 22. Short list of measures by service type

Service Type	Recovery	Cognition and Emotion		Functioning	Social Inclusion	Quality of Life	Experience of Service	Multidimensional
Counselling—face-to-face	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Counselling, support, information & referral—telephone	1-5?[1]	6-9		X	X	20-22	23-26	28
Counselling, support, information & referral—online	1-5?	6		X	X	20-22	23-26	28
Self-help—online	1-5?	6		X	X	20-22	23-26	28
Group support activities	1-5?	6-9		X	17-19?	20-22	23-26	28
Mutual support and self-help	1-5?	6-9?		X	17-19?	20-22	23-26	28
Staffed residential services	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Personalised support—linked to housing	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Personalised support—other	1-5	6-9	10-12?	13-16	17-19	20-22	23-26	27-31
Family & carer support	X	6-12		X	X	20-22	23-25	X
Individual advocacy	X	X		X	X	X	X	X
Care co-ordination	1-5?	6-9?	10-12?	13-16?	17-19?	20-22?	23-26	27-31
Service integration infrastructure	X	X		X	X	X	X	X
Education, employment & training	1-5?	6-9?	10-12?	13-16?	17-19	20-22?	23-26	X
Sector development and representation	X	X		X	X	X	X	X
Mental health promotion	X	X		X	X	X	X	X
Mental illness prevention	X	X		X	X	X	X	X

Factors that may influence measure selection include program design, age, Aboriginal and Torres Strait Islander status, Cultural and Linguistic Diversity, and intellectual disability.

---

[1] Question mark indicates that for some programs these measures may be suitable but this is yet to be determined.

References

1. Commonwealth of Australia, *Fourth national mental health plan – An agenda for collaborative government action in mental health 2009–2014*. 2009, Commonwealth of Australia: Canberra.
2. Productivity Commission, *Annex J: Not-for-profit sector feedback: government funded services, Contribution of the Not-for-Profit Sector Research Report*. 2010, Canberra.
3. Australian Institute of Health and Welfare, *Mental Health Non-Government Organisation Data Development Project, Version 1*. 2011, Australian Institute of Health and Welfare: Canberra.
4. Australian Mental Health Outcomes and Classification Network, *Review of Recovery Measures, Version 1.01*. 2010, Australian Mental Health Outcomes and Classification Network: Sydney.
5. Counselling), B.B.A.f., *Counselling-Definition of terms in use with expansion and rationale (Information Sheet 1)* in *Rugby*, B.A.f. Counselling., Editor. 1986.
6. Division, M.H.a.A.C.S., *Standards for psychiatric disability rehabilitation and support services*. 2003, Department of Human Services: Victoria.
7. Solomon, P., *Peer support/ peer provided services underlying processes, benefits and critical ingredients.* Psychiatric Rehabilitation Journal, 2004. **27**(4): p. 392-401.
8. Communities, D.o. *Supporting recovery: Mental health community services plan 2011-2017*. 2011 3 April 2013]; Available from: http://www.communities.qld.gov.au/resources/disability/support-services/our-services/mental-health/documents/mental-health-community-services-plan.pdf.
9. 2012., M.A. *Family and Carer Mental Health Program (NSW)*. 2012 3 April 2013]; Available from: http://www.missionaustralia.com.au/community-services/3114-family-carer-mental-health-nsw.
10. Australia, C.o. *National Mental Health Reform—Ministerial Statement*. 2012 3 April 2013]; Available from: http://www.health.gov.au/internet/budget/publishing.nsf/Content/849AC423397634B7CA25789C001FE0AA/$File/DHA%20Ministerial.PDF.
11. FaHCSIA (Department of Families, H., Community Services and Indigenous Affairs) *National Disability Advocacy Program*. 2011 3 April 2013]; Available from: http://www.fahcsia.gov.au/sa/disability/progserv/providers/NationalDisabilityAdvocacyProgram/Pages/TypesDisabilityAdvocacy.aspx#3.
12. COAG., *National action plan on mental health 2006-2011*. 2006, Council of Australian Governments.: Canberra.
13. Green G, et al., *A mental health clients perspective to stigmatisation.* Journal of Mental Health 2003. **12**: p. 223-234.
14. Australian Health Ministers, *Fourth National Mental Health Plan*. 2009, Commonwealth Department of Health and Family Services: Canberra.
15. Andrews, G., L. Peters, and M. Teesson, *Measurement of Consumer Outcome in Mental Health: A Report to the National Mental Health Information Strategy Committee*. 1994, Clinical Research Unit for Anxiety Disorders: Sydney.
16. Bickman, L., Nurcombe, B., Townsend, C., Belle, M., Schut, J., and Karver, M., *Consumer Measurement Systems in Child and Adolescent Mental Health*. 1998, Canberra, ACT: Department of Health and Family Services.

17. Slade, M., *What outcomes to measure in routine mental health services, and how to assess them: a systematic review.* Australian and New Zealand Journal of Psychiatry 2002. **36**(6): p. 743-53.
18. Cella, D., et al., *The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years.* Medical Care, 2007. **45**(5 Suppl 1): p. S3-S11.
19. National Institute for Mental Health in England, B.a.T.L.S.o.M.a.D., *Outcomes Compendium: Helping You Select the Right Tools for Best Mental Health Care Practice in Your Field*. 2008, Department of Health: London.
20. Hampson, M., et al., *Occasional Paper OP78, Outcome measures recommended for use in adult psychiatry*. 2011, Royal College of Psychiatrists: London.
21. Age, F.o.t.P.o.O., *Occasional Paper OP86, Individual patient outcome measures recommended for use in older people's mental health*. 2012, Royal College of Psychiatrists London.
22. Penrose-Wall, J., et al., *Mapping the difference we make. Non-government organisations use of rotuine consumer outcome evaluation in providing mental health care in NSW*. 2006, Mental Health Coordinating Council of NSW: Sydney.
23. Community Mental Health Australia, *Taking our place-Community Mental Health Australia: Working together to improve mental health in the community,*. 2012, Community Mental Health Australia: Sydney.
24. Slade, M., *Routine outcome assessment in mental health services.* Psychol Med, 2002. **32**(8): p. 1339-43.
25. Snyder, C., et al., *Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations.* Quality of Life Research, 2012. **21**(8): p. 1305-1314.
26. Crawford, M.J., et al., *Selecting outcome measures in mental health: the views of service users.* Journal of Mental Health, 2011. **20**(4): p. 336-346.
27. Greenhalgh, J., et al., *Reviewing and selecting outcome measures for use in routine practice.* Journal of Evaluation in Clinical Practice, 1998. **4**(4): p. 339-350.
28. Cohen, J.A., *A coefficent of agreement for nominal scales.* Education and Psychological Measurement, 1960. **20**(1): p. 37-46.
29. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**: p. 159-174.
30. Streiner, D.L., *Starting at the beginning: An introduction to coefficient alpha and internal consistency.* Journal of Personality Assessment, 2003. **80**(1): p. 99-103.
31. Ralph, R.O., K. Kidder, and D. Phillips, *Can We Measure Recovery? A Compendium of Recovery and Recovery-related Instruments*. 2000, Human Services Research Institute: Cambridge, MA.
32. Farkas, M., *Personal communication*. 2009.
33. Campbell-Orde T, et al., *Measuring the Promise: A Compendium of Recovery Measures*. Vol. II. 2005, Cambridge, Massachusetts: The Evaluation Center @ HSRI.
34. Giffort, D., et al., *Recovery Assessment Scale*. 1995, Illinois Department of Mental Health: Chicago, IL.
35. Corrigan, P.W., et al., *Recovery as a psychological construct.* Community Mental Health Journal, 1999. **35**(3): p. 231-239.
36. Chiba, R., Y. Miyamoto, and N. Kawakami, *Reliability and validity of the Japanese version of the Recovery Assessment Scale (RAS) for people with chronic mental illness: Scale development.* International Journal of Nursing Studies, 2010. **47**(3): p. 314-322.
37. Murnen, S.K. and L. Smolak, *Agreement with Recovery Attitudes Scale*. 1996.
38. Hopper, K., et al., *Rochester Recovery Inquiry*. 1996, Center for the Study of Public Issues in Mental Health: Orangeburg, NY.

39. Bloom, B.L. and A. Miller, *The Consumer Recovery Outcomes System (CROS 3.0): Assessing Clinical Status and Progress in Persons with Severe and Persistent Mental Illness*. 2004, CROS, LLC / Colorado Health Networks: Colorado Springs, CO.
40. Ensfield, L.B., *The Personal Vision of Recovery Questionnaire: The Development of a Consumer-derived Scale*. 1998, University of Cincinnati: Cincinnati, OH.
41. Ensfield, L.B., et al., *Personal Vision of Recovery Questionnaire*. 1998, University of Cincinnati: Cincinnati, OH.
42. Heil, J. and L.K. Johnson, *Recovery Interview*. 1998, Institute for Local Government Administration and Rural Development: Athens, OH.
43. Borkin, J.R., *Recovery Attitudes Questionnaire: Development and Evaluation.* Psychiatric rehabilitation journal, 2000. **24**(2): p. 95-103.
44. Young, S.L. and W.A. Bullock, *The Mental Health Recovery Measure*. 2003, University of Toledo: Toledo, OH.
45. Campbell-Orde, T., et al., *Measuring the Promise: A Compendium of Recovery Measures*. 2000, Human Services Research Institute: Cambridge, MA.
46. Silver, T., et al., *Impact of teaching mental health best practices and recovery processes on constituent populations of the mental health system*, in *New Research in Mental Health*, D. Roth, Editor. 2002, Ohio Department of Mental Health: Columbus, OH.
47. Leavy, R.L., et al., *Predictors of subjective quality of life in mental health consumers: Baseline results*, in *New Research in Mental Health*, D. Roth, Editor. 2002, Ohio Department of Mental Health: Columbus, OH.
48. Mueser, K.T., et al., *The Illness Management and Recovery (IMR) Scales (Client and Clinician Versions)*. 2004, New Hampshire-Dartmouth Psychiatric Research Center: Concord, NH.
49. Ohio Department of Mental Health, *Procedural Manual (6th Edition Revised). The Ohio Mental Health Consumer Outcomes System, 2004*. 2004, Ohio Department of Mental Health.
50. Campbell, J., et al., *Peer Outcomes Protocol Questionnaire*. 2004, University of Illinois: Chicago, IL.
51. Campbell, J., et al., *Peer Outcomes Protocol (POP): Administration Manual*. 2004, University of Illinois: Chicago, IL.
52. Campbell, J., et al., *Peer Outcomes Protocol (POP): Psychometric Properties of the POP*. 2004, University of Illinois: Chicago, IL.
53. Ralph, R.O., *At the individual level: A personal measure of recovery*, in *NASMHPD/NTAC e-Report on Recovery (Fall 2004)*. 2004.
54. Resnick, S.G., et al., *An empirical conceptualization of the recovery orientation.* Schizophrenia Research, 2005. **75**(1): p. 119-128.
55. Jerrell, J.M., V.C. Cousins, and K.M. Roberts, *Psychometrics of the Recovery Process Inventory.* Journal of behavioral health services and research, 2006. **33**(4): p. 464-473.
56. Cuijpers, P. and H. Stam, *Burnout among relatives of psychiatric patients attending psychoeducational support groups. .* Psychiatric Services., 2000. **51**(3): p. 375-379.
57. Beeble, M.L. and D.A. Salem, *Understanding the phases of recovery from serious mental illness: The roles of referent and expert power in a mutual-help setting.* Journal of Community Psychology, 2009. **37**(2): p. 249-267.
58. Drapalski AL, et al., *Assessing Recovery of People With Serious Mental Illness: Development of a New Scale.* Psychiatric Services, 2012. **63**(1): p. 48-53.
59. MacKeith, J. and S. Burns, *Mental Health Recovery Star*. 2008, Mental Health Providers Forum and Triangle Consulting: London.
60. Neil ST, et al., *The questionnaire about the process of recovery (QPR): A measurement tool developed in collaboration with service users.* Psychosis, 2009. **1**(2): p. 145-155.
61. Haddock G, et al., *The Subjective Experiences of Psychosis Scale (SEPS): Psychometric evaluation of a scale to assess outcome in psychosis.* Schizophrenia Research, 2011. **133**: p. 244-249.

62. Andresen, R., P. Caputi, and L. Oades, *Stages of recovery instrument: Development of a measure of recovery from serious mental illness.* Australian and New Zealand Journal of Psychiatry, 2006. **40**: p. 972-980.
63. Corrigan, P.W. and S.M. Phelan, *Social support and recovery in people with serious mental illnesses.* Community Mental Health Journal, 2004. **40**(6): p. 513-523.
64. Corrigan, P.W., et al., *Examining the factor structure of the Recovery Assessment Scale.* Schizophrenia Bulletin, 2004. **30**(4): p. 1035-1041.
65. Flinn, S., *Reliability and Validity of the Recovery Assessment Scale for Consumers with Severe Mental Illness Living in Group Home Settings*. 2005, Kent State University: Kent, OH.
66. McNaught, M., et al., *Testing the validity of the Recovery Assessment Scale using an Australian sample.* Australian and New Zealand Journal of Psychiatry, 2007. **41**(5): p. 450-457.
67. Andresen R, Caputi P, and Oades LG, *Do clinical outcome measures assess consumer-defined recovery?* Psychiatry Research, 2010. **177**: p. 309-317.
68. Corrigan PW and Phelan SM, *Social Support and Recovery in People with Serious Mental Illnesses.* Community Mental Health Journal, 2004. **40**(6): p. 513-523.
69. McColl, M., et al., *The Community Integration Measure: Development and preliminary validation.* Archives of Physical Medicine and Rehabilitation, 2001. **82**: p. 429-434.
70. Weeks G, Slade M, and Hayward M, *A UK validation of the Stages of Recovery Instrument.* International Journal of Social Psychiatry, 2011. **57**: p. 446-454.
71. Lloyd C, King R, and Moore L, *Subjective and Objective Indicators of Recovery in Severe Mental Illness: a Cross-Sectional Study.* International Journal of Social Psychiatry, 2010. **56**: p. 220-229.
72. Corrigan PW, et al., *Examining the Factor Structure of the Recovery Assessment Scale.* Schizophrenia Bulletin, 2004. **30**(4): p. 1035-1041.
73. Law H, et al., *Recovery from psychosis: a user informed review of self-report instruments for measuring recovery.* Journal of Mental Health, 2012. **21**(2): p. 193-208.
74. Miller A, *Consumer Recovery Outcomes System (CROS 3.0)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 25-31.
75. Eisen, S.V., et al., *The revised Behavior and Symptom Identification Scale (BASIS-R).* Medical care, 2004. **42**: p. 1230-1241.
76. New York Crisis Hostel Project, *Crisis Hostel Healing Scale*, in *Can We Measure Recovery? A Compendium of Recovery and Recovery-related Instruments*. 1998, The Evaluation Center @ HSRI: Cambridge, MA.
77. Bullock WA, *Mental Health Recovery Measure (MHRM)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 32-35.
78. Bullock WA, *The Mental Health Recovery Measure (MHRM): updated normative data and psychometric properties*. 2009, University of Toledo.
79. Office of Research and Program Evaluation, *Ohio Mental Health Consumer Outcomes System - Adult Consumer Form*. 2000, Ohio Department of Mental Health: Columbus, Ohio.
80. Andresen, R., P. Caputi, and L.G. Oades, *Do clinical outcome measures assess consumer-defined recovery?* Psychiatry Research, 2010. **177**(3): p. 309-317.
81. Cavelti M, et al., *Assessing recovery from schizophrenia as an individual process. A review of self-report instruments.* European Psychiatry, 2012. **27**: p. 19-32.
82. Young, S.L. and D.S. Ensing, *Exploring recovery from the perspective of people with psychiatric disabilities.* Psychiatric rehabilitation journal, 1999. **22**(3): p. 291.
83. Silver T, et al., *Reciprocal Support Scale*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 57-59.

84. Leavy RL, et al., *Relationships and Activities that Facilitate Recovery Survey (RAFRS)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 68-70.

85. Hasson-Ohayon, I., D. Roe, and S. Kravetz, *The psychometric properties of the illness management and recovery scale: Client and clinician versions.* Psychiatry Research, 2007. **160**: p. 228-235.

86. Salyers, M.P., et al., *Measuring illness management outcomes: A psychometric study of clinician and consumer rating scales for illness self management and recovery.* Community Mental Health Journal, 2007. **43**(5): p. 459-480.

87. Ridgway P. and Press A., *Assessing the recovery-orientation of your mental health program: a user's guide for the Recovery-Enhancing Environment Scale (REE)*. 2004, University of Kansas, School of Social WElfare, Office of Mental Health Training and Resarch: Larewnce, Kansas.

88. Sklar M, et al., *The Psychometric properties of the Illness Management and Recovery scale in a large American public health system.* Psychiatry Research, 2012. **199**: p. 220-227.

89. Fardig R, et al., *A Randomized Controlled Trial of the Illness Management and Recovery Program for Persons With Schizophrenia.* Psychiatric Services, 2011. **62**(6): p. 606-612.

90. Mueser KT, et al., *The Illness Management and Recovery Program: Rationale, Development and Preliminary Findings.* Schizophrenia Bulletin, 2006. **32**(1): p. 32-43.

91. Roth D, *Ohio Mental Health Consumer Outcomes System (Ohio Outcomes System)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 42-49.

92. Cook JA, et al., *Peer Outcomes Protocol (POP)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 50-56.

93. Ralph RO, *Recovery Measurement Tool Version 4.0 (RMT)*, in *Measuring the Promise: A Compendium of Recovery Measures*, Campbell-Orde T, et al., Editors. 2005, The Evaluation Center @ HSRI: Cambridge, Massachusetts. p. 65-67.

94. Fisher DG, et al., *Psychometric Properties of an Assessment for Mental Health Recovery Programs.* Community Ment Health J, 2009. **45**: p. 246-250.

95. Miller L, et al., *Patterns of Recovery from Severe Mental Illness: A Pilot Study of Outcomes.* Community Ment Health J, 2010. **46**: p. 177-187.

96. Pilon D and Ragins M, *Milestones of Recovery Scale (MORS)*. 2007.

97. Ahmed AO, et al., *A psychometric study of recovery among Certified Peer Specialists.* Psychiatry Research, 2013. **Epub ahead of print. doi: 10.1016/j.psychres.2013.01.011**.

98. Killaspy H, et al., *Psychometric Properties of the Mental Health Recovery Star.* The British Journal of Psychiatry, 2012. **201**: p. 65-70.

99. Dickens G, et al., *Recovery Star: validating user recovery.* The Psychiatrist, 2012. **36**: p. 45-50.

100. Onifade Y, *The Mental Health Recovery Star.* Mental Health and Social Inclusion, 2011. **15**(2): p. 78-87.

101. Stanley S and Weleminsky J, *Recovery Star: Online Data Report*. 2010.

102. Morrison AP, et al., *Cognitive therapy for people with a schizophrenia spectrum diagnosis not taking antipsychotic medication: an exploratory trial* Psychological Medicine, 2012. **42**: p. 1049-1056.

103. Song L-Y and Hsu S-T, *The Development of the Stages of Recovery Scale for Persons with Persistent Mental Illness.* Research on Social Work Practice, 2011. **21**: p. 572-581.

104. Kessler, R.C., et al., *Short screening scales to monitor population prevalences and trends in non-specific psychological distress.* Psychological Medicine, 2002. **32**(6): p. 959-976.

105. Veit, C.T. and J.E. Ware, Jr., *The structure of psychological distress and well-being in general populations.* J Consult Clin Psychol, 1983. **51**(5): p. 730-42.

106. Eisen, S.V., *Behavior and Symptom Identification Scale (BASIS-32)*, in *Outcomes Assessment in Clinical Practice*, L.I. Sederer and B. Dickey, Editors. 1996, Williams and Wilkins: Baltimore.
107. Hoffmann, F.L., K. Capelli, and X. Mastrianni, *Measuring treatment outcome for adults and adolescents: reliability and validity of BASIS-32.* Journal of Mental Health Administration, 1997. **24**(3): p. 316-31.
108. Klinkenberg, W.D., D.W. Cho, and B. Vieweg, *Reliability and validity of the interview and self-report versions of the BASIS-32.* Psychiatric Services, 1998. **49**(9): p. 1229-31.
109. Russo, J., et al., *The relationship of patient-administered outcome assessments to quality of life and physician ratings: Validity of the BASIS-32.* Journal of Mental Health Administration, 1997. **24**(2): p. 200-214.
110. Eisen, S.V., H.S. Leff, and E. Schaefer, *Implementing outcome systems: lessons from a test of the BASIS-32 and the SF-36.* Journal of behavioral health services and research, 1999. **26**(1): p. 18-27.
111. Eisen, S.V., et al., *Assessing behavioral health outcomes in outpatient programs: reliability and validity of the BASIS-32.* Journal of behavioral health services and research, 1999. **26**(1): p. 5-17.
112. Piccinelli, M., et al., *Validity and test-retest reliability of the italian version of the 12-item General Health Questionnaire in general practice: A comparison between three scoring methods.* Comprehensive Psychiatry, 1993. **34**(3): p. 198-205.
113. Beck, A.T., et al., *An Inventory for Measuring Clinical Anxiety: Psychometric Properties.* Journal of Consulting & Clinical Psychology, 1988. **56**(6): p. 893-897.
114. Hamilton, M., *A rating scale for depression.* Journal of Neurology, Neurosurgery & Psychiatry, 1960. **23**: p. 56-61.
115. Hamilton, M., *The assessment of anxiety states by rating.* British Journal of Medical Psychology, 1959. **32**: p. 50-55.
116. Spielberger, C.D., *Manual for the State–Trait Anxiety Inventory (Form Y)*. 1983, Mind Garden: Palo Alto, CA: .
117. Cox, J.L., J.M. Holden, and R. Sagovsky, *Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale.* British Journal of Psychiatry, 1987. **150**: p. 782-6.
118. Brink, T.L., et al., *Screening Tests for Geriatric Depression.* Clinical Gerontologist, 1982. **1**(1): p. 37-43.
119. Sheikh, J.I. and J.A. Yesavage, *Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version.* Clinical Gerontologist, 1986. **5**(1-2): p. 165-173.
120. Aikman, G.G. and M.E. Oehlert, *Geriatric Depression Scale.* Clinical Gerontologist, 2001. **22**(3-4): p. 63-70.
121. Derogatis, L.R., K. Rickels, and A.F. Rock, *The SCL-90 and the MMPI: a step in the validation of a new self-report scale.* British Journal of Psychiatry, 1976. **128**: p. 280-9.
122. Derogatis, L.R., et al., *The Hopkins Symptom Checklist (HSCL): a self-report inventory.* Behvioural Science 1974. **19**(1-15).
123. Derogatis, L.R., *SCL-90-R: Symptom Checklist-90-R*. 1994, NCS Pearson: Minnesota
124. Bonicatto, S., et al., *Validity and reliability of Symptom Checklist '90 (SCL90) in an Argentine population sample.* Social Psychiatry and Psychiatric Epidemiology, 1997. **32**(6): p. 332-338.
125. Goldberg, D.P. and V. Hillier, *A scaled version of the General Health Questionnaire.* Psychological Medicine, 1979. **9**(1): p. 139-145.
126. Goldberg, D.P., et al., *A comparison of two psychiatric screening tests.* The British Journal of Psychiatry, 1976. **129**: p. 61-67.
127. Goldberg, D. and P. Williams, *A user's guide to the General Health Questionnair*. 1991, NFER-Nelson: Windsor.
128. Campbell, A., J. Walker, and G. Farrell, *Confirmatory Factor Analysis of the GHQ-12: Can I See that Again?* Australian and New Zealand Journal of Psychiatry, 2003. **37**(4): p. 475-483.

129. Penninkilampi-Kerola, V., J. Miettunen, and H. Ebeling, *A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample.* Scandinavian Journal of Psychology, 2006. **47**(5): p. 431-440.

130. Clarke, D.M. and D.P. Mckenzie, *An Examination of the Efficiency of the 12-Item SPHERE Questionnaire as a Screening Instrument for Common Mental Disorders in Primary Care.* Australian and New Zealand Journal of Psychiatry, 2003. **37**(2): p. 236-239.

131. Hickie, I.B., et al., *Development of a simple screening tool for common mental disorders in general practice.* Medical Journal of Australia, 2001. **175 Suppl**: p. S10-7.

132. Gaynes, B.N., et al., *Feasibility and diagnostic validity of the M-3 checklist: a brief, self-rated screen for depressive, bipolar, anxiety, and post-traumatic stress disorders in primary care.* Annals of Family Medicine, 2010. **8**(2): p. 160-9.

133. Glazer, W.M., *Tools help family doctors treat anxiety. Behavioral health experts validate screening tool and treatment model for anxiety disorders.* Behavioral Healthcare, 2010. **30**(7): p. 38-9.

134. Reynolds, W.M., *Suicidal Ideation Questionnaire.* . 1987, Psychological Assessment Resources.: Odessa, FL:.

135. Davis, J.M., *Suicidal Ideation Questionnaire.* Journal of Psychoeducational Assessment, 1992. **10**(3): p. 298-301.

136. Posner, K., et al., *The Columbia–Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults.* American Journal of Psychiatry, 2011. **168**(12): p. 1266-1277.

137. Prigerson, H.G., et al., *Inventory of complicated grief: A scale to measure maladaptive symptoms of loss.* Psychiatry Research, 1995. **59**(1–2): p. 65-79.

138. Jacobs, S., C. Mazure, and H. Prigerson, *Diagnostic criteria for traumatic grief.* Death Studies, 2000. **24**(3): p. 185-199.

139. Andrykowski, M.A., et al., *Posttraumatic Stress Disorder After Treatment for Breast Cancer: Prevalence of Diagnosis and Use of the PTSD Checklist-Civilian Version (PCL-C) as a Screening Instrument.* Journal of Consulting & Clinical Psychology, 1998. **66**(3): p. 586-590.

140. Weathers. *PTSD Checklist*. 1993; Available from: http://www.ptsd.va.gov/professional/pages/assessments/ptsd-checklist.asp.

141. Bernstein, E.M. and F.W. Putnam, *Development, reliability, and validity of a dissociation scale.* Journal of Nervous & Mental Disease. **174**(12): p. 727-35.

142. Carlson, E.B., et al., *Validity of the Dissociative Experiences Scale in screening for multiple personality disorder: a multicenter study.* American Journal of Psychiatry, 1993. **150**(7): p. 1030-1036.

143. Goodman, R., *The Strengths and Difficulties Questionnaire: A research note.* Journal of Child Psychology and Psychiatry and Allied Disciplines, 1997. **38**(5): p. 581-586.

144. Goodman, R., *Psychometric properties of the Strengths and Difficulties Questionnaire.* Journal of the American Academy of Child and Adolescent Psychiatry, 2001. **40**(11): p. 1337-1345.

145. Achenbach, T.M. and C.S. Edelbrock, *Manual for the child behaviour checklist and revised behaviour profile*. 1983, University of Vermont Department of Psychiatry: Burlington.

146. Reddy, L., *The Devereux Early Childhood Assessment.* Canadian Journal of School Psychology, 2007. **22**(1): p. 121-127.

147. Lien, M.T. and J.S. Carlson, *Psychometric Properties of the Devereux Early Childhood Assessment in a Head Start Sample.* Journal of Psychoeducational Assessment, 2009. **27**(5): p. 386-396.

148. Silverman, W.K. and W.B. Nelles, *The Anxiety Disorders Interview Schedule for Children.* Journal of the American Academy of Child & Adolescent Psychiatry, 1988. **27**(6): p. 772-778.

149. Silverman WK and A. AM, *The Anxiety Disorders Interview Schedule for Children for DSM-IV: (Child and Parent Versions)*. 1996, Psychological Corporation San Antonio, TX.
150. Loyd, B.H. and R.R. Abidin, *Revision of the Parenting Stress Index.* Journal of Pediatric Psychology, 1985. **10**(2): p. 169-177.
151. Abidin, R.R., *Parenting Stress Index-Manual*. 1983, Pediatric Psychology Press.: Charlottesville, VA: .
152. Constantino, J.N. and C.P. Gruber, *Social Responsiveness Scale (SRS)*. 2005, Western Psychological Services.: Los Angeles, CA:.
153. Kuhl, J., L. Jarkon-Horlick, and R. Morrissey, *Measuring Barriers to Help-Seeking Behavior in Adolescents.* Journal of Youth and Adolescence, 1997. **26**(6): p. 637-650.
154. Wilson, C.J., et al., *Measuring Help-Seeking Intentions: Properties of the General Help-Seeking Questionnaire.* Canadian Journal of Counselling, 2005. **39**(1): p. 15-28.
155. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *Mini-mental state: A practical method for grading the cognitive state of patients for the clinician.* Journal of Psychiatric Research, 1975. **12**(3): p. 189-198.
156. Roth, M., et al., *CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia.* British Journal of Psychiatry, 1986. **149**: p. 698-709.
157. Hodkinson, H., *Evaluation of a mental test score for the assessment of mental impairment in the elderly.* Age and Ageing, 1972. **1**: p. 233-238.
158. Lindeboom, J., et al., *Some psychometric properties of the CAMCOG.* Psychological Medicine, 1993. **23**(01): p. 213-219.
159. Storey, J.E., et al., *The Rowland Universal Dementia Assessment Scale (RUDAS): a multicultural cognitive assessment scale.* International Psychogeriatrics, 2004. **16**(01): p. 13-31.
160. Tennant, R., et al., *The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation.* Health & Quality of Life Outcomes, 2007. **5**: p. 63.
161. Stewart-Brown, S., et al., *Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey.* Health & Quality of Life Outcomes, 2009. **7**: p. 1-8.
162. Bordin, E.S., *Of human bonds that bind or free.*, in *Society for Psychotherapy Research*. 1980: Pacific Grove, CA.
163. Horvath, A.O. and L.S. Greenberg, *Development and Validation of the Working Alliance Inventory.* Journal of Counseling Psychology, 1989. **36**(2): p. 223-233.
164. Tracey, T.J. and A.M. Kokotovic, *Factor structure of the Working Alliance Inventory.* Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1989. **1**(3): p. 207-210.
165. Rogers, S., et al., *A consumer-constructed scale to measure empowerment among users of mental health services.* Psychiatric Services, 1997. **48**(8): p. 1042-1047.
166. Carver, C., *You want to measure coping but your protocol' too long: Consider the brief cope.* International Journal of Behavioral Medicine, 1997. **4**(1): p. 92-100.
167. Costa, P., T, Jr., and R. McCrae, R, *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor model inventory (NEO-FFI) professional manual.* 1992, Psychological Assessment Resources: Odessa, FL.
168. Deane, F.P., C.J. Wilson, and J. Ciarrochi, *Suicidal ideation and help-negation: Not just hopelessness or prior help.* Journal of Clinical Psychology, 2001. **57**(7): p. 901-914.
169. Zarit, S.H., K.E. Reever, and J. Bach-Peterson, *Relatives of the Impaired Elderly: Correlates of Feelings of Burden.* The Gerontologist, 1980. **20**(6): p. 649-655.
170. Matthey, S. and C. Ross-Hamid, *Repeat testing on the Edinburgh Depression Scale and the HADS-A in pregnancy: Differentiating between transient and enduring distress.* Journal of Affective Disorders, 2012. **141**(2–3): p. 213-221.

171. Andrews, G. and T. Slade, *Interpreting scores on the Kessler Psychological Distress Scale (K10).* Australian and New Zealand Journal of Public Health, 2001. **25**(6): p. 494-497.
172. Kessler, R.C., et al., *Screening for serious mental illness in the general population.* Archives of General Psychiatry, 2003. **60**(2): p. 184-189.
173. Brooks, R.T., J. Beard, and Z. Steel, *Factor Structure and Interpretation of the K10.* Psychological Assessment, 2006. **18**(1): p. 62-70.
174. CATI Technical Reference Group, *Population Health Monitoring and Surveillance: Question Development Field Testing - Field Test 2 Report: Alcohol Consumption, Cardiovascular Disease and Tobacco Consumption*. 2004, CATI Technical Reference Group: Canberra.
175. Murugesan, G., et al., *Inpatient psychosocial rehabilitation in rural NSW: assessment of clinically significant change for people with severe mental illness.* Aust N Z J Psychiatry, 2007. **41**(4): p. 343-50.
176. Australian Bureau of Statistics, *Mental Health and Wellbeing: Profile of Adults, Australia*. 1998, Australian Bureau of Statistics: Canberra.
177. Australian Bureau of Statistics, *National Health Survey: Mental Health*. 2003, Australian Bureau of Statistics: Canberra.
178. **Dal Grande, E.**, **A. Taylor**, and **D. Wilson**, *South Australian Health and Wellbeing Survey*. 2002, South Australian Department of Human Services: Adelaide.
179. d'Espaignet, E., M. Measey, and E. Dal Grande, *Northern Territory Health and Wellbeing Survey: Non-Indigenous Population*. 2002, Department of Health and Community Services: Darwin.
180. Victorian Government Department of Human Services, *Victorian Population Health Survey 2001*. 2002, Victorian Government Department of Human Services: Melbourne.
181. Saunders, D. and A. Daly, *2000 Collaborative Health and Wellbeing Survey: Psychological Distress in the Western Australian Population*. 2001, Health Department of Western Australia: Perth.
182. New South Wales Department of Health, *New South Wales Adult Health Survey, 2002.* NSW Public Health Bulletin Supplement, 2003. **14**(S4).
183. New South Wales Department of Health, *New South Wales Adult Health Survey, 2003.* NSW Public Health Bulletin Supplement, 2004. **15**(S4).
184. Pirkis, J., et al., *Better outcomes in mental healthcare?* Primary Care Mental Health, In press.
185. Australian Divisions of General Practice, *Familiarisation Training GP and Practice Manual, Second Edition*. 2003, Australian Divisions of General Practice: Canberra.
186. Kornblith, A.B., et al., *The impact of docetaxel, estramustine, and low dose hydrocortisone on the quality of life of men with hormone refractory prostate cancer and their partners: A feasibility study.* Annals of Oncology, 2001. **12**(5): p. 633-41.
187. Kornblith, A.B., et al., *Social support as a buffer to the psychological impact of stressful life events in women with breast cancer.* Cancer, 2001. **91**(2): p. 443-54.
188. Kornblith, A.B., et al., *Impact of azacytidine on the quality of life of patients with myelodysplastic syndrome treated in a randomized phase III trial: A Cancer and Leukemia Group B study.* Journal of Clinical Oncology, 2002. **20**(10): p. 2441-52.
189. Manne, S. and R. Schnoll, *Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the Mental Health Inventory.* Psychological Assessment, 2001. **13**(1): p. 99-109.
190. Zika, S. and K. Chamberlain, *On the relation between meaning in life and psychological well-being.* British Journal of Psychology, 1992. **83**(1): p. 133-145.
191. Stedman, T., et al., *Measuring Consumer Outcomes In Mental Health: Field Testing of Selected Measures of Consumer Outcome in Mental Health*. 1997, Department of Health and Family Services: Canberra.
192. Cohen, L., et al., *The effects of type of surgery and time on psychological adjustment in women after breast cancer treatment.* Annals of Surgical Oncology, 2000. **7**(6): p. 427-34.

193. Florian, V. and Y. Drory, *Mental Health Inventory (MHI): Psychometric properties and normative data in the Israeli population.* Psychologia: Israel Journal of Psychology, 1990. **2**: p. 26-35.
194. Tanaka, J.S. and G.J. Huba, *Structures of psychological distress: Testing confirmatory hierarchical models.* Journal of Consulting and Clinical Psychology, 1984. **52**(4): p. 719-721.
195. Huebeck, B.G. and J.T. Neill, *Confirmatory factor analysis and reliability of the Mental Health Inventory for Australian adolescents.* Psychological Reports, 2000. **87**(2): p. 431-440.
196. Ostroff, J.S., et al., *Use of the Mental Health Inventory with adolescents: A secondary analysis of the Rand Health Insurance Study.* Psychological Assessment, 1996. **8**(1): p. 105-107.
197. Zautra, A.J., C.A. Guarnaccia, and J.W. Reich, *Factor structure of mental health measures for older adults.* Journal of Consulting and Clinical Psychology, 1988. **56**(4): p. 514-519.
198. Moran, T. and M. O'Hara, *A partner-rating scale of postpartum depression: The Edinburgh Postnatal Depression Scale - Partner (EPDS-P).* Archives of Women's Mental Health, 2006. **9**(4): p. 173-180.
199. Graham, C., et al., *Victorian Mental Health Outcomes Measurement Strategy: Consumer Perspectives on Future Directions for Outcome Self-Assessment: Report of the Consumer Consultation Project*. 2001, Department of Human Services, Victoria.
200. Lecomte, T., et al., *Further validation of the Client Assessment of Strengths Interests and Goals.* Schizophrenia Research, 2004. **66**(1): p. 59-70.
201. Wallace, C.J., et al., *CASIG: A consumer-centered assessment for planning individualized treatment and evaluating program outcomes.* Schizophrenia Research, 2001. **50**(1-2): p. 105-19.
202. Doerfler, L., M. Addis, and P. Moran, *Evaluating mental health outcomes in an inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32.* The Journal of Behavioral Health Services & Research, 2002. **29**(4): p. 394-403.
203. Eisen, S.V., M. Wilcox, and E. Schaefer, *Use of BASIS-32 for Outcome Assessment of Recipients of Outpatient Mental Health Services*. 1997, McLean Hospital: Belmont, MA. p. 5-17.
204. Eisen, S.V., M.C. Grob, and A.A. Klein, *BASIS: The development of a self-report measure for psychiatric inpatient evaluation.* Psychiatric hospital, 1986. **17**(4): p. 165-71.
205. Trauer, T. and G. Tobias, *The Camberwell Assessment of Need and Behaviour and Symptom Identification Scale as routine outcome measures in a psychiatric disability rehabilitation and support service.* Community Ment Health J, 2004. **40**(3): p. 211-21.
206. Dornelas, E.A., et al., *Validity of a brief measure of post-hospital adjustment for psychiatric patients.* Comprehensive Psychiatry, 2001. **42**(5): p. 410-415.
207. Sousa, S.A. and R. Frazier, *A nursing tool for adherence and recovery in psychosis: A pilot study.* Journal of psychosocial nursing and mental health services, 2004. **42**(3): p. 28-36.
208. Wilkins, L.P. and M.B. White, *Interrater reliability and concurrent validity of the Global Assessment of Relational Functioning (GARF) Scale using a card sort method: A pilot study.* Family Therapy, 2001. **28**(3): p. 157-170.
209. Chow, J.C.-C., L.R. Snowden, and W. McConnell, *A confirmatory factor analysis of the BASIS-32 in racial and ethnic samples.* Journal of behavioral health services and research, 2001. **28**(4): p. 400-411.
210. Eisen, S.V., D.L. Dill, and M.C. Grob, *Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation.* Hospital and Community Psychiatry, 1994. **45**(3): p. 242-7.
211. Eisen, S.V., M.C. Grob, and D.L. Dill, *Substance abuse in an inpatient population.* McLean Hospital Journal, 1989. **14**: p. 1-22.
212. Pallant, J.F., R.L. Miller, and A. Tennant, *Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis.* BMC Psychiatry, 2006. **6**: p. 28.

213. Crawford, J.R. and J.D. Henry, *The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample.* The British Journal of Clinical Psychology, 2003. **42**: p. 111-31.
214. Campbell, A., B. Hayes, and B. Buckby, *Aboriginal and Torres Strait Islander women's experience when interacting with the Edinburgh Postnatal Depression Scale: A brief note.* Australian Journal of Rural Health, 2008. **16**(3): p. 124-131.
215. Bergink, V., et al., *Validation of the Edinburgh Depression Scale during pregnancy.* Journal of Psychosomatic Research, 2011. **70**(4): p. 385-389.
216. Yadavaia, J.E. and S.C. Hayes, *Acceptance and Commitment Therapy for Self-Stigma Around Sexual Orientation: A Multiple Baseline Evaluation.* Cognitive and Behavioral Practice, 2012. **19**(4): p. 545-559.
217. Page, A.C., G.R. Hooke, and D.L. Morrison, *Psychometric properties of the Depression Anxiety Stress Scales (DASS) in depressed clinical samples.* British Journal of Clinical Psychology, 2007. **46**(3): p. 283-297.
218. Ng, F., et al., *The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure.* Acta Neuropsychiatrica, 2007. **19**(5): p. 304-310.
219. Beck, A.T., et al., *Comparison of Beck Depression Inventories-IA and -II in Psychiatric Outpatients.* Journal of Personality Assessment, 1996. **67**(3): p. 588.
220. Lako, I.M., et al., *A systematic review of instruments to measure depressive symptoms in patients with schizophrenia.* Journal of Affective Disorders, 2012. **140**(1): p. 38-47.
221. Osman, A., et al., *Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples.* Journal of Clinical Psychology, 2008. **64**(1): p. 83-102.
222. Segal, D.L., et al., *Psychometric Properties of the Beck Depression Inventory—II (BDI-II) Among Community-Dwelling Older Adults.* Behavior Modification, 2008. **32**(1): p. 3-20.
223. Cattell, R.B. and H.E. P. Cattell, *Personality Structure and the New Fifth Edition of the 16PF.* Educational and Psychological Measurement, 1995. **55**(6): p. 926-937.
224. Reynolds, W.M., *Adolescent psychopathology scale: Psychometric and technical manual.* 1998, Psychological Assessment Resources: Odessa, FL.
225. Al-Musawi, N.m.M., *Psychometric properties of the Beck Depression Inventory-II with university students in Bahrain.* Journal of Personality Assessment, 2001. **77**(3): p. 568-579.
226. Aben, I., et al., *Validity of the beck depression inventory, hospital anxiety and depression scale, SCL-90, and hamilton depression rating scale as screening instruments for depression in stroke patients.* Psychosomatics, 2002. **43**(5): p. 386-93.
227. Gallagher, D., G. Nies, and L.W. Thompson, *Reliability of the Beck Depression Inventory with older adults.* Journal of Consulting & Clinical Psychology, 1982. **50**(1): p. 152-153.
228. Sprinkle, S.D., et al., *Criterion Validity, Severity Cut Scores, and Test-Retest Reliability of the Beck Depression Inventory-II in a University Counseling Center Sample.* Journal of Counseling Psychology, 2002. **49**(3): p. 381-385.
229. Schneibel, R., et al., *Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the Beck Depression Inventory in depressed inpatients.* Psychiatry Research, 2012. **198**(1): p. 62-67.
230. Hiroe, T., et al., *Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression.* Psychiatry Research, 2005. **135**(3): p. 229-235.
231. Leentjens, A.F.G., et al., *The validity of the Beck Depression Inventory as a screening and diagnostic instrument for depression in patients with Parkinson's disease.* Movement Disorders, 2000. **15**(6): p. 1221-1224.
232. Craven, J.L., G.M. Rodin, and C. Littlefield, *The Beck Depression Inventory as a screening device for major depression in renal dialysis patients.* International Journal of Psychiatry in Medicine, 1988. **18**(4): p. 365-74.

233. Lustman, P.J.P., et al., *Screening for Depression in Diabetes Using the Beck Depression Inventory.* Psychosomatic Medicine January/February, 1997. **59**(1): p. 24-31.
234. Strober, M., J. Green, and G. Carlson, *Utility of the Beck Depression Inventory with psychiatrically hospitalized adolescents.* Journal of Consulting & Clinical Psychology, 1981. **49**(3): p. 482-483.
235. Beck, A.T. and R.A. Steer, *Relationship between the beck anxiety inventory and the Hamilton anxiety rating scale with anxious outpatients.* Journal of Anxiety Disorders, 1991. **5**(3): p. 213-223.
236. Leyfer, O.T., J.L. Ruberg, and J. Woodruff-Borden, *Examination of the utility of the Beck Anxiety Inventory and its factors as a screener for anxiety disorders.* Journal of Anxiety Disorders, 2006. **20**(4): p. 444-458.
237. Hewitt, P.L. and G. Norton, *The Beck Anxiety Inventory: A psychometric analysis.* Psychological Assessment, 1993. **5**(4): p. 408-412.
238. De Ayala, R.J., D.J. Vonderharr-Carlson, and D. Kim, *Assessing the Reliability of the Beck Anxiety Inventory Scores.* Educational and Psychological Measurement, 2005. **65**(5): p. 742-756.
239. Hedman, E., et al., *Cost-effectiveness and long-term effectiveness of Internet-based cognitive behaviour therapy for severe health anxiety.* Psychological Medicine, 2013. **43**(2): p. 363-374.
240. Jolly, J.B., et al., *The utility of the Beck anxiety inventory with inpatient adolescents.* Journal of Anxiety Disorders, 1993. **7**(2): p. 95-106.
241. Kabacoff, R.I., et al., *Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the state-trait anxiety inventory with older adult psychiatric outpatients.* Journal of Anxiety Disorders, 1997. **11**(1): p. 33-47.
242. Heo, M., C.F. Murphy, and B.S. Meyers, *Relationship between the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale in depressed elderly: A meta-analysis.* The American Journal of Geriatric Psychiatry, 2007. **15**(10): p. 899-905.
243. Bagby, R., et al., *The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight?* The American Journal of Psychiatry, 2004. **161**(12): p. 2163-77.
244. Laenen, A., et al., *Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scales.* Journal of Psychiatric Research, 2009. **43**(7): p. 730-738.
245. Gonzalez-Pinto, A., et al., *Validity and reliability of the Hamilton Depression Rating Scale (5 items) for manic and mixed bipolar disorders.* Journal of Nervous and Mental Disease, 2009. **197**(9): p. 682-686.
246. Bech, P., *Fifty Years with the Hamilton Scales for Anxiety and Depression.* Psychotherapy and Psychosomatics, 2009. **78**(4): p. 202-11.
247. Kobak, K.A., et al., *Sources of unreliability in depression ratings.* Journal of Clinical Psychopharmacology, 2009. **29**(1): p. 82-85.
248. Lipman , R.S., et al., *Pharmacotherapy of anxiety and depression.* Psychopharmacology Bulletin, 1981. **17**: p. 91-103.
249. Mondolo, F., et al., *Evaluation of anxiety in Parkinson's disease with some commonly used rating scales.* Neurological Sciences, 2007. **28**(5): p. 270-275.
250. Clark, D.B. and J.E. Donovan, *Reliability and validity of the Hamilton Anxiety Rating Scale in an adolescent sample.* Journal of the American Academy of Child & Adolescent Psychiatry, 1994. **33**(3): p. 354-360.
251. Therrien, Z. and J. Hunsley, *Assessment of anxiety in older adults: A systematic review of commonly used measures.* Aging & Mental Health, 2012. **16**(1): p. 1-16.
252. Maier, W., et al., *The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders.* Journal of Affective Disorders, 1988. **14**(1): p. 61-68.
253. Therrien, Z. and J. Hunsley, *Assessment of anxiety in older adults: A systematic review of commonly used measures.* Aging & Mental Health, 2011. **16**(1): p. 1-16.

254. Bieling, P.J., M.M. Antony, and R.P. Swinson, *The State--Trait Anxiety Inventory, Trait version: structure and content re-examined.* Behaviour Research and Therapy, 1998. **36**(7–8): p. 777-788.
255. Figueiredo, B., et al., *Partner relationships during the transition to parenthood.* Journal of Reproductive and Infant Psychology, 2008. **26**(2): p. 99-107.
256. Kessler, R.C., P. Barker, and L.J. Colpe, *Screening for serious mental illness in the general population.* Archives of General Psychiatry, 2003. **60**(2): p. 184-189.
257. Julian, L.J., *Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A).* Arthritis Care & Research, 2011. **63**(S11): p. S467-S472.
258. Victorian Government. Mental Health Services, *Safety and Quality in Victoria's Specialist Public Mental Health System - C&C Experience*. 2008, Victorian Government: Melbourne.
259. van Boeijen, C.A., et al., *Treatment of anxiety disorders in primary care practice: a randomised controlled trial.* British Journal of General Practice, 2005. **55**(519): p. 763-9.
260. Gros, D.F., L.J. Simms, and M.M. Antony, *Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads.* Behavior Therapy, 2010. **41**(3): p. 277-284.
261. E, W.J. *SF-36 Health Survey Update from www.sf-36.org*. 2012 12 December 2012]; Available from: http://www.sf-36.org/tools/sf36.shtml
262. Cutcliffe, J.R. and P. Goward, *Mental health nurses and qualitative research methods: a mutual attraction?* Journal of Advanced Nursing, 2000. **31**(3): p. 590-598.
263. Rossi, V. and G. Pourtois, *Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: a comparative review.* Anxiety, Stress & Coping, 2011. **25**(6): p. 603-645.
264. Logsdon, M.C., W.M. Usui, and M. Nering, *Validation of Edinburgh postnatal depression scale for adolescent mothers.* Archives of Women's Mental Health, 2009. **12**(6): p. 433-40.
265. Potvin, O., et al., *Norms and associated factors of the STAI-Y State anxiety inventory in older adults: results from the PAQUID study.* International Psychogeriatrics, 2011. **23**(06): p. 869-879.
266. Gros, D.F., et al., *Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI).* Psychological Assessment, 2007. **19**(4): p. 369-381.
267. Virella, B., C. Arbona, and D.M. Novy, *Psychometric properties and Factor Structure of the Spanish Version of the State--Trait Anxiety Inventory.* Journal of Personality Assessment, 1994. **63**(3): p. 401.
268. Guedeney, N. and J. Fermanian, *Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results about use and psychometric properties.* European Psychiatry, 1998. **13**(2): p. 83-89.
269. Garcia, K.S., et al., *Repetitive transcranial magnetic stimulation treats postpartum depression.* Brain Stimulation, 2010. **3**(1): p. 36-41.
270. Newham, J.J., et al., *State–trait anxiety inventory (STAI) scores during pregnancy following intervention with complementary therapies.* Journal of Affective Disorders, 2012. **142**(1–3): p. 22-30.
271. Dennis, R., S. Boddington, and N. Funnell, *Self-report measures of anxiety: Are they suitable for older adults?* Aging & Mental Health, 2007. **11**(6): p. 668-677.
272. Zung, W.K., *A self-rating depression scale.* Archives of General Psychiatry, 1965. **12**(1): p. 63-70.
273. Lucas-Carrasco, R., *Spanish version of the Geriatric Depression Scale: reliability and validity in persons with mild–moderate dementia.* International Psychogeriatrics, 2012. **24**(Special Issue 08): p. 1284-1290.

274. Mahoney, F.I. and D.W. Barthel, *Functional evaluation: The Barthel Index: A simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill.* Maryland State Medical Journal, 1965. **14**: p. 61-65.

275. Alexopoulos, G.S., et al., *Cornell Scale for Depression in Dementia.* Biological Psychiatry, 1988. **23**(3): p. 271-284.

276. Centre for Participant Reported Outcomes, *About the Australian WHOQoL-Bref*. 2007, University of Melbourne.

277. Pedraza, O., et al., *Internal Consistency and Test-Retest Stability of the Geriatric Depression Scale-Short Form in African American Older Adults.* Journal of Psychopathology and Behavioral Assessment, 2009. **31**(4): p. 412-416.

278. Martens, M.P., et al., *Analyzing Reliability of Change in Depression Among Persons With Rheumatoid Arthritis.* Arthritis & Rheumatism: Arthritis Care & Research, 2005. **53**(6): p. 973-978.

279. Rapp, S.R., et al., *Validation of a cognitive assessment battery administered over the telephone.* Journal of the American Geriatrics Society, 2012. **60**(9): p. 1616-1623.

280. Evans, C., F. Margison, and M. Barkham, *The contribution of reliable and clinically significant change methods to evidence-based mental health.* Evidence Based Mental Health, 1998. **1**(3): p. 70-72.

281. DeJong, A. and I.W. Molenaar, *An application of Mokken's model for stochastic, cumulative scaling in psychiatric research.* Journal of Psychiatric Research, 1987. **21**(2): p. 137-149.

282. da Rocha, N.S., et al., *An introduction to Rasch analysis for Psychiatric practice and research.* Journal of Psychiatric Research, 2013. **47**(2): p. 141-148.

283. Olsen, L.R., E.L. Mortensen, and P. Bech, *The SCL-90 and SCL-90R versions validated by item response models in a Danish community sample.* Acta Psychiatrica Scandinavica, 2004. **110**(3): p. 225-229.

284. Rauter, U.K., C.E. Leonard, and C.P. Swett, *SCL-90-R factor structure in an acute, involuntary, adult psychiatric inpatient sample.* Journal of Clinical Psychology, 1996. **52**(6): p. 625-629.

285. Bjørkly, S., *SCL-90-R profiles in a sample of severely violent psychiatric inpatients.* Aggressive Behavior, 2002. **28**(6): p. 446-457.

286. Hafkenscheid, A., *Psychometric evaluation of the symptom checklist (SCL-90) in psychiatric inpatients.* Personality and Individual Differences, 1993. **14**(6): p. 751-756.

287. Øiesvold, T., T. Bakkejord, and J.A. Sexton, *Concurrent validity of the Health of the Nation Outcome Scales compared with a patient-derived measure, the Symptom Checklist-90-Revised in out-patient clinics.* Psychiatry Research, 2011. **187**(1–2): p. 297-300.

288. Schauenburg, H. and M. Strack, *Measuring psychotherapeutic change with the symptom checklist SCL 90 R.* Psychotherapy and Psychosomatics, 1999. **68**(4): p. 199-206.

289. van Straten, A., P. Cuijpers, and N. Smits, *Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: Randomized controlled trial.* Journal of Medical Internet Research, 2008. **10**(1): p. 80-89.

290. Schmitz, N., et al., *Comparison of the standard and the computerized versions of the Symptom Check List (SCL-90-R): a randomized trial.* Acta Psychiatrica Scandinavica, 2000. **102**(2): p. 147-152.

291. Dale, B., U. Soderhamn, and O. Soderhamn, *Psychometric properties of the Norwegian version of the General Health Questionnaire (GHQ-30) among older people living at home.* Psychology Research and Behavior Management, 2012. **5**: p. 151-157.

292. Mann, R.E., et al., *Estimating the prevalence of anxiety and mood disorders in an adolescent general population: An evaluation of the GHQ12.* International Journal of Mental Health and Addiction, 2011. **9**(4): p. 410-420.

293. Campbell, A. and S. Knowles, *A confirmatory factor analysis of the GHQ12 using a large Australian sample.* European Journal of Psychological Assessment, 2007. **23**(1): p. 2-8.

294. Quek, K.F., et al., *Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study.* Psychiatry and Clinical Neurosciences, 2001. **55**(5): p. 509-513.
295. Vallejo, M.A., et al., *Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the general health questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Tevised (SLR-90-R).* Journal of Medical Internet Research, 2007. **9**(1): p. 1-10.
296. Tait, R.J., G.K. Hulse, and S.I. Robertson, *A review of the validity of the General Health Questionnaire in adolescent populations.* Australian and New Zealand Journal of Psychiatry, 2002. **36**(4): p. 550-557.
297. Donath, S., *The validity of the 12-item General Health Questionnaire in Australia: A comparison between three scoring methods.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(2): p. 231-235.
298. Clover, K., et al., *Concurrent validity of the PSYCH-6, a very short scale for detecting anxiety and depression, among oncology outpatients.* Australian & New Zealand Journal of Psychiatry, 2009. **43**(7): p. 682-8.
299. Zigmond, A.S. and R.P. Snaith, *The Hospital Anxiety and Depression Scale.* Acta Psychiatrica Scandinavica, 1983. **67**(6): p. 361-370.
300. Bennett, B., et al., *Fatigue and psychological distress--exploring the relationship in women treated for breast cancer.* European Journal of Cancer, 2004. **40**(11): p. 1689-95.
301. Yellen, S.B., et al., *Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system.* Journal of Pain and Symptom Management, 1997. **13**(2): p. 63-74.
302. Graham, P.H., et al., *Inhalation aromatherapy during radiotherapy: results of a placebo-controlled double-blind randomized trial.* Journal of Clinical Oncology. **21**(12): p. 2372-6.
303. Goldstein, D., et al., *Fatigue states after cancer treatment occur both in association with, and independent of, mood disorder: a longitudinal study.* BMC Cancer. **6**: p. 240.
304. Hickie, I., et al., *Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study.* BMJ, 2006. **333**(7568): p. 575.
305. Pinto, A., M.A. Whisman, and K.J.M. McCoy, *Suicidal Ideation in Adolescents: Psychometric Properties of the Suicidal Ideation Questionnaire in a Clinical Sample.* Psychological Assessment, 1997. **9**(1): p. 63-66.
306. Horon, R., et al., *A Study of the Use and Interpretation of Standardized Suicide Risk Assessment: Measures within a Psychiatrically Hospitalized Correctional Population.* Suicide and Life-Threatening Behavior, 2013. **43**(1): p. 17-38.
307. Kaminer, Y., et al., *Suicidal Ideation among Adolescents with Alcohol Use Disorders during Treatment and Aftercare.* The American Journal on Addictions, 2006. **15**: p. s43-s49.
308. Reynolds, W.M. and J.J. Mazza, *Assessment of suicidal ideation in inner-city children and young adolescents: Reliability and validity of the Suicidal Ideation Questionnaire-JR.* School Psychology Review, 1999. **28**(1): p. 17-30.
309. Asarnow, J.R., et al., *Treatment of selective serotonin reuptake inhibitor-resistant depression in adolescents: Predictors and moderators of treatment response.* Journal of the American Academy of Child & Adolescent Psychiatry, 2009. **48**(3): p. 330-339.
310. Diamond, G.S., et al., *Attachment-based family therapy for adolescents with suicidal ideation: A randomized controlled trial.* Journal of the American Academy of Child & Adolescent Psychiatry, 2010. **49**(2): p. 122-131.
311. Mundt, J.C., et al., *Feasibility and validation of a computer-automated Columbia-Suicide severity rating scale using interactive voice response technology.* Journal of Psychiatric Research, 2010. **44**(16): p. 1224-1228.
312. Beck, A.T., M. Kovacs, and A. Weissman, *Assessment of suicidal intention: The Scale for Suicide Ideation.* Journal of Consulting and Clinical Psychology, 1979. **47**(2): p. 343-352.

313. Chappell, P., et al., *Initial Validity and Reliability Data on the Columbia–Suicide Severity Rating Scale.* American Journal of Psychiatry, 2012. **169**(6): p. 662-663.
314. Wilson, S., *The Validation of the Texas Revised Inventory of Grief on an Older Latino Sample.* Journal of Social Work in End-Of-Life & Palliative Care, 2007. **2**(4): p. 33-60.
315. Jacobs, S., et al., *The measurement of grief: bereaved versus non-bereaved.* The Hospice Journal 1987. **2**: p. 21-36.
316. Wang, Y., et al., *Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale.* International Journal of Nursing Studies, 2009. **46**(6): p. 813-823.
317. O'Connor, M., et al., *A confirmatory factor analysis of combined models of the Harvard Trauma Questionnaire and the Inventory of Complicated Grief-Revised: Are we measuring complicated grief or posttraumatic stress?* Journal of Anxiety Disorders, 2010. **24**(7): p. 672-679.
318. Ott, C., *The impact of comlicated grief on mental and physical health at various points in the bereavement process.* Death Studies, 2003. **27**(3): p. 249-272.
319. Boelen, P.A., L. Keijsers, and M.A. van den Hout, *The role of self-concept clarity in prolonged grief disorder.* Journal of Nervous and Mental Disease, 2012. **200**(1): p. 56-62.
320. Boelen, P.A. and H. Hoijtink, *An Item Response Theory Analysis of a Measure of Complicated Grief.* Death Studies, 2009. **33**(2): p. 101-129.
321. Meert, K.L., et al., *Follow-up study of complicated grief among parents eighteen months after a child's death in the pediatric intensive care unit.* Journal of Palliative Medicine, 2011. **14**(2): p. 207-214.
322. Shear K, F.E.H.P.R.R.C.F., *Treatment of complicated grief: A randomized controlled trial.* JAMA, 2005. **293**(21): p. 2601-2608.
323. Pini, S., et al., *The relationship between adult separation anxiety disorder and complicated grief in a cohort of 454 outpatients with mood and anxiety disorders.* Journal of Affective Disorders, 2012. **143**(1–3): p. 64-68.
324. Ruggiero, K.J., et al., *Psychometric properties of the PTSD checklist—civilian version.* Journal of Traumatic Stress, 2003. **16**(5): p. 495-502.
325. Horowitz, M., N. Wilner, and W. Alvarez, *Impact of Event Scale: a measure of subjective stress.* Psychosomatic Medicine, 1979. **41**(3): p. 209-18.
326. Vreven, D.L., et al., *The civilian version of the Mississippi Scale: A psychometric evaluation.* Journal of Traumatic Stress, 1995. **8**: p. 91 - 109.
327. Blake, D.D., et al., *A clinician rating scale for assessing current and lifetime PTSD: the CAPS-1.* The Behavior Therapist, 1990. **13**: p. 187-188.
328. Pietrzak, R.H., et al., *Diagnostic utility and factor structure of the PTSD Checklist in older adults.* International Psychogeriatrics, 2012. **24**(10): p. 1684-1696.
329. Hem, C., et al., *The Norwegian version of the PTSD Checklist (PCL): Construct validity in a community sample of 2004 tsunami survivors.* Nordic Journal Of Psychiatry, 2012. **66**(5): p. 355-359.
330. Blanchard, E.B., et al., *Psychometric properties of the PTSD checklist (PCL).* Behaviour Research and Therapy, 1996. **34**(8): p. 669-673.
331. Forbes, D., M. Creamer, and D. Biddle, *The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD.* Behaviour Research and Therapy, 2001. **39**(8): p. 977-986.
332. Conybeare, D., et al., *The PTSD Checklist—Civilian Version: Reliability, Validity, and Factor Structure in a Nonclinical Sample.* Journal of Clinical Psychology, 2012. **68**(6): p. 699-713.
333. Carison, E.B., et al., *Validity of the Dissociative Experiences Scale in screening for multiple personality disorder: a multicenter study.* American Journal of Psychiatry, 1993. **150**(7): p. 1030-1036.
334. Simeon, D., D.J. Stein, and E. Hollander, *Depersonalization disorder and self-injurious behavior.* Journal of Clinical Psychiatry, 1995. **56**(Suppl 4): p. 36-39.

335. Blevins, C.A., F.W. Weathers, and E.A. Mason, *Construct Validity of Three Depersonalization Measures in Trauma-Exposed College Students.* Journal of Trauma & Dissociation, 2012. **13**(5): p. 539-553.
336. Douglas, A.N., *Racial and Ethnic Differences in Dissociation: An Examination of the Dissociative Experiences Scale in a Nonclinical Population.* Journal of Trauma & Dissociation, 2009. **10**(1): p. 24-37.
337. Dubester, K.A. and B.G. Braun, *Psychometric properties of the Dissociative Experiences Scale.* Journal of Nervous & Mental Disease, 1995. **183**(4): p. 231-5.
338. Chard, K.M., *An evaluation of cognitive processing therapy for the treatment of posttraumatic stress disorder related to childhood sexual abuse.* Journal of Consulting & Clinical Psychology, 2005. **73**(5): p. 965-71.
339. Thabet, A.A., D. Stretch, and P. Vostanis, *Child mental health problems in Arab children: Application of the Strengths and Difficulties Questionnaire.* International Journal of Social Psychiatry, 2000. **46**(4): p. 266-280.
340. Muris, P., C. Meesters, and F. van den Berg, *The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents.* European child and adolescent psychiatry, 2003. **12**(1): p. 1-8.
341. Smedje, H., et al., *Psychometric properties of a Swedish version of the 'Strengths and Difficulties Questionnaire'.* European child and adolescent psychiatry, 1999. **8**(2): p. 63-70.
342. Hawes, D.J. and M.R. Dadds, *Australian data and psychometric properties of the Strengths and Difficulties Questionnaire.* Australian New Zealand Journal of Psychiatry, 2004. **38**(8): p. 644-651.
343. Koskelainen, M., A. Sourander, and M. Vauras, *Self-reported strengths and difficulties in a community sample of Finnish adolescents.* European child and adolescent psychiatry, 2001. **10**(3): p. 180-185.
344. Ronning, J.A., et al., *The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples.* European Child Adolescent Psychiatry, 2004. **13**(2): p. 73-82.
345. van Widenfelt, B.M., et al., *Dutch version of the Strengths and Difficulties Questionnaire (SDQ).* European child and adolescent psychiatry, 2003. **12**(6): p. 281-289.
346. Koskelainen, M., A. Sourander, and A. Kalijonen, *The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents.* European child and adolescent psychiatry, 2000. **9**(4): p. 277-284.
347. Malmberg, M., A.M. Rydell, and H. Smedje, *Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe).* Nordic journal of psychiatry, 2003. **57**(5): p. 357-63.
348. Mellor, D., *Furthering the use of the Strengths and Difficulties Questionnaire: Reliability with younger child respondents.* Psychological Assessment, 2004. **16**(4): p. 396-401.
349. Mellor, D., *Normative data for the strengths and difficulties questionaire in Australia.* Australian Psychologist, 2005. **40**(3): p. 215-222.
350. Goodman, R., H. Meltzer, and V. Bailey, *The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version.* European child and adolescent psychiatry, 1998. **7**(3): p. 125-130.
351. Goodman, R., H. Meltzer, and V. Bailey, *The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version.* International Review of Psychiatry, 2003. **15**(1-2): p. 173-177.
352. Goodman, R., *The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden.* Journal of Child Psychology and Psychiatry and Allied Disciplines, 1999. **40**(5): p. 791-799.

353. Mathai, J., P. Anderson, and A. Bourne, *Use of the Strengths and Difficulties Questionnaire as an outcome measure in a child and adolescent mental health service.* Australasian Psychiatry, 2003. **11**: p. 334-337.
354. Goodman, R. and S. Scott, *Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful?* Journal of Abnormal Child Psychology, 1999. **27**(1): p. 17-24.
355. Doerfler, L., D. Connor, and P. Toscano, Jr., *The CBCL Bipolar Profile and Attention, Mood, and Behavior Dysregulation.* Journal of Child and Family Studies, 2011. **20**(5): p. 545-553.
356. Seligman, L.D., et al., *The Utility of Measures of Child and Adolescent Anxiety: A Meta-Analytic Review of the Revised Children's Manifest Anxiety Scale, the State–Trait Anxiety Inventory for Children, and the Child Behavior Checklist.* Journal of Clinical Child & Adolescent Psychology, 2004. **33**(3): p. 557-565.
357. Ang, R.P., et al., *Examining the criterion validity of CBCL and TRF problem scales and items in a large Singapore sample.* Child Psychiatry and Human Development, 2012. **43**(1): p. 70-86.
358. Nakamura, B.J., et al., *A Psychometric Analysis of the Child Behavior Checklist DSM-Oriented Scales.* Journal of Psychopathology and Behavioral Assessment, 2009. **31**(3): p. 178-189.
359. Chorpita, B., et al., *Assessment of Tripartite Factors of Emotion in Children and Adolescents I: Structural Validity and Normative Data of an Affect and Arousal Scale.* Journal of Psychopathology and Behavioral Assessment, 2000. **22**(2): p. 141-160.
360. Chorpita, B.F., C.E. Moffitt, and J. Gray, *Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample.* Behaviour Research and Therapy, 2005. **43**(3): p. 309-322.
361. Garton, A.F., S.R. Zubrick, and S.R. Silburn, *The Western Australian child health survey: a pilot study.*
362. Zubrick, S.R., et al., *Western Australian Child Health Survey: Education, health and competence.* . 1997, Australian Bureau of Statistics and the TVW Telethon Institute for Child Health Research: Perth, Western Australia.
363. Tan, T.X., *Two-Year Follow-up of Girls Adopted from China: Continuity and Change in Behavioural Adjustment.* Child and Adolescent Mental Health, 2011. **16**(1): p. 14-21.
364. Willemen, A.M., et al., *Change in psychopathology in referred children: the role of life events and perceived stress.* Journal of Child Psychology and Psychiatry, 2008. **49**(11): p. 1175-1183.
365. Starling, J., J.M. Rey, and J.M. Simpson, *Depressive symptoms and suicidal behaviour: Changes with time in an adolescent clinic cohort. [References].*
366. Lipman, E.L., et al., *Evaluation of a community-based program for young boys at-risk of antisocial behaviour: Results and issues.* Journal of the Canadian Academy of Child and Adolescent Psychiatry / Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent, 2008. **17**(1): p. 12-19.
367. Weisz, J.R., et al., *Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy.* Journal of Consulting and Clinical Psychology, 2011. **79**(3): p. 369-380.
368. Nolan, T.M., et al., *Child Behaviour Checklist classification of behaviour disorder.* Journal of Paediatrics and Child Health, 1996. **32**(5): p. 405-411.
369. Siddons, H. and S. Lancaster, *An Overview of the Use of the Child Behavior Checklist within Australia*. 2004, Australian Council of Edicational Research: Melbourne.
370. LeBuffe, P.A. and V.B. Shapiro, *Lending "Strength" to the Assessment of Preschool Social-Emotional Health.* California School Psychologist, 2004. **9**: p. 51-61.
371. Ogg, J.A., et al., *Factor structure and invariance across gender of the Devereux Early Childhood Assessment protective factor scale.* School Psychology Quarterly, 2010. **25**(2): p. 107-118.

372. Jaberg, P.E., D.J. Dixon, and G.M. Weis, *Replication Evidence in Support of the Psychometric Properties of the Devereux Early Childhood Assessment.* Canadian Journal of School Psychology, 2009. **24**(2): p. 158-166.
373. Crane, J., M.S. Mincic, and A. Winsler, *Parent–Teacher Agreement and Reliability on the Devereux Early Childhood Assessment (DECA) in English and Spanish for Ethnically Diverse Children Living in Poverty.* Early Education & Development, 2011. **22**(3): p. 520-547.
374. Achenbach, T.M., S.H. McConaughy, and C.T. Howell, *Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity.* Psychological Bulletin, 1987. **101**(2): p. 213-232.
375. Wilson, K.R., S.S. Havighurst, and A.E. Harley, *Tuning in to Kids: An effectiveness trial of a parenting program targeting emotion socialization of preschoolers.* Journal of Family Psychology, 2012. **26**(1): p. 56-65.
376. Goelman, H., *Three complementary community-based approaches to the early identification of young children at risk for developmental delays/disorders.* Infants & Young Children, 2008. **21**(4): p. 306-323.
377. Anderson, S. and T. Ollendick, *Diagnosing Oppositional Defiant Disorder Using the Anxiety Disorders Interview Schedule for DSM-IV: Parent Version and the Diagnostic Interview Schedule for Children.* Journal of Psychopathology and Behavioral Assessment, 2012. **34**(4): p. 467-475.
378. Shaffer, D., et al., *NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses.* Journal of the American Academy of Child & Adolescent Psychiatry, 2000. **39**(1): p. 28-38.
379. Lyneham, H.J., M.J. Abbott, and R.M. Rapee, *Interrater reliability of the Anxiety Disorders Interview Schedule for DSM-IV: child and parent version.* Journal of the American Academy of Child & Adolescent Psychiatry. **46**(6): p. 731-6.
380. Grills, A.E. and T.H. Ollendick, *Multiple informant agreement and the anxiety disorders interview schedule for parents and children.* Journal of the American Academy of Child & Adolescent Psychiatry. **42**(1): p. 30-40.
381. Storch, E.A., et al., *Multiple informant agreement on the anxiety disorders interview schedule in youth with autism spectrum disorders.* Journal of Child & Adolescent Psychopharmacology. **22**(4): p. 292-9.
382. Brown-Jacobsen, A.M., D.P. Wallace, and S.P.H. Whiteside, *Multimethod, multi-informant agreement, and positive predictive value in the identification of child anxiety disorders using the SCAS and ADIS-C.* Assessment. **18**(3): p. 382-92.
383. Silverman, W.K., L.M. Saavedra, and A.A. Pina, *Test-retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for DSM-IV: child and parent versions.* Journal of the American Academy of Child & Adolescent Psychiatry, 2001. **40**(8): p. 937-44.
384. Jarrett, M.A. and T.H. Ollendick, *Treatment of comorbid attention-deficit/hyperactivity disorder and anxiety in children: A multiple baseline design analysis.* Journal of Consulting and Clinical Psychology, 2012. **80**(2): p. 239-244.
385. van der Leeden, A.J.M., et al., *Stepped Care Cognitive Behavioural Therapy for Children with Anxiety Disorders: A New Treatment Approach.* Behavioural and Cognitive Psychotherapy, 2011. **39**(01): p. 55-75.
386. Cobham, V.E., *Do anxiety-disordered children need to come into the clinic for efficacious treatment?* Journal of Consulting and Clinical Psychology, 2012. **80**(3): p. 465-476.
387. McKelvey, L.M., et al., *Validity of the short form of the parenting stress index for fathers of toddlers.* Journal of Child and Family Studies, 2009. **18**(1): p. 102-111.
388. Haskett, M.E., et al., *Factor Structure and Validity of the Parenting Stress Index-Short Form.* Journal of Clinical Child & Adolescent Psychology, 2006. **35**(2): p. 302-312.

389. Costa, N., et al., *Parenting Stress and Childhood Psychopathology: An Examination of Specificity to Internalizing and Externalizing Symptoms.* Journal of Psychopathology and Behavioral Assessment, 2006. **28**(2): p. 113-122.
390. Gerdes, A.C., L.M. Haack, and B.W. Schneider, *Parental Functioning in Families of Children With ADHD: Evidence for Behavioral Parent Training and Importance of Clinically Meaningful Change.* Journal of Attention Disorders, 2012. **16**(2): p. 147-156.
391. Hibel, L.C., E. Mercado, and J.M. Trumbell, *Parenting stressors and morning cortisol in a sample of working mothers.* Journal of Family Psychology, 2012. **26**(5): p. 738-746.
392. Zaidman-Zait, A., et al., *Factor analysis of the Parenting Stress Index-Short Form with parents of young children with autism spectrum disorders.* Autism Research, 2011. **4**(5): p. 336-346.
393. Sheppard, M., P. McDonald, and P. Welbourne, *The Parent Concerns Questionnaire and Parenting Stress Index: comparison of two Common Assessment Framework-compatible assessment instruments.* Child & Family Social Work, 2010. **15**(3): p. 345-356.
394. Leekam, S., et al., *Repetitive behaviours in typically developing 2-year-olds.* Journal of Child Psychology and Psychiatry, 2007. **48**(11): p. 1131-1138.
395. Wigham, S., et al., *The reliability and validity of the Social Responsiveness Scale in a UK general child population.* Research in Developmental Disabilities, 2012. **33**(3): p. 944-950.
396. Wang, J., et al., *Assessing Autistic Traits in a Taiwan Preschool Population: Cross-Cultural Validation of the Social Responsiveness Scale (SRS).* Journal of Autism and Developmental Disorders, 2012. **42**(11): p. 2450-2459.
397. Pine, E., et al., *Quantitative assessment of autistic symptomatology in preschoolers.* Autism, 2006. **10**(4): p. 344-352.
398. Bolte, S., F. Poustka, and J.N. Constantino, *Assessing autistic traits: Cross-cultural validation of the Social Responsiveness Scale (SRS).* Autism Research, 2008. **1**(6): p. 354-363.
399. White, S., et al., *Randomized Controlled Trial: Multimodal Anxiety and Social Skill Intervention for Adolescents with Autism Spectrum Disorder.* Journal of Autism and Developmental Disorders, 2013. **43**(2): p. 382-394.
400. Yui, K., et al., *Effects of large doses of arachidonic acid added to docosahexaenoic acid on social impairment in individuals with autism spectrum disorders: A double-blind, placebo-controlled, randomized trial.* Journal of Clinical Psychopharmacology, 2012. **32**(2): p. 200-206.
401. Booker, K.W. and L. Starling, **Test Review: Social Responsiveness Scale by J. N. Constantino and C. P. Gruber.* Assessment for Effective Intervention, 2011. **36**(3): p. 192-194.
402. Cunningham, A., *Measuring Change in Social Interaction Skills of Young Children with Autism.* Journal of Autism and Developmental Disorders, 2012. **42**(4): p. 593-605.
403. Wilson, C.J. and F.P. Deane, *Brief report: Need for autonomy and other perceived barriers relating to adolescents' intentions to seek professional mental health care.* Journal of Adolescence, 2012. **35**(1): p. 233-237.
404. Wilson, C.J., F.P. Deane, and J. Ciarrochi, *Can hopelessness and adolescents' beliefs and attitudes about seeking help account for help negation?* Journal of Clinical Psychology, 2005. **61**(12): p. 1525-1539.
405. Winkel-Witlox, A.C.M.T., et al., *Efficient screening of cognitive dysfunction in stroke patients: Comparison between the CAMCOG and the R-CAMCOG, Mini Mental State Examination and Functional Independence Measure-cognition score.* Disability & Rehabilitation, 2008. **30**(18): p. 1386-1391.
406. Granger, C.V., et al., *Performance profiles of the functional independence measure.* American Journal of Physical Medicine and Rehabilitation 1993. **72**: p. 84 - 89.
407. de Koning, I., et al., *A Short Screening Instrument for Poststroke Dementia: The R-CAMCOG.* Stroke: Journal of the American Heart Association, 200. **31**(7): p. 1502-1508.

408. Gallagher, D., et al., *Detecting prodromal Alzheimer's disease in mild cognitive impairment: utility of the CAMCOG and other neuropsychological predictors.* International Journal of Geriatric Psychiatry, 2010. **25**(12): p. 1280-1287.
409. Lolk, A., et al., *CAMCOG as a screening instrument for dementia: the Odense Study.* Acta Psychiatrica Scandinavica, 2000. **102**(5): p. 331-335.
410. Gatten, S.L., *Clinical differentiation of mental disorders in the elderly: Validation of the CAMDEX*. 1993, Ball State University: United States -- Indiana. p. 80-80 p.
411. Huppert, F.A., et al., *Psychometric properties of the CAMCOG and its efficacy in the diagnosis of dementia.* Aging, Neuropsychology, and Cognition, 1996. **3**(3): p. 201-214.
412. Paradela, E.M.P., C.d.S. Lopes, and R.A. Lourenco, *Reliability of the Brazilian version of the Cambridge Cognitive Examination Revised CAMCOG-R.* Arquivos de Neuro-Psiquiatria. **67**(2B): p. 439-44.
413. Stratford, J.A., et al., *A memory clinic at a geriatric hospital: a report on 577 patients assessed with the CAMDEX over 9 years.* Australian and New Zealand Journal of Psychiatry, 2003. **37**(3): p. 319-326.
414. Cullum, S., et al., *Decline across different domains of cognitive function in normal ageing: results of a longitudinal population-based study using CAMCOG.* International Journal of Geriatric Psychiatry, 2000. **15**(9): p. 853-862.
415. Heinik, J., et al., *The Cambridge cognitive examination (CAMCOG): validation of the Hebrew version in elderly demented patients.* International Journal of Geriatric Psychiatry, 1999. **14**(12): p. 1006-1013.
416. Heinik, J., *Effects of Trihexyphenidyl on MMSE and CAMCOG Scores of Medicated Elderly Patients With Schizophrenia.* International Psychogeriatric, 1998. **10**: p. 103 - 108.
417. Hvas, A.-M., et al., *No effect of vitamin B-12 treatment on cognitive function and depression: a randomized placebo controlled study.* Journal of Affective Disorders, 2004. **81**(3): p. 269-273.
418. Galea, M. and M. Woodward, *Mini-Mental State Examination (MMSE).* Australian Journal of Physiotherapy, 2005. **51**(3): p. 198.
419. Schmand, B., et al., *Relation between education and dementia: The role of test bias revisited.* Journal of Neurology, Neurosurgery & Psychiatry, 1995. **59**(2): p. 170-174.
420. Shiroky, J.S., et al., *Can You Have Dementia With an MMSE Score of 30?* American Journal of Alzheimer's Disease and Other Dementias, 2007. **22**(5): p. 406-415.
421. Rosen, W.G., R.C. Mohs, and K.L. Davis, *A new rating scale for Alzheimer's disease.* The American Journal of Psychiatry, 1984. **141**(11): p. 1356-1364.
422. Wouters, H., et al., *Three sides of the same coin: measuring global cognitive impairment with the MMSE, ADAS-cog and CAMCOG.* International Journal of Geriatric Psychiatry, 2010. **25**(8): p. 770-779.
423. Holzer, C.E., et al., *An epidemiologic assessment of cognitive impairment in a community population.* Research in Community & Mental Health, 1984. **4**: p. 3-32.
424. Foreman, M.D., *Reliability and validity of mental status questionnaires in elderly hospitalized patients.* Nursing Research, 1987. **36**(4): p. 216-220.
425. Tombaugh, T., et al., *Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A psychometric comparison and normative data.* Psychological Assessment, 1996. **8**(1): p. 48-59.
426. Mitrushina, M. and P. Satz, *Reliability and validity of the mini-mental state exam in neurologically intact elderly.* Journal of Clinical Psychology, 1991. **47**(4): p. 537-543.
427. Lopez, M.N., et al., *Psychometric Properties of the Folstein Mini-Mental State Examination.* Assessment, 2005. **12**(2): p. 137-144.
428. Hensel, A., M.C. Angermeyer, and S.G. Riedel-Heller, *Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination.* Journal of Neurology, Neurosurgery & Psychiatry, 2007. **78**(12): p. 1298-1303.

429. Nielsen, T., et al., *Cognitive testing in non-demented Turkish immigrants-Comparison of the RUDAS and the MMSE.* Scandinavian Journal of Psychology, 2012. **53**(6): p. 455-460.
430. Milne, A., et al., *Screening for dementia in primary care: a review of the use, efficacy and quality of measures.* International Psychogeriatrics, 2008. **20**: p. 911-926.
431. Cherbuin, N., K.J. Anstey, and D.M. Lipnicki, *Screening for dementia: a review of self- and informant-assessment instruments.* International Psychogeriatrics, 2008. **20**(03): p. 431-458.
432. Basic, D., et al., *Rowland Universal Dementia Assessment Scale, Mini-Mental State Examination and General Practitioner Assessment of Cognition in a multicultural cohort of community-dwelling older persons with early dementia.* Australian Psychologist, 2009. **44**(1): p. 40-53.
433. Wong, L., et al., *Reliability of the Rowland Universal Dementia Assessment Scale (RUDAS) via video conferencing.* International Journal of Geriatric Psychiatry, 2011. **26**(9): p. 988-989.
434. Jorm, A.F. and A.E. Korten, *Assessment of cognitive decline in the elderly by informant interview.* British Journal of Psychiatry, 1988. **159**: p. 209-213.
435. Goncalves, D.C., et al., *Case finding in dementia: Comparative utility of three brief instruments in the memory clinic setting.* International Psychogeriatrics, 2011. **23**(5): p. 788-796.
436. Rowland, J.T., et al., *The Rowland Universal Dementia Assessment Scale (RUDAS) and the Folstein MMSE in a multicultural cohort of elderly persons.* International Psychogeriatrics, 2006. **18**(1): p. 111-120.
437. Iype, T., et al., *Usefulness of the Rowland Universal Dementia Assessment Scale in South India.* Journal of Neurology, Neurosurgery & Psychiatry, 2006. **77**(4): p. 513-514.
438. Basic, D., et al., *The Validity of the Rowland Universal Dementia Assessment Scale (RUDAS) in a Multicultural Cohort of Community-dwelling Older Persons With Early Dementia.* Alzheimer Disease & Associated Disorders April/June, 2009. **23**(2): p. 124-129.
439. Wong, L., et al., *The Rowland Universal Dementia Assessment Scale (RUDAS) as a reliable screening tool for dementia when administered via videoconferencing in elderly post-acute hospital patients.* Journal Of Telemedicine And Telecare, 2012. **18**(3): p. 176-179.
440. Bartram, D.J., et al., *Validation of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) as an overall indicator of population mental health and well-being in the UK veterinary profession.* The Veterinary Journal, 2011. **187**(3): p. 397-398.
441. Stewart-Brown, S., et al., *Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey.* Health & Quality of Life Outcomes, 2009. **7**: p. 15.
442. Odou, N. and D.A. Vella-Brodrick, *The efficacy of positive psychology interventions to increase well-being and the role of mental imagery ability.* Social Indicators Research, 2013. **110**(1): p. 111-129.
443. Beckham, A.J., T. Greene, and S. Meltzer-Brody, *A pilot study of heart rate variability biofeedback therapy in the treatment of perinatal depression on a specialized perinatal psychiatry inpatient unit.* Archives of Women's Mental Health, 2013. **16**(1): p. 59-65.
444. Collins, J., et al., *Counselling in the workplace: How time-limited counselling can effect change in well-being.* Counselling and Psychotherapy Research, 2012. **12**(2): p. 84-92.
445. Jacobson, N.S., & Truax, P., *Clinical significance: a statistical approach to defining meaningful change in psychotherapy research.* Journal of Consulting and Clinical Psychology, 1991. **59**(1): p. 12-19.
446. Gremigni, P. and S. Stewart-Brown, *Measuring mental well-being: Italian validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS).* Giornale Italiano di Psicologia, 2011. **38**(2): p. 485-505.
447. Corbett, J., et al., *The Scottish Health Survey 2008*. 2010, United Kingdom: Scottish Government.: Edinburgh.

448. Catty, J., H. Winfield, and S. Clement, *The therapeutic relationship in secondary mental health care: a conceptual review of measures.* Acta Psychiatrica Scandinavica, 2007. **116**(4): p. 238-252.

449. Luborsky, L., et al., *The revised Helping Alliance questionnaire (HAq-II): Psychometric properties.* Journal of Psychotherapy Practice & Research, 1996. **5**(3): p. 260-271.

450. Munder, T., et al., *Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients.* Clinical Psychology & Psychotherapy, 2010. **17**(3): p. 231-239.

451. Busseri, M.A. and J.D. Tyler, *Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form.* Psychological Assessment, 2003. **15**(2): p. 193-197.

452. Bachelor, A., *Clients' and Therapists' Views of the Therapeutic Alliance: Similarities, Differences and Relationship to Therapy Outcome.* Clinical Psychology & Psychotherapy, 2013. **20**(2): p. 118-135.

453. Ross, E.C., D.L.L. Polaschek, and M. Wilson, *Shifting Perspectives: A Confirmatory Factor Analysis of the Working Alliance Inventory (Short Form) With High-Risk Violent Offenders.* International Journal of Offender Therapy and Comparative Criminology, 2011. **55**(8): p. 1308-1323.

454. Lorentzen, S., et al., *Psychodynamic group psychotherapy: impact of group length and therapist professional characteristics on development of therapeutic alliance.* Clinical Psychology & Psychotherapy, 2012. **19**(5): p. 420-33.

455. Owen, J., et al., *Alliance in action: A new measure of clients' perceptions of therapists' alliance activity.* Psychotherapy Research, 2012. **23**(1): p. 67-77.

456. Rogers, E., R. Ralph, and M. Ralzer, *Validating the Empowerment Scale With a Multisite Sample of Consumers of Mental Health Services.* Psychiatric Services, 2010. **61**(9): p. 933-6.

457. Herth, K., *Abbreviated instrument to measure hope: development and psychometric evaluation.* Journal of Advanced Nursing, 1992. **17**(10): p. 1251-1259.

458. Segal, S.P., C. Silverman, and T. Temkin, *Measuring empowerment in client-run self-help agencies.* Community Mental Health Journal, 1995. **31**(3): p. 215-227.

459. Lehman, A.F., *A Quality of Life Interview for the chronically mentally ill.* Evaluation and Program Planning, 1988. **11**(1): p. 51-62.

460. Shern, D.L., et al., *Client Outcomes II: Longitudinal Client Data from the Colorado Treatment Outcome Study.* The Milbank Quarterly, 1994. **72**(1): p. 123-148.

461. Corrigan, P.W., et al., *The construct validity of empowerment among consumers of mental health services.* Schizophrenia Research, 1999. **38**(1): p. 77-84.

462. Sarason, I.G., et al., *A Brief Measure of Social Support: Practical and Theoretical Implications.* Journal of Social and Personal Relationships, 1987. **4**(4): p. 497-510.

463. Robins, R.W., H.M. Hendin, and K.H. Trzesniewski, *Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale.* Personality and Social Psychology Bulletin, 2001. **27**(2): p. 151-161.

464. Lukoff, D., R.P. Liberman, and K.H. Nuechterlein, *Manual for the expanded Brief Psychiatric Rating Scale (BPRS).* Schizophrenia Bulletin, 1986. **12**: p. 594-602.

465. Corrigan, P.W., B. Buican, and S. McCraken, *The Needs and Resources Assessment interview for severely mentally ill adults.* Psychiatric Services, 1995. **46**(5): p. 504-505.

466. Chou, K.-R., et al., *Psychosocial rehabilitation activities, empowerment, and quality of community-based life for people with schizophrenia.* Archives of Psychiatric Nursing, 2012. **26**(4): p. 285-94.

467. Cooper, C., C. Katona, and G. Livingston, *Validity and reliability of the brief COPE in carers of people with dementia: The LASER-AD study.* Journal of Nervous and Mental Disease, 2008. **196**(11): p. 838-843.

468. Kapsou, M., et al., *Dimensionality of Coping: An Empirical Contribution to the Construct Validation of the Brief-COPE with a Greek-speaking Sample.* Journal of Health Psychology, 2010. **15**(2): p. 215-229.
469. Perczek, R., et al., *Coping, Mood, and Aspects of Personality in Spanish Translation and Evidence of Convergence With English Versions.* Journal of Personality Assessment, 2000. **74**(1): p. 63-87.
470. Coomber, K. and R. King, *A longitudinal examination of burden and psychological distress in carers of people with an eating disorder.* Social Psychiatry and Psychiatric Epidemiology, 2013. **48**(1): p. 163-171.
471. Carver, C.S., M.F. Scheier, and J.K. Weintraub, *Assessing Coping Strategies: A Theoretically Based Approach.* Journal of Personality & Social Psychology, 1989. **56**(2): p. 267-283.
472. Snell, D.L., et al., *Factor structure of the Brief COPE in people with mild traumatic brain injury.* The Journal of Head Trauma Rehabilitation, 2011. **26**(6): p. 468-477.
473. Yusoff, M.S.B., *A multicenter study on validity of the 30-items Brief COPE in identifying coping strategies among medical students.* International Medical Journal, 2010. **17**(4): p. 249-253.
474. Piedmont, R.L. and J.W. Ciarrocchi, *The utility of the Revised NEO Personality Inventory in an outpatient, drug rehabilitation context.* Psychology of Addictive Behaviors, 1999. **13**(3): p. 213-226.
475. Källmen, H., P. Wennberg, and H. Bergman, *Psychometric properties and norm data of the Swedish version of the NEO-PI-R.* Nordic Journal Of Psychiatry, 2011. **65**(5): p. 311-314.
476. McCrae, R.R., et al., *Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity.* Personality and Social Psychology Review, 2011. **15**(1): p. 28-50.
477. Samuel, D.B., A.J. Connolly, and S.A. Ball, *The Convergent and Concurrent Validity of Trait-Based Prototype Assessment of Personality Disorder Categories in Homeless Persons.* Assessment, 2012. **19**(3): p. 287-298.
478. Aboaja, A., C. Duggan, and B. Park, *An exploratory analysis of the NEO-FFI and DSM personality disorders using multivariate canonical correlation.* Personality and Mental Health, 2011. **5**(1): p. 1-11.
479. Costa, P.T. and R.R. McCrae, *Normal personality assessment in clinical practice: The NEO Personality Inventory.* Psychological Assessment, 1992. **4**(1): p. 5-13.
480. Murray, G., et al., *NEO Five-Factor Inventory Scores: Psychometric Properties in a Community Sample.* Measurement and Evaluation in Counseling and Development, 2003. **36**(3): p. 140-149.
481. Aluja, A., et al., *Comparison of the NEO-FFI, the NEO-FFI-R and an alternative short version of the NEO-PI-R (NEO-60) in Swiss and Spanish samples.* Personality and Individual Differences, 2005. **38**(3): p. 591-604.
482. Pocnet, C., et al., *Personality Changes in Patients With Beginning Alzheimer Disease.* Canadian Journal of Psychiatry, 2011. **56**(7): p. 408-17.
483. Detrick, P. and J.T. Chibnall, *Revised NEO Personality Inventory Normative Data for Police Officer Selection.* Psychological Services.
484. Wilson, C.J., et al., *The effects of need for autonomy and preference for seeking help from informal sources on emerging adults' intentions to access mental health services for common mental disorders and suicidal thoughts.* Advances in Mental Health, 2011. **10**(1): p. 29-38.
485. Cuijpers, P. and H. Stam, *Burnout among relatives of psychiatric patients attending psychoeducational support groups.* Psychiatric Services, 2000. **51**(3): p. 375-379.
486. Dare, A., et al., *Carer Outcome Measurement in Mental Health Services: Scoping the Field*. 2008, Australian Mental Health OUtcomes and Classification Network: Sydney.
487. Schene, A.H., *Objective and subjective burdens of family burden.* Social Psychiatry and Psychiatric Epidemiology, 1990. **25**: p. 289-297.

488. van Wijngaarden, B., et al., *Caregiving in schizophrenia: Development, internal consiconsistency and reliability of the Involvement Evaluation Questionnaire - European Version: EPSILON Study 4.* British Journal of Psychiatry, 2000. **177**(39): p. s21-27.
489. Schene, A.H. and B. van Winjgaarden, *The Involvement Evaluation Questionnaire*. 1992, Department of Psychiatry, University of Amsterdam: Amsterdam.
490. Reinhard, S.C., et al., *Burden Assessment Scale for families of the seriously mentally ill.* Evaluation and Program Planning, 1994. **17**(3): p. 261-269.
491. Foldemo, A., et al., *Quality of life and burden in parents of outpatients with schizophrenia.* Social Psychiatry and Psychiatric Epidemiology, 2005. **40**: p. 133-138.
492. Ivarsson, A.-B., B. Sidenvall, and M. Carlsson, *The factor structure of the Burden Assessment Scale and the perceived burden of caregivers for individuals with severe mental disorders.* Scandinavian Journal of Caring Sciences, 2004. **18**(4): p. 396-401.
493. Bachner, Y.G. and N. O'Rourke, *Reliability generalization of responses by care providers to the Zarit Burden Interview.* Aging & Mental Health, 2007. **11**(6): p. 678-685.
494. Gallagher, D., et al., *Dependence and Caregiver Burden in Alzheimer's Disease and Mild Cognitive Impairment.* American Journal of Alzheimer's Disease and Other Dementias, 2011. **26**(2): p. 110-114.
495. Knight, B.G., et al., *A Sociocultural Stress and Coping Model for Mental Health Outcomes Among African American Caregivers in Southern California.* The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 2000. **55**(3): p. P142-P150.
496. Lam, L.C.W., et al., *A randomized controlled trial to examine the effectiveness of case management model for community dwelling older persons with mild dementia in Hong Kong.* International Journal of Geriatric Psychiatry, 2010. **25**(4): p. 395-402.
497. Brouwer, W., et al., *The CarerQol instrument: A new instrument to measure carer-related quality of life of informal caregivers for use in economic evaluations.* Quality of Life Research, 2006. **15**: p. 1005-1021.
498. Hoefman, R.J., et al., *A new test of the construct validity of the CarerQol instrument: measuring the impact of informal care giving.* Quality of Life Research, 2011. **20**(6): p. 875-887.
499. Fossey, E.M. and C.A. Harvey, *A conceptual review of functioning: implications for the development of consumer outcome measures.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(1): p. 91-98.
500. Burns, T. and D. Patrick, *Social functioning as an outcome measure in schizophrenia studies.* Acta Psychiatrica Scandinavica, 2007. **116**(6): p. 403-418.
501. Brissos, S., et al., *The importance of measuring psychosocial functioning in schizophrenia.* Annals of General Psychiatry, 2011. **10**(art. no.18).
502. Peuskens, J. and P. Gorwood, *How are we assessing functioning in schizophrenia? A need for a consensus approach.* European Psychiatry, 2012. **27**: p. 391-395.
503. Brissos, S., et al., *The Portuguese version of the Personal and Social Performance Scale (PSP): Reliability, validity and relationship with cognitive measures in hospitalized and community schizophrenia patients.* Social Psychiatry and Psychiatric Epidemiology, 2012. **47**(7): p. 1077-1086.
504. Figueira, M. and S. Brissos, *Measuring psychosocial outcomes in schizophrenia patients.* Current Opinion in Psychiatry, 2011. **24**: p. 91-99.
505. Rosen, A., D. Hadzi-Pavlovic, and G. Parker, *The Life Skills Profile: A measure assessing function and disability in schizophrenia.* Schizophrenia Bulletin, 1989. **15**(2): p. 325-337.
506. Buckingham, W., et al., *Developing a Casemix Classification for Mental Health Services. Volume 2: Resource Materials*. 1998, Commonwealth Department of Health and Family Services: Canberra.
507. Rosen, A., et al., *Development of a brief form of the Life Skills Profile: The LSP-20.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(5): p. 677-83.

508. Parker, G., et al., *The Life Skills Profile: Psychometric properties of a measure assessing function and disability in schizophrenia.* Acta Psychiatrica Scandinavica, 1991. **83**(2): p. 145-152.
509. Buckingham, W., et al., *Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report*. 1998, Commonwealth Department of Health and Family Services: Canberra.
510. Burgess, P., et al., *Developing a casemix classification for specialist mental health services.* Casemix, 1999. **1**(4): p. 4 - 20.
511. Trauer, T., R.A. Duckmanton, and E. Chiu, *The Life Skills Profile: A study of its psychometric properties.* Australian and New Zealand Journal of Psychiatry, 1995. **29**(3): p. 492-499.
512. Bream, V. and A. Buchanan, *Distress among children whose separated or divorced parents cannot agree arrangements for them.* British Journal of Social Work, 2003. **33**(2): p. 227-238.
513. Vázquez Morejón, A.J. and R. Jiménez Ga-Bóveda, *Social Functioning Scale: new contributions concerning its psychometric characteristics in a Spanish adaptation.* Psychiatry Research, 2000. **93**(3): p. 247-256.
514. Birchwood, M., et al., *The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients.* British Journal of Psychiatry. **157**: p. 853-9.
515. Leifker, F., et al., *Validating measure of real-world outcome: the results of the VALERO expert survey and RAND panel.* Schizophrenia Bulletin, 2011. **37**(2): p. 334-343.
516. Mundt, J.C., et al., *The Work and Social Adjustment Scale: a simple measure of impairment in functioning.* British Journal of Psychiatry, 2002. **180**: p. 461-4.
517. Morosini, P.L., et al., *Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social funtioning.* Acta Psychiatrica Scandinavica, 2000. **101**(4): p. 323-329.
518. Jones, S., et al., *A brieft mental heatlh outcome scale-reliability and validity of the Global Assessment of Functioning (GAF).* British Journal of Psychiatry, 1995. **166**: p. 654-659.
519. Aas, I., *Guidelines for rating Global Assessment of Functioning (GAF).* Annals of General Psychiatry, 2011. **10**(2): p. 1-11.
520. Aas, I., *Global Assessment of Functioning (GAF): Properties and frontier of current knowledge.* Annals of General Psychiatry, 2010. **9**(art. no. 20): p. 1-11.
521. Tyrer, P., et al., *The Social Functioning Questionnaire: A Rapid and Robust Measure of Perceived Functioning.* International Journal of Social Psychiatry, 2005. **51**(3): p. 265-275.
522. Uston, T., et al., *Developing the World Health Organization Disability Assessment Schedule 2.0.* Bulletin World Health Organization, 2010. **88**: p. 815-823.
523. Barker, S., et al., *A Community Ability Scale for Chronically Mentally Ill consumers" Part I. Reliability and Validity.* Community Mental Health Journal, 1994. **30**(4): p. 363-383.
524. Donnelly, C. and A. Carswell, *Individualized outcome measures: A review of the literature.* Canadian Journal of Occupational Therapy, 2002. **69**(2): p. 84-94.
525. Endicott, J., et al., *The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance.* Archives of General Psychiatry, 1976. **33**.
526. Schaffer, D., Gould, M.S., Brasic, J., et al, *A children's global assessment scale (CGAS).* Archives of General Psychiatry, 1983. **40**: p. 1228-1231.
527. Bird, H.R., et al., *Impairment in the epidemiological measurement of childhood psychopathology in the community.* Journal of the American Academy of Child and Adolescent Psychiatry, 1990. **29**(5): p. 796-803.
528. Weissman, M.M., V. Warner, and M. Fendrich, *Applying impairment criteria to children's psychiatric diagnosis.* Journal of the American Academy of Child and Adolescent Psychiatry, 1990. **29**(5): p. 789-795.
529. Glascoe, F., *A method for deciding how to respond to parents' concerns about development and behaviour.* Ambulatory Child Health, 1999. **5**: p. 197-208.

530. Glascoe, F., *Collaborating with Parents: Using Parents' Evluations of Developmental Status to Detect and Address Developmental and Behavioral Problems*. 1998, Nashville, TN: Ellsworth and Vandermeer Press Ltd.

531. Frances Page, G., *Parents' evaluation of developmental status: How well do parents' concerns identify children with behavioral and emotional problems?* Clinical Pediatrics, 2003. **42**(2): p. 133-8.

532. Coghlan, D., J.S. Kiing, and M. Wake, *Parents' Evaluation of Developmental Status in the Australian day-care setting: developmental concerns of parents and carers.* Journal of Paediatrics & Child Health, 2003. **39**(1): p. 49-54.

533. Ng, W., et al., *Measuring the Prevalence of Children at Risk Using Parents' Evaluation of Developmental Status in a Telephone Survey.* Child indicators research, 2010. **3**(2): p. 167.

534. Voluse, A.C., et al., *Psychometric properties of the Drug Use Disorders Identification Test (DUDIT) with substance abusers in outpatient and residential treatment.* Addictive Behaviors, 2012. **37**(1): p. 36-41.

535. Berman, A.H., et al., *Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample.* European Addiction Research, 2005. **11**(1): p. 22-31.

536. Berman, A.H., et al., *The self-report Drug Use Disorders Identification Test—Extended (DUDIT-E): Reliability, validity, and motivational index.* Journal of Substance Abuse Treatment, 2007. **32**(4): p. 357-369.

537. Cassidy, C.M.M., N. Schmitz, and A. Malla, *Validation of the Alcohol Use Disorders Identification Test and the Drug Abuse Screening Test in First Episode Psychosis.* Canadian Journal of Psychiatry, 2008. **53**(1): p. 26-33.

538. Saunders, J., et al., *Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption - II.* Addiction, 1993. **88**: p. 791 - 804.

539. Leifker, F., et al., *Validating Measures of Real-World Outcome: The Results of the VALERO Expert Survey and RAND panel.* Schizophrenia Bulletin, 2013. **37**(2): p. 334-343.

540. Schneider, L. and E. Struening, *SLOF: a behavioral rating scale for assessing the mentally ill.* Social Work Research & Abstracts, 1983. **19**(3): p. 9-21.

541. Wallace, C., et al., *The Independent Living Skills Survey: A Comprehensive Measure of the Community Functioning of Severley and Persistently Mentally Ill Individuals.* Schizophrenia Bulletin, 2000. **26**(3): p. 631-658.

542. Wykes, T. and E. Sturt, *The Measurement of Social Behaviour in Psychiatric Patients: An Assessment of the Reliability and Validity of the SBS Schedule.* British Journal of Psychiatry, 1986. **148**: p. 1-11.

543. Hodges, K., A. Doucette-Gates, and Q. Liao, *The Relationship Between the Child and Adolescent Functional Assessment Scale (CAFAS) and Indicators of Functioning.* Journal of Child and Family Studies, 1999. **8**(1): p. 109-122.

544. Bates, M., *The Child and Adolescent Functional Assessment Scale (CAFAS): Review and Current Status.* Clinical Child and Family Psychology Review, 2001. **4**(1): p. 63-84.

545. Steinhausen, H.-C. and C.W. Metzke, *Global measures of impairment in children and adolescents: Results from a Swiss community survey.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(3): p. 282-286.

546. Singer, J., S. Eack, and C. Greeno, *The Columbia Impairment Scale: Factor Analysis Using a Community Mental Health Sample.* Research on Social Work Practice, 2011. **21**(4): p. 458-468.

547. Andrews, G., et al., *Follow-up of community placement in the chronic mentally ill in New South Wales.* Hospital and Community Psychiatry, 1990. **41**: p. 184-188.

548. Wooff, D., et al., *Correlates of stress in carers.* Journal of Mental Health, 2003. **12**(1): p. 29-40.

549. Trauer, T. and K. Eagar, *New Zealand Mental Health Consumers and their Outcomes*. 2004, Health Research Council of New Zealand: Auckland.
550. Dickinson, D. and R.D. Coursey, *Independence and overlap among neurocognitive correlates of community functioning in schizophrenia.* Schizophrenia Research, 2002. **56**(1-2): p. 161-170.
551. Parker, G., et al., *Assessing outcome in community mental health patients: A comparative analysis of measures.* International Journal of Social Psychiatry, 2002. **48**(1): p. 11-19.
552. Simon, A.E., et al., *Dysexecutive syndrome and social adjustment in schizophrenia.* Australian and New Zealand Journal of Psychiatry, 2003. **37**(3): p. 340-346.
553. Norman, R.M.G., et al., *The relationship of symptoms and level of functioning in schizophrenia to general wellbeing and the Quality of Life Scale.* Acta Psychiatrica Scandinavica, 2000. **102**(4): p. 303-309.
554. Trauer, T., R.A. Duckmanton, and E. Chiu, *Estimation of costs of public psychiatric treatment.* Psychiatric Services, 1998. **49**(4): p. 440-442.
555. Trauer, T., R.A. Duckmanton, and E. Chiu, *The assessment of clinically significant change using the Life Skills Profile.* Aust N Z J Psychiatry, 1997. **31**(2): p. 257-63.
556. Keller, S. and R. Hayes, *The relationship between the Allen Cognitive Level Test and the Life Skills Profile.* American Journal of Occupational Therapy, 1998. **52**(10): p. 851-856.
557. Kirkby, K., et al., *A survey of social outcome in schizophrenia in Tasmania.* Australian and New Zealand Journal of Psychiatry, 1997. **31**(3): p. 405-10.
558. Browne, G. and M. Courtney, *Measuring the impact of housing on people with schizophrenia.* Nursing and health sciences, 2004. **6**(1): p. 37-44.
559. Parker, G. and D. Hadzi-Pavlovic, *The capacity of a measure of disability (the LSP) to predict hospital readmission in those with schizophrenia.* Psychological Medicine, 1995. **25**(1): p. 157-163.
560. Preston, N.J., *Predicting community survival in early psychosis and schizophrenia populations after receiving intensive case management.* Australian and New Zealand Journal of Psychiatry, 2000. **34**(1): p. 122-8.
561. Ballesteros, R.J., et al., *Assessment of the Life Skills Profile and the Brief Psychiatric Rating Scale as predictors of the length of psychiatric hospitalization.* Actas Esplanolas de Psiquiatria, 2002. **30**: p. 225-232.
562. Kisely, S., N. Preston, and M. Rooney, *Pathways and outcomes of psychiatric care: Does it depend on who you are, or what you've got?* Australian and New Zealand Journal of Psychiatry, 2000. **34**(6): p. 1009-14.
563. Sanderson, K., et al., *Cost-effectiveness of Intensive Case Management for People with Serious Mental Illness*. 1996, South Eastern Sydney Area Mental Health Service: Sydney.
564. Johnston, S., et al., *Intensive case management: A cost-effectiveness analysis.* Australian and New Zealand Journal of Psychiatry, 1998. **32**(4): p. 551-9.
565. Craig, T., et al., *The consumer-employee as a member of a Mental Health Assertive Outreach Team: (1) Clinical and social outcomes.* Journal of Mental Health, 2004. **13**(1): p. 59-69.
566. Rosen, A. and M. Teesson, *Does case management work? The evidence and the abuse of evidence-based medicine.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(6): p. 731-746.
567. Hamernik, E. and K.I. Pakenham, *Assertive community treatment for persons with severe mental disorders: A controlled treatment outcome study.* Behaviour Change, 1999. **16**(4): p. 259-268.
568. Hambridge, J. and A. Rosen, *Assertive community treatment for the seriously mentally ill in suburban Sydney: A programme description and evaluation.* Australian and New Zealand Journal of Psychiatry, 1994. **28**: p. 438-445.
569. Ford, R., et al., *Intensive case management for people with serious mental illness - Site 2: Clinical and social outcome.* Journal of Mental Health, 1997. **6**(2): p. 181-190.

570. Ford, R., et al., *Maintaining contact with people with severe mental illness: 5-year follow-up of assertive outreach.* Social Psychiatry and Psychiatric Epidemiology, 2001. **36**(9): p. 444-447.

571. Stedman, T., et al., *The perceived utility of six selected measures of consumer outcomes proposed for routine use in Australian mental health services.* Australian and New Zealand Journal of Psychiatry, 2000. **34**(5): p. 842-9.

572. Vaskinn, A., et al., *Sex Differences in Neuropsychological Performance and Social Functioning in Schizophrenia and Bipolar Disorder.* Neuropsychology, 2011. **25**(4): p. 499-510.

573. Hellvin, T., et al., *Validation of the Norwegian version of the Social Functioning Scale (SFS) for schizophrenia and bipolar disorder.* Scandinavian Journal of Psychology, 2010. **51**: p. 525-533.

574. Hwan Jang, J., et al., *Longitudinal patterns of social functioning and conversion to psychosis in subjects at ultra-high risk.* Australian and New Zealand Journal of Psychiatry, 2011. **45**: p. 763-770.

575. Jolley, S., et al., *A pilot validation study of a new measure of activity in psychosis.* Social Psychiatry and Psychiatric Epidemiology, 2005. **40**: p. 905-911.

576. Simonsen, C., et al., *Psychosocial function in schizophrenia and bipolar disorder: Relationship to neurocognition and clinical symptoms.* Journal of International Neuropsychological Society, 2010. **16**: p. 771-783.

577. Dickerson, F., et al., *Outpatients with schizophrenia and bipolar I disorder: So they differ in their cognitive and social functioning?* Psychiatry Research, 2001. **102**: p. 21-27.

578. Barrowclough, C. and N. Tarrier, *Social Functioning in schizophrenic patients. The effects of expressed emotion and family intervention.* Social Psychiatry and Psychiatric Epidemiology, 1990. **25**(3): p. 125-129.

579. Rogers, K., et al., *The British Sign Language Versions of the Patient Health Questionnaire, the Generalized Anxiety Disorder 7-item Scale, and the Work and Social Adjustment Scale.* Journal of Deaf Studies and Deaf Education, 2013. **18**(1): p. 110-122.

580. Mataix-Cols, D., et al., *Reliability and validity of the Work and Social Adjustment Scale in phobic disorders.* Comprehensive Psychiatry, 2005. **46**(3): p. 223-8.

581. Jansson-Frojmark, M., *The Work and Social Adjustment Scale as a Measure of Dysfunction in Chronic Insominia: Reliability and Validity.* Behavioural and Cognitive Psycotherapy, 2013. **Feb 12 Epub** p. 1-13.

582. Tchanturia, K., et al., *Work and Social Adjustment in patients with anorexia nervosa.* Comprehensive Psychiatry

2013. **54**(1): p. 41-45.

583. Cella, M., M. Sharpe, and T. Chalder, *Measuring disability in patients with chronic fatigue syndrome: reliability and validity of the Work and Social Adjustment Scale.* Journal of Psychosomatic Research, 2011. **71**: p. 124-128.

584. Meyer, B., et al., *Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial.* Journal of Medical Internet Research, 2009. **11**: p. e15.

585. Kawata, A. and D. Revicki, *Psychometric properties of the Personal and Social Performance scale (PSP) among individuals with schizophrenia living in the community.* Quality of Life Research, 2008. **17**: p. 1247 - 1256.

586. Juckel, G., et al., *Validation of the Personal and Social Performance (PSP) Scale in a German sample of acutely ill patients with schizophrenia.* Schizophrenia Research, 2008. **104**: p. 287-293.

587. Apiquian, R., et al., *Validity of the Spanish version of the Personal and Social Performance scale in schizophrenia.* Schizophrenia Research, 2009. **112**(1-3): p. 181-186.

588. Tianmei, S., et al., *The Chinese version of the Personal and Social Performance Scale (PSP):validity and reliability.* Psychiatry Research, 2011. **185**: p. 275-279.

589. Schaub, D., et al., *The illness and everyday living: close interplay of psychopathological syndromes and psychosocial functioning in chronic schizophrenia.* European archives of psychiatry and clinical neuroscience, 2011. **261**: p. 85-93.
590. Nasrallah, H., P. Morosini, and D. Gagnon, *Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia.* Psychiatry Research, 2008. **161**: p. 213-224.
591. Patrick, D., et al., *Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia.* Current Medical Research and Opinion, 2009. **25**(2): p. 325 - 338.
592. Nicholl, D., et al., *Personal and social functioning in schizophrenia: defining a clinically meaningful measure of maintenance in relapse prevention.* Current Medical Research and Opinion, 2010. **26**(6): p. 1471-1484.
593. Molodynski, A., et al., *Clinical and research experience with the personal and social performance scale and suggested modifications.* Acta Psychiatrica Scandinavica, 2012. **126**(4): p. 300-301.
594. Bacon, S., M. Collins, and E. Plake, *Does the Global Assessment of Funcitoning Assess Functioning?* Journal of Mental Health Counseling, 2002. **24**(3): p. 202-212.
595. Pederson, G. and S. Karterud, *The symptom and functioning dimensions of the Global Assessment of Functioning (GAF) scale.* Comprehensive Psychiatry, 2012. **53**: p. 292-298.
596. Hilsenroth, M., et al., *Reliability and Validity of DSM-IV.* American Journal of Psychiatry, 2000. **157**: p. 1858-1863.
597. Patterson, D. and L. Myung-Shin, *Field Trial of the Global Assessment of Functioning Scale - Modified.* American Journal of Psychiatry, 1995. **152**(9): p. 1386-1388.
598. Hall, R., *Global Assessment of Functioning: A modified scale.* Psychosomatics, 1995. **36**(3): p. 267-275.
599. Startup, M., M. Jackson, and S. Bendix, *The concurrent validity of the Global Assessment of Functioning (GAF).* British Journal of Clinical Psychology, 2002. **41**: p. 417-422.
600. Grootenboer, E., et al., *Reliability and validity of the Global Assessment of Functioning Scale in clinical outpatients with depressive disorders.* Journal Evaluation in Clinical Practice, 2011. **18**: p. 502-507.
601. Roy-Byrne, P., et al., *Evidence for Limited Validity of the Revised Global Assessment of Functioning Scale.* Psychiatric Services, 1996. **47**(9): p. 864-866.
602. Moos, R., A. Nichol, and B. Moos, *Global Assessment of Functioning Ratings and the Allocation and Outcomes of Mental Health Services.* Psychiatric Services, 2002. **53**(6): p. 730-737.
603. Salvi, G., M. Leese, and M. Slade, *Routine use of mental health outcome assessments: choosing the measure.* Br J Psychiatry, 2005. **186**: p. 146-52.
604. Greenberg, G. and R. Rosenheck, *Using the GAF as a national mental health outcome measure in the department of veterens affairs.* Psychiatric Services, 2005. **56**: p. 420-426.
605. Tungstrom, S., P. Soderberg, and B.A. Armelius, *Relationship between the Global Assessment of Functioning and other DSM axes in routine clinical work.* Psychiatr Serv, 2005. **56**(4): p. 439-43.
606. Hay, P., et al., *A Two-Year Follow-Up Study and Prospective Evlaution of the DSM-IV Axis V.* Psychiatric Services, 2003. **54**(7): p. 1028-1030.
607. Gordon, R., P. Jardiolin, and K. Gordon, *Predicting length of hospital stay of psychiatric patients.* American Journal of Psychiatry, 1985. **142**: p. 235-237.
608. Soderberg, P., S. Tungstrom, and B.A. Armelius, *Reliability of global assessment of functioning ratings made by clinical psychiatric staff.* Psychiatr Serv, 2005. **56**(4): p. 434-8.
609. Vatnaland, T., et al., *Are GAF scores reliable in routine clincial use?* Acta Psychiatrica Scandinavica, 2007. **115**: p. 326-330.

610. Loevdahl, H. and S. Friss, *Routine evaluation of mental health: reliable informatin or worthless 'guesstimates'?* Acta Psychiatrica Scandinavica, 1996. **93**(2): p. 125-128.
611. Schennach-Wolff, R., et al., *Defining and predicting functional outcome in schizophrenia and schizophrenia spectrum disorders.* Schizophrenia Research, 2009: p. 210-217.
612. Burlingame, G.M., et al., *Selection of outcome assessment instruments for inpatients with severe and persistent mental illness.* Psychiatr Serv, 2005. **56**(4): p. 444-51.
613. Piersma, H. and J. Boes, *The GAF and Psychiatric Outcome: A Descriptive Report.* Community Mental Health Journal, 1997. **33**(1): p. 35 -41.
614. Carlson, J., et al., *Evaluating a Measure of Social Health Derived from Two Mental Health Recovery Measures: The California of Life(CA-QOL) and Mental Health Statistics Improvment Program Consumer Survey (MHSIP).* Community Mental Health Journal, 2011. **47**: p. 454-462.
615. Tyrer, P., et al., *A pilot study of the effects of early intervention on clinical symptoms and social functioning in psychiatric emergencies.* Irish Journal of Psychological Medicine, 1990. **7**(132-134).
616. Newton-Howes, G., P. Tyrer, and T. Weaver, *Social Functioning of Patients With Personality Disorder in Secondary Care.* Psychiatric Services, 2008. **59**(9): p. 1033-7.
617. Seivewright, H., P. Tyrer, and T. Johnson, *Persistent social dysfunction in anxious and dperessed patients with personality disorder.* Acta Psychiatrica Scandinavica, 2004. **109**: p. 104-109.
618. Nur, U., et al., *Social function, clinical symptoms and personality disturbance.* Irish Journal of Psychological Medicine, 2004. **21**(1): p. 18-21.
619. Tyrer, P., et al., *The effect of personality disorder on clinical outcome, social networks and adjustment: a controlled clinical trial of psychiatric emergencies.* Psychological Medicine, 1994. **24**(3): p. 731-740.
620. Judge, R., et al., *Discontinuation symptoms: comparison of brief interruptions in fluoxetine and paroxetine treatment.* International Clinical Psychopharmacology, 2002. **17**: p. 217-225.
621. Barrett, B., et al., *An examination of the factors that influence costs in medical patients with health anxiety.* Journal of Psychosomatic Research, 2012. **73**(1): p. 59-62.
622. Ro, E. and A. Clark, *Psychosocial Functioning in the Context of Diagnosis: Assessment and Theoretical Issues.* Psychological Assessment, 2009. **21**(3): p. 313-324.
623. Serfaty, M., et al., *Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: A randomized controlled trial.* Archives of General Psychiatry, 2009. **66**(12): p. 1332-1340.
624. Tyrer, P., et al., *Randomized controlled trial of brief congnitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: the POPMACT study.* Psychological Medicine, 2003. **33**(6): p. 969-976.
625. Huband, N., et al., *Social problem-solving plus psychoeducation for adults with personality disorder: pragmatic randomised controlled trial.* British Journal of Psychiatry, 2007. **190**: p. 307-13.
626. Chopra, P., J. Couper, and H. Herrman, *The assessment of patients with long-term psychotic disorders: Application of the WHO Disability Assessment Schedule II.* Australian and New Zealand Journal of Psychiatry, 2004. **38**(9): p. 753-759.
627. Chopra, P., H. Herrman, and G. Kennedy, *Comparison of disability and quality of life measures in patients with long-term psychotic disorders and patients with multiple sclerosis: an application of the WHO Disability Assessment schedule II and WHO Quality of Life-BREF.* International Journal of Rehabilitation Research, 2008. **31**(2): p. 141-149.
628. Guilera, G., et al., *Utility of the World Health Organization Disability Assessment Schedule II in schizophrenia.* Schizophrenia Research, 2012. **138**: p. 240-247.
629. Hu, L., y.-L. Zang, and N. Li, *The applicability of the WHODAS 2.0 in adolescents in China.* Journal of Clinical Nursing, 2012. **21**: p. 2438-2451.

630. Garin, O., et al., *Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases.* Health and Quality of Life Outcomes, 2010. **8**(51): p. 1 - 15.
631. Zhao, H., et al., *Activity limitation and participation restrictions of breast cancer patients receiving chemotherapy: psychometric properties and validation of the Chinese version of the WHODAS 2.0.* Quality of Life Research, 2013. **22**: p. 897-906.
632. Chwastiak, L. and M. Von Korff, *Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting.* Journal of Clinical Epidemiology, 2003. **56**: p. 507-514.
633. Posl, M., A. Cieza, and G. Stucki, *Psychometric properties of the WHODASII in rehabilitation patients.* Quality of Life Research, 2007. **16**: p. 1521-1531.
634. Hudson, M., et al., *Quality of life in systemic sclerosis:psychometirc properties of the World Health Organization Disability Assessment Schedule II.* Arthritis and Rheumatism, 2008. **59**(59): p. 270-278.
635. Dan, A., et al., *A comparative study on quality of life of patients of schizophrenia with and without depression.* Psychiatry Research, 2011. **189**: p. 185-189.
636. McKibbon, C., T. Patterson, and D. Jeste, *Assessing Disability in Older Patients With Schizophrenia. Results from WHODAS-II.* Journal of Nervous and Mental Disease, 2004. **192**(6): p. 405-413.
637. Luciano, J., et al., *Utility of the twelve-item World Health Organization Disability Assessment Schedule II (WHO-DAS II) for discriminating depression "caseness" and severity in Spanish primary care patients.* Quality of Life Research, 2010. **19**: p. 97-101.
638. Baron, M., et al., *The clinimetric properties of the World Health Organization Assessment Scedule II in ealy inflammatory arthritis.* Arthritis and Rheumatism, 2008. **59**: p. 382-390.
639. Kim, J., et al., *Physical health, depression and cognitive function as correlates of diability in an older Korean population.* International Journal of Geriatric Psychiatry, 2005. **20**: p. 160-167.
640. Chavez, L., et al., *Psychometric properties of the Spanish version of two mental healht outcome measures:World health Organization Disability Assessment Scheule II and Lehmans's Quality of Life Interview.* Mental Health Services Research, 2005. **7**: p. 145-159.
641. Noonan, V., et al., *A review of participation instruments based on the International Classification of Functioning, Disability and Health.* Disability and Rehabilitation, 2009. **31**(23): p. 1883-1901.
642. Chisolm, T., et al., *The WHO-DAS II: psychometirc properties in the measurement of functional health status in adults with aquired hearing loss.* Trends in amplification, 2005. **9**: p. 111-126.
643. Perini, S., T. Slade, and G. Andrews, *Generic effectiveness measures: Sensitivity to symptom change in anxiety disorders.* Journal of Affective Disorders, 2006. **90**(2-3): p. 123-130.
644. Pyne, J., et al., *Comparing the sensitivity of generic effectiveness measures with symptom improvement in persons with schizophrenia.* Medical Care, 2003. **41**: p. 208-217.
645. Federici, S., F. Meloni, and A. Presit, *International Literature Review on WHODAS II.* Life Span and Disability, 2009. **12**(1): p. 83-110.
646. Buist-Bouwman, M., et al., *Psychometric properites of the World Health Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders.* International Journal of Methods in Psychiatric Research, 2008. **17**(4): p. 185-197.
647. Kulnik, S. and D. Nikoletou, *WHODAS 2.0 in community rehabiltiation: a qualitative investigation into the validity of a generic patient-reported measure of disability.* Disability and Rehabilitation. **2013**(Epub): p. 1-9.
648. Hendryx, M., et al., *A test of the Reliability and Validity of the Multnomah Community Ability Scale.* Community Mental Health Journal, 2001. **37**(2): p. 157-168.

649. Corbiere, M., et al., *Factor structure of the Multnomah Community Ability Scale.* Journal of Nervous & Mental Disease, 2002. **190**(6): p. 399-406.

650. Bassani, D., et al., *Factor structure of the Multnomah Community Ability Scale - longitudinal analysis.* Psychiatry Research, 2009. **167**: p. 178-189.

651. O'Malia, L., et al., *A level-of-functioning self-report measure for consumers with severe mental illness.* Psychiatric Services, 2002. **53**(3): p. 326-331.

652. Brown, E., et al., *Clinician ratings vs. global ratings of symptom severity: a comparison of symptom measures in the bipolar disorder module, phaseII, Texas medication Algorithm Project* Psychiatry Research, 2003. **117**(2).

653. Trauer, T., *Symptom Severity and Personal Functioning Among Patients with Schizophrenia Discharged from Long-Term Hospital Care into the Community.* Community Mental Health Journal, 2001. **37**(2): p. 145-155.

654. Lavin, D. and P. Ryan, *Using quantative research to measure recovery outcomes and correlates.* Irish Journal of Psychological Medicine, 2012. **29**(3): p. 157-162.

655. Shinn, A., et al., *Auditory hallucinations in a cross-diagnostic sample of psychotic disorder patients: A descriptive, cross-sectional study.* Comprehensive Psychiatry, 2012. **53**(6): p. 718-726.

656. Barker, S., et al., *A Community Ability Scale for Chronically Mentally Ill Consumers: Part II. Applications.* Community Mental Health Journal, 1994. **30**(5): p. 459-472.

657. Prouteau, A., et al., *Cognitive predictors of psychosocial functioning outcome in schizophrenia: A follow-up study of subjects participating in a rehabiltiation program.* Schizophrenia Research, 2005. **77**: p. 343-353.

658. Zani, B., et al., *Statewide Replicaiton of Predictive Validation for the Multnomah Community Ability Scale.* Community Mental Health Journal, 1999. **35**(3): p. 223-229.

659. Milton, C., et al., *Continuity of care and health care costs among persons with severe mental illness.* Psychiatric Services, 2005. **56**(9): p. 1070-1076.

660. Dickerson, F.B., *Assessing clinical outcomes: the community functioning of persons with serious mental illness.* Psychiatric Services, 1997. **48**: p. 897-902.

661. Hopkins, M. and N. Ramsundar, *Which factors predict case management services and how do these services relate to client outcomes?* Psychiatric Rehabilitation Journal, 2006. **29**(3): p. 219-222.

662. McDevitt, J., et al., *A walking program for outpatients in psychiatric rehabilitation: Pilot study.* Biological Research for Nursing, 2005. **7**(2): p. 87-97.

663. Adair, C., et al., *Continuity of care and health outcomes amongst persons with severe mental illness.* Psychiatric Services, 2005. **56**(9): p. 1061-1069.

664. McColl, M., et al., *Validity and community utility of the Canadian Occupational Performance Measure.* Canadian Journal of Occupational Therapy, 2000. **67**(1): p. 22-29.

665. Stuber, C. and D. Nelson, *Convergent Validity of Three Occupational Self-Assessments.* Physical and Occupational Therapy in Geriatrics, 2010. **28**(1): p. 13-21.

666. Harper, K., C. Stalker, and G. Templeton, *The Use and Validity of the Canadian Occupational Performance Measure in a Posttraumatic Stress Program.* OTJR Occupation, Participation and Health, 2006. **26**(2): p. 45-55.

667. Samsonraj, R., M. Loughran, and J. Secker, *Evaluating outcomes of therapies offered by occupational therapists in adult mental health.* Journal of Mental Health 2012. **21**(6): p. 531-538.

668. McNulty, M. and A. Beplat, *The Validity of Using the Canadian Occupational Performance Measure with Older Adults with and without Depressive Symptoms.* Physical and Occupational Therapy in Geriatrics, 2008. **27**(1): p. 1-15.

669. Pan, A., L. Chung, and H. Hsin, G, *Reliability and validity of the Canadian Occupational Performance Measure for clients with psychiatric disorders in Taiwan.* Occupational Therapy International, 2003. **10**(4): p. 269-277.

670. Doig, E., et al., *Clinical utility of the combined use of the Canadian occupational performance measure and goal attainment scaling.* American Journal of Occupational Therapy, 2010. **64**(6): p. 904-914.
671. Bird, H.R., et al., *Further measures of the psychometric properties of the Children's Global Assessment Scale.* Archives of General Psychiatry, 1987. **44**(9): p. 821-824.
672. Green, B., et al., *The Children's Global Assessment Scale in clinical practice: an empirical evaluation.* J Am Acad Child Adolesc Psychiatry, 1994. **33**(8): p. 1158-64.
673. Sourander, A. and J. Piha, *Parent, teacher and clinical ratings on admission to child psychiatric inpatient treatment: A study of cross-informant correlations using the CBCL, Rutter scales and the CGAS.* Nordic journal of psychiatry, 1997. **51**(5): p. 365-370.
674. Sourander, A., et al., *Short-term child psychiatric inpatient treatment. Place of residence as one-year outcome measure.* European child and adolescent psychiatry, 1996. **5**(1): p. 38-43.
675. Sourander, A., H. Helenius, and J. Piha, *Child psychiatric short-term inpatient treatment: CGAS as follow-up measure.* Child Psychiatry and Human Development, 1996. **27**(2): p. 93-104.
676. Shaffer, D., et al., *A children's global assessment scale (CGAS).* Arch Gen Psychiatry, 1983. **40**(11): p. 1228-31.
677. Dyrborg, J., et al., *The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice: Substance and reliability as judged by intraclass correlations.* European child and adolescent psychiatry, 2000. **9**(3): p. 195-201.
678. Hodges, K. and J. Gust, *Measures of impairment for children and adolescents.* Journal of Mental Health Administration, 1995. **22**(4): p. 403-413.
679. Glascoe, F., *Parents' Evalutation of Developmental Status: How well do parents' concerns identify children with behavioral and emotional problems?* Clinical Pediatrics, 2003. **42**(2): p. 133-138.
680. Jones, D., P. Trudinger, and M. Crawford, *Intelligence and achievement of children referred following sexual abuse.* Journal of Paediatrics and Child Health, 2004. **40**(8): p. 455-460.
681. Glascoe, F., et al., *Can developmental-behavioral screening tests detect children with autism spectrum disorder>.* Clinical Pediatrics, 2007. **46**: p. 801-805.
682. Sices, L., et al., *PEDS and ASQ Develeopomental Screening Tests May Not Identify the Same Children.* Pediatrics, 2009. **124**(4): p. e640-e647.
683. Glascoe, F., *Do parents discuss concerns about children's development with health care providers?* Ambulatory Child Health, 1997. **2**: p. 349-356.
684. Kiing, J.S.H., et al., *Interpreting parents' concerns about their children's development with the Parents Evaluation of Developmental Status: culture matters.* Journal of Developmental & Behavioral Pediatrics, 2012. **33**(2): p. 179-83.
685. Radecki, L., et al., *Trends in the Use of Standardized Tools for Developmental Screening in Early Childhood: 2002-2009.* Pediatrics, 2011. **128**(1): p. 14-19.
686. Schonwald, A., et al., *Routine Developmental Screening Implemented in Urban Primary Care Settings: More Evidence of Feasbility and Effectiveness.* Pediatrics, 2009. **123**(2): p. 660-668.
687. Davies, S. and H. Feeney, *A pilot of the parent's evaluation of developmental status tool.* Community Practitioner, 2009. **82**(7): p. 29-31.
688. Kling, J., et al., *Interpreting Parents' Concerns About Their Children's Development With the Parents Evaluation of Developmental Status: Culture Matters.* Journal of Developmental and Behavioural Pediatrics, 2012. **33**(2): p. 179-183.
689. Durbeej, N., et al., *Validation of the Alcohol Use Disorders Identification Test and the Drug Use Disorders Identification Test in a Swedish sample of suspected offenders with signs of mental health problems: Results from the Mental Disorder, Substance Abuse and Crime study.* Journal of Substance Abuse Treatment, 2010. **39**(4): p. 364-377.
690. Nesvag, R., et al., *The use of screening instruments for detecting alcohol and other drug use disorders in first-episode psychosis.* Psychiatry Research, 2010. **177**: p. 228-234.

691. Sinadinovic, K., et al., *Internet-based assessment and self-monitoring of problematic alcohol and drug use.* Addictive Behaviors, 2010. **35**: p. 464-470.
692. Landheim, A., K. Bakken, and P. Vaglum, *Impact of comorbid psychiatric disorders on hte outcom of substance abusers: a six year prospective follow-up in two Norwegian countries.* BMC Psychiatry, 2006. **6**: p. 44.
693. Hallinan, P., et al., *Factor structure and validity of the alcohol use disorders indentification test (AUDIT) in a sample of mentally disordered offenders.* Journal of Forensic Psychiatry and Psychology, 2011. **22**(4): p. 586-602.
694. Shelvin, M. and G. Smith, *The factor strucure and concurrent validity of the alcohol use disorder identification test based on a nationally representative UK sample.* Alcohol and Alcoholism, 2007. **42**(6): p. 582-587.
695. Donovan, D., S. Doyle, and D. Kivlahan, *The factor structure of the alcohol use disorders identification test (AUDIT).* Journal of Studies on Alcohol and Drugs, 2007. **68**: p. 474-479.
696. Shields, A., et al., *An examination of the factor structure of the Alcohol Use DIsorders Identification Test in two high-risk samples.* Substance Use & Misuse, 2004. **39**: p. 1162-1182.
697. Skipsey, K., J. Burleson, and H. Kranzler, *Utility of the AUDIT for identification of hazardous or harmful drinking in drug-dependent patients.* Drug and Alcohol Dependence, 1997. **45**: p. 157-163.
698. Rist, F., A. Glockner-Rist, and R. Demmel, *The Alcohol Use Disorders Identification Test revisited: Establishing its structure using nonlinear factor analysis and identifying subgroups of respondents using latent class factor analysis.* Drug and Alcohol Dependence, 2009. **100**: p. 71-82.
699. O'Hare, T., et al., *Validating the Alcohol Use Disorders Identification Test with Persons Who Have a Serious Mental Illness.* Research on Social Work Practice, 2004. **14**(1): p. 36-42.
700. Gache, P., et al., *The alcohol use disorders identificaiton test (AUDIT) as a screening tool for excessive drinking in primary care: reliability and validity of a french version.* Alcoholism: Clinical and Experimental Research, 2005. **29**(11): p. 2001-2007.
701. Philpot, M., et al., *Screening for problem drinking in older people referred to a mental health service: A comparison of CAGE and AUDIT* Aging and Mental Health, 2003. **7**(3): p. 171-175.
702. Dybek, I., et al., *The reliability and validity of hte the Alcohol Used DIsoreder Identification (AUDIT) in a German general practice population sample.* Journal of Studies on Alcohol 2006. **67**(3): p. 473-481.
703. Moussas, G., et al., *The Alcohol Use Disorders Identification Test (AUDIT): Reliability and validity of the Greek version.* Annals of General Psychiatry, 2009. **8**: p. art.no.11.
704. Depp, C., et al., *Social competence and observer-rated social functioning in bipolar disorder.* Bipolar Disorders, 2010. **12**(8): p. 843-850.
705. Sabbag, S., et al., *Predictors of the accuracy of self assessment of everyday functioning in people with schizophrenia.* Schizophrenia Research, 2012. **137**: p. 190-195.
706. Gupta, M., et al., *Functional outcomes in schizophrenia: Understanding the competence-performance discrepancy.* Journal of Psychiatric Research, 2012. **46**: p. 205-211.
707. Harvey, P., et al., *Validating the Measurement of Real-World Functional Outcomes: Phase 1 Results of the VALERO study.* American Journal of Psychiatry, 2011. **168**: p. 1195-1201.
708. Sabbag, S., et al., *Assessing everyday functioning in schizophrenia: Not all informants seem equally informative.* Schizophrenia Research, 2011. **131**: p. 250- 255.
709. Hjarthag, F., et al., *Family burden and functional assessment in the Swedish CLIPS-study: do staff and relatives agree on individuals with psychotic disorders' functional status?* Social Psychiatry and Psychiatric Epidemiology, 2012. **47**(581-587): p. 581-587.
710. Bowie, C., et al., *Self-assessment of functional status in schizophrenia.* Journal of Psychiatric Research, 2007. **41**: p. 1012-1018.

711. Cardenas, k., et al., *When functional capacity and Real-World Functioning Converge: The Role of Self Efficacy.* Schizophrenia Bulletin, 2012. **39**(4): p. 908-916.

712. Dogan, S., et al., *Psychosocial approaches in outpatients with schizophrenia.* Psychiatric Rehabilitation Journal, 2004. **27**(3): p. 279-282.

713. Cyr, M., et al., *Assessment of Independent Living Skills for Psychotic Patients.* Journal of Nervous and Mental Disease, 1994. **182**(2): p. 91-97.

714. Aubin, G., et al., *Daily Activities, cognition and community functioning in persons with schizophrenia.* Schizophrenia Research, 2009. **107**: p. 313-318.

715. Perivoliotis, D., E. Granholm, and T. Patterson, *Psychosocial functioning on the Independent Living Skills Survey in older outpatients with schizophrenia.* Schizophrenia Research, 2004. **69**: p. 307-316.

716. Wallace, C., *Functional Assessment in Rehabilitation.* Schizophrenia Bulletin, 1986. **12**: p. 604-630.

717. Bystritsky, A., et al., *Social Functioning and Quality of Life Comparisons between Obsessive-Compulsive and Schizophrenic Disorders.* Depression and Anxiety, 2001. **14**: p. 214-218.

718. Liberman, R., et al., *Skills Training versus Psychosocial Occupational Therapy for Persons with Persistent Schizophrenia.* American Journal of Psychiatry, 1998. **155**: p. 1087-1091.

719. Granholm, E., et al., *A Randomized, Controlled Trial of Cognitive Behavioral Social Skills Training for Middle-Aged and Older Outpatients With Chronic Schizophrenia.* American Journal of Psychiatry, 2005. **162**(3): p. 520-529.

720. Menditto, A., et al., *Functional Assessment of Independent Living Skills.* Psychiatric Rehabilitation Skills, 1999. **3**(2): p. 200-219.

721. Mausbach, B., et al., *A Review of Instruments for Measuring Functional Recovery in Those Diagnosed With Psychosis.* Schizophrenia Bulletin, 2009. **35**(2): p. 307-318.

722. Curson, D., et al., *Four behavioural syndromes of schizophrenia: a replicaiton in a second inner-London epidemiological sample.* Schizophrenia Research, 1999. **37**: p. 165-176.

723. Lima, L., et al., *The measurement of social disablement and assessment of psychometric properties of the social behaviour schedule (SBS-BR) in 881 Brazilian long-stay psychiatric patients.* International Journal of Social Psychiatry, 2006. **52**(2): p. 101-109.

724. Harvey, C., et al., *Four Behavioural Syndromes of Schizophrenia.* British Journal of Psychiatry, 1996. **168**(5): p. 562-570.

725. Sturt, E. and T. Wykes, *Assessment schedules for chronic psychiatric problems.* Psychological Medicine, 1987. **17**(2): p. 485-493.

726. Poyurovsky, M., et al., *Obsessive-compulsive disorder in hospitalized patients with chronic schizophrenia.* Psychiatry Research, 2001. **102**: p. 49-57.

727. Allan, S. and I. McGonagle, *A comparison of HoNOS with the Social Behaviour Schedule in three settings.* Journal of Mental Health, 1997. **6**(2): p. 117-124.

728. Salvador-Carulla, L., et al., *A reliablity study of the Spanish version of the Social Behaviour Schedule (SBS) in a population of adults with learning disabilities.* Journal of Intellectual Disability Research, 1998. **42**: p. 22-28.

729. Grinshpoon, A., Y. Naisberg, and A. Weizman, *A Six-Month Outcome of Long-Stay Inpatients Resettled in a Hostel.* Psychiatric Rehabilitation Journal, 2006. **30**(2): p. 89-95.

730. Wykes, T., et al., *What are the effects of group cognitive behaviour therapy for voices? A randomised controlled trial.* Schizophrenia Research, 2005. **77**: p. 201-210.

731. Macpherson, R., et al., *Outcome of clinical risk management in the Gloucester rehabilitation service.* Psychiatric Bulletin, 2002. **26**: p. 449-452.

732. Hodges, K. and M. Wong, *Psychometric Characteristics of a Multidimensional Measure to Assess Impairment: The Child and Adolescent Functional Assessment Scale.* Journal of Child and Family Studies, 1996. **5**(4): p. 445-467.

733. Rosenblatt, A. and J. Rosenblatt, *Assessing the Effectiveness of Care for Youth with Severe Emotional Disturbances: Is There Agreeement Between Popular Outcome Measures?* Journal of Behavioral Health Services & Research, 2002. **29**(3): p. 259-273.
734. Francis, S., C. Ebesutani, and B. Chorpita, *Differences in Levels of Functional Impairment and Rates of Serious Emotional Disturbance Between Youth With Internalizing and Externalizing Disorders When Using the CAFAS or GAF to assess Functional Impairment.* Journal of Emotional and Behavioural Disorders, 2012. **20**(4): p. 226-240.
735. Quist, R. and D. Matshazi, *The Child and Adolescent Functional Assessment Scale (CAFAS): A Dynamic Predictor of Juvenile Recidivism.* Adolescence, 2000. **35**(137): p. 181-192.
736. Hodges, K., A. Doucette-Gates, and C.-S. Kim, *Predicting Service Utilization with the Child and Adolescent Functional Assessment Scale in a Sample of Youths with Serious Emotional Disturbance Served by Center for Mental Health Services - Funded Demonstrations.* Journal of Behavioral Health Services & Research, 2000. **27**(1): p. 47-59.
737. Roy, K., et al., *Measuring Treatment Outcome for Children with Serious Emotional Disturbances: Discriminant Validity and Clinical Significance of the Child and Adolescent Functioning Assessment Scale.* Journal of Child and Family Studies, 2008. **17**: p. 232-240.
738. Hodges, K. and M. Wong, *Use of the Child and Adolescent Functional Assessment Scale to predict service utilization and cost.* Journal of Health Administration, 1997. **24**(278-290).
739. Hodges, K. and J. Wotring, *The Role of Monitoring Outcomes in Initiating Implementation of Evidence-Based Treatments at the State Level.* Psychiatric Services, 2004. **55**(4): p. 396-400.
740. Hodges, K., M. Wong, and M. Latessa, *Use of the Child and Adolescent Functional Assessment Scale (CAFAS) as an Outcome Measure in Clinical Settings.* Journal of Behavioral Health Services & Research, 1998. **25**(3): p. 325-336.
741. Boydell, K., et al., *A Feasibility Study to Assess Service Providers' Perspective Regarding the Use of the Child and Adolescent Functional Assessment Scale in Ontario.* Journal of Behavioral Health Services & Research, 2005. **32**(1): p. 105-109.
742. Shaffer, D., C. Lucas, and J. Richters, eds. *Diagnostic Assessment in Child and Adolescent Psychopathology*. 1999, The Guilford Press: New York.
743. Bird, H.R., et al., *The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents.* International Journal of Methods in Psychiatric Research, 1993. **3**(3): p. 167-176.
744. Bird, H.R., et al., *Global measures of impairment for epidemiologic and clinical use with children and adolescents.* International Journal of Methods in Psychiatric Research, 1996. **6**(4): p. 295-307.
745. Harris, E., R. Canning, and K. Kelleher, *A Comparison of Measures of Adjustment, Symptoms, and Impairment among Children with Chronic Medical Conditions.* Journal of the American Academy of Child & Adolescent Psychiatry, 1996. **35**(8): p. 1025-1032.
746. Hamilton, J. and J. Bridge, *Outcome at 6 months for 50 Adolescents With Major Depression Treated in a Health Maintenance Organization.* Journal of the American Academy of Child & Adolescent Psychiatry, 1999. **38**(11): p. 1340-1346.
747. Goodman, S.H., et al., *Major depression and dysthymia in children and adolescents: discriminant validity and differential consequences in a community sample.* Journal of the American Academy of Child and Adolescent Psychiatry, 2000. **39**(6): p. 761-70.
748. Fiks, A., et al., *Shared Decision Making and Behavioural Impariment: a national study among children with special health care needs.* BMC Pediatrics, 2012. **12**(153).
749. Flisher, A., et al., *Psychosocial Characteristics of Physically Abused Children and Adolescents.* Journal of the American Academy of Child & Adolescent Psychiatry, 1997. **36**(1): p. 123-131.
750. Gould, M., et al., *Seeking Help From the Internet During Adolescence.* Journal of the American Academy of Child & Adolescent Psychiatry, 2002. **41**(10): p. 1182-1189.
751. Bastiaens, L. and C. Dello Stritto, *Quality of Life of Children and Adolescents During Psychiatric Treatment in a Community Mental Health Setting.* Psychiatry, 2005. **2**(8).

752. Huxley, P., et al., *Development of a social inclusion index to capture subjective and objective life domains (Phase II): Psychometric development study.* Health Technology Assessment, 2012. **16**(1): p. 1-248.

753. Marino-Francis, F. and A. Worrall-Davies, *Development and validation of a social inclusion questionnaire to evaluate the impact of attending a moderised mental health day service.* Mental Health Review Journal, 2010. **15**(1): p. 37-47.

754. Berry, H., B. Rodgers, and K. Dear, *Preliminary development and validation of an Australian community participation questionnaire: Types of participation and associations with distress in a coastal community.* Social Science and Medicine, 2007. **64**(8): p. 1719-1737.

755. Dorer, G., P. Harries, and L. Marston, *Measuring social inclusion: a staff survey of mental health service users' participation in community occupations.* British Journal of Occupational Therapy, 2009. **72**(12): p. 520-530.

756. Ramon, S., et al., *Towards social inclusion through lifelong learning in mental health: Analysis of change in the lives of the EMILIA Project service users.* International Journal of Social Psychiatry, 2009. **57**(3): p. 211-223.

757. Secker, J., et al., *Development of a measure of social inclusion for arts and mental health project participants.* Journal of Mental Health, 2009. **18**(1): p. 65-72.

758. Hacking, S. and P. Bates, *The Inclusion Web: A tool for person-centred planning and service evaluation.* Mental Health Review Journal, 2008. **13**(2): p. 4-15.

759. Lloyd, C., et al., *Reliability of a composite measure of social inclusion for people with psychiatric disabilities.* Australian Occupational Therapy Journal, 2008. **55**: p. 47-56.

760. Stickley, T. and R. Shaw, *Evaluating social inclusion.* Mental Health Practice, 2006. **9**(10): p. 14-20.

761. Stewart, G., et al., *A brief measure of vocational activity and community participation: Development and reliability of the Activity and Participation Questionnaire.* Australian and New Zealand Journal of Psychiatry, 2010. **44**(3): p. 258-266.

762. Harris, M., et al., *Reliability and validity of a measure of role functioning among people with psychiatric disabilities.* Australian Occupational Therapy Journal, 2011. **58**(3): p. 203-208.

763. Waghorn, G., D. Chant, and R. King, *Work-related self-efficacy among community residents with psychiatric disabilities.* Psychiatric rehabilitation journal, 2005. **29**(2): p. 105-113.

764. Webber, M. and P. Huxley, *Measuring access to social capital: The validity and reliability of the Resource Generator-UK and its association with common mental disorder.* Social Science and Medicine, 2007. **65**(3): p. 481-492.

765. Core System Group, *CORE System (Information Management) Handbook*. 1998, Core System Group, University of Leeds: Leeds.

766. Schafer, T., *Empowerment: Towards a participatory model for the evaluation of the empowering therapeutic environment.* Mental Health Care, 2000. **3**(7): p. 233-237.

767. Test MA, et al., *Construct validity of a measure of subjective satisfaction with life of adults with serious mental illness.* Psychiatric Services, 2005. **56**(3): p. 292-300.

768. Luk AL, *Investigating the long-term effects of a psychiatric rehabilitation programme for persons with serious mental illness in the community: a follow-up study.* Journal of Clinical Nursing, 2011. **20**: p. 2712-2720.

769. Walters LH and Klein AE, *A cross-validated investigation of the Crumbaugh Purpose-in-Life Test.* Educational and Psychological Measurement, 1980. **40**: p. 1065-1071.

770. Lehman AF, *A quality of life interview for the chronically mentally ill.* Evaluation and Program Planning, 1988. **11**: p. 51-62.

771. Carlson JA, et al., *Evaluating a measure of social health derived from two mental health recovery measures: The California Quality of Life (CA-QOL) and Mental Health Statistics Improvement Program consumer survey (MHSIP).* Community Mental Health Journal, 2011. **47**: p. 454-462.

772. Greenley JR, Greenberg JS, and Brown R, *Measuring Quality of Life: a new and practical survey instrument.* Social Work, 1997. **42**(3): p. 244-254.
773. Dazord A, et al., *Quality of Life Assessment in Psychiatry: The Subjective Quality of Life Profile (SQLP) - First Results of a New Instrument.* Community Mental Health Journal, 1998. **34**(5): p. 525-535.
774. Ware, J., *The SF-36 Survey*, in *Quality of llife and pharmacoeconomics in clinical trials*, B. Spilker, Editor. 1990, Lippincott-Raven Publishers: Philadelphia. p. 337-345.
775. Law CW, et al., *Impact of untreated psychosis on quality of life in patients with first-episode scizophrenia.* Quality of Life Research, 2005. **14**: p. 1803-1811.
776. Swan A, Watson HJ, and Nathan PR, *Quality of life in depression: an important outcome measure in an outpatient cognitive-behavioural therapy group programme?* Clinical Psychology and Psychotherapy, 2009. **16**: p. 485-496.
777. Murphy N and Cutts H, *Can the introduction of a quality of life tool affect individual professional practice and the quality of care planning in a community mental health team?* Journal of Psychiatric and Mental Health Nursing, 2009. **16**: p. 941-946.
778. Lancon C, et al., *Validation of the brief Quality of Life Interview (QoLI): I. Internal structure.* Encephale, 2000. **26**(4): p. 11-16.
779. Ruggeri M, et al., *Determinants of subjective quality of life in patients attending community-based mental health services. The South-Verona Outcome Project 5.* Acta Psychiatrica Scandinavica, 2002. **105**: p. 131-140.
780. Ferrans, C., Powers, MJ., *Quality of life index: development and psychometric properties.* Advances in Nursing Science, 1985. **8**(1): p. 15-24.
781. Cohen CI, et al., *Assessing quality of life in older persons with schizophrenia.* The American Journal of Geriatric Psychiatry, 2003. **11**(6): p. 658-666.
782. Baker, F., Intagliata, J., *Quality of life in the evaluation of community support systems.* Evaluation and Program Planning, 1982. **5**: p. 69-79.
783. Heinrichs DW, Hanlon TE, and J. Carpenter WT, *The Quality of Life Scale: An Instrument for Rating the Schizophrenic Deficit Syndrome.* Schizophrenia Bulletin, 1984. **10**(3): p. 388-398.
784. Diaz P, et al., *An evaluation of psychometric properties of the client's questionnaire of the Wisconsin Quality of Life Index - Canadian version (CaW-QLI).* Quality of Life Research, 1999. **8**: p. 509-514.
785. Lubin B and Van Whitlock R, *Psychometric properties of the Brief Life Satisfaction Scales.* Journal of Clinical Psychology, 2004. **60**(1): p. 11-27.
786. Diener E, et al., *The Satisfaction With Life Scale.* Journal of Personality Assessment, 1985. **49**(1): p. 71-75.
787. International Wellbeing Group. *Personal Wellbeing Index: 5th Edition*. 2013; Available from: http://www.deakin.edu.au/research/acqol/instruments/wellbeing-index/index.php.
788. Council on Quality & Leadership. *Personal outcome measures*. 2000; Available from: http://www.thecouncil.org/Personal_Outcome_Measures.aspx.
789. McCormack, B. and M. Farrell, *Translating quality of life into service action: use of personal outcome measures in the Republic of Ireland.* British Journal of Learning Disabilities, 2009. **37**(4): p. 300-307.
790. Lee KK, et al., *Longitudinal invariance of the Satisfaction with Life Scale for individuals with schizophrenia.* Research on Social Work Practice, 2010. **20**(2): p. 234-241.
791. Stevanovic D, *Quality of Life Enjoyment and Satisfaction Questionnaire - short form for quality of life assessments in clinical practice: a psychometric study.* Journal of Psychiatric and Mental Health Nursing, 2011. **18**: p. 744-750.
792. Endicott, J., et al., *Quality of Life Enjoyment and Satisfaction Questionnaire: A new measure.* Psychopharmacology Bulletin, 1993. **29**(2): p. 321 - 326.
793. Priebe S, et al., *Application and results of the Manchester Short Assessment of Quality of Life (MANSA).* International Journal of Social Psychiatry, 1999. **45**(1): p. 7-12.

794. Oliver JPJ, et al., *Measuring the quality of life of severly mentally ill people using the Lancashire Quality of Life Profile.* Social Psychiatry & Psychiatric Epidemiology, 1997. **32**: p. 76-83.

795. Ferrans, C.E. and M.J. Powers, *Quality of life index: development and psychometric properties.* Advances in Nursing Science, 1985. **8**(1): p. 15-24.

796. Ferrans, C. and M. Powers, *Psychometric Assessment of the Quality of Life Index.* Research in Nursing and Health, 1992. **15**: p. 29-38.

797. Ferrans, C., *Development of a quality of life index for patients with cancer.* Oncology Nursing Forum, 1990. **17**(3): p. 15-19.

798. Carlson, J., Ochoa, S., Haro, JM., Escartin, G., Ahuir, M., Gutierrez-Zotes, A., Salamero, M., Valero, J., Canizares, S., Bernado, M., Canete, J., Gallo, P., *Adaptation and validation of the quality-of-life scale Satisfaction with the Satisfaction wit Life Domains Scale by Baker and Intagliata.* Comprehensive Psychiatry, 2009. **50**: p. 76-80.

799. Ritsner, M., et al., *Condensed version of the Quality of Life Scale for schizophrenia for use in outcome studies.* Psychiatry Research, 2005. **135**(1): p. 65-75.

800. Cramer, J.A., et al., *Quality of Life in Schizophrenia: A Comparison of Instruments.* Schizophrenia Bulletin, 2000. **26**(3): p. 659-666.

801. Meltzer, H., et al., *Effects of six months of clozapine treatment on the quality of life of chronic schizophrenic patients.* Hospital and Community Psychiatry, 1990. **41**: p. 892-897.

802. Pavot, W. and E. Diener, *Review of the satisfaction with life scale* Psychological Assessment, 1993. **5**(2): p. 164-172.

803. Cummins, R., et al., *Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index.* Social Indicators Research, 2003. **64**(2): p. 159-190.

804. Malins, G., K. Oakley, and R. Doyle, *Building consumers in to service evaluation: development of the MH-CoPES Framework in New South Wales.* Australasian Psychiatry: Bulletin Of Royal Australian And New Zealand College Of Psychiatrists, 2011. **19**(4): p. 360-363.

805. Jenkinson, C., et al., *Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care.* Quality & Safety in Health Care, 2002. **11**(4): p. 335-9.

806. Jenkinson, C., A. Coulter, and S. Bruster, *The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries.* International Journal of Quality in Health Care, 2002. **14**: p. 353-358.

807. Victorian Government. Mental Health Services. *Safety and Quality in Victoria's Specialist Public Mental Health System: C&C Experience.* 2008; Available from: http://www.health.vic.gov.au/mentalhealth/quality/consumer.htm.

808. Oltedal, S., A. Garratt, and J.O. Johannessen, *Psychiatric outpatients' experiences with specialized health care delivery. A Norwegian national survey.* Journal of Mental Health, 2007. **16**(2): p. 271-279.

809. Jerrell, J., *Psychometrics of the MHSIP Adult Consumer Survey.* The Journal of Behavioral Health Services & Research, 2006. **33**(4): p. 483-488.

810. Gianfrancesco, P., *Tools for reviewing Australian mental health services, Volume 2: Review Guide*. 1998, Australian Health Ministers Advisory Committee National Mental Health Working Group, Commonwealth Department of Health and Family Services.

811. Helfinger, C., *Measuring service system coordination in managed mental health care for children and youth.* Evaluation and Program Planning, 1996. **19**(2): p. 155-163.

812. Wickizer, T., et al., *Activating communities for health promotion: A process evaluation method.* American Journal of Public Health, 1993. **83**(4): p. 561-567.

813. Browne, G., Roberts, J., Gafni, A., Byrne, C., Kertyzia, J., Loney, P., *Conceptualizing and validating the human services integration measure.* International Journal of Integrated Care, 2004. **4**(19 May): p. 1-9.

814. Oakley, K., G. Malins, and R. Doyle, *The MH-CoPES Framework and Questionnaires ready for statewide implementation: Final Report of the MH-CoPES Stage 2 Project: APPENDICES*. 2011, NSW Consumer Advisory Group – Mental Health Inc.: Sydney.
815. Oakley, K., G. Malins, and R. Doyle, *The MH-CoPES Framework and Questionnaires ready for statewide implementation Final Report of MH-CoPES Stage 2 Project*. 2011, NSW Consumer Advisory Group – Mental Health Inc.: Sydney.
816. Garratt, A., et al., *The Psychiatric Out-Patient Experiences Questionnaire (POPEQ): data quality, reliability and validity in patients attending 90 Norwegian clinics.* Nordic Journal Of Psychiatry, 2006. **60**(2): p. 89-96.
817. Bjørngaard, J.H., et al., *Patients' Experiences and Clinicians' Ratings of the Quality of Outpatient Teams in Psychiatric Care Units in Norway.* Psychiatric Services, 2007. **58**(8): p. 1102-7.
818. Eisen, S., et al., *Toward a national consumer survey: Evalution of the CABHS and MHSIP instruments.* The Journal of Behavioral Health Services and Research, 2001. **28**(3): p. 347-369.
819. Jerrell, J.M., *Psychometrics of the MHSIP Adult Consumer Survey.* Journal of behavioral health services and research, 2006. **33**(4): p. 483-488.
820. Ganju V, Wackwitz J, and Trabin T, *The Mental Health Statistics Improvement Program (MHSIP) Consumer survey*. 1998: Submitted to the Committee on Performance Measurement, NCQA.
821. Lakeman, R., *Family and carer participation in mental health care: perspectives of consumers and carers in hospital and home care settings.* Journal of Psychiatric & Mental Health Nursing, 2008. **15**(3): p. 203-11.
822. Queensland Government, *Disability and mental health service users and carers satisfaction survey* 2009, Department of Communities: Brisbane
823. Fox D, Holder J, and Netten A, *Personal Social Services Survey of Adult Carers in England - 2009-10: Survey Development Project, Technical Report.* 2010, Personal Social Services Research Unit: United Kingdom.
824. The NHS Information Centre Social Care, *Personal Social Services Survey of Adult Carers in England - 2009-20*. 2010, The Health and Social Care Information Centre: United Kingdom.
825. Mental Health Council of Australia, *Adversity to Advocacy: The Lives and Hopes of Mental Health Carers*. 2010, Mental Health Council of Australia: Canberra
826. Weiss, E., Anderson R, Lasker R., *Findings from the national study of partnership functioning: report to the parterships that participated.* 2003.
827. Hebert, R., Veil, A., Raiche, M., Dubuc, N., Tousignant, M., The Prisma-Estrie Group,, *Evaluation of the implementation of PRISMA, a coodination-type integrated service delivery system for fraill older people in Quebec.* Journal of Integrated Care, 2008. **16**(6): p. 4-14.
828. Andresen, R., P. Caputi, and L.G. Oades, *Interrater reliability of the Camberwell Assessment of Need Short Appraisal Schedule.* Australian & New Zealand Journal of Psychiatry, 2000. **34**(5): p. 856-61.
829. Trauer, T., G. Tobias, and Slade M, *Development and evaluation of a patient-rated version of the Camberwell Assessment of Need Short Appraisal Scale (CANSAS-P).* Community Mental Health Journal, 2008. **44**(2): p. 113-124.
830. Clarke, S., et al., *The role of symptom distress and goal attainment in promoting aspects of psychological recovery for consumers with enduring mental illness.* Journal of Mental Health, 2009. **18**(5): p. 389-397.
831. Oades, L., Crowe, T, Deane, FP, in *The Oxford Handbook of Happiness*, S. David, I. Boniwell, and A.C. Ayers, Editors. 2013, Oxford University Press: Oxford, UK. p. 1050-1066.
832. Clarke, S., et al., *Collaborative goal technology: Theory and Practice.* Psychiatric Rehabilitation Journal, 2006. **30**: p. 129-136.

833. Wing, J.K., et al., *Health of the Nation Outcome Scales (HoNOS). Research and development.* British Journal of Psychiatry, 1998. **172**: p. 11-8.
834. Wing, J.K., P. Lelliott, and A.S. Beevor, *Progress on HoNOS.* British Journal of Psychiatry, 2000. **176**: p. 392-393.
835. Wing, J., R. Curtis, and A. Beevor, *Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet.* British J Psychiatry, 1999. **174**: p. 432 - 434.
836. Hunter, J., I. Higginson, and E. Garralda, *Sytematic literature review: Outcome measures for child and adolescent mental health services.* Journal of public health medicine, 1996. **18**(2): p. 197-206.
837. Jensen, P.S., K. Hoagwood, and T. Petti, *Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model.* Journal of the American Academy of Child and Adolescent Psychiatry, 1996. **35**: p. 1064-1077.
838. Gowers, S.G., et al., *Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for children and Adolescents (HoNOSCA).* British Journal of Psychiatry, 1999. **174**: p. 413-6.
839. Gowers, S., et al., *The Health of the Nation Outcome Scales for Child and Adolescent Mental Health (HoNOSCA).* Child Psychology and Psychiatry Review, 2000. **5**(2): p. 50-56.
840. Gowers, S.G., et al., *Health of the Nation Outcome Scales: Trainer's Guide*. 1997, The University of Liverpool: Liverpool.
841. Gowers, S.G., et al., *Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). Glossary for HoNOSCA score sheet.* British Journal of Psychiatry, 1999. **174**: p. 428-31.
842. Shergill, S.S., et al., *The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly.* Journal of Mental Health (UK), 1999. **8**(5): p. 511-521.
843. Ashaye, K., et al., *Do the Health of the Nation Outcome Scales predict outcome in the elderly mentally ill? A 1-year follow-up study.* Journal of Mental Health (UK), 1999. **8**(6): p. 615-620.
844. Burns, A., et al., *Health of the Nation Outcome Scales for Elderly People (HoNOS 65+) - Glossary for HoNOS 65+score sheet.* British Journal of Psychiatry, 1999. **174**: p. 435-438.
845. Miller, S.D., et al., *The partners for change outcome management system.* Journal of Clinical Psychology, 2005. **61**(2): p. 199-208.
846. Campbell, A. and S. Hemsley, *Outcome Rating Scale and Session Rating Scale in psychological practice: clinical utility of ultra-brief measures.* Clinical Psychologist, 2009. **13**(1): p. 1-9.
847. Timimi, S., et al., *Outcome Orientated Child and Adolescent Mental Health Services (OO-CAMHS): A whole service model.* Clinical Child Psychology and Psychiatry, 2013. **18**(2): p. 169-184.
848. Miller, S., D., et al., *Using Formal Client Feedback to Improve Retention and Outcome: Making Ongoing, Real-time Assessment Feasible.* Journal of Brief Therapy, 2006. **5**(1): p. 2-23.
849. Miller, S.D., et al., *The Outcome Rating Scale: A Preliminary Study of the Reliability, Validity,and Feasibility of a Brief Visual Analog Measure.* Journal of Brief Therapy, 2003. **2**(2): p. 91 -100.
850. Bringhurst, D.L., et al., *The Reliability and Validity of the Outcome Rating Scale: A Replication Study of a Brief Clinical Measure.* The Journal of Brief Therapy 2006. **5**(1): p. 23 - 30.
851. Reininghaus U, et al. *The validity of patient- and clinician-rated measures of needs and the therapeutic relationship in psychosis: A pooled anlaysis*. Psychiatry Research, 2013. **Epub ahead of print**, DOI: http://dx.doi.org/10.1016/j.psychres.2013.01.013i.
852. Priebe S and Gruyters T, *Patients' assessment of treatment prediting outcome.* Journal of Nervous and Mental Disease, 1995. **181**: p. 552-557.
853. Phelan M, et al., *The Camberwell Assessment of Need: the validity and reliability of an instrument to assess the needs of people with severe mental illness.* British Journal of Psychiatry, 1995. **167**: p. 589-595.

854. Andresen R, Caputi P, and L. Oades, *Interrater Reliability of the Camberwell Assessment of Need Short Appraisal Schedule.* Australian and New Zealand Journal of Psychiatry, 2000. **34**: p. 856-861.
855. Hansson L, Bjorkman T, and Svensson B, *The assessment of needs in psychiatric patients: interrater reliability of the Swedish version of the Camberwell Assessment of Needs instrument and results from a cross-sectional study.* Acta Psychiatr Scand, 1995. **92**: p. 285-293.
856. Orrell, M., et al., *Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community.* British Journal of Psychiatry, 1999. **174**: p. 409-12.
857. McClelland, R., et al., *Validation of an outcome scale for use in adult psychiatric practice.* Quality in Health Care, 2000. **9**(2): p. 98-105.
858. Pilon, D. and M. Ragins, *Milestones of Recovery Scale (MORS)*. n.d.
859. Browne, S., M. Doran, and S. McGauran, *Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population.* Irish Journal of Psychological Medicine, 2000. **17**(1): p. 17-19.
860. Amin, S., et al., *Evaluating the Health of the Nation Outcome Scales. Reliability and validity in a three-year follow-up of first-onset psychosis.* British Journal of Psychiatry, 1999. **174**: p. 399-403.
861. Rees, A., A. Richards, and D.A. Shapiro, *Utility of the HoNOS in measuring change in a community mental health care population.* Journal of Mental Health, 2004. **13**(3): p. 295-304.
862. Bebbington, P., et al., *Validation of the Health of the Nation Outcome Scales.* British Journal of Psychiatry, 1999. **174**: p. 389-94.
863. Hope, J.D., T. Trauer, and N.A. Keks, *Reliability, validity and utility of the Health of the Nation Outcomes Scale (HoNOS) in Australian adult psychiatric services.* Schizophrenia Research, 1998. **29**(1,2): p. 9-10.
864. Trauer, T., *The subscale structure of the Health of the Nation Outcome Scales (HoNOS).* Journal of Mental Health, 1999. **8**(5): p. 499-509.
865. Page, A.C., G.R. Hooke, and E.M. Rutherford, *Measuring mental health outcomes in a private psychiatric clinic: Health of the Nation Outcome Scales and Medical Outcomes Short Form SF-36.* Aust N Z J Psychiatry, 2001. **35**(3): p. 377-81.
866. Trauer, T., *Comment.* Australian and New Zealand Journal of Psychiatry, 2000. **34**(3): p. 520-521.
867. Brooks, R., *The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures.* Australian and New Zealand Journal of Psychiatry, 2000. **34**(3): p. 504-11; discussion 520-1.
868. Trauer, T., et al., *Health of the Nation Outcome Scales. Results of the Victorian field trial.* British Journal of Psychiatry, 1999. **174**: p. 380-8.
869. Bonsack, C., F. Borgeat, and A. Lesage, *Measuring patients' problems severity and outcomes in a psychiatric sector: A field study with the French version of the Health of Nation Outcome Scales (HoNOS-F)/Mesurer la severite des problemes des patients et leur evolution dans un secteur psychiatrique: Une etude sur le terrain du Health of Nation Outcome Scales en francais: (HoNOS-F).* Annales Medico-Psychologiques, 2002. **160**(7): p. 483-488.
870. Gallagher, J. and M. Teesson, *Measuring disability, need and outcome in Australian community mental health services.* Aust N Z J Psychiatry, 2000. **34**(5): p. 850-5.
871. Ashaye, O., G. Mathew, and M. Dhadphale, *A comparison of older longstay psychiatric and learning disability inpatients using the Health of the Nation Outcome Scales.* Int J Geriatr Psychiatry, 1997. **12**(5): p. 548-52.
872. Sharma, V.K., G. Wilkinson, and S. Fear, *Health of the Nation Outcome Scales: A case study in general psychiatry.* British Journal of Psychiatry, 1999. **174**: p. 395-8.
873. Stafrace, S., *Doubts about HoNOS.* Aust N Z J Psychiatry, 2002. **36**(2): p. 270.

874. Stein, G.S., *Usefulness of the Health of the Nation Outcome Scales.* Br J Psychiatry, 1999. **174**: p. 375-7.
875. Gilbody, S.M., A.O. House, and T.A. Sheldon, *Psychiatrists in the UK do not use outcomes measures. National survey.* Br J Psychiatry, 2002. **180**: p. 101-3.
876. Yates, P., M.E. Garralda, and I. Higginson, *Paddington Complexity Scale and Health of the Nation Outcome Scales for Children and Adolescents.* British Journal of Psychiatry, 1999. **174**: p. 417-23.
877. Bilenberg, N., *Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA): Results of a Danish field trial.* European child and adolescent psychiatry, 2003. **12**(6): p. 298-302.
878. Gowers, S., et al., *Use of a routine, self-report outcome measure (HoNOSCA-SR) in two adolescent mental services.* British Journal of Psychiatry, 2002. **180**(3): p. 266-269.
879. Garralda, M.E., P. Yates, and I. Higginson, *Child and adolescent mental health service use. HoNOSCA as an outcome measure.* British Journal of Psychiatry, 2000. **177**: p. 52-8.
880. Brann, P., *Routine Outcome Measurement in Child/Adolescent Mental Health: HoNOSCA - Valid Enough? Feasible Enough? PhD Thesis*. 2006, Department of Psychology, Monash University: Melbourne.
881. Brann, P., G. Coleman, and E. Luk, *Routine outcome measurement in a child and adolescent mental health service: An evaluation of HoNOSCA.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(3): p. 370-376.
882. Brouwer, W.B.F., et al., *The CarerQol instrument: A new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations.* Quality of Life Research, 2006. **15**(6): p. 1005-1021.
883. Garralda, E. and P. Yates, *HoNOSCA: Uses and limitations.* Child Psychology and Psychiatry Review, 2000. **5**(3): p. 131-132.
884. Manderson, J. and N. McCune, *The use of HoNOSCA in a child and adolescent mental health service.* Irish Journal of Psychological Medicine, 2003. **20**(2): p. 52-55.
885. Mozley, C.G., et al., *'Not knowing where I am doesn't mean I don't know what I like': Cognitive impairment and quality of life responses in elderly people.* International Journal of Geriatric Psychiatry, 1999. **14**(9): p. 776-83.
886. Spear, J., et al., *Does the HoNOS 65+ meet the criteria for a clinical outcome indicator for mental health services for older people?* Int J Geriatr Psychiatry, 2002. **17**(3): p. 226-30.
887. Bagley, H., et al., *Recognition of depression by staff in nursing and residential homes.* Journal of Clinical Nursing, 2000. **9**(3): p. 445-50.
888. Allen, L., et al., *Experience and application of HoNOS65+.* Psychiatric Bulletin, 1999. **23**(4): p. 203-206.
889. Reilly, S., et al., *The use of assessment scales in Old Age Psychiatry Services in England and Northern Ireland.* Aging Mental Health, 2004. **8**(3): p. 249-255.
890. Miller, S., et al., *The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of brief visual analog measure.* Journal of Brief Therapy, 2003. **2**(2): p. 91-100.
891. Bringhurst Major, D., et al., *The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure.* Journal of Brief Therapy, 2006. **5**(1): p. 23-30.
892. Antony, M., et al., *Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress scales in clinical groups and a community sample.* Psychological Assessment, 1998. **10**: p. 176-181.
893. Burckhardt, S. and K. Anderson, *The quality of life scale (QOLS): Reliability, validity and utilisation.* Health and Quality of Life Outcomes, 2003. **1**(60).
894. Horvath, A. and L. Greenberg, *Development and validation of the Working Alliance Inventory.* Journal of Counselling Psychology, 1989. **36**: p. 223-233.
895. Sundet, R., *Therapist perspectives on the use of feedback on process and outcome: patient-focused research in practice.* Canadian psychology, 2012. **53**(2): p. 122-130.