Author/s:
Hagiliassis, N;Koritsas, S;Cuzzillo, C

Title:
Measuring Goal Realization Associated With Disability Services and Supports: Initial Evidence for a New Tool

Date:
2020-03-01

Citation:
Hagiliassis, N., Koritsas, S. & Cuzzillo, C. (2020). Measuring Goal Realization Associated With Disability Services and Supports: Initial Evidence for a New Tool. Journal of Policy and Practice in Intellectual Disabilities, 17 (1), pp.4-12. https://doi.org/10.1111/jppi.12291.

Hagiliassis Nick (Orcid ID: 0000-0002-7075-2553)

JPPID-18-0011 "Measuring Goal Realisation associated with Disability Services and Supports: Initial Evidence for a new tool."

Journal of Policy and Practice in Intellectual Disabilities.

Response to reviewers October 2018

Reviewer 1	
1. More details about the changes made to create the SF-MOSS are needed.	**We have added the following to the manuscript:** P. 18. The original version of the tool, the Measuring Outcomes in Services and Supports (MOSS) was developed in a disability service provider setting, to measure the extent to which a person's goals are realised in the course of service delivery. It was designed along the following criteria (1) evaluates outcomes (the impact of a service or support), rather than outputs (the services delivered by an agency) or process (the way in which services are provided) (2) is flexible to goals covering a range of life areas (e.g., wellbeing, social, economic), (3) is versatile across areas of services and supports (e.g., therapy, supported accommodation, community inclusion), (4) can be accessed by people with intellectual disability through its design and administration characteristics (e.g., using simple visual scaling), (5) reduces reliance on proxy reporting but where required, ensures this occurs from the person's perspective and potential sources of bias are understood, (6) has demonstrable evidence, (7) is achievable for service providers to administer in day-to-day practice, (8) allows for aggregating up for generating comparative data, (9) uses 'performance' and 'satisfaction (with performance)' as indices of measurement, and (10) enquires about contextual factors (enablers and barriers) that influence personal outcomes. p.18 An evaluation of the original tool revealed a number of shortcomings (Quilliam, Wilson, Hagiliassis & Nicola-Richmond, 2010). From goals extracted from 73 completed forms, the far majority (70%) of information was provided by proxies (i.e. service providers, family members) - only

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jppi.12291

	28% of responses to the MOSS were informed by the person with a disability independently. Originally planned for use in therapy services, it was felt to be therapy-centric and to include non-essential data fields (e.g., questions about legal issues). Resulting changes were improvements in easy-language wording, recasting the tool for use across any service, removing the need to provide superfluous information, simplification of visual scaling, and improvements in format/layout. The resultant version, the SF-MOSS (Hagiliassis, Nicola-Richmond, Wilson & Mackay, 2014), is the tool that is subject to the present pilot study. Added reference/s: Quilliam, C., Wilson, E., Hagiliassis, N. & Nicola-Richmond, K. (2010). Measurement of Services and Supports (MOSS): Project Overview and Findings. Melbourne: Deakin University.
2. Also, the specific relationship to GAS would be useful. To what degree did this inform the SF-MOSS. What adaptions were required or necessary versus just using GAS directly with the person with a disability. And, there is use of the GAS, particularly in self-determination research, including student with disability self-report, that might be useful to review.	**We have added the following to the manuscript:** p.19. Notwithstanding its limitations, the GAS has a number of procedures that are generalised to the SF-MOSS. Each generates objectives prior to onset of the service or 'intervention'. Like the GAS, the SF-MOSS reflects actual improvement in a person's functional ability relating to aspects of his or her physical, social and attitudinal world and can be adapted to any life domain thereby covering aspects of life as a whole (Bouwens, van Heugten, & Verhey 2009). The SF-MOSS however is composed of 'performance' and 'satisfaction' indices, taking into consideration both the performance itself and the individual's reported satisfaction with performance of self-selected goals. Another key distinction is in accessibility design and useability characteristics. The GAS has been used widely for measuring the outcomes of personalized interventions delivered in mental health and educational settings (e.g., the educational programs of students with autism; Ruble, McGrew & Toland, 2012) there have been limited applications of GAS in intellectual disability services (Chapman, et al. (2006). Chapman et al (2006) concluded that the GAS system does not seem to enable its use in an empowering way with people with intellectual disabilities (e.g., staff completing the forms at

	their desks, most interviewees felt GAS was too complicated to be shared with service recipients). Further, the GAS data is not easily comparable across individuals and therefore cannot be easily aggregated into service group data (Wilson, Jenkin & Campain, 2011). The SF-MOSS strives to address these limitations through its design characteristics, described below. p.32. Further research is also required comparing the SF-MOSS with other outcome measurement tools – particularly the GAS being the most widely used tool for evaluating individual progress towards goals – for useability by service providers and accessibility by people with disability. Added reference/s: Wilson, E.; Jenkin, E. & Campain, R. (2011). *Outcome Measurement of Community Based Mental Health Services in Western Australia: Literature and Concept Summary.* Melbourne: Inclusion Matters. Bouwens, S. F., van Heugten, C. M., & Verhey, F. R. (2009). The practical use of goal attainment scaling for people with acquired brain injury who receive cognitive rehabilitation. Clinical Rehabilitation, 23(4), 310–320. https://doi.org/10.1177/0269215508101744 Ruble, L., McGrew, J. H., & Toland, M. D. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of autism and developmental disorders*, *42*(9), 1974-83.
3. What justification was used to allow for proxy reporting? How are concerns with proxy reporting that are frequently identified when comparing self- vs. proxy-data addressed?	**We draw attention to the point that despite our best efforts, there will be tools that remain inaccessible to a cohort of people with disabilities. The manuscript already highlights this point:** Wilson et al. (2013) highlight the importance of striving to obtain the views of people with a disability directly and developing tools using accessible design principles (e.g., simplified language and layout) so that the life experiences and opinions of people with disability are captured. There

	are limits to this as there will remain some people with more severe disability unable to access tools despite there being reasonable accommodations, but self-report remains the "ideal" (Wilson et al., 20113). **In addition, we have added the following to the manuscript:** p.20. A determination of when to use proxy reporting is based on judgement of the tool user. Currently, there are no steadfast criteria for when the need for proxy reporting is triggered, the administration manual (and parallel administration video) of the SF-MOSS includes guidelines as to when proxy report may be justified. In essence, it is the principle of 'last resort' that is advocated (Wilson et al. 2013) with users encouraged to consider what procedures can be applied (e.g., communication supports) to overcome a person's functional impairments, in order to increase a person's access to the tool. p.29. Proxy ratings should be treated as an opinion or a partial view on the extent to which the person has realised their goal (White-Koning et al., 2005). But they should not be used interchangeably with self-ratings, because of the limits of agreement between these two reporting methods (Chen, Hsieh, Mao, & Huang, 2007). p.29. As a minimum, we suggest the simple precaution of recording who is providing the information, so that inferences are made in the full knowledge of whose responses they reflect. The SF-MOSS documents this information source, providing data transparency. p.30. Some quality of life tools (e.g., Personal Wellbeing Index, Intellectual Disability, Cummins & Lau, 2010) designed for people with intellectual disability incorporate pre-testing to ensure competence in responding to questions. However, this procedure lacks social desirability (Wilson et al., 2013) and increases administration time (National Disability Services, 2012) and for these reasons does not form a part of SF-MOSS administration.

	Added reference/s: Chen, M.-H., Hsieh, C.-L., Mao, H.-F., & Huang, S.-L. (2007). Differences between patient and proxy reports in the assessment of disability after stroke. Clinical Rehabilitation, 21(4), 351–356. https://doi.org/10.1177/0269215507072544
4. What training /qualifications does the staff person need to administer?	**We have added the following to the manuscript:** p.22. Whilst the administrator requires no minimum qualifications, they are to be experienced in communicating with people with a disability, especially in an interview context, and capable of supporting people with a disability to set goals.
5. What does the beginning and end of service delivery mean? Is this a specific set of services aligned with the goal or service more generally? If more generally, when would services end? **And, how is it determined which services have an impact.**	**We have added the following to the manuscript:** p.20. People specify a goal and the service that will help achieve that goal. People rate their achievement of their goal when that specific service first commences or in the case where the service has already started, at a "pre" or baseline wave of a service delivery. People re-rate their goal when the service finishes (e.g., service disengagement) or at a service review point (e.g., scheduled plan review). At follow up administration, the goal is restated using the original wording. In this way, the person is attuned to the specific goal being targeted by the particular service.
6. More detail is needed on what is meant by the resonance and conveyance of meaning in relation to the Likert scale, beyond just a citation. Also, why was continuous scaling not considered, or using concrete anchors like in GAS, particularly for performance?	**We have added the following to the manuscript:** p.21. Speech pathologists with expertise in easy-language suggested wording for the stem questions and Likert scale (this was through a professional plain-language translation service that also employs people with low English literacy to provide feedback on language, layout, images and readability). After trialling the tool with people with intellectual disability, minor semantic changes were made. For example, people suggested the term "happy" be offered as an alternative to "satisfied" (as in "How satisfied/happy are you with the way it works for you now?"). This wording appeared to resonate with people with intellectual disability more than any other. The Personal Wellbeing Index Intellectual Disability (Cummins

	& Lau, 2010) similarly replaces 'satisfaction' (as used in the general adult population version) with 'happiness' because it is more likely within the language experience of people with intellectual disability. A number of visual scaling formats were considered (e.g., 11-point). In the end, a 5-point response Likert scale was opted for that included pictorial representations of response alternatives. This was determined as cognitively simplest for participants with intellectual disability, whilst having the number of response alternatives set at 5 (versus for example 3) increases the amount of information collected (Hartley & Maclean, 2006). Hartley and Maclean (2006) suggest that as many as five response alternatives can be used in Likert-type scales for people with intellectual disability without decreases in response rates. Added reference/s: Hartley, S.L. & Maclean, W. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. Journal of intellectual disability research. 50. 813-27. 10.1111/j.1365-2788.2006.00844.x.
7. Is satisfaction in relation to the goal or the services? If the goal, why was the wording chosen? If it is a goal, it is necessarily "working." Also, the use of happiness should be explained as this is a different construct from goal attainment.	**We have added the following to the manuscript:** p.20. Measurement indices are performance ("How well does it work for you now?") and satisfaction with this level of performance ("How happy are you with the way it works for you now?") using 5-point Likert scales. See also Reviewer 1 Response 6 (above)
8. How were the disability service provider organizations selected, how many staff participated? It is stated they are from large organizations, but only 49 surveys were returned, so what this a low response rate? If so, what concerns might there be about how representativeness of perspectives.	**We have added the following to the manuscript:** p.24. These were organisations who had started embedding outcomes measurement in practice and the SF-MOSS specifically. The first agency (from which the majority of participants for the face validity component were drawn) had been implementing the SF-MOSS for some years, the second (which provided the smaller sample) for approximately 6 months. p. 26. Although response rate could not be estimated, the size of these organisations would suggest the number of

	participants was fewer than anticipated. p. 32. We caution that our conclusions are drawn based on the small sample size of a pilot study using data from two targeted agencies. **This issue is also addressed in Reviewer 2, Revision 30.**
9. More detail is needed on how the tool is introduced to adults with disabilities and what administration looks like	**We have included the following in the manuscript:** p. 22. An administration guide is provided in written and DVD format which covers person-centred goal setting, considerations for use of proxy reporting, and has specific instructions for administration, scoring and interpretation. **Further details about administration are not given in the manuscript to keep the paper within word-count limits. However, this can be expanded if the reviewer feels it is required.**
10. Was any analysis undertaken of differences based on proxy responses?	**No, this level of analysis was not undertaken, there was an insufficient sample size to allow this.** **We have included the following in the manuscript:** p.32 The level of agreement between self- and proxy-report ratings on the same goal outcomes also warrants investigation, with previous research on such tools suggesting a discord between these two rating sources.
11. More justification for why the Shogren et al. framework is useful for content validity is needed. Also the specific steps undertaken to do the content analysis of the goals should be further specified.	**We have included the following in the manuscript:** p. 23. This procedure was drawn from the method used by Joyce, Rockwood and Mate-Kole (1994) in their investigation of the content validity of the GAS. Joyce et al compared problem areas identified through the GAS scaling with a list of areas identified in the literature as being particularly important in the rehabilitation of brain injury patients. These 18 areas included communication, memory, mobility and social skills. They observed that goals were set in all areas but one (vocation) and concluded that because it represents goals of all types from those identified as important in the literature, the GAS has adequate content validity. A similar procedure

	was used in the present study where goals extracted from SF-MOSS forms were compared against Shogren et al's (2015) framework for cataloguing personal outcomes in intellectual disability (based on The Council on Quality Leadership and the National Core Indicators outcomes models). There are a number of analytical frameworks that could be used to capture the focus of goals. Shogren et al.'s framework was selected because it provides also a framing for what might be the influencers of these outcomes. In the present study, enablers and barriers (as specified in completed SF-MOSS forms) to goal realisation were also analysed using the Shogren et al. framework, using a similar rationale for the relevance of this method to content validity. p. 25. Goals, enablers and barriers identified on SF-MOSS forms were entered onto a database. These were categorised thematically according to the named analytical framework by the first researcher and independently confirmed for accuracy by the second researcher. Given that interpretation of the focus of goals is potentially influenced by the subjective viewpoints of researchers, results were compared and discussed, where points of clarification were required. Goals were assigned to categories within the frameworks where both researchers had identified these categories.
12. The Discussion does a nice job of addressing implications and research directions, but I would like to see more specific content on how this could be utilized in practice, and further changes and considerations that might need to be made to promote adoption and use.	**We have included the following in the manuscript:** p. 31. An online version of the tool is being developed to promote access and utilisation. The availability of video based training for its administration is also an advantage, with workforce training demands being an increasing consideration for service providers.
13. Reviewer 2	
14. On page 1, the tool is described as new however, according to the authors, a previous version (non-short form) has already been published, as has this short form version (2014). The authors do not make clear in the manuscript how this paper is different from the 2014 publication. Nor do they discuss how the short form differs from the	See Reviewer 1, Response 1

original version, nor what evidence is there to suggest that the original version requires reducing.	
15. Page 3: The point about the need for agencies to provide targeted services is well made. This has ethical implications, both in terms of service user experience and in ensuring effective and efficient spending of public finances. However, I think more discussion about the link between this and current policy and practice trends would enhance this paper's relevance to this journal's scope	**We have included the following in the manuscript:** p. 15. Disability policy places significant emphasis on individual outcomes. Service providers are increasingly expected to demonstr5ate outcomes associated with their services and supports. Collecting outcomes data enables providers to deliver focused services targeting the person's goals and show evidence for what changes as a result (Chapman, Burton, Hunt & Reeves, 2006; Hagiliassis, Koritsas & Cuzzillo, 2017). Funding bodies and people with disability exercising greater choice and control places greater emphasis on the need for monitoring of individual outcomes to ensure accountability and responsiveness to consumer preferences (National Disability Services, 2012).
16. P.3 Line 30 ", there are few outcomes tools that can be readily accomplished in service delivery settings" - I don't think the word accomplished is correct here. Do you mean administered?	**We have included the following in the manuscript:** p. 15. There are few outcome tools that can be easily administered in service settings that providers can utilise across a range of situations, service types and client groups.
17. Page 4 Line 18/19 repeated use of the word perceived in one sentence doesn't read well.	p.16. Corrected.
18. The need for simplicity of administration is well made here. The lack of ID-specific outcome measures which are sensitive to change is becoming increasingly acknowledged and is an urgent topic for attention in all areas of ID research, from service provision to intervention trials. These wider implications could be flagged up in this paper.	**We have included the following in the manuscript:** p.16. Whilst proxy report is considered reliable for reporting objective measurements, it is not necessarily equivalent to a person's own perception of his or her life (Stancliffe, 2000). Wilson et al. (2013) highlight the importance of striving to obtain the views of people with a disability directly and developing tools using accessible design principles (e.g., simplified language and layout) so that the life experiences and opinions of people with disability are captured. People self-reporting, or communicating directly, about their own views is a fundamental principle underlying human rights which, beyond providing information about personal outcomes, increases understanding of the views and needs of people with a disability and their capacity to influence research,

	policy making, and advocacy (Wilson et al, 2013).
19. The date of the Wilson reference needs to be corrected.	p.16. Corrected
20. Page 5 Participants: It is not clear from this section how many staff were invited, thus the percentage recruitment. It is not discussed how many staff came from each of the two services. I assume from the manuscript that staff from one of the services had been using the measure for a while and therefore had direct experience with it, where as those in the other service had not. May this group difference not effect face validity assessments? Also, why were the interviews not administered at both centres when assessing content validity?	See Reviewer 2 Response 30 See Reviewer 2 Response 28 for reasoning
21. Methods section: There is a lot of detail here about the original version of the interview. I think this detail and historical context is more relevant in the introduction and may better help set the context for the need for a short form version.	See Reviewer 1 Response 1 **The authors feel that this information sits better with the method, and that this might disrupt the manuscript and make the introduction bulky. However, if the reviewer feels this is essential, we are able to make this change.**
22. Page 6 The measure is administered by a staff member to a PWID. Thus it is better seen as a structured Interview and not a questionnaire. However, no details are provided regarding instructions to staff about levels of interaction, prompting,follow-up/alternative version questions etc. This is important as PWID may have vastly different levels of cognitive ability.	See Reviewer 1 Response 9
23. A good strength here is that PWID were involved in developing the wording that goes along with the visual response options. It is good that there are three versions which are used depending upon a person's preference. When the authors use the word preference, may this be associated with ability level? Are the three options linked to differing ability levels? If so, are there differing options for question wording depending on ability levels?	**We have included the following in the manuscript:** p. 22. Visual scale selection is a personal preference i.e., a person with a disability may choose to use one format over the other. Whilst a person's preference may be linked to what they understand (and therefore their cognitive level), there is no ability-based criteria for choice of visual scales.
24. Page 7 line 37 Can you state clearly that the language used was in fact assessed and passed by	See Reviewer 1 Response 6

Easy English specialists? From where? Can you confirm if the language used is pitched or matched in any way to the three likert scale response versions?	**The same wording is used across the three visual response options – just the visual layout differs**.
25. Page 8 Line 30 It is not clear how the authors are defining what is a "relevant" outcome measurement framework. Which one did they use? This paragraph is not clearly written.	See Reviewer 1 Response 11
26. Line 42 "The rationale behind this method is that if a measurement procedure has adequate content validity, then it should be relevant and representative of the construct that it will be used to measure" . It is not clear how the word 'relevant' is being used here – relevant to what? Is this term redundant or relating to something that requires being more clearly stated.	See Reviewer 1 Response 11 **This whole section we feel has been strengthened.** **The term "relevant" is redundant and has been removed.**
27. Page 9: Procedure section. Which ethics committee approved the study?	**We have included the following in the manuscript** p. 24. The research was approved by an ethics committee registered with National Health and Medical Research Council. **The specific committee is not named because it means that it will no longer be blinded.**
28. The authors do not provide a rationale for why staff members from two services rated the interview schedule for face validity but only PWID from one service completed the forms. See my comments above regarding the participants section.	**We have included the following in the manuscript** p. 24. People with disability from one of the service providers participated in the content validity component of the research. These participants were drawn from the agency who had been implementing the SF-MOSS for some months. It was not possible to recruit from both agencies because of research funding limitations.
29. Line 27 If PWID were given an information sheet about the study and then were able to give informed consent to allowing their forms to be analysed then this would imply a certain level of cognitive ability. However, 26% of the PWID were unable to complete the interview independently. How then were they able to provide informed consent independently?	p.24. Consent to participate in the research was provided by the person themselves or the person responsible. **This is in accordance with common practice in Australia for the recruitment of people with disabilities for research.** **The manuscript has been revised to clarify this.**

30. Page 10 Results section: I think it would be useful to know how many staff participants came from each service, as there is a clear difference between the groups – i.e. one group has direct experience of using the interview and the other does not. It would be useful to know there were any differences between the groups, in terms of face validity ratings.	**We have included the following in the manuscript** p. 26. Staff from one organisation comprised 39 whilst from the other 10. The former had been implementing the SF-MOSS in their practice for some years, that latter some months. There were no obvious differences in face validity ratings across the two organisations – for example 89% from the first organisation and 90% from the second organisation rated 'agree' or 'strongly agree' to the question "I understand what the SF MOSS is asking me to do".
31. There were more managers than clinical staff in the sample. Are managers more likely to complete this interview with people with ID, or are clinical/support staff? If the latter is the case, then does this have implications for the results? 12% of staff said the purpose of the interview was unclear – which grade(s) were these staff from? This would be useful to know.	**The tool is designed for use by a range of disability services staff, not just clinical/support staff, across a range of situations, service types and client groups. The question raised by the reviewer was not addressed in the present research given its pilot study nature. The question warrants further attention and has been flagged as a direction of future research in the manuscript.** **We have included the following in the manuscript** p.32. The adequacy of the tool for staff in their various roles (e.g., service managers compared with clinical staff) also requires further analysis, the sample size was insufficient to meaningfully address this question. **We have included the following in the manuscript** p. 26. Only a small proportion of responses (12%; 3 managers/coordinators, 1 planner, 1 administrator, 1 unspecified) suggested the purpose of the tool was unclear (e.g., "Don't know") or something different to what it was intended to measure (e.g., "quality of life").
32. Page 13 Discussion: The authors propose that 65% of the PWID were able to make ratings independently. The measure is not a questionnaire, it is an interview schedule which by it's nature involves an interaction between the	**We have included the following in the manuscript** p. 29. In these (65% of) cases, staff reported that they administered the tool using information from people with disability directly, without the need for information from

interviewer and interviewee. The authors do not provide any detail about instructions to the interviewer regarding levels of prompting or providing alternative question wording. Therefore we know that 65% of interviews were completed with staff but we can not definitively say to which degree the PWID rated each of their goal attainments, nor listed and rated enablers and barriers, independently.	additional sources or reliance on proxy report.
33. The authors themselves note the potential for bias which proxy reports. This may be particularly important if staff are providing proxy ratings for assessing the impact of the service that they themselves are providing. If proxy responses are being given then perhaps staff should be asked to provide details of concrete examples to evidence their assertions of outcome.	**The reviewer raises a valid point, and gives a practical suggestion on how to attenuate for this factor. The researchers will consider this as an area of focus for our future work on the tool and its research.**
34. Page 14 The authors seem to be defining content validity as the extent to which the goals listed were categorised similarly to a Shogrens framework. However, as I read the manuscript, the purpose of the authors' interview schedule is to measure attainment of personal goals. It is not a taxonomy for categorising goals. Therefore, surely content validity here would relate to the extent to which the interview allows for the goals to be identified, recorded, progress to be measured and contextual influences to be identified. This has nothing to do with which categories were identified in the pilot study. The authors may see some relevance in Patrick et al. (2011) who provide an ISPOR Taskforce report in the journal Value in Health. They discuss content validity in developing a tool for measuring patient reported outcomes in the medical field.	See Reviewer 1 Response 11 **This procedure was drawn from the method used by Joyce, Rockwood and Mate-Kole (1994) in their investigation of the content validity of the GAS. Yes, the reviewer's observations appear accurate - we similarly interpret content validity as the versatility of the tool to incorporate goals of all types as per the method used by Joyce, Rockwood and Mate-Kole (1994). The justification for the content validity approach has been strengthened.** **The suggestion of the Patrick et al reference is useful.**
35. The paper raises the importance of the need to develop ID outcome measures, particularly ones that are sensitive to change. The relevance of this to other aspects of ID research, e.g. clinical trials, could be highlighted. Similarly, the discussion section would benefit from highlighting the	**We have included the following in the manuscript** (bearing in mind need to stay within word limit) p. 31. The paper raises the importance of the need to develop outcome measures that are effective achieving self-report by people with disability. The policy

potential impact of interview schedules such as the MOSS, and how this relates to policy and practice – this is not well developed.	environment recognises that giving people with disability choice and control over their supports can help to improve their outcomes. In implementing different dimensions of choice and control, service providers are expected to provide every opportunity for people to communicate their views and experiences directly. Outcome tools, such as the SF-MOSS, that seek the subjective experience of people with disabilities bring us closer to this objective. The application of self-report to other fields, such as research, policy making and advocacy, is a complex and growing area of study and has much potential (Wilson et al., 2013).
36. My last point is one of curiosity. The authors do not mention in their manuscript about the Outcomes and Impact Scale- Revised. This measure is also reported on the SCOPE website along with the SF-MOSS and there is overlap between the teams who developed both measures. It may be interesting to know how the current interview schedule differs from the OIS-R.	**Both tools (and others) are from The Scope Outcomes Project which aims to capture reliability and validity of surveys to measure outcomes in the disability sector. The SF-MOSS is a goal realisation measure, the Outcomes and Impact Scale - Revised evaluates the impact of disability services on the person with disability across 9 life domains. Both measure personal outcomes, but have a somewhat different focus and application. Not all tools could be covered in this manuscript (beyond the scope of the paper) and information on these tools has been published elsewhere.**

REVISED MANUSCRIPT

Measuring Goal Realisation associated with Disability Services and Supports: Initial Evidence for a new tool.

Background: There are few outcome measurement tools that can be accessed directly by people with intellectual disability and that can be readily used in service delivery settings. Developed to address these limitations, the Measuring Outcomes in Services and Supports – Short Form is a measure of the extent to which a person's goals are realised in the context of service and supports delivery.

Specific Aims: To establish initial evidence for the face and content validity of the tool, an essential step in the first stage of developing evaluation tools.

Method: Face validity was explored via an on-line survey eliciting information from an 'expert group' about the purpose, adequacy, clarity and attractiveness of the tool. Content validity was examined by comparing goals for people with disability against a framework for categorising personal outcomes and their influencers, to gauge the tool's relevance and representation for measuring whole of life disability outcomes.

Findings: Participants gave positive support overall for aspects of face validity surveyed. Suggestions for further improvements were provided, most of which were accommodated. Other issues raised reflected themes general to the field of outcomes measurement in disability populations (e.g., complexity for people with intellectual disability), with strategies suggested for

attenuating their impact. Participant goals were representative of outcome domains reflecting the current policy environment in disability, as were their influencing factors.

Discussion: With limited tools to measure goal realisation for people accessing a range of disability supports, the tool has acceptable face validity and content validity and is sufficiently practical for use in disability services.

Keywords: personal outcomes, goal realisation, disability, face validity, content validity

INTRODUCTION

Disability policy places significant emphasis on individual outcomes. Service providers are increasingly expected to demonstrate outcomes associated with their services and supports. Collecting outcomes data enables providers to deliver focused services targeting the person's goals and show evidence for what changes as a result (Chapman, Burton, Hunt & Reeves, 2006; Hagiliassis, Koritsas & Cuzzillo, 2017). Funding bodies and people with disability exercising greater choice and control places greater emphasis on the need for monitoring of individual outcomes to ensure accountability and responsiveness to consumer preferences (National Disability Services, 2012).

There are limited tools however that have the ability to express change brought about by services and supports that can be accessed directly by people with disability, particularly people with intellectual disability. Most tools require sufficient competence to understand and respond to complex terminology and procedures. For example, The Canadian Occupational Performance Measure (2005) measures occupational outcomes for people with disabilities but is problematic in that it includes an inaccessible format (i.e. lack of pictures or images) and a complex numerical rating scale (Quilliam and Wilson, 2011). Further, there are few outcome tools that can be easily administered in service settings - tools that are brief, easy to use, avoid administrative burden, and do not require specific training - which providers can utilise across a range of situations, service types and client groups. For example, the Personal Outcome Measures (Gardner & Carran, 2005) are designed specifically for broad use across disability

services and professional service disciplines. However, the tool is lengthy (i.e., requires attention to 21 pre-determined sub-domains), requires training, and is administratively intensive which is not feasible for many providers (Quilliam & Wilson, 2011).

Goal realisation refers to the achievement of something desired or anticipated that occurs in the course of service or support delivery (referred to as an 'intervention' in traditional goal attainment literature e.g., Turner-Stokes, 2009). The most widely used tool for evaluating individual progress towards goals is Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968) which has documented validity, reliability, and sensitivity properties (Hurn, Kneebone & Cropley, 2006; Chapman et al., 2006). It has been surprisingly under-utilised in intellectual disability services (Chapman et al, 2006). Partly this is attributable to its administrative steps, the most demanding of which is specifying a continuum of possible outcomes, that is, assigning a criterion for what constitutes an expected outcome (‘anchoring point’) and then a range of possibilities from best possible outcome, to worst possible outcome (Schlosser, 2004). The level of goal attainment is evaluated at the end of intervention. The system is illustrated in Table 1. Chapman et al, (2006) describe the implementation of GAS in the context of an intellectual disability agency. They report benefits and drawbacks of GAS as perceived by staff, the most concerning of which being that very little evaluation of individual goals occurred with service users directly. In the majority of cases, GAS was used by staff in isolation because of their hesitancy to involve service users in such a complex process. In addition, staff described the tool as “time-consuming” as well as reporting difficulties completing forms, including scaling goals.

INSERT TABLE 1 ABOUT HERE

Collecting subjective data from people with intellectual disability poses a number of measurement challenges, which in practical terms, defaults to a reliance on information from carers, referred to as proxy report. Whilst proxy report is considered reliable for reporting objective measurements, it is not necessarily equivalent to a person's own perception of his or her life (Stancliffe, 2000). Wilson et al. (2013) highlight the importance of striving to obtain the views of people with a disability directly and developing tools using accessible design principles (e.g., simplified language and layout) so that the life experiences and opinions of people with disability are captured. They add that people self-reporting, or communicating directly, about their own views is a fundamental principle underlying human rights which, beyond providing information about personal outcomes, increases understanding of the views and needs of people with a disability and their capacity to influence research, policy making, and advocacy. There are limits to this as there will remain some people with more severe disability unable to access tools despite there being reasonable accommodations, but self-report remains the "ideal" (Wilson et al., 2013). A concurrent issue faced by disability service providers is their capacity to administer outcome tools feasibly, such as within existing time, workload and training allocations, with the ease of use of any outcome measurement system playing a vital part in the success of implementation (Chapman et al., 2006; Quilliam & Wilson, 2011). In this study, we report initial evidence for the face and content validity of a tool, the Measuring Outcomes in Services and

Supports – Short Form (SF-MOSS), developed to address the limitations in access by people with intellectual disability and ease of use by service providers.

METHOD

Participants

Participants were staff from two major disability service providers in Australia. Adults with disability were recruited from one of these agencies, which provides services to people with intellectual and physical disability.

Method

The original version of the tool, the Measuring Outcomes in Services and Supports (MOSS) was developed in a disability service provider setting, to measure the extent to which a person's goals are realised in the course of service delivery. It was designed along the following criteria (1) evaluates outcomes (the impact of a service or support), rather than outputs (the services delivered by an agency) or process (the way in which services are provided) (2) is flexible to goals covering a range of life areas (e.g., wellbeing, social, economic), (3) is versatile across areas of services and supports (e.g., therapy, supported accommodation, community inclusion), (4) can be accessed by people with intellectual disability through its design and administration characteristics (e.g., using simple visual scaling), (5) reduces reliance on proxy reporting but where required, ensures this occurs from the person's perspective and potential sources of bias are understood, (6) has demonstrable evidence, (7) is achievable for service providers to administer in day-to-day practice, (8) allows for aggregating up for generating comparative data, (9) uses 'performance' and 'satisfaction' (satisfaction with performance) as indices of measurement, and (10) enquires about contextual factors (enablers and barriers) that influence personal outcomes.

An evaluation of the original tool revealed a number of shortcomings (Quilliam, Wilson, Hagiliassis & Nicola-Richmond, 2010). From goals extracted from 73 completed forms, the far majority (70%) of information was provided by proxies (i.e. service providers, family members) - only 28% of responses to the MOSS were informed by the person with a disability

independently. Originally planned for use in therapy services, it was felt to be therapy-centric and to include non-essential data fields (e.g., questions about legal issues). Resulting changes were improvements in easy-language wording, recasting the tool for use across any service, removing the need to provide superfluous information, simplification of visual scaling, and improvements in format/layout. The resultant version, the SF-MOSS (Hagiliassis, Nicola-Richmond, Wilson & Mackay, 2014), is the tool that is subject to the present pilot study.

Notwithstanding its limitations, the GAS has a number of procedures that are generalised to the SF-MOSS. Each generates objectives prior to onset of the service or 'intervention'. Like the GAS, the SF-MOSS reflects actual improvement in a person's functional ability relating to aspects of his or her physical, social and attitudinal world and can be adapted to any life domain thereby covering aspects of life as a whole (Bouwens, van Heugten, & Verhey 2009). The SF-MOSS however is composed of 'performance' and 'satisfaction' indices, taking into consideration both the performance itself and the individual's reported satisfaction with performance of self-selected goals. Another key distinction is in accessibility design and useability characteristics. The GAS has been used widely for measuring the outcomes of personalized interventions delivered in mental health and educational settings (e.g., educational programs of students with autism; Ruble, McGrew & Toland, 2012) there have been limited applications of GAS in intellectual disability services (Chapman, et al. (2006). Chapman et al (2006) concluded that the GAS system does not seem to enable its use in an empowering way with people with intellectual disabilities (e.g., staff completing the forms at their desks, most

interviewees feeling GAS was too complicated to be shared with service recipients). Further, the GAS data is not easily comparable across individuals and therefore cannot be easily aggregated into service group data (Wilson, Jenkin & Campain, 2011). The SF-MOSS strives to address these limitations through its design characteristics, described below.

The SF-MOSS is administered by a staff person directly with the person with disability in the form of an interview. For people who still find the administration characteristics too complex, a carer can assist the person provide information or, as a last resort, act as a proxy with carers completing the survey on the person's behalf, from the viewpoint of the person. A determination of when to use proxy reporting is based on the judgement of the tool user. Currently, there are no steadfast criteria for when the need for proxy reporting is triggered, the administration manual (and parallel administration video) of the SF-MOSS includes guidelines as to when proxy report may be justified. In essence, it is the principle of 'last resort' that is advocated (Wilson et al. 2013) with users encouraged to consider what procedures can be applied (e.g., communication supports) to overcome a person's functional impairments, in order to increase a person's access to the tool. The source of the information (person independently, with assistance or proxy) is recorded on the SF-MOSS form so that the potential biases of proxy reporting are considered when interpreting data.

People specify a goal and the service that will help achieve that goal. This is entered onto the SF-MOSS form, in the person's own words as far as is possible (e.g., "I want to visit mum in the country. Every 4–6 weeks."). People rate their achievement of their goal when the specific

service first commences or in the case where the service has already started, at a “pre” or baseline wave of a service delivery. People re-rate their goal when the service finishes (e.g., service disengagement) or at a service review point (e.g., scheduled plan review). At follow up administration, the goal is restated using the original wording. In this way, the person is attuned to the specific goal being targeted by the particular service.

Measurement indices are performance (“How well does it work for you now?”) and satisfaction with this level of performance (“How satisfied/happy are you with the way it works for you now?”) using 5-point Likert scales. Ordered choices for performance are “Very bad” (1), “A little bad” (2), “Not good or bad” (3), “A little good” (4), “Very good” (5). For satisfaction, they are “Very dissatisfied (unhappy)” (1), “A little dissatisfied (unhappy)” (2), “Not satisfied (happy) or dissatisfied (unhappy)” (3), “A little satisfied (happy)” (4), “Very satisfied (happy)” (5).

Speech pathologists with expertise in easy-language suggested wording for the stem questions and Likert scales (this was through a professional plain-language translation service that also employs people with low English literacy to provide feedback on language, layout, images and readability). After trialling the tool with people with intellectual disability, minor semantic changes were made. For example, people suggested the term “happy” be offered as an alternative to “satisfied” (as in “How satisfied/happy are you with the way it works for you now?”). This wording appeared to resonate with people with intellectual disability more than any other. The Personal Wellbeing Index Intellectual Disability (Cummins & Lau, 2010) similarly replaces ‘satisfaction’ (as used in the general adult population version) with ‘happiness’ because it is

more likely within the language experience of people with intellectual disability. A number of visual scaling formats were considered (e.g., 11-point). In the end, a 5-point response Likert scale was opted for that included pictorial representations of response alternatives. This was determined as cognitively simplest for participants with intellectual disability, whilst having the number of response alternatives set at 5 (versus for example 3) increases the amount of information collected (Hartley & Maclean, 2006). Hartley and Maclean (2006) suggest that as many as five response alternatives can be used in Likert-type scales for people with intellectual disability without decreases in response rates.

There are three pictorial representations of the Likert scales that can be used. The use of visual scaling is to combat possible acquiescence and to overcome communication problems, with respondents having the option to point to the symbol that best describes how they feel. An example of a visual scale is given in Figure 1. This particular scale is modelled on the response format used in the Glasgow Depression Scale for people with a Learning Disability (Cuthill, Espie & Cooper, 2002) which has the option of presenting items in two stages: the first requiring (in the case of performance) the person to indicate whether their performance of a goal is 'good/bad', and the second requiring an indication of the degree of this (e.g., if the person answers 'good' they are then asked ‘very good’ or ‘a little good'). It means that the person can engage with a 5-point Likert scale, but only needing to make a two-choice response at each step. Visual scale selection is a personal preference i.e., a person with a disability may choose to use

one format over the other. Whilst a person's preference may be linked to what they understand (and therefore their cognitive level), there is no ability-based criteria for choice of visual scales.

INSERT FIGURE 1 ABOUT HERE

The tool yields a performance and satisfaction change score in either a positive, negative or neutral direction. Data can be aggregated up for group or service-level comparative purposes (e.g., proportion of people achieving positive, negative or neutral change). When administered at the end of service delivery, respondents are asked about whether there were any factors that enabled or impeded progress toward their goal. The role of these potential influencing variables are asked from the perspective of the person with a disability, but can be asked separately from the viewpoint of carers and service providers.

The tool takes an average of 10-15 minutes to administer at each time point. Whilst the administrator requires no minimum qualifications, they are expected to be experienced in communicating with people with a disability, especially in an interview context, and capable of supporting people with a disability to set goals. An administration guide is provided in written and video format which covers person-centred goal setting, considerations for use of proxy reporting, and has specific instructions for administration, scoring and interpretation.

For the current research, face validity was explored via an on-line survey eliciting information about the purpose, adequacy, clarity and attractiveness of the tool. Participants rated the extent to which the SF-MOSS achieved these aspects using a 5-point Likert scale ranging from "Strongly

Agree" to "Strongly Disagree". Two open-ended question were also included to further check their understanding of the purpose of the tool and elicit information about any improvements that could be made. Basic demographic data was also collected (i.e., role, years of experience in disability, gender).

For content validity, the SF-MOSS was administered to people with disability in one disability provider and data relating to goals, enablers and barriers were extracted from completed forms. Goals were compared with a relevant outcomes measurement framework for categorising disability policy goals and personal outcomes, and for understanding the contextual factors influencing personal outcomes in disability. This procedure was drawn from the method used by Joyce, Rockwood and Mate-Kole (1994) in their investigation of the content validity of the GAS. Joyce et al compared problem areas identified through the GAS scaling with a list of areas identified in the literature as being particularly important in the rehabilitation of brain injury patients. These 18 areas included communication, memory, mobility and social skills. They observed that goals were set in all areas but one (vocation) and concluded that because it represents goals of all types from those identified as important in the literature, the GAS has adequate content validity. A similar procedure was used in the present study where goals extracted from SF-MOSS forms were compared against Shogren et al's (2015) framework for cataloguing personal outcomes in intellectual disability (based on The Council on Quality Leadership and the National Core Indicators outcomes models). There are a number of analytical frameworks that could be used to capture the focus of goals. Shogren et al.'s framework was

selected because it provides also a framing for what might be the influencers of these outcomes. In the present study, enablers and barriers (as specified in completed SF-MOSS forms) to goal realisation were also analysed using the Shogren et al. framework, using a similar rationale for the relevance of this method to content validity.

Procedure

The research was approved by an ethics committee registered with National Health and Medical Research Council.

Disability staff from two large disability service providers were invited to participate in the face validity component of the research. These were organisations who had started embedding outcomes measurement in practice and the SF-MOSS specifically. The first agency (from which the majority of participants for the face validity component were drawn) had been implementing the SF-MOSS for some years, the second (which provided the smaller sample) for approximately 6 months. Explanatory statements were distributed to staff by managers on behalf of the researchers via email, including a link to the on-line survey, as well as a PDF copy of the SF-MOSS from which to base their review. Participation was voluntary and anonymous and submission of completed surveys implied consent.

People with disability from one of the service providers participated in the content validity component of the research. These participants were drawn from the agency who had been implementing the SF-MOSS for 6 months. It was not possible to recruit from both agencies

because of research funding limitations. Consent to participate in the research was provided by the person themselves or the person responsible. Before the administration of the SF-MOSS, the person with disability was given an explanatory statement in plain English asking for a non-identifiable copy of their completed SF-MOSS. The researchers were given a non-identifiable copy of the form, unless potential participants told staff that they did not want to participate (i.e., opt-out).

Data analysis

Descriptive statistics and thematic analyses were used to analyse face validity data. For content validity, goals, enablers and barriers as identified by people with disability were compared against Shogren, Luckasson and Schalock's (2015) framework for understanding disability policy goals and personal outcomes and their influencer variables. Goals, enablers and barriers identified on SF-MOSS forms were entered onto a database. These were categorised thematically according to the named analytical framework by the first researcher and independently confirmed for accuracy by the second researcher. Given that interpretation of the focus of goals is potentially influenced by the subjective viewpoints of researchers, results were compared and discussed, where points of clarification were required. Goals were assigned to categories within the frameworks where both researchers had identified these categories.

RESULTS

Face Validity

Forty-nine staff completed the survey comprising managers/coordinators (n=22), therapists (n=14), staff in support roles e.g., direct care worker, planner, trainer (n=9), administration assistant (n=1) and unspecified (n=3). There were 40 females and 9 males. They had worked in the disability sector for an average of 10.6 (SD=6.5) years and hence were considered sufficiently familiar with the target audience to determine how practical and relevant the tool is for the group. Although response rate could not be estimated, the size of these organisations would suggest the number of participants was fewer than anticipated. Staff from one organisation comprised 39 whilst from the other 10. The former had been implementing the SF-MOSS in their practice for some years, that latter some months. There were no obvious differences in face validity ratings across the two organisations, for example 89% from the first organisation and 90% from the second organisation rated 'agree' or 'strongly agree' to the question "I understand what the SF MOSS is asking me to do".

Table 2 presents the participant responses about the purpose of the tool, coded into seven broad themes, with illustrative quotes. Overall, participants were able to articulate the purpose of the tool adequately and accurately. Only a small proportion of responses (12%; 3 managers/coordinators, 1 planner, 1 administrator, 1 unspecified) suggested the purpose of the tool was unclear (e.g., "Don't know") or something different to what it was intended to measure (e.g., "quality of life").

INSERT TABLE 2 ABOUT HERE

Table 3 reports the percentage of participants who agreed or strongly agreed with each of the five statements about the adequacy, clarity and attractiveness of the tool, and shows that the majority responded positively.

INSERT TABLE 3 ABOUT HERE

Twenty-five suggestions were provided about improving the tool. These broadly surround seven themes, presented in Table 4 along with illustrative examples. The most common issues identified were in relation to further reducing complexity for use with people with intellectual disability and clearer terminology and wording, particularly in relation to use of the word "it" when asking about performance (i.e. "how well does *it* work for you now?"). In addition, general issues relating to outcomes measurement but not necessarily specific to the SF-MOSS were named, such as respondents potentially judging satisfaction based on their current mood.

INSERT TABLE 4 ABOUT HERE

Content Validity

The SF-MOSS was responded to by people with disability independently (65%), with assistance (10%), and by proxy (16%), the remainder were unspecified. From 19 completed SF-MOSS forms, 29 goals were documented (some participants had multiple goals). Table 5 shows that goals were spread across the personal outcome domains and domain indicators of the Shogren et al. (2015) framework. The majority could be categorised under the domains of inclusion in society and community life (9), education/life-long learning (8) and well-being (6). Two

additional domains apply to this framework, self-determination (e.g., freedom to make choices, participation in decisions), and citizenship (e.g. rights, freedom of expression), which were interpreted in the present study as implicit to all goals.

INSERT TABLE 5 ABOUT HERE

Factors reported as enablers and barriers to goal realisation, as perceived by the person with a disability, significant others (e.g., family) and service providers, are presented in Table 6. The table reveals the most commonly reported enabler to goal realisation was having support from an agency/service provider (17), and the main impediments were lack of control (7), lack of choice and opportunity (7), as well as the time taken (delays) to realise goals (7).

INSERT TABLE 6 ABOUT HERE

DISCUSSION

The pilot study provides initial evidence for the validity of the SF-MOSS, a tool to measure progress towards goal realisation for people with disability. Participants gave positive to strongly positive support for all aspects of face validity surveyed, suggesting that those familiar with the target population find the tool acceptable and understandable. From the range of suggestions given by participants for further improvement of the tool, most were readily accommodated in its current iteration. These were clearer terminology and wording (especially the use of the word "it" when asking about performance i.e. "How well does it work for you now?" has been

changed to “Think about your goal (restate goal). How well (good or bad) do you do this now?”), improved format and clearer administration instructions.

Other issues raised apply to any system of outcome measurement and are not particular to the SF-MOSS. Participants raised the possibility that in using the SF-MOSS, respondents might rate their satisfaction based on their current mood or focussing on a specific aspect of service delivery that is salient at the time. Similar concerns have been expressed in the literature regarding quality of life judgments, with several research findings suggesting in typical assessment situations current mood/momentarily salient information has negligible effects on life satisfaction ratings (Schimmack, 2006). Nevertheless, Schimmack (2006) suggests it is possible to take precautions to reduce the chance of distorted life satisfaction judgments, such as retesting at the same time point and observing whether the same result is produced. Retesting is not part of the SF-MOSS tool’s standard administration and as such, its impact on useability has not been explored. Users of the tool would need to determine whether retesting is justified or not, this extra step would be expected to add administrative burden however.

The most frequently named area for improvement of the SF-MOSS was further reducing complexity for people with intellectual disability. As described, the SF-MOSS was developed to address the limitations in access by people with intellectual disability. But whilst it attenuates this issue, it does not remove it entirely. Sixty-five percent of people with disability were able to make ratings independently, suggesting that in the majority of cases the tool achieved an acceptable method of assessment. In these cases, staff reported that they administered the tool

using information from people with disability directly, without the need for information from additional sources or reliance on proxy report. However, the tool remained beyond the capacity of 16% of people with disability who reverted to proxy report. Wilson et al. (2013) stress that self-rating is the ideal but acknowledge that the nature of impairment is an insurmountable barrier to self-report for some people whose disability is more severe in nature, making the need for proxy report inevitable. The findings in the present study concur with this view.

The general consensus from disability outcomes literature is that proxy ratings should be interpreted cautiously. This is because they do not necessarily equate to the person's own perspectives, life standards or expectations (Kroll 2011; Stancliffe, 2000; Wilson et al., 2013). The difference between proxy and person responses may reflect biased or erroneous information or genuine difference in opinion between two parties on the same issue (White-Koning et al., 2005). Proxy ratings should be treated as an opinion or a partial view on the extent to which the person has realised their goal (White-Koning et al., 2005). But they should not be used interchangeably with self-ratings, because of the limits of agreement between these two reporting methods (Chen, Hsieh, Mao, & Huang, 2007). As a minimum, we suggest the simple precaution of recording who is providing the information, so that inferences are made in the full knowledge of whose responses they reflect. The SF-MOSS documents this information source, providing data transparency. Other suggestions are obtaining ratings from several proxies for an overall rating or training for proxy raters to reduce inaccuracy or bias (Bailey & Simeonsson, 1988; White-Koning et al., 2005). Some quality of life tools (e.g., Personal Wellbeing Index,

Intellectual Disability, Cummins & Lau, 2010) designed for people with intellectual disability incorporate pre-testing to ensure competence in responding to questions. However, this procedure lacks social desirability (Wilson et al., 2013) and increases administration time (National Disability Services, 2012) and for these reasons does not form a part of SF-MOSS administration.

Evidence for content validity emerged in that participant goals mapped adequately to Shogren's (2015) categorisation of outcomes in disability. In order that an outcome tool be valid, it must adequately represent the construct it proclaims to measure. For the SF-MOSS, we would expect the tool to have versatility for measuring progress towards goals in virtually any of the major life areas. According to Quilliam and Wilson (2011), the current policy environment in disability requires a focus on whole-of-life outcomes for people with disabilities generated through individual supports. Outcome measurement instruments must engage with a broad range of outcome areas (for example, across the Articles of the UN Convention on the Rights of Persons with Disabilities) (Quilliam and Wilson, 2011). Few tools enable this breadth and the SF-MOSS appears to achieve this.

When enablers and barriers to goal realisation were analysed, these were also found to be representative of the contextual factors that research suggests act as independent or intervening variables in personal outcomes (Shogren et al., 2015). The foremost determinants of success were having access to services and supports, opportunities for decision making, choice, empowerment, and control, adequate finances access/planning and adequate time to achieve

goals. Quilliam and Wilson (2011) reinforce the importance of "collecting information about the ingredients of change or areas of service improvement needed to increase positive outcomes" (p.25). Their review of disability outcome tools reveals that most tools do not provide mechanisms to collect this data. In the present study, the SF-MOSS was observed to be useful in capturing this data from which service providers can target strategies to enhance personal outcomes (Shogren et al., 2015).

The tool is sufficiently practical for use in disability services, issues raised regarding administration time were minimal. Service providers require practice tools that are brief, easily to use and have minimal training requirements, so as they can be readily accomplished in the settings they are intended to be used (National Disability Services, 2012). Chapman et al, (2006) suggest that 'ease of use' plays a key role in the successful adoption of an outcome measurement system, one of the crucial barriers to uptake of the GAS being that staff perceived the tool as "time-consuming". Increasingly as service providers are expected to achieve efficiencies across their services, the need for valid outcomes tools that can be implemented routinely with minimal impact on staffing and systems resources becomes a priority. An online version of the tool is being developed to promote access and utilisation. The availability of video based training for its administration is also an advantage, with workforce training demands being an increasing consideration for service providers.

More broadly, the paper raises the importance of the need to develop outcome measures that are effective achieving self-report by people with disability. The policy environment recognises that

giving people with disability choice and control over their supports can help to improve their outcomes. In implementing different dimensions of choice and control, service providers are expected to provide every opportunity for people to communicate their views and experiences directly. Outcome tools, such as the SF-MOSS, that seek the subjective experience of people with disabilities bring us closer to this objective. The application of self-report to other fields, such as research, policy making and advocacy, is a complex but growing area of study that has much potential (Wilson et al., 2013).

Limitations

The current study provides initial evidence for the face and content validity of the SF-MOSS. We caution that our conclusions are drawn based on the small sample size of a pilot investigation using data from two targeted agencies. The study must be followed by additional testing with an expanded sample to examine the psychometric properties of the tool more fully. This includes its correlation and comparison of useability with other valid criterion measures (e.g., GAS), its sensitivity to detect change, and the degree to which test results are consistent over time (test-retest reliability). Further research is also required comparing the SF-MOSS with other outcome measurement tools – particularly the GAS being the most widely used tool for evaluating individual progress towards goals – for useability by service providers and accessibility by people with disability. The level of agreement between self- and proxy-report ratings on the same goal outcomes also warrants investigation, with previous research on such tools suggesting a discord between these two rating sources. The adequacy of the tool for staff in their various

roles (e.g., service managers compared with clinical staff) also requires further analysis, the sample size was insufficient to meaningfully address this question. Among the considerations for future iterations of the tool are procedures for bringing the tool closer the accessibility benchmarks outlined in this paper. In its present format, the SF-MOSS only identifies the direction of change in goal realisation but not its statistical significance.

Conclusion

With limited tools to measure goal realisation for people accessing a range of disability supports, the SF-MOSS has acceptable face validity and content validity as assessed by an expert group and an analysis of goals respectively. The tool shows promise for tracking the outcomes resulting from services and supports and for service users to monitor change associated with the services they receive.

REFERENCES

Bailey, D., & Simeonsson, R. (1988). Investigation of use of goal attainment scaling to evaluate individual progress of clients with severe and profound mental retardation. *Mental Retardation, 26*, 289-295.

Bouwens, S. F., van Heugten, C. M., & Verhey, F. R. (2009). The practical use of goal attainment scaling for people with acquired brain injury who receive cognitive rehabilitation. *Clinical Rehabilitation*, 23(4), 310–320. https://doi.org/10.1177/0269215508101744

Buntinx, W.H.E., & Schalock, R. (2010). Models of disability, quality of life, and individualized. supports: Implications for professional practice in intellectual disability. *Journal of Policy and Practice in Intellectual Disabilities*, *7*(4), 283–294.

Chapman, M., Burton, M., Hunt, V., & Reeves, D. (2006). Implementation of Goal Attainment Scaling in Community Intellectual Disability Services. *Journal of Policy and Practice in Intellectual Disabilities*, *3*: 119–128.

Chen, M.-H., Hsieh, C.-L., Mao, H.-F., & Huang, S.-L. (2007). Differences between patient and proxy reports in the assessment of disability after stroke. *Clinical Rehabilitation*, 21(4), 351–356.

Cummins, R.A., & Lau, A.L.D. (2010). *Personal Wellbeing Index – Intellectual Disability. Australian Centre on Quality of Life.* Melbourne: Deakin University.

Cuthill, F.M., Espie, C.A., & Cooper, S.A. (2003). Development and psychometric properties of the Glasgow Depression Scale for people with a learning disability. *British Journal of Psychiatry, 182*, 347–353.

Gardner, J.F., & Carran, D.T. (2005). Attainment of personal outcomes by people with developmental disabilities, *Mental Retardation*, *43*(3), 157-174.

Hagiliassis, N. Nicola-Richmond, K., Wilson, E., & Mackay, A. (2014). *Measuring Outcomes in Services and Supports – Short Form (SF-MOSS).* Melbourne: Scope.

Hartley, S.L., & Maclean, W. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. Journal of intellectual disability research. 50. 813-27.

Hurn, J., Kneebone, I., & Cropley, M. (2006). Goal setting as an outcome measure: A systematic review. *Clinical Rehabilitation*, *20*, 756-772.

Joyce, B.M., Rockwood, K.J., & Mate-Kole, C.C. (1994). Use of goal attainment scaling in brain injury in a rehabilitation hospital. *American Journal of Physical Medicine and Rehabilitation*, *73*(1), 10–15.

Kiresuk, T.J., & Sherman, R.E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health, 4*, 443-453.

Koritsas, S., Hagiliassis, N., & Cuzzillo, C. (2017). The Outcomes and Impact Scale – Revised: the psychometric properties of a scale assessing the impact of service provision. *Journal of Intellectual Disability Research, 61*,450–460.

Kroll, T. (2011). Designing mixed methods studies in health-related research with people with disabilities. *International Journal of Multiple Research Approaches, 5*(1), 64-75.

National Disability Services. (2012). Measuring Outcomes for People with Disability. Retrieved from http://www.asksource.info/resources/measuring-outcomes-people-disability

Nicola-Richmond, K., Hagiliassis, N., Mackay, A., & Wilson, E. (2008). *Measuring the Outcomes of Therapy and Psychology Services for People with a Disability*. Brisbane, Australia: AusACPDM Conference.

Nicola-Richmond, K., Wilson, E., Hagiliassis, N., & Mackay, A. (2011). *Summary Sheet – Measuring Outcomes in Services and Supports (MOSS) tool.* Melbourne: Scope.

Quilliam, C., Wilson, E., Hagiliassis, N. & Nicola-Richmond, K. (2010). *Measurement of Services and Supports (MOSS): Project Overview and Findings.* Melbourne: Deakin University.

Quilliam, C. & Wilson, E. (2011). *Literature Review - Outcomes measurement in disability services: a review of policy contexts, measurement approaches and selected measurement tools.* Melbourne: Deakin University.

Ruble, L., McGrew, J. H., & Toland, M. D. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of autism and developmental disorders,* 42(9), 1974-83.

Schimmack, U. (2006). Internal and External Determinants of Subjective Wellbeing: Review and Policy Implications. In Ng, Y.-K., & Ho, L. S. (Eds.), *Happiness and public policy: Theory, case studies and implications*. New York: Palgrave Macmillan.

Schlosser, R.W. (2004). Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *Journal of Communication Disorders, 37,* 217-239.

Shogren, K.A., Luckasson, R., & Schalock, R.L. (2015). Using Context as an Integrative Framework to Align Policy Goals, Supports, and Outcomes in Intellectual Disability. *Intellectual and Developmental Disabilities, 53*(5), 367-376.

Stancliffe, R. (2000). Proxy Respondents and Quality of Life. *Evaluation and Program Planning. 23,* 89-93.

The Canadian Occupational Performance Measure. (2005). Retrieved from http://www.caot.ca/copm/description.html

Turner-Stokes, L. (2009). Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clinical Rehabilitation, 23*(4), 362-70.

White-Koning, M., Arnaud, C., Bourdet-Loubère, S., Bazex, H., Colver, A., & Grandjean, H. (2005). Subjective quality of life in children with intellectual impairment - how can it be assessed? *Developmental Medicine & Child Neurology, 47,* 281–285.

Wilson, E., Campain, R., Moore, M., Hagiliassis, N., McGillivray, J., Gottliebson, D., … Graffam, J. (2013). An accessible survey method: increasing the participation of people with a disability in large sample social research, *Telecommunications Journal of Australia*. *63*(2), 24.1-24.13.

Wilson, E., Jenkin, E., & Campain, R. (2011). *Outcome Measurement of Community Based Mental Health Services in Western Australia: Literature and Concept Summary*. Melbourne: Inclusion Matters.

TABLES AND FIGURES

TABLE 1: Example individual goal and continuum of possible outcomes using GAS (from Schlosser, 2004)

GOAL: When talking with a responsive peer, Josh will take three conversational turns using his communication book or natural gestures.	
+2 BEST EXPECTED	Answers at least 4 questions during reading and informal conversation
+1 MORE THAN EXPECTED	Answers 3 questions during reading AND informal conversation
0 EXPECTED	Answers 3 questions during reading OR informal conversation
-1 LESS THAN EXPECTED	Answers 2 questions during reading OR informal conversation
-2 WORST EXPECTED	Answers 1 or less questions during reading OR informal conversation

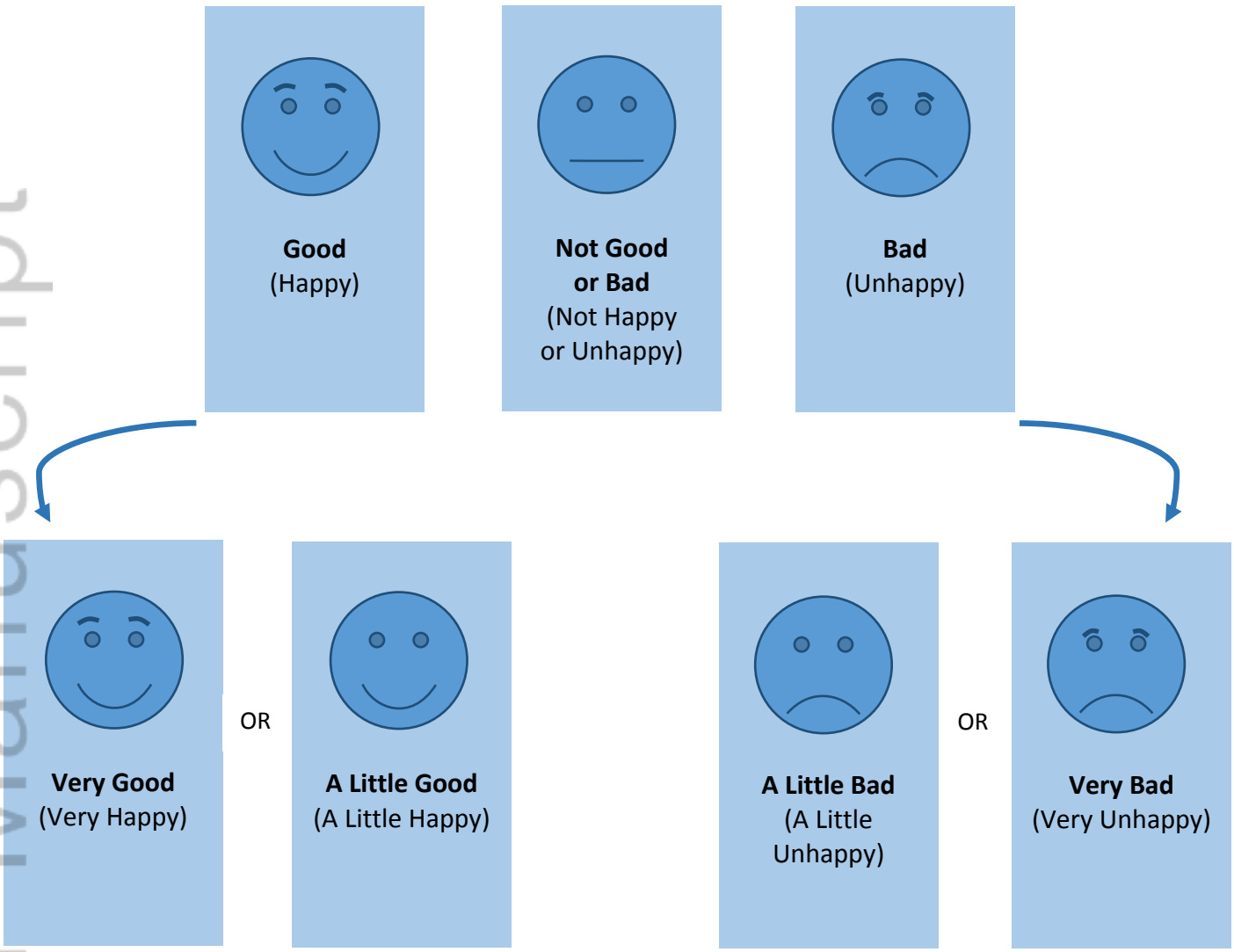

Figure 1. Example of a visual scale used with the SF-MOSS

Table 2: Purpose of the SF-MOSS: Themes and their frequency of occurrence in participants' comments

Purpose	Example quotes	Frequency of occurrence
Measures dimensions of performance and/or satisfaction	"Measures a client or caregivers performance at achieving a goal and their satisfaction with the goal"	20
Measures goals and outcomes generally	"How the client's goal is progressing"	14
Measures change as a result of a specific service	"How much of a difference a service has made in relation to a goal"	9
Identifies barriers and enablers to goal realization	"What they felt enabled/prevented this goal from being achieved"	8
Measures from a client perspective	"Goal from the point of view of the client"	4
Purpose unclear	"I am actually not sure at all, I have never heard of it before"	4
Measures other construct/s	"Measures improvement in the quality of life of the client to carry out their life"	4

Table 3: Percentage of participants who agreed or strongly agreed with various statements about the characteristics of the SF-MOSS.

Question	Agreement
I understand what the SF-MOSS is asking me to do	92%
Other staff involved in service provision will be able to answer these questions	86%
The tool adequately measures goals and progress towards goals	82%
The questions are clear and easy to understand	74%
The layout/formatting of the SF-MOSS is easy to read and follow	74%

Table 4: Suggested improvements to SF-MOSS: Themes and their frequency of occurrence in participants' comments

Themes	Example	Frequency of occurrence
Reducing complexity for use with people with intellectual disability	"Will a person with a disability understand what pre and post mean?"	8
Clearer terminology and wording (especially the use of the word "it" when asking about performance i.e. "how well does *it* work for you now?")	"I find the use of the word *it* confusing when being asked "how well does *it* work…?""	4
General issues relating to outcomes measurement (not the SF-MOSS specifically)	"Depending on how the client feels on the day, that will influence how they rate their satisfaction"	4
Improving format (especially being able to accommodate for multiple goals)	"Being able to rate more than one goal per form would be useful and more time efficient"	3

Clearer instructions for administration	"I feel there needs to be a simple explanation of what to do (on the tool itself, not just the administration manual)"	2
Concerns about administration time required	"Needs to be shorter"	2
Clarity of purpose	"Measuring outcomes of what?"	2

Table 5: Participant goals categorised according to Schogren et al.,'s (2015) Personal Outcome Domains and Domain Indicators.

Domain	Indicator (example participant goal, paraphrased)	Number of Goals
Inclusion in Society and Community Life	Community living (e.g., maintain stable living arrangements, live independently)	9
	Community inclusion (e.g., access and be part of community)	
Education/Life-Long Learning	Vocational education (e.g., complete my certificate)	8
	Personal growth and development (e.g., improve communication skills)	
Well-Being	Physical well-being (e.g., be able to self-transfer for personal care)	6
	Health and wellness (e.g., improve my fitness)	
Human Relationships	Meaningful relations (e.g., assist family members)	4
	Social networks (e.g., maintain social networks)	
Productivity	Economic self-sufficiency (e.g., improve financial skills)	2
	Meaningful engagement in activities (e.g., improve cooking skills)	

Table 6: Frequency of enablers and barriers influencing goal realisation from person with a disability, other (e.g., family member) and service provider

	Total	**Person with disability**	**Other**	**Support Provider**
Enablers				
Services support	**17**	9	4	4
Decision making, empowerment, control	**6**	2	2	2
Finances access, planning	**6**	4		2
Family involvement, friendships	**4**	1	2	1
Choices, opportunities	**3**	2		1
Effective goal setting	**3**	3		
Assistive technology	**3**	1		2
Accessible information	**2**	2		
Mainstream services access	**1**			1
Self-care skills	**1**	1		
Experiential learning	**1**			1
Community access, inclusion, participation	**1**	1		
Functional training, education	**1**			1
Barriers				
Control over environment (lack of)	**7**	5		2
Choices, opportunities (lack of)	**7**	4	1	2

Time taken to achieve goals	**7**	3	1	3
Carer/family stress	**6**	1	4	1
Finances access, planning (lack of)	**4**	2		2
Decision making, empowerment, control (lack of)	**4**	2		2
Physical ill-health	**3**	3		
Social exclusion	**2**	1		1
Unforeseen changes	**1**			1
Poor service coordination	**1**			1
Community negative attitudes	**1**		1	
Poor practices	**1**			1
Assistive technology (lack of)	**1**	1		