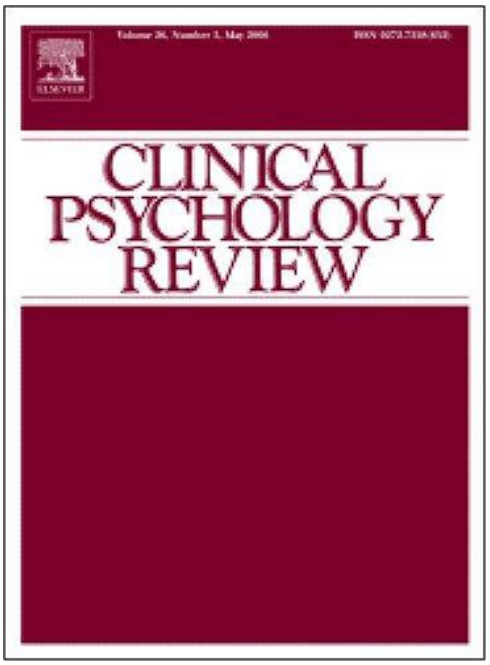
Accepted Manuscript: Measurement tools for mental health problems and mental well-being in people with severe or profound intellectual disabilities: A systematic review



DOI:	https://doi.org/10.1016/j.cpr.2017.08.006
Please cite as:	Flynn, S., Vereenooghe, L., Hastings, R. P., Adams, D., Cooper, S. A., Gore, N., Hatton, C., Hood, K., Jahoda, A., Langdon, P. E., McNamara, R., Oliver, C., Roy, A., Totsika, V., & Waite, J. (2017). Measurement tools for mental health problems and mental well-being in people with severe or profound intellectual disabilities: A systematic review. Clinical Psychology Review.

Measurement tools for mental health problems and mental well-being in people with severe or profound intellectual disabilities: A systematic review

Ms Samantha Flynn[1], Dr Leen Vereenooghe[2], Prof. Richard P Hastings[1], Dr Dawn Adams[3], Prof. Sally-Ann Cooper[4], Dr Nick Gore[5], Prof. Chris Hatton[6], Prof. Kerry Hood[7], Prof. Andrew Jahoda[4], Dr Peter E Langdon[5], Dr Rachel McNamara[7], Prof. Chris Oliver[8], Dr Ashok Roy[9], Dr Vasiliki Totsika[1] and Dr Jane Waite[8]

[1] CEDAR, University of Warwick, UK

[2] Fakultät für Psychologie und Sportwissenschaft, Universität Bielefeld, Germany

[3] Autism Centre of Excellence, Griffith University, Australia

[4] Institute of Health and Wellbeing, University of Glasgow, UK

[5] Tizard Centre, University of Kent, UK

[6] Faculty of Health and Medicine, Lancaster University, UK

[7] Centre for Trials Research, Cardiff University, UK

[8] School of Psychology, University of Birmingham, UK

[9] Coventry and Warwickshire Partnership NHS Trust, UK

Corresponding author:

Ms Samantha Flynn (s.flynn.1@warwick.ac.uk)

CEDAR, University of Warwick, New Education Building, Westwood Campus, University of Warwick, Coventry, CV4 7AL

Abstract

Mental health problems affect people with intellectual disabilities (ID) at rates similar to or in excess of the non-ID population. People with severe ID are likely to have persistent mental health problems. In this systematic review (PROSPERO 2015:CRD42015024469), we identify and evaluate the methodological quality of available measures of mental health problems or well-being in individuals with severe or profound ID. Electronic searches of ten databases identified relevant publications. Two reviewers independently reviewed titles and abstracts of retrieved records (n=41,232) and full-text articles (n=573). Data were extracted and the quality of included papers was appraised. Thirty-two papers reporting on 12 measures were included. Nine measures addressed a broad spectrum of mental health problems, and were largely observational. One physiological measure of well-being was included. The Aberrant Behavior Checklist, Diagnostic Assessment for the Severely Handicapped Scale-II and Mood, Interest and Pleasure Questionnaire are reliable measures in this population. However, the psychometric properties of six other measures were only considered within a single study – indicating a lack of research replication. Few mental health measures are available for people with severe or profound ID, particularly lacking are tools measuring well-being. Assessment methods that do not rely on proxy reports should be explored further.

Keywords: intellectual disabilities; mental health; mental illness; psychiatric disorder; mental well-being; measurement

Background

Children and adolescents with intellectual disabilities (ID) are 4-5 times more likely to present with symptoms of diagnosable mental health problems in comparison with children who do not have ID (Emerson & Hatton, 2007). Similarly, adults with ID are at increased risk of mental health problems (Cooper et al., 2007a), with a recent UK based cohort study reporting that between 21 and 34% of participants had a mental health problem over the course of the study (Sheehan et al., 2015). There is evidence that the incidence of severe mental health problems is heightened in people with ID compared to people without ID (cf. Sheehan et al., 2015 with Hardoon et al., 2013). Within the population of people with ID, there is some indication that those with severe ID are at a greater risk of mental health problems (Cooper et al., 2007a, 2007b; Hove & Havik, 2010; Smiley et al., 2007). Furthermore, longitudinal data from a cohort of Australian children and adolescents with ID over a period of 14 years using the Developmental Behaviour Checklist (Einfeld & Tonge, 1992; 1995) suggest that mental health problems are more likely to persist for people with severe or profound ID, whereas mental health problems for people with mild ID may decrease over time (Einfeld et al., 2006).

Assessing mental health problems and mental well-being in people with severe or profound ID presents multiple challenges. Methodological challenges include people with severe or profound ID having difficulties with self-report, difficulties labelling and communicating emotions (Adams & Oliver, 2011) – although this is not unique to people with severe to profound ID, and is a challenge for people with mild and moderate ID too (Mellor & Dagnan, 2005) – concern about the accuracy of proxy reports (Emerson, Felce & Stancliffe, 2013), and the process of diagnostic overshadowing whereby symptoms of a mental health problem are ascribed to the person's ID or another co-morbid problem rather than being recognized as a mental health problem (Deb et al., 2001). Measurement of mental health problems and mental well-being in this population is also conceptually difficult as some mental health problems may have overlaps with behavioural problems (Marston et al., 1997; Ross & Oliver, 2002; Hayes et al., 2011), and some standard diagnostic criteria may be appropriate for people with mild or moderate ID (e.g., negative cognitions), but less so for people with severe or profound ID (Evans, Cotton, Einfeld & Florio, 1999).

To examine mental health problems and mental well-being in people with severe ID in the context of research and clinical practice, robust measurement tools are needed with supporting data specifically with the severe/profound ID population. Previous systematic reviews have explored the assessment of depression (Walton & Kerr, 2016; Hermans & Evenhuis, 2010; Perez-Achiaga, Nelson & Hassiotis, 2009) and anxiety (Reardon, Gray & Melvin, 2015; Hermans et al., 2011) in people with ID. However, these reviews did not focus on people with severe or profound ID and did not extend across the lifespan. All five of these systematic reviews did present some quality appraisal of the identified mental health measurement tools. Additionally, a non-systematic narrative review by Matson et al. (2012) offered a representation of broad mental health measures across the lifespan; no quality appraisal was undertaken within this review, and evidence pertaining specifically to people with severe or profound ID was not included. Moreover, Matson et al.'s review exclusively reported evidence for rating scale instruments, excluding other methods of assessment. Additionally, none of the aforementioned reviews included both measures of mental health problems and mental well-being.

Considering the identified gaps in current knowledge, the main questions for this systematic review were: (a) What are the tools available to measure mental health problems and mental well-being in children and adults with severe ID, and (b) What is the methodological quality of these measurement tools?

Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed for all stages of this systematic review. The protocol for this systematic review was registered with PROSPERO, an international register for systematic reviews with health related outcomes (PROSPERO 2015:CRD42015024469).

Search strategy

Online databases (CINAHL, ERIC, EMBASE, MEDLINE, ASSIA, PsycINFO, PsycTESTS, CENTRAL, and the Social Sciences and Science Citation Indices) were searched during July 2015. Search strings were based on those of Vereenooghe and Langdon (2013) and included terms

pertaining to ID, mental health and well-being, and the evaluation of measures (see Table 1 for an example search string, and Appendix 1 for the full search strings for each database).

Searches for a related systematic review exploring interventions for mental health problems in people with severe and profound ID (Vereenooghe et al., submitted) were performed at the same time, with results from both reviews being pooled initially in case of relevant papers being retrieved from the other search. Forwards and backwards searches were undertaken by hand searching the reference lists, and 'cited by' records of all included papers – a cut-off of September 2016 was imposed for retrieving papers from these searches. As an additional search strategy, citation searches for studies which reported on the identified measurement tools were undertaken to ensure that no potentially eligible study was missed.

Study selection

One reviewer (LV) screened all titles and abstracts for broad relevance, and a random sample of 20% was double reviewed by a second reviewer (SF); this is an accepted practice when a review is large and resources limited (Petticrew & Roberts, 2006). All full-text papers were independently double reviewed (LV and SF).

Papers were included within this review if they met the following inclusion criteria: (a) at least 70% of the sample in the study were reported as having severe or profound ID (although in some senses an arbitrary criterion, this was to ensure that there was a majority of people with severe or profound ID in the study samples) or the data for participants with severe or profound ID were reported separately, (b) an original study, presenting quantitative or psychometric outcome data, (c) the study focused on the development, adaptation, or evaluation of a measure of mental health or well-being, and (d) the study was reported in English, Dutch, French or German. A later adjustment to the review protocol was made to exclude records from screening (n=106 records) which were published before 1980, coinciding with the publication of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III; APA, 1980). This assured a minimal level of consistency in the recognition and diagnosis of mental health problems. It is likely that there would have been a delay between the publication of the DSM-III and its first use in published research, but

searches back to 1980 were essential to ensure that no potentially relevant studies were missed. No inclusion restrictions were placed on the age or gender of participants.

The definition of mental and physical components of mental, behavioural, and physical health problems into mutually exclusive categories can be challenging, particularly as some components are symptomatic of multiple disorders, and there is a high probability of comorbidity between certain problems. For the purpose of this systematic review, the inclusion criteria for mental health problems was derived from the ICD-10. Eligible mental health problems, and their key diagnostic symptoms, were in the following classifications: (a) F20-29: schizophrenia, schizotypal and delusional disorders; (b) F30-39: mood (affective) disorders; (c) F40-48: neurotic, stress-related and somatoform disorders; (d) F60-69: disorders of adult personality and behaviour; and, (e) F90-97: behavioural and emotional disorders with onset usually occurring in childhood and adolescence. As such, organic mental disorders, disorders due to psychoactive substance abuse, behavioural syndromes associated with physiological disturbances and physical factors, intellectual disability, disorders of psychological development (e.g. childhood autism and specific developmental disorders), as well as other behavioural and emotional disorders with onset usually occurring in childhood and adolescence which are not within F90-97 (e.g. pica, stereotyped movement disorder) were not considered to be eligible for inclusion in this review. Mental well-being is "…a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community." (WHO, 2014). By including mental well-being, as well as mental health problems, it was possible that more positively formulated titles and abstracts which did not focus on any specific diagnostic category would be identified. The categorisation of measures of mental health problems or mental well-being was reached by team consensus by applying the aforementioned definitions.

[TABLE 1 ABOUT HERE]

Data extraction and synthesis

Data were extracted into a table format by one reviewer (SF), and were checked for accuracy by a second reviewer (LV). Extracted data included participant demographics, details about the instrument (target, subscales, scoring, interpretation, etc.), and information about the data analyses. Data were summarised for all the studies reporting on each measurement tool, and a narrative synthesis of the methodological quality of each measure was completed.

Quality appraisals

The Characteristics of Assessment Instructions for Psychiatric Disorders in Persons with Intellectual Developmental Disorders (CAPs-IDD; Zeilinger et al., 2013) was used to guide quality appraisal, and eventual synthesis of the data by a first reviewer (SF). CAPs-IDD appraisal was subsequently checked by a second reviewer (LV) for accuracy. The CAPs-IDD is designed to evaluate the quality and psychometric properties of mental health tools for people with ID (Zeilinger et al., 2013). This tool combines typical criteria about the validity and reliability of measures generally with criteria about face validity and other issues specific to people with ID. As such, it can be used as a general quality tool for assessing measures across the ID field. The addition of mental well-being in the *concepts to be measured* section ensured that this data was being captured within the tool. Dimensions of the CAPs-IDD include: test development and theoretical foundation, measurement characteristics, validity (criterion, content, construct, face), reliability (internal consistency, reliability, measurement error), objectivity of application, objectivity of interpretation, norming, and fairness, and feasibility. All studies pertaining to each individual measurement tool were included in the assessment of each measure, allowing the authors to establish the weight of evidence for each measure in turn.

Measures were rated on a four point scale (++ excellent; + good; - fair; -- poor) using standard interpretation guidelines (Landis & Koch, 1977; Cicchetti & Sparrow, 1981; Cicchetti, 1994; Hinkle, Wiersma & Jurs, 2002); this can be found in Table 2.

[TABLE 2 ABOUT HERE]

To enable a summary of the overall quality of included measures, we considered measures to have good evidence of methodological quality if they had: (a) at least two dimensions of reliability with a ++ rating and more than one supporting study, and (b) at least one dimension of validity with a + rating or higher and more than one supporting study.

Results

Electronic searches resulted in 41,232 records being identified (see Figure 1). Following de-duplication the titles and abstracts of 24,883 papers were screened for relevance by a reviewer (LV) and a random sample of 20% was double reviewed by a second reviewer (SF). The double reviewing of 4,792 records resulted in 4,784 agreements (99.83%) and a Kappa coefficient of 0.91. All disagreements were resolved by LV and SF through discussion of each paper in turn.

From the title and abstract screening, 573 papers were selected for full-text screening. All full-text papers were independently double reviewed (LV and SF) with a total of 10 disagreements being discussed between the reviewers and resolved (Kappa coefficient for initial agreement = 0.83). There were no cases where a third reviewer was required to resolve disagreements. The main cause of disagreement was where it was unclear in the paper as to whether a sufficient percentage of the sample had severe or profound ID. Of the 10 disagreements, two were included in the final review (Esbensen et al., 2003; Janssen & Maes, 2013) as they presented some data separately for people with severe or profound ID.

Forwards and backwards searches of the reference lists, and 'cited by' records of all included papers identified three additional papers which underwent full inclusion assessment. Because they fulfilled the eligibility criteria, they were all included in the review. Two doctoral theses were identified through searches for studies reporting on identified measurement tools. At the end of this process, 32 papers were included, describing a total of 12 mental health measurement tools.

[FIGURE 1 ABOUT HERE]

We focus on the mental health measures in terms of the evidence only for people with severe to profound ID. Some measurement tools may have additional supporting data for populations of individuals with mild to moderate ID, but this information is not included in the current review.

Mental health and well-being measurement tools

Overall, 32 papers reporting on 12 tools for measuring mental health problems, and mental well-being in adults with severe or profound ID were identified (Table 3). No included studies pertained to the measurement of mental health problems or well-being in children or adolescents with severe or profound ID.

[TABLE 3 ABOUT HERE]

Included measurement tools were intended to screen for a broad spectrum of problems, with the exceptions being the Autism Spectrum Disorders-Comorbidity for Adults [ASD-CA; Matson, Terlonge & Gonzalez, 2006] (LoVullo & Matson, 2009; Matson & Boisjoli, 2008) and the Physiological Measure of the Subjective Well-Being of Persons With Profound Intellectual and Multiple Disabilities [Physiological Measure of Subjective Well-Being] (Vos et al., 2010) which were concerned with well-being, along with the Depression Scale for Severe Disability [DEPRESSED] (Cooper, 2007), which focussed on symptoms of depression, and the Mood, Interest and Pleasure Questionnaire [MIPQ; Ross & Oliver, 2002; 2003] (Petry et al., 2010) for mood. The frequency, severity, or duration of target behaviours were most often used to measure mental health problems.

All measures were evaluated in the English language, with six being validated in a second language, including: Dutch (Aberrant Behavior Checklist [ABC; Aman & Singh, 1985]: Petry et al., 2010; Anxiety, Depression and Mood Scale [ADAMS; Esbensen et al., 2003]: Hermans et al, 2012; Mini Psychiatric Assessment Schedule for Adults with a Developmental Disability [Mini PAS-ADD; Prosser et al., 1998]: Janssen & Maes, 2013; MIPQ: Petry et al., 2010), Japanese (ABC: Ono, 1996), and Spanish (Diagnostic Assessment for the Severely Handicapped Scale-II [DASH-II; Matson, 1995]: Vargas-Vargas et al., 2015). Seventeen studies were undertaken in the USA (Bamburg et al.,

2001; Bihm & Poindexter, 1991; Cooper, 2007; Esbensen et al., 2003; LoVullo & Matson, 2009; Matson et al., 1991a; Matson et al., 1991b; Matson et al., 1999; Matson & Smiroldo, 1997; Matson & Boisjoli, 2008; Paclawskyj et al., 1997; Redding, 1997; Rojahn et al., 2003; Sevin et al., 1995; Sturmey, Burcham & Perkins, 1995; Sturmey et al., 1996; Sturmey, Matson & Lott, 2004). One paper each, reported on samples from the USA and New Zealand (Aman et al., 1987a), Japan (Ono, 1996), Norway (Myrbakk & von Tetzchner, 2008), Spain (Vargas-Vargas et al., 2015), and China (Liu et al., 2007). Two papers reported research undertaken in the Netherlands (Dumont et al., 2014; Hermans et al., 2012), three papers were from Belgium (Janssen & Maes, 2013; Petry et al., 2010; Vos et al., 2010), and four were from the UK (Hill, Powlitch & Furniss, 2008; Moss et al., 1998; Prosser et al., 1998; Ross & Oliver, 2003).

All identified measures were for use with adults, and despite searching for measures specifically aimed at children, none were found. Four measures were developed with the intention of being used explicitly with people with severe or profound ID: the Diagnostic Assessment for the Severely Handicapped Scale [DASH; Matson et al., 1991a] (Matson et al., 1991a; Matson et al., 1991b; Redding, 1997; Sevin et al., 1995), DASH-II (Bamburg et al., 2001; Matson & Smiroldo, 1997; Matson et al., 1999; Myrbakk & von Tetzchner, 2008; Paclawskyj et al., 1997; Sturmey, Matson & Lott, 2004; Vargas-Vargas et al., 2015), DEPRESSED (Cooper, 2007), and the Physiological Measure of Subjective Well-being (Vos et al., 2010). It was unclear whether the remaining measures had been adapted for use with people with severe or profound ID or designed primarily to be used across a wider range of ID.

Twenty of the included papers evaluated the DASH (Matson et al., 1991a; Matson et al., 1991b; Redding, 1997; Sevin et al., 1995), DASH-II (Bamburg et al., 2001; Matson & Smiroldo, 1997; Matson et al., 1999; Myrbakk & von Tetzchner, 2008; Paclawskyj et al., 1997; Sturmey, Matson & Lott, 2004; Vargas-Vargas et al., 2015) or the ABC (Aman et al., 1987a; Aman et al., 1987b; Bihm & Poindexter, 1991; Dumont et al., 2014; Hill, Powlitch & Furniss, 2008; Ono, 1996; Paclawskyj et al., 1997; Petry et al., 2010; Rojahn et al., 2003; Ross & Oliver, 2003). Nine of the other measures were evaluated in only one study each with adults with severe to profound ID (DEPRESSED: Cooper, 2007; Interact Short Form: Liu et al., 2007; Psychiatric Assessment Schedule

for Adults with a Developmental Disability Checklist [PAS-ADD]: Moss et al., 1998; Physiological Measure of Subjective Well-being: Vos et al., 2010) or in only two studies each (ADAMS: Esbensen et al., 2003; Hermans et al., 2012; ASD-CA: LoVullo & Matson, 2009; Matson & Boisjoli, 2008; Mini PAS-ADD: Janssen & Maes, 2013; Prosser et al., 1998; MIPQ: Petry et al., 2010; Ross & Oliver, 2003; Reiss Screen for Maladaptive Behaviour [Reiss Screen; Reiss, 1988]: Sturmey, Burcham & Perkins, 1995; Sturmey et al., 1996). For three measures, the second study to evaluate the tool focused on a translated version of the original measure (ADAMS: Hermans et al., 2012; Mini PAS-ADD: Janssen & Maes, 2013; MIPQ: Petry et al., 2010). For ten of the measures, the researcher by whom it was developed was typically involved in its evaluation (ABC: Aman et al., 1987a; Aman et al., 1987b; Rojahn et al., 2003; ADAMS: Esbensen et al., 2003; ASD-CA (LoVullo & Matson, 2009; Matson & Boisjoli, 2008; DEPRESSED: Cooper, 2007; DASH: Matson et al., 1991a; Matson et al., 1991b; Sevin et al., 1995; DASH-II: Bamburg et al., 2001; Matson & Smiroldo, 1997; Matson et al., 1999; Paclawskyj et al., 1997; Sturmey, Matson & Lott, 2004; Mini PAS-ADD: Prosser et al., 1998; MIPQ: Ross & Oliver, 2003; PAS-ADD: Moss et al., 1998; Physiological Measure of Subjective Well-being: Vos et al., 2010).

Considering the participant samples, 17 of the studies included a mixed sample of people with ID and presented separate data for people with severe or profound ID (Aman et al., 1987a; Aman et al., 1987b; Bihm & Poindexter, 1991; Dumont et al., 2014; Ono, 1996; Paclawskyj et al., 1997; Rojahn et al., 2003; Esbensen et al., 2003; Hermans et al., 2012; LoVullo & Matson, 2009; Matson & Boisjoli, 2008; Myrbakk & von Tetzchner, 2008; Janssen & Maes, 2013; Prosser et al., 1998; Moss et al., 1998; Sturmey, Burcham & Perkins, 1995; Sturmey et al., 1996). The remaining 15 studies involved only participants with severe or profound ID (Hill, Powlitch & Furniss, 2008; Petry et al., 2010; Ross & Oliver, 2003; Cooper, 2007; Matson et al., 1991a; Matson et al., 1991b; Redding, 1997; Sevin et al., 1995; Bamburg et al., 2001; Matson & Smiroldo, 1997; Matson et al., 1999; Sturmey, Matson & Lott, 2004; Vargas-Vargas, 2015; Liu et al., 2007; Vos et al., 2010). All but one measure drew on informant (proxy) reports, including: carer, direct support staff, nurses, and other relatives. These proxy reports were based on the unstructured observation of behaviours (i.e., knowing the

person whose behaviours are being rated) or on clinical impressions. Only one measure used physiological indicators (Vos et al., 2010).

Only the ABC, DASH, DASH-II, Interact Short Form, MIPQ, and the Physiological Measure of Subjective Well-being measures reported on the time frame for measurement (over the past one month, two weeks, two weeks, present, two weeks, and present respectively). All other measures lacked clarity in terms of a timeframe for ratings.

Methodological quality of mental health measures

The CAPs-IDD quality assessment instrument (Zeilinger et al., 2013) was used to summarise the psychometric properties of all included measures, and their translated language variants. Whilst the CAPs-IDD does not provide a summary score for the measures, it does enable researchers to summarise methodological quality. A coded summary table of the methodological quality of each measure is presented in Table 4. Where source papers presented conflicting data regarding the strength of psychometric properties for a measure, measures were coded with consideration to all included papers.

[TABLE 4 ABOUT HERE]

Ten of the measures reported having some theoretical, expert, or classification model basis for the development of the tool, and in seven cases multiple methods of development were described. Measures were reported as being developed with guidance from a version of the International Classification of Diseases (ICD) (ASD-CA, DEPRESSED, Mini PAS-ADD), the Diagnostic and Statistical Manual of Mental Disorders (DSM) (ADAMS, ASD-CA, DASH, DASH-II, MIPQ), were guided by published literature and/or previous scales (ADAMS, ASD-CA, DEPRESSED, DASH, DASH-II, MIPQ, PAS-ADD, Mini PAS-ADD), or were guided by clinical experience (ADAMS, DEPRESSED). The development of four measurement tools was guided by a factor analysis of a pool of items (ABC, ADAMS, MIPQ, Reiss Screen). Neither the Interact Short Form nor the Physiological

Measure of Subjective Well-being were reported to have been developed using any theoretical or other methods.

Included papers did not contain information about sample norms, and six of the included measures were reported to have cut-off data: ADAMS (Hermans et al., 2012), ASD-CA (LoVullo et al., 2009), DASH (Sevin et al., 1995), DASH-II (Bamburg et al., 2001; Matson & Smiroldo, 1997; Myrbakk & von Tetzchner, 2008; Vargas-Vargas et al., 2015), mini PAS-ADD (Janssen et al., 2013; Prosser et al., 1998), PAS-ADD (Moss et al., 1998).

All except the Physiological Measure of Subjective Well-being were supported by evidence regarding internal consistency and inter-rater reliability. Whilst this would have been less relevant for the physiological measures, it would have been possible to undertake inter-rater reliability assessments for the behavioural observations. The internal consistency of the DASH was considered to be poor, based on the data presented in two studies. However, in the DASH-II this appears to have been rectified as the five papers reporting on this psychometric property indicated good internal consistency. Similarly, improvements in inter-rater reliability were evident from the DASH to the DASH-II. Where reported, the test-retest reliability of the measures were relatively strong and indicated consistency over time.

In the case of the validity of measures, little evidence was presented, particularly for criterion and content validity. Ten measures had some data supporting their construct validity from 13 papers: 11 undertook a Factor Analysis (Aman et al., 1987b; Bihm & Poindexter, 1991; Liu et al., 2007; Matson et al., 1991; Matson & Boisjoli, 2008; Moss et al., 1998; Ono, 1996; Petry et al., 2010; Redding, 1997; Sturmey et al., 1996; 2004), one study measured correspondence with the DSM-IV-TR (Cooper, 2007), and one reported correspondence between similar constructs in two measures – the ABC and DASH-II (Paclawskyj et al., 1997). Where available, validity data were encouraging especially for the DASH-II, Interact Short Form, and MIPQ. It is notable, however, that each psychometric property of six measures was only supported by a single research study.

Considering convergent validity, the ABC and DASH-II were evaluated in relation to other measures. The ABC total scores showed some strong correlations with: the Behavioural Problems Inventory (BPI; Rojahn et al., 2001 – although note that this is a measure of challenging behaviour not

mental health problems) total scores – both for the English (r=.37-.78: Dumont et al., 2014; Hill, Powlitch & Furniss, 2008; Rojahn et al., 2003) and Dutch (r=.78: Petry et al., 2010) versions. ABC scores also correlated with the DASH-II total scores (r=0.75: Paclawskyj et al., 1997). Correlations between the ABC and the MIPQ subscales varied (r=-.00 (ABC excessive speech subscale and MIPQ mood subscale: which would not be expected to correlate) to -.63 (ABC lethargy, social withdrawal subscale and MIPQ interest and pleasure subscale: which would be the expected direction of the correlation): Ross & Oliver, 2003). The DASH-II correlated well with DSM-IV clinical criteria (r=.94: Matson & Smiroldo, 1997), and also had some good correlations with the Matson Evaluation of Social Skills for Individuals with Severe Mental Retardation (MESSIER; Matson, 1995) subscales (r=-.00 (DASH-II sleep disorder subscale and MESSIER positive non-verbal subscale) to -.63 (DASH-II language disorder subscale and MESSIER negative verbal subscale): Sturmey et al., 2004), and the Mood subscale correlated well with the DEPRESSED (r=.79: Cooper, 2007). Correlations between the PAS-ADD and psychiatrists' diagnoses were variable (56% detection success rate for mild disorders and 92% for severe disorders: Moss et al., 1998). However, correlations were stronger between the Mini PAS-ADD and psychiatrists' diagnoses (91% detection success rate: Prosser et al., 1998). For the DASH depression subscale, six out of 15 core depression items were endorsed by more than 45% of the depressed sample (Matson et al., 1999). Finally, Bamburg et al. (2001) reported significantly higher reports of frequency and severity for the group with a diagnosis of schizophrenia, than for the groups who did not have a diagnosis of schizophrenia on the DASH-II.

Based on the summary CAPs-IDD criteria for overall quality (see Method), only three measures (the ABC, DASH-II, and MIPQ) were considered to have a reliably good level of methodological quality for use with individuals who have severe to profound ID to assess mental health problems (e.g. mood/mood disorders: ABC, DASH-II, MIPQ; and anxiety, and schizophrenia: DASH-II). Table 5 includes a complete list of subscales for these three tools and the mental health problems they purport to measure (only domains which were eligible for inclusion in this review have been included).

[TABLE 5 ABOUT HERE]

The Irritability subscale of the ABC was found to be reliable, with good to excellent internal consistency for the English (α=.89-.92: Bihm & Poindexter, 1991; Paclawskyj et al., 1997), Dutch (α=.88: Petry et al., 2010), and Japanese (α=.92: Ono, 1996) versions. The Irritability subscale of the Japanese ABC also had good inter-rater reliability (r=.78: Ono, 1996) and test-retest reliability (r=.90: Ono, 1996). Similarly, the Lethargy subscale had excellent internal consistency for the English (α=.91-92: Bihm & Poindexter, 1991; Paclawskyj et al., 1997), Dutch (α=.90: Petry et al., 2010), and Japanese (α=.95: Ono, 1996) versions. The Lethargy subscale of the Japanese ABC also had good inter-rater reliability (r=.68: Ono, 1996) and test-retest reliability (r=.85: Ono, 1996). The ABC Irritability subscale has been validated against the MIPQ Mood subscale (r=-.30: Ross & Oliver, 2003), the DASH-II Mood Disorders (r=.56: Paclawskyj et al., 1997), and Mania (r=.63: Paclawskyj et al., 1997) subscales. The Lethargy subscale of the ABC has also been validated against the MIPQ Mood (r=-.40: Ross & Oliver, 2003), and Interest and Pleasure (r=-.63: Ross & Oliver, 2003) subscales. None of the other subscales map onto the ICD-10 criteria used as eligibility criteria in this review.

The DASH-II Anxiety Disorders subscale was shown to have poor internal consistency (α=.44-.54: Paclawskyj et al., 1997; Myrbakk & von Tetzchner, 2008). The Mood Disorders subscale also has acceptable to poor internal consistency (α=.53-.70: Paclawskyj et al., 1997; Myrbakk & von Tetzchner, 2008), and the Mania subscale has been shown to have a questionable to excellent level of internal consistency (α=.61-.97: Paclawskyj et al., 1997, Matson & Smiroldo, 1997). Finally, the Schizophrenia subscale has a poor level of internal consistency (α=.53: Paclawskyj et al., 1997). The reliability of the Spanish version of the DASH-II has also been reported, with the Mood Disorders subscale having a questionable internal consistency (α=.69), excellent test-retest reliability (r=.93), and good inter-rater reliability (r=.63) (Vargas-Vargas et al., 2015). The Spanish Anxiety Disorders subscale also had good test-retest reliability (r=.78), inter-rater reliability (r=.79), and questionable internal consistency (α=.61) (Vargas-Vargas et al., 2015). The Mania subscale on the Spanish version of the DASH-II had acceptable internal consistency (α=.73), and strong inter-rater reliability (r=.75) and test-retest reliability (r=.95) (Vargas-Vargas et al., 2015). Lastly, the Schizophrenia subscale had

unacceptable internal consistency (α=.48), and a good level of inter-rater reliability (r=.71) and test-retest reliability (r=.90) (Vargas-Vargas et al., 2015). Only the Mood Disorders, and Mania subscales have been found to be successfully validated against other measures: both subscales have been validated against the ABC Irritability subscale (as outlined above). None of the other subscales aligned with the eligibility criteria for this systematic review.

The MIPQ Mood subscale had good internal consistency (α=.89), and good test-retest reliability (r=.69), and inter-rater reliability (r=.90) (Ross & Oliver, 2003). Similarly the MIPQ Interest and Pleasure subscale was found to have excellent internal consistency (α=.90), and good test-retest reliability (r=.76), and inter-rater reliability (r=.84) (Ross & Oliver, 2003). The Dutch version of the MIPQ also reported high reliability for the Mood and Interest subscales (Petry et al., 2010). The Mood subscale has been validated against the ABC Irritability and Lethargy subscales; as has the Interest and Pleasure subscale with the ABC Lethargy subscale – as outlined above.

The ADAMS and DEPRESSED both had at least two dimensions of reliability with a ++ rating, and one dimension of validity with a + rating. However, neither measure had more than one supporting study for any dimension of reliability or validity, so they had not been independently tested.

Discussion

Thirty-two papers reporting on 12 different measures of mental health problems (and specific dimensions of mental health problems – see Table 2) and well-being were included within this review. Twenty of the studies were undertaken on the ABC, the DASH and the DASH-II. However the evidence for other measures is limited. Only one measure specifically focuses on well-being, rather than mental health problems (Physiological Measure of Subjective Well-being), but this was deemed to be a methodologically weaker measure. Based on our interpretation of the CAPs-IDD quality appraisal, the ABC, DASH-II, and MIPQ were deemed to have a good level of support regarding methodological quality for use with individuals with severe to profound ID. Considering the best currently available evidence, as reported in this systematic review, the ABC is a reliable measure which could be useful for monitoring symptomology of depression in people with severe or profound

ID. The ABC Irritability subscale has been found to correspond with the DASH-II Mania subscale, and the ABC Lethargy subscale is negatively correlated with the MIPQ Interest and Pleasure subscale; indicating some degree of convergent validity between the specific domains. The DASH-II is a reliably strong measure of mental health problems, and can reliably be used to assess mood disorders (depression and mania), anxiety, and schizophrenia in people with severe to profound ID. Data for other dimensions were not included, as they did not align with our eligibility criteria for this review. The MIPQ can reliably measure mood in people with severe or profound ID, and the developers note that it could be used to measure depression (Ross & Oliver, 2003), as indicated by a negative correlation with the ABC Lethargy subscale.

As might have been expected, measurement methods were observational, based on proxy reports about people with severe ID. Reliability data for these measures were not always available, or were typically only evidenced within a single research study. Although the quality of evidence for the psychometric properties of the measures included in this review is generally positive, a priority for future research is further testing of existing measures. Robust measures are crucial given that they may be used to inform the health plans of people with severe or profound ID (Eng, Addison & Ring, 2013), and not all mental health treatments are risk free (e.g., anti-psychotic medication: de Kuijper et al, 2013).

A further limitation of the existing research base is not simply the lack of data, but that the published data were often reported by the developers of each measure, and most often these were from the USA only. There is, therefore, a lack of independent testing of the measures. This does give rise to the possibility of some unintentional biases within the evaluations of these measures. For example, it is possible that the measures have been evaluated within the same or similar sub-populations of people with severe ID, such as those living in larger scale residential settings, and the research evidence may not be more widely generalizable.

To date, we could find only one non-proxy report measure: the Physiological Measure of Subjective Well-being. This measure had only preliminary psychometric data but may be worth exploring in further research. Additionally, physiological measurement might prove to be problematic in practice, as the technology used (multiple sensors encased in a vest) might not be suitable for all

people with severe or profound ID such as those who have additional physical health problems or physical disabilities. Regarding proxy, or informant, reports, Perkins concluded in her 2003 literature review that proxy reports were more reliable when considering objective measurements, rather than subjective ones. This could have implications for mental health problems and well-being, and it is possible that informants may be able to more reliably identify symptoms of more easily behaviourally recognisable mental health problems (e.g., anxiety and depression) than symptoms of more complex problems (e.g., psychosis), as well as subjective well-being (Vos et al., 2010). This would likely due to the complex nature of mental health problems such as psychosis in relation to the methodological and conceptual challenges of assessment of people with severe or profound ID, as outlined earlier (e.g. Emerson, Felce & Stancliffe, 2013; Deb et al., 2001; Marston et al., 1997; Ross & Oliver, 2002; Hayes et al., 2011; Evans, Cotton, Einfeld & Florio, 1999). As proxy measurements of mental health problems in people with severe to profound ID typically depend on ratings of behaviour, it is possible that mental health problems will not be reliably identified, instead being attributed to a behavioural problem, or to the person's ID diagnosis. Diagnostic overshadowing, such as this, is a major concern when considering physical health inequalities in people with ID (Emerson & Baines, 2010), and may be of equal concern for mental health problems as it could lead to delays in diagnosis and in accessing appropriate treatment.

Only the ADAMS, ASD-CA, DASH, DASH-II, mini PAS-ADD, and PAS-ADD were reported as having cut-off data, and norming data for the measures was not presented in any of the included papers. Without this information, it is not possible to determine norms and the point of clinical significance for the measures; thus presenting a limitation for the practical application of measures of mental health or well-being for people with severe or profound ID.

Finally, there were no studies meeting the inclusion criteria that evaluated measures for children or adolescents with severe or profound ID. Although there were several screened studies including some children with severe or profound ID, the proportion of this sub-group within the sample was insufficient or no separate data were reported for the severe-profound ID sub-group. For example, there were 10 papers reporting on the Developmental Behaviour Checklist (DBC; Einfeld & Tonge, 1992) screened for inclusion. None of these included more than 70% of individuals with

severe or profound ID, or indeed reported these data separately. There are candidate mental health problems measures for children with ID that could be tested rapidly with children with severe to profound ID and this is a priority for future research. In addition to the DBC, other potentially useful measures include: the Nisonger Child Behavior Rating Form (NCBRF; Aman, Tassé, Rojahn, & Hammer, 1996) and the Reiss Screen for use with children (Reiss & Velenti-Hein, 1994). Evidence for the DBC, for example, supports its validity and reliability in mixed ID populations and includes some evaluations independent of the developing team (Dekker, Nunn & Koot, 2002; Hastings et al., 2001; Einfeld & Tonge, 1995).

The current review is novel in its scope and addresses a critical review question for a vulnerable and under-served population. By undertaking double review at every stage, and adhering to PRISMA guidelines, it was possible to ensure that the methodology of this systematic review was stringent; methodological rigour is also demonstrated through the strong agreement resulting from the double review process. The consideration of the overall methodological quality of the identified measures enables researchers and clinicians alike to better understand the evidence for each measure when undertaken with a person with severe or profound ID. Although not originally intended, the lack of included studies focused on children with severe to profound ID means that the present review is limited to adults. As identified in the registered protocol, the findings of the current review are limited to the definition of mental health and well-being as described in the method, thus conditions related to psychoactive substance abuse, behavioural disorders, and organic disorders are not represented in this review. Such comorbidities are common among people with ID, and the understanding of them is valuable, however they did not fall within the scope of this review. Finally, by only including studies published after 1980 it is possible that some relevant data were excluded from this review. However, through initial screening of studies published pre-1980 by both reviewers it was established that no relevant articles had been excluded due to the implementation of this additional exclusion criterion.

Future research should seek to improve measures of mental health problems and mental well-being in both children and adults with severe or profound ID. Most of the included measures were guided by diagnostic criteria, clinical experience, or by previous measures. It is possible that a bottom-up approach to measure development would be helpful in people with severe to profound ID.

Thus, first carrying out careful observations to identify candidate indicators of mental health problems (or well-being) specifically in this population and then testing the identified items (i.e., rather than relying on existing diagnostic constructs). In general, the measures reviewed here rarely included any reference to the theoretical process of their development and so more theory-building is also needed.

Conclusions

Mental health measures are available for people with severe or profound ID. However, more research must be conducted into their validity and reliability before they can be justifiably recommended for use in practice. Further development of non-proxy report methods is also an important focus for future research adopting, for example, the use of physiological data. Further work is needed to establish the methodological quality of mental health measures specifically for children with severe or profound ID, although several candidate measures do already exist. Based on the evidence currently available, the ABC and DASH-II are both reliable measures of mood disorders, the DASH-II is also a reliable measure for anxiety and schizophrenia, and the MIPQ is a reliably good measure of mood in people with severe or profound ID.

Acknowledgements

We would like to express our gratitude to Prof Nigel Beail, Dr Umesh Chauhan, Prof Michael Kerr and Dr Howard Ring for their contributions to the development of the research proposal.

Funding

This review was funded by the Baily Thomas Charitable Fund (Reference number: TRUST/RNA/AC/SG/3543/6297), and was sponsored by the University of Warwick (Reference number: REGO-2015-1605).

References

Adams, D., & Oliver, C. (2011). The expression and assessment of emotions and internal states in individuals with severe or profound intellectual disabilities. *Clinical psychology review, 31*(3), 293-306.

Aman, M.G. & Singh, N.N. (1985). *The Aberrant Behaviour Checklist.* East Aurora, New York: Slossen.

Aman, M. G., Richmond, G., Stewart, A. W., Bell, J. C., & Kissel, R. C. (1987a). The aberrant behavior checklist: factor structure and the effect of subject variables in American and New Zealand facilities. *American Journal of Mental Deficiency, 91*(6), 570-578.

Aman, M. G., Singh, N. N., & Turbott, S. H. (1987b). Reliability of the Aberrant Behavior Checklist and the effect of variations in instructions. *American Journal of Mental Deficiency*, *92*(2), 237-240.

Aman, M. G., Tassé, M. J., Rojahn, J., & Hammer, D. (1996). The Nisonger CBRF: A child behavior rating form for children with developmental disabilities. *Research in Developmental Disabilities, 17,* 41–57. doi: 10.1016/0891-4222(95)00039-9

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Bamburg, J. W., Cherry, K. E., Matson, J. L., & Penn, D. (2001). Assessment of schizophrenia in persons with severe and profound mental retardation using the Diagnostic Assessment for the Severely Handicapped-II (DASH-II). *Journal of Developmental and Physical Disabilities*, *13*(4), 319-331. doi: 10.1023/A:1012218611103

Bihm, E. M. & Poindexter, A. R. (1991). Cross-validation of the factor structure of the Aberrant Behavior Checklist for persons with mental retardation. *American Journal of Mental Retardation, 96*(2), 209-211.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284. doi: 10.1037/1040-3590.6.4.284

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*.

Cooper, C. (2007). *The depression scale for severe disability: A diagnostic tool for the assessment of depression in adults with severe and profound intellectual disabilities*. (Unpublished doctoral thesis). Louisiana State University and Agricultural and Mechanical College.

Cooper, S-A., Smiley, E., Morrison, J., Williamson, A., & Allan, L. (2007a). Mental ill-health in adults with intellectual disabilities: prevalence and associated factors. *British Journal of Psychiatry, 190*, 27-35. doi: 10.1192/bjp.bp.106.022483

Cooper S.-A., Smiley, E., Finlayson, A., Jackson, L., Allan, L., Williamson, A., Mantry, D. & Morrison, J. (2007b). Prevalence, incidence and factors predictive of mental ill-health in adults with profound intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 20*, 493-501. doi: 10.1111/j.1468-3148.2007.00401.x

de Kuijper, G., Mulder, H., Evenhuis, H., Scholte, F., Visser, F., & Hoekstra, P. J. (2013). Determinants of physical health parameters in individuals with intellectual disability who use long-term antipsychotics. *Research in Developmental Disabilities*, *34*(9), 2799-2809. doi: 10.1016/j.ridd.2013.05.016

Deb, S., Thomas, M. & Bright, C. (2001). Mental disorder in adults with intellectual disability. Prevalence of functional psychiatric illness among community-based population between aged 16 and 64 years. *Journal of Intellectual Disability Research, 45*, 495–505.

Dekker, M. C., Nunn, R., & Koot, H. M. (2002). Psychometric properties of the revised Developmental Behaviour Checklist scales in Dutch children with intellectual disability.

Journal of Intellectual Disability Research, 46(1), 61-75. doi: 10.1046/j.1365-2788.2002.00353.x

Dumont, E., Kroes, D., Korzilius, H., Didden, R., & Rojahn, J. (2014). Psychometric properties of a Dutch version of the Behavior Problems Inventory-01 (BPI-01). *Research in Developmental Disabilities, 35*(3), 603-610. doi: 10.1016/j.ridd.2014.01.003

Einfeld, S. L., & Tonge, B. J. (1992). *Manual for the Developmental Behaviour Checklist.* Clayton, Melbourne and Sydney: Monash University for Developmental Psychiatry and School of Psychiatry, University of New South Wales.

Einfeld, S. L., & Tonge, B. J. (1995). The Developmental Behavior Checklist: The development and validation of an instrument to assess behavioral and emotional disturbance in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders, 25*(2), 81-104. doi: 10.1007/BF02178498

Einfeld, S. L., Piccinin, A. M., Mackinnon, A., Hofer, S. M., Taffe, J., Gray, K. M., Bontempo, D. E., Hoffman, L. R., Parmenter, T., & Tonge, B. J. (2006). Psychopathology in young people with intellectual disability. *Journal of the American Medical Association, 296,* 1981-1989. doi: 10.1001/jama.296.16.1981

Emerson, E., & Hatton, C. (2007). Mental health of children and adolescents with intellectual disabilities in Britain. *British Journal of Psychiatry, 191,* 493–499. doi: 10.1192/bjp.bp.107.038729

Emerson, E., Felce, D., & Stancliffe, R. J. (2013). Issues concerning self-report data and population-based data sets involving people with intellectual disabilities. *Intellectual and developmental disabilities, 51*(5), 333-348.

Eng, B., Addison, P., & Ring, H. (2013). A guide to intellectual disability psychiatry assessments in the community. *Advances in Psychiatric Treatment, 19*(6), 429-436. doi: 10.1192/apt.bp.113.011213

Esbensen, A. J., Rojahn, J., Aman, M. G., & Ruedrich, S. (2003). Reliability and validity of an assessment instrument for anxiety, depression, and mood among individuals with mental retardation. *Journal of Autism and Developmental Disorders*, *33*(6), 617-629. doi: 10.1023/B:JADD.0000005999.27178.55

Evans, K. M., Cotton, M. M., Einfeld, S. L., & Florio, T. (1999). Assessment of depression in adults with severe or profound intellectual disability. *Journal of Intellectual and Developmental Disability*, *24*(2), 147-160.

Hardoon, S., Hayes, J. F., Blackburn, R., Petersen, I., Walters, K., Nazareth, I., & Osborn, D. P. (2013). Recording of severe mental illness in United Kingdom primary care, 2000–2010. *PLoS One*, *8*(12), e82365. doi: 10.1371/journal.pone.0082365

Hastings, R. P., Brown, T., Mount, R. H., & Cormack, K. M. (2001). Exploration of psychometric properties of the Developmental Behavior Checklist. *Journal of Autism and Developmental Disorders*, *31*(4), 423-431. doi: 10.1023/A:1010668703948

Hayes, S., McGuire, B., O'neill, M., Oliver, C., & Morrison, T. (2011). Low mood and challenging behaviour in people with severe and profound intellectual disabilities. *Journal of Intellectual Disability Research*, *55*(2), 182-189.

Hermans, H., & Evenhuis, H. M. (2010). Characteristics of instruments screening for depression in adults with intellectual disabilities: Systematic review. *Research in Developmental Disabilities*, *31*, 1109-1120. doi: 10.1016/j.ridd.2010.04.023

Hermans, H., van der Pas, F. H., & Evenhuis, H. M. (2011). Instruments assessing anxiety in adults with intellectual disabilities: A systematic review. *Research in Developmental Disabilities*, *32*, 861-870. doi: 10.1016/j.ridd.2011.01.034

Hermans, H., Jelluma, N., van der Pas, F. H., & Evenhuis, H. M. (2012). Feasibility, reliability and validity of the Dutch translation of the Anxiety, Depression And Mood Scale in older adults

with intellectual disabilities. *Research in Developmental Disabilities*, *33*(2), 315-323. doi: 10.1016/j.ridd.2011.09.018

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2002). Applied statistics for the behavioural sciences (5th Edition).

Hill, J., Powlitch, S., & Furniss, F. (2008). Convergent validity of the aberrant behavior checklist and behavior problems inventory with people with complex needs. *Research in Developmental Disabilities*, *29*(1), 45-60. doi: 10.1016/j.ridd.2006.10.002

Hove, O., & Havik, O. E. (2010). Developmental level and other factors associated with symptoms of mental disorders and problem behaviour in adults with intellectual disabilities living in the community, *Social Psychiatry and Psychiatric Epidemiology*, *45*, 105-113. doi: 10.1007/s00127-009-0046-0

Janssen, R. & Maes, B (2013). Psychometric evaluation of a Dutch version of the Mini PAS-ADD for assessing psychiatric disorders in adults with different levels of intellectual disability. *Journal of Intellectual Disability Research*, *57*(8), 689-702. doi: 10.1111/j.1365-2788.2012.01544.x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Liu, K. P. Y., Lee, T., Yan, A., Siu, C. W. M., Choy, F. W. Y., Leung, K. L. K., Siu, T. Y., & Kwan, A. C. S. (2007). Use of the Interact Short Form as a tool to evaluate emotion of people with profound intellectual disabilities. *Journal of Intellectual Disability Research*, *51*(11), 884-891. doi: 10.1111/j.1365-2788.2007.00961.x

LoVullo, S. V. & Matson, J. L. (2009). Comorbid psychopathology in adults with Autism Spectrum Disorders and intellectual disabilities. *Research in Developmental Disabilities*, *30*(6), 1288-1296. doi: 10.1016/j.ridd.2009.05.004

Marston, G. M., Perry, D. W., & Roy, A. (1997). Manifestations of depression in people with intellectual disability. *Journal of Intellectual Disability Research, 41*(6), 476-480.

Matson, J. L. (1995). *The Matson Evaluation of Social Skills for Individuals with Severe Retardation*. Baton Rouge, LA: Scientific Publishers, Inc.

Matson, J. L., Coe, D. A., Gardner, W. I., & Sovner, R. (1991a). A factor analytic study of the Diagnostic Assessment for the Severely Handicapped Scale. *Journal of Nervous and Mental Disease, 179*(9), 553-557.

Matson, J. L., Gardner, W. I., Coe, D. A., & Sovner, R. (1991b). A scale for evaluating emotional disorders in severely and profoundly mentally retarded persons. Development of the Diagnostic Assessment for the Severely Handicapped (DASH) scale. *The British Journal of Psychiatry, 159*(3), 404-409. doi: 10.1192/bjp.159.3.404

Matson, J. L. & Smiroldo, B. B. (1997). Validity of the mania subscale of the Diagnostic Assessment for the Severely Handicapped-II (DASH-II). *Research in Developmental Disabilities, 18*(3), 221-225. doi: 10.1016/S0891-4222(97)00005-X

Matson, J. L., Rush, K. S., Hamilton, M., Anderson, S. J., Bamburg, J. W., Baglio, C. S., Williams, D., & Kirkpatrick–Sanchez, S. (1999). Characteristics of depression as assessed by the Diagnostic Assessment for the Severely Handicapped-II (DASH-II). *Research in Developmental Disabilities, 20*(4), 305-313. doi: 10.1016/S0891-4222(99)00012-8

Matson, J.L., Terlonge, C. & González, M.L. (2006). *Autism Spectrum Disorders- Comorbidity-Adult Version.* Baton Rouge, La. Disability Consultants, LLC.

Matson, J. L. & Boisjoli, J. A. (2008). Autism spectrum disorders in adults with intellectual disability and comorbid psychopathology: Scale development and reliability of the ASD-CA. *Research in Autism Spectrum Disorders, 2*(2), 276-287. doi: 10.1016/j.rasd.2007.07.002

Matson, J. L., Belva, B. C., Hattier, M. A., & Matson, M. L. (2012). Scaling methods to measure psychopathology in persons with intellectual disabilities. *Research in Developmental Disabilities, 33*, 549-562. doi: 10.1016/j.ridd.2011.10.023

Mellor, K., & Dagnan, D. (2005). Exploring the concept of alexithymia in the lives of people with learning disabilities. *Journal of Intellectual Disabilities, 9*(3), 229-239.

Moss, S., Prosser, H., Costello, H., Simpson, N., Patel, P., Rowe, S., Turner, S., & Hatton, C. (1998). Reliability and validity of the PAS-ADD Checklist for detecting psychiatric disorders in adults with intellectual disability. *Journal of Intellectual Disability Research, 42*(2), 173-183. doi: 10.1046/j.1365-2788.1998.00116.x

Myrbakk, E., & von Tetzchner, S. (2008). Screening individuals with intellectual disability for psychiatric disorders: Comparison of four measures. *American Journal on Mental Retardation, 113*(1), 54-70. doi: 10.1352/0895-8017(2008)113[54:SIWIDF]2.0.CO;2

Ono, Y. (1996). Factor validity and reliability for the Aberrant Behavior Checklist-Community in a Japanese population with mental retardation. *Research in Developmental Disabilities, 17*(4), 303-309. doi: 10.1016/0891-4222(96)00015-7

Paclawskyj, T. R., Matson, J. L., Bamburg, J. W., & Baglio, C. S. (1997). A comparison of the Diagnostic Assessment for the Severely Handicapped-II (DASH-II) and the Aberrant Behavior Checklist (ABC). *Research in Developmental Disabilities, 18*(4), 289-298.

Perez-Achiaga, N., Nelson, S., & Hassiotis, A. (2009). Instruments for the detection of depressive symptoms in people with intellectual disabilities: A systematic review. *Journal of Intellectual Disabilities, 13*, 55-76. doi: 10.1177/1744629509104487

Petry, K., Kuppens, S., Vos, P., & Maes, B. (2010). Psychometric evaluation of the Dutch version of the Mood, Interest and Pleasure Questionnaire (MIPQ). *Research in Developmental Disabilities, 31*(6), 1652-1658. doi: 10.1016/j.ridd.2010.04.011

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences. A practical guide*. London: Blackwell Publishing.

Prosser, H., Moss, S., Costello, H., Simpson, N., Patel, P., & Rowe, S. (1998). Reliability and validity of the Mini PAS-ADD for assessing psychiatric disorders in adults with intellectual disability. *Journal of Intellectual Disability Research, 42*(4), 264-272. doi: 10.1046/j.1365-2788.1998.00146.x

Reardon, T. C., Gray, K. M., & Melvin, G. A. (2015). Anxiety disorders in children and adolescents with intellectual disability: Prevalence and assessment. *Research in developmental disabilities, 36*, 175-190.

Redding, M. M. (1997). *Construct and criterion-related validity of the Diagnostic Assessment for the Severely Handicapped (DASH) scale.* (Unpublished doctoral thesis). Mississippi State University.

Reiss, S. (1988). Test manual for the Reiss Screen for maladaptive behavior. *International Diagnostic Systems Inc., Worthington, OH*.

Reiss, S., & Valenti-Hein, D. (1994). Development of a psychopathology rating scale for children with mental retardation. *Journal of Consulting and Clinical Psychology, 62*, 28–33. doi: 10.1037/0022-006X.62.1.28

Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2001). The Behavior Problems Inventory: An instrument for the assessment of self-injury, stereotyped behavior and aggression/destruction in individuals with developmental disabilities. *Journal of Autism and Developmental Disorders, 31*, 577–588.

Rojahn, J., Aman, M. G., Matson, J. L., & Mayville, E. (2003). The aberrant behavior checklist and the behavior problems inventory: Convergent and divergent validity. *Research in Developmental Disabilities, 24*(5), 391-404. doi: 10.1016/S0891-4222(03)00055-6

Ross, E., & Oliver, C. (2002). The relationship between levels of mood, interest and pleasure and 'challenging behaviour'in adults with severe and profound intellectual disability. *Journal of Intellectual Disability Research*, *46*(3), 191-197.

Ross, E., & Oliver, C. (2003). Preliminary analysis of the psychometric properties of the Mood, Interest & Pleasure Questionnaire (MIPQ) for adults with severe and profound learning disabilities. *British Journal of Clinical Psychology*, *42*(1), 81-93. doi: 10.1348/014466503762842039

Sevin, J. A., Matson, J. L., Williams, D., & Kirkpatrick-Sanchez, S. (1995). Reliability of emotional problems with the Diagnostic Assessment for the Severely Handicapped (DASH). *British Journal of Clinical Psychology*, *34*, 93-94. doi: 10.1111/j.2044-8260.1995.tb01442.x

Sheehan, R., Hassiotis, A., Walters, K., Osborn, D., Strydom, A., & Horsfall, L. (2015). Mental illness, challenging behaviour, and psychotropic drug prescribing in people with intellectual disability: UK population based cohort study. doi: 10.1136/bmj.h4326

Smiley, E., Cooper, S-A., Finlayson, J., Jackson, A., Allan, L., Mantry, D., McGrother, C., McConnachie, A., & Morrison, J. (2007). Incidence and predictors of mental ill-health in adults with intellectual disabilities: Prospective study. *British Journal of Psychiatry*, *191*, 313-319. doi: 10.1192/bjp.bp.106.031104

Sturmey, P., Burcham, K. J., & Perkins, T. S. (1995). The Reiss Screen for Maladaptive Behaviour: its reliability and internal consistencies. *Journal of Intellectual Disability Research*, *39*(3), 191-195. doi: 10.1111/j.1365-2788.1995.tb00500.x

Sturmey, P., Jamieson, J., Burcham, J., Shaw, B., & Bertman, L. (1996). The factor structure of the Reiss Screen for Maladaptive Behaviors in institutional and community populations. *Research in Developmental Disabilities*, *17*(4), 285-291. doi: 10.1016/0891-4222(96)00013-3

Sturmey, P., Matson, J. L., & Lott, J. D. (2004). The factor structure of the DASH-II. *Journal of Developmental and Physical Disabilities*, *16*(3), 247-255. doi: 10.1023/B:JODD.0000032300.05833.d1

Vargas-Vargas, C., Rafanell, A., Montalvo, D., Estarlich, M., Pomarol-Clotet, E., & Sarró, S. (2015). Validity and reliability of the Spanish version of the diagnostic assessment for the severely handicapped (DASH-II). *Research in Developmental Disabilities*, *36*, 537-542. doi: 10.1016/j.ridd.2014.10.034

Vereenooghe, L., & Langdon, P. E. (2013). Psychological therapies for people with intellectual disabilities: A systematic review and meta-analysis. *Research in Developmental Disabilities*, *34*, 4085-4102. doi: 10.1016/j.ridd.2013.08.030

Vereenooghe, L., Flynn, S., Hastings, R.P., Adams, D., Chauhan, U., Cooper, S-A., Gore, N., Hatton, C., Hood, K., Jahoda, A., Langdon, P.E., McNamara, R., Oliver, C., Roy, A., Totsika, V., & Waite, J. (submitted). Interventions for mental health problems in children and adults with severe intellectual disabilities: A systematic review. *Journal of Intellectual Disabilities Research.*

Vos, P., De Cock, P., Petry, K., Van Den Noortgate, W., & Maes, B. (2010). Do You Know What I Feel? A First Step Towards a Physiological Measure of the Subjective Well-Being of Persons With Profound Intellectual and Multiple Disabilities. *Journal of Applied Research in Intellectual Disabilities*, *23*(4), 366-378. doi: 10.1111/j.1468-3148.2010.00553.x

Walton, C. A. C., & Kerr, M. (2016). Prader Willi syndrome: systematic review of the prevalence and nature of presentation of unipolar depression. *Advances in Mental Health and Intellectual Disabilities*, *10*(3). doi: 10.1108/AMHID-08-2015-0037

World Health Organisation. (2014). Mental health: a state of well-being. Available at: http://www.who.int/features/factfiles/mental_health/en/ (Accessed 23 June 2017).

Zeilinger, E. L., Nader, I. W., Brehmer-Rinderer, B., Koller, I., & Weber, G. (2013). CAPs-IDD: Characteristics of Assessment Instruments for Psychiatric Disorders in Persons with Intellectual Developmental Disorders. *Journal of Intellectual Disability Research*, *57*, 737-746. doi: 10.1111/jir.12003

Table 1. Example search strategy (for ProQuest databases)*

Psychological Assessments	Mental Health/Well-being	Intellectual Disabilities
(MJSUB.EXACT("Psychological Assessment") OR TI(clinical NEAR/2 (outcome* OR diagnosis OR evaluat*)) OR TI(assess* OR index* OR instrument* OR interview* OR inventor* OR item* OR measure* OR subscale* OR scale* OR screen* OR tool* OR survey* OR self-report* OR test*form OR observ* OR rating* OR rated OR score*) OR TI(validat* OR validity OR reliab* OR accura* OR sensitive* OR specific* OR predictab*))	(SU.EXACT.EXPLODE("Mental Disorders") OR (TI(mental* NEAR/2 (ill* OR well-being OR health* OR disease* OR disorder* OR abnormal* OR patholog* OR problem* OR condition*)) OR AB(mental* NEAR/2 (ill* OR well-being OR health* OR disease* OR disorder* OR abnormal* OR patholog* OR problem* OR condition*))) OR (TI(psych* NEAR/2 (ill* OR well-being OR health* OR disease* OR disorder* OR abnormal* OR patholog* OR problem* OR condition*)) OR AB(psych* NEAR/2 (ill* OR well-being OR health* OR disease* OR disorder* OR abnormal* OR patholog* OR problem* OR condition*))) OR (SU.EXACT.EXPLODE("Depression (Emotion)") OR SU.EXACT.EXPLODE("Anxiety Disorders") OR SU.EXACT.EXPLODE("Personality Disorders")) OR (TI(anger NEAR/3 (problem* OR disorder*)) OR AB(anger NEAR/3 (problem* OR disorder*))) OR (TI(anxiet* OR anxious* OR "gad" OR phobia OR phobic OR traum* OR	(SU.EXACT.EXPLODE("Intellectual Development Disorder") OR (TI(mental* NEAR/3 (disab* OR impair* OR handicap* OR subnormal* OR deficien* OR retard*)) OR AB(mental* NEAR/3 (disab* OR impair* OR handicap* OR subnormal* OR deficien* OR retard*))) OR (TI(learning NEAR/3 (disab* OR impair* OR difficult* OR disorder*)) OR AB(learning NEAR/3 (disab* OR impair* OR difficult* OR disorder*))) OR (TI(moron* OR imbecile* OR feeble*minded OR subnormal* OR retard*) OR AB(moron* OR imbecile* OR feeble*minded OR subnormal* OR retard*)) OR (TI(intellectual* NEAR/3 (disab* OR impair* OR handicap* OR disorder* OR subnormal* OR deficien*)) OR AB(intellectual* NEAR/3 (disab* OR impair* OR handicap* OR disorder* OR subnormal* OR deficien*))) OR (TI((low*functioning OR severe) NEAR/3 autis*) OR AB((low*functioning OR severe) NEAR/3 autis*)) OR (TI("Smith-Magenis" OR Rett* OR "Lesch-Nyhan" OR "Prader-

	post*traumatic) OR AB(anxiet* OR anxious* OR "gad" OR phobia OR phobic OR traum* OR post*traumatic)) OR (TI(depress* NEAR/2 (disorder* OR symptom* OR behavio* OR thought*)) OR AB(depress* NEAR/2 (disorder* OR symptom* OR behavio* OR thought*))) OR (TI(dysthymi* OR dysphori* OR melancholy* OR schizophren* OR schizoaffective OR dementia OR psychosis OR psychotic OR alcoholism OR addiction OR obsessive-compulsive) OR AB(dysthymi* OR dysphori* OR melancholy* OR schizophren* OR schizoaffective OR dementia OR psychosis OR psychotic OR alcoholism OR addiction OR obsessive-compulsive)) OR (TI((psychological OR psychosocial) NEAR/2 function*) OR AB((psychological OR psychosocial) NEAR/2 function*)) OR (TI(well-being OR "quality of life") OR AB(well-being OR "quality of life")))	Willi" OR Angelman OR "fragile X" OR "Cri-du-chat" OR "Cornelia de Lange" OR "de Lange" OR "Rubinstein-Taybi" OR velocardiofacial OR DiGeorge OR "22q11.2" OR (Down* NEAR/2 syndrome)) OR AB("Smith-Magenis" OR Rett* OR "Lesch-Nyhan" OR "Prader-Willi" OR Angelman OR "fragile X" OR "Cri-du-chat" OR "Cornelia de Lange" OR "de Lange" OR "Rubinstein-Taybi" OR velocardiofacial OR DiGeorge OR "22q11.2" OR (Down* NEAR/2 syndrome))))

*Search strings were connected by the Boolean Operator “AND”

Table 2. Interpretation guidance used to determine psychometric quality of included measures

	Range	Our rating	Interpretation guidance source
Cohen's kappa	0.75 – 1	++	Landis and Koch (1977)
	0.61 – 0.75	+	
	0.41- 0.6	-	
	<0 – 0.4	--	
Correlation coefficient	0.7 – 1	++	Hinkle, Wiersma and Jurs (2002)
	0.5 – 0.7	+	
	0.3 – 0.5	-	
	0 – 0.3	--	
Cronbach's alpha	0.9 – 1	++	Cicchetti (1994)
	0.8 – 0.9	+	
	0.7 – 0.8	-	
	0.0 – 0.7	--	
Intra-class correlations	0.7 – 1	++	Cicchetti and Sparrow (1981)
	0.5 – 0.69	+	
	0.49 – 0.3	-	
	<0.29	--	

Table 3. Descriptions of assessments and demographic data from all studies

Assessment characteristics		Study characteristics		
Purpose and composition	**Administration and scoring**	**Study**	**Sample**	**ID Severity**
Aberrant Behaviour Checklist (ABC; Aman et al., 1985)				
Evaluation	Third party	Aman et al. (1987a)	N = 531; 61.4% male; Age: 33.5 (12.5); 8.4% psychosis; Country: USA and New Zealand	6.8% moderate; 27.3% severe; 67.2% profound
Broad spectrum of disorders	Behavioural observation			
5 Factors, 58 items: Irritability (15), Lethargy, Social Withdrawal (16), Stereotypic Behaviour (7), Hyperactivity, Non-compliance (16), Inappropriate speech (4)	5-7 minutes Likert (0, 1, 2): Severity, Frequency, Real time frequency	Aman et al. (1987b)	*Sample 1a:* N = 38; 100% male; Age: 19 (4.4)	*Sample 1a:* 18% moderate; 82% severe
			Sample 1b: N = 39; 100% male; Age: 25.2 (3.2)	*Sample 1b:* 11% moderate; 79% severe; 10% profound
			Sample 2: N = 42; 100% male; Age: 35.1 (11.8) Country: not reported	*Sample 2:*4% severe; 96% profound
		Bihm & Poindexter (1991)	N = 470; 53% male; Age: 27.07 (8.72); Country: USA	7% moderate; 21% severe; 72% profound
		Dumont et al. (2014)	N = 195; 57% male; Age: 41.3 (15.3); Country: The Netherlands	11% moderate; 46% severe; 43% profound
		Hill, Powlitch & Furniss (2008)	N = 69; 84% male; Age: 18.42; Country: UK	100% severe
		Ono (1996) [Japanese]	N = 322; 58.4% male; Age: 29.79 (12.45); Country: Japan	30% moderate; 48% severe; 22% profound

		Paclawskyj et al. (1997)	N = 233; 55.4% male; Age range: 0-71+; Country: USA	2.2% mild; 5.2% moderate; 15.1% severe; 75.4% profound; 2.1% unknown
		Petry et al. (2010) [Dutch]	N = 360; 50.8% male; Age: 42.2 (12.9); 9.6% autism; 24.9% mental health problems; Country: Belgium	100% severe or profound
		Rojahn et al. (2003)	N = 226; 55.8% male; Age range: 20-91; 17.3% Stereotyped Movement Disorder; 4.9% Bipolar disorder; 4.4% Pervasive Developmental Disorder; 4.4% autism; Country: USA	3.1% mild; 7.1% moderate; 18.1% severe; 61.9% profound; 9.7% unspecified
		Ross & Oliver (2003)	N = 53; 60.9% male; Age: 39.36 (9.9); 3.8% autism; 18.9% mental health problems; Country: UK	100% severe or profound
Anxiety, Depression and Mood Scale (ADAMS; Esbensen et al., 2003)				
Evaluation	Third party	Esbensen et al. (2003)	*Sample 1:* N = 265; 51.9% male; Age: 39.2 (11.3)	*Sample 1:* 4.8% borderline; 24.1% mild; 23.7% moderate; 15.4% severe; 25.6% profound
Broad spectrum of disorders	Behavioural observation			
5 subscales, 28 items: Manic/hyperactive behaviour (5), Depressed mood (7), Social avoidance (7), General anxiety (7), Obsessive/compulsive behaviour (3)	Likert (0-3): Frequency, Severity		*Sample 2:* N = 268; 53.2% male; Age: 39 (13)	*Sample 2:* 4.5% borderline; 20.7% mild; 28.6% moderate; 15.8% severe; 25.2% profound
			Sample 3: N = 129; 52.3% male; 42.4 (12.7)	*Sample 3:* 2.4% borderline; 10.3% mild; 20.6% moderate; 19.8% severe; 46% profound

			Country: USA	
		Hermans et al. (2012) [Dutch]	N = 975; 51% male; Age: 62.2 (8.1); Country: The Netherlands	3.2% borderline; 20.6% mild; 47.9% moderate; 16.9% severe; 9.1% profound
Autism Spectrum Disorders-Comorbidity for Adults (ASD-CA; Matson, Terlonge & Gonzalez, 2006)				
Screening	Third party	LoVullo & Matson (2009)	*Sample 1:* N = 151; 53% male; Age: 55 (14.1)	*Sample 1:* 4% unspecified; 1.3% mild; 7.9% moderate; 23.8% severe; 62.9% profound
Broad spectrum of disorders	Behavioural observation			
5 subscales, 37 items: Anxiety/Repetitive Behaviours, Conduct Problems, Irritability/Behavioural Excesses, Attention/Hyperactivity/ Impulsivity, Depressive Symptoms	Likert (0-1): Severity		*Sample 2:* N = 120; 55.8% male; Age: 49 (11.5); 100% autism	*Sample 2:* 3.3% unspecified; 1.7% moderate; 5% severe; 90% profound
			Sample 3: N = 42; 61.9% male; Age: 48 (11.8); 100% autism and Axis I diagnosis	*Sample 3:* 7.1% unspecified; 4.8% moderate; 2.4% severe; 85.7% profound
			Country: USA	
		Matson & Boisjoli (2008)	N = 169; 57.4% male; Age: 48.59; Country: USA	4.7% unspecified; 2.4% moderate; 4.1% severe; 88.8% profound
Depression Scale for Severe Disability (DEPRESSED; Cooper, 2007)				
Developed for severe ID	Third party	Cooper (2007) [unpublished doctoral thesis]	N = 144; 54.2% male; Age: 51.6 (13.3); 54.9% Axis I diagnosis; Country: USA	38.2% severe; 61.8% profound
Screening	Behavioural observation			
Depression	15 minutes			

4 factors, 20 items: Sleep, Mood, Skills, Motor	Likert (0-2): Severity, Frequency, Duration			
Diagnostic Assessment for the Severely Handicapped Scale (DASH; Matson et al., 1991a)				
Developed for severe ID	Third party	Matson et al. (1991a)	N = 506; 51% male; Age: 37.7; Country: USA	37% severe; 63% profound
Screening	Behavioural observation	Matson et al. (1991b)	N = 506; 51% male; Age: 37.7 (14.8); Country: USA	32.3% severe; 62.7% profound
Broad spectrum of disorders	Likert (0-2): Severity, Frequency, Duration	Redding (1997) [unpublished doctoral thesis]	N = 475; 57.7% male; Age range: 7-83; 31.8% Axis I or II diagnosis; Country: USA	18.3% severe; 81.7% profound
13 subscales, 84 items: anxiety, mood disorder-depression, mood disorder-mania, autism, schizophrenia, stereotypies/tics, self-injurious behaviours, elimination disorders, eating disorders, sleep disorders, sexual disorders, organic syndromes, impulse control and miscellaneous problems.		Sevin et al. (1995)	N = 658; 58% male; Age: 39; Country: USA	33% severe; 67% profound
Diagnostic Assessment for the Severely Handicapped Scale-II (DASH-II; Matson, 1995)				
Developed for severe ID	Third party	Bamburg et al. (2001)	*Sample 1:* N = 20; 55% male; Age: 47.7; 100% schizophrenia	*Total sample:* 82% severe; 18% profound
Screening	Behavioural observation 25 minutes			
Broad spectrum of disorders	Likert (0-2): Severity, Frequency, Duration		*Sample 2:* N = 20; 55% male; Age: 43.7	
13 subscales, 84 items: Anxiety, Depression, Mania, PDD/Autism, Schizophrenia,			*Sample 3:* N = 20; 55% male; Age: 45.1	

Stereotypies, Self-Injury, Elimination, Eating, Sleep, Sexual, Organic, Impulse Control			Country: USA	
		Matson & Smiroldo (1997)	N = 44; 40.9% male; Age: 44 (13.44); 50% bipolar; Country: USA	27.3% severe; 72.7% profound
		Matson et al. (1999)	N = 57; 43.9% male; Age range: 22-79; 31.6% depression; 33.3% autism; Country: USA	100% severe or profound
		Myrbakk & von Tetzchner (2008)	N = 126; 54.7% male; Age: 39; Country: Norway	13.5% moderate; 48.4% severe; 38.1% profound
		Paclawskyj et al. (1997)	N = 233; 55.4% male; Age range: 0-71+; Country: USA	2.2% mild; 5.2% moderate; 15.1% severe; 75.4% profound; 2.1% unknown
		Sturmey, Matson & Lott (2004)	N = 451; 59% male; Age: 48 (15); Country: USA	11% severe; 89% profound
		Vargas-Vargas et al. (2015) [Spanish]	N = 83; 100% female; Age: 53.92 (10.89); Country: Spain	47% severe; 53% profound
Interact Short Form (Baker & Dowling, 1995)				
Screening	Third party	Liu et al. (2007)	N = 75; 53.3% male; Age: 40.5 (13); Country: China	100% profound
Broad spectrum of disorders	Behavioural observation			
6 subscales, 12 items: Mood, speech, relating to others, relating to the environment, need for prompting, stimulation level	Likert (1-5): Frequency			
Mini Psychiatric Assessment Schedule for Adults with a Developmental Disability (Mini PAS-ADD; Prosser et al., 1997)				

Screening Broad spectrum of disorders *86 items, 7 subscales*: depression, anxiety and phobias, mania, OCD, psychosis, unspecified disorder (including dementia), PDD	Third party Clinical opinion Likert (4 point scale): Severity	Janssen & Maes (2013) [Dutch]	*Sample 1:* N = 377; 60% male; Age range <25 - >65; 8% depression; 3% anxiety disorder; 3% (hypo)mania; 5% OCD; 9% psychosis; 14% autism; 6% personality disorder; 2% dementia	*Sample 1:* 22% mild; 44% moderate; 24% severe; 9% profound; 1% missing data
			Sample 2: N = 90; 56% male; Age range: >25 - <65; 11% depression; 1% anxiety disorder; 3% (hypo)mania; 6% OCD; 15% psychosis; 16% autism; 20% personality disorder; 2% dementia	*Sample 2:* 56% mild; 34% moderate; 6% severe; 3% profound; 1% missing data
		Prosser et al. (1998)	Country: Belgium N = 68; 62% male; 57% Psychiatric condition covered by PAS-ADD; 31% Psychiatric condition not fully covered by PAS-ADD; Country: UK	Mean IQ = 35 (Range – 17-48)
Mood, Interest and Pleasure Questionnaire (MIPQ; Ross & Oliver, 2002; 2003)				
Screening Mood *2 subscales, 25 items*:	Third party Behavioural observation Likert (5 point scale): Severity	Petry et al. (2010) [Dutch]	N = 360; 50.8% male; Age: 42.2 (12.9); 9.6% autism; 24.9% mental health problems; Country: Belgium	100% severe or profound
		Ross & Oliver (2003)	N = 53; 60.9% male; Age: 39.36 (9.9); 3.8% autism;	100% severe or profound

mood (12), interest & pleasure (13)			18.9% mental health problems; Country: UK	
Psychiatric Assessment Schedule for Adults with a Developmental Disability Checklist (PAS-ADD; Moss et al., 1996)				
Screening	Third party	Moss et al. (1998)	N = 66; Age range: 16-69; Country: UK	Mean IQ = 30.5 (Range = 14-47)
Broad spectrum of disorders	Behavioural observation			
5 subscales, 27 items: depression, anxiety, psychosis, dementia, autism	Likert (4 point scale)			
Physiological Measure of the Subjective Well-Being of Persons With Profound Intellectual and Multiple Disabilities (Vos et al., 2010)				
Screening	Third party	Vos et al. (2010)	N = 3; 66.7% male; Age: 32.7; Country: Belgium	100% profound
Well-being	Behavioural observation and physiology			
Physiological (respiration, heart rate, movement, skin conductance) and observation	5 point coding system: Severity			
Reiss Screen for Maladaptive Behaviour (Reiss, 1988)				
Screening	Third party	Sturmey, Burcham & Perkins (1995)	N = 60; 55% male; Age: adults; Country: USA	16.7% mild; 11.7% moderate; 30% severe; 41.6%
Broad spectrum of disorders	Behavioural observation	Sturmey et al. (1996)	*Sample 1:* N = 102; 53% male; Median age: 38	*Sample 1:* 2% mild, 8% moderate, 16% severe, 64% profound
8 scales, 38 items: Aggressive Behaviour, Autism, Psychosis, Paranoia, Depression (Behavioural), Depression (Physical), Dependent Personality	Likert (3 point scale): Severity		*Sample 2:* N = 71; 76% male; Age: 39 (13.1) Country: USA	*Sample 2*: 7% mild, 20% moderate, 25% severe, 48% profound

Disorder, Avoidant Personality Disorder	

Age: mean age in years (Standard Deviation)

Table 4. Strength of the psychometric properties of each assessment

Assessment	Reliability			Validity		
	Internal consistency	Test-retest	Inter-rater reliability	Criterion	Content	Construct
ABC	++ (5)	++ (2)	+ (2)	+ (1)		+ (9)
ADAMS	++ (1)	+ (1)	++ (1)	+ (1)		
ASD-CA	+ (1)	+ (1)	- (1)		+ (1)	+ (2)
DEPRESSED	++ (1)	++ (1)	++ (1)			+ (1)
DASH	-- (2)	++ (1)	+ (2)	- (1)		+ (2)
DASH-II	+ (5)	++ (3)	++ (3)	++ (3)	+ (1)	++ (5)
Interact Short Form	++ (1)		+ (1)		++ (1)	++ (1)
Mini PAS-ADD	+ (2)		+ (2)	++ (1)		
MIPQ	++ (2)	+ (2)	++ (2)		++ (1)	+ (2)
PAS-ADD	+ (1)		+ (1)			++ (1)
Physiological Measure of the Subjective Well-Being of Persons with Profound and Multiple Disabilities						+ (1)
Reiss Screen for Maladaptive Behaviours	+ (1)	+ (1)	+ (1)			++ (1)

++ Excellent; + Good; - Fair; -- Poor.

Numbers in parentheses indicate the number of studies which reported on a given psychometric property.

Table 5. Domains of mental health problems/well-being covered by measures rated as having a reliably good level of methodological quality for use with individuals who have severe to profound ID

Measure	**Subscale**	**Mental health problem it maps on to**	**Reliability data available**	**Validity data available**
ABC	Irritability, agitation, crying	Mood disorders	Y	Y
	Lethargy, social withdrawal	Mood disorders	Y	Y
	Stereotypic behaviour	Not included in this review	N/A	N/A
	Hyperactivity, non-compliance	Not included in this review	N/A	N/A
	Inappropriate speech	Not included in this review	N/A	N/A
DASH-II	Impulse control	Not included in this review	N/A	N/A
	Organic problems	Not included in this review	N/A	N/A
	Anxiety	Anxiety disorder	Y	N
	Mood disorders	Mood disorders	Y	Y
	Mania	Mania (mood disorders)	Y	Y
	Pervasive Developmental Disorder-Autism	Not included in this review	N/A	N/A
	Schizophrenia	Schizophrenia	Y	N

	Stereotypies	Not included in this review	N/A	N/A
	Self-injurious behaviour	Not included in this review	N/A	N/A
	Elimination disorders	Not included in this review	N/A	N/A
	Eating disorders	Not included in this review	N/A	N/A
	Sleep disorders	Not included in this review	N/A	N/A
	Sexual disorders	Not included in this review	N/A	N/A
MIPQ	Mood	Mood disorders	Y	Y
	Interest/pleasure	Well-being	Y	Y

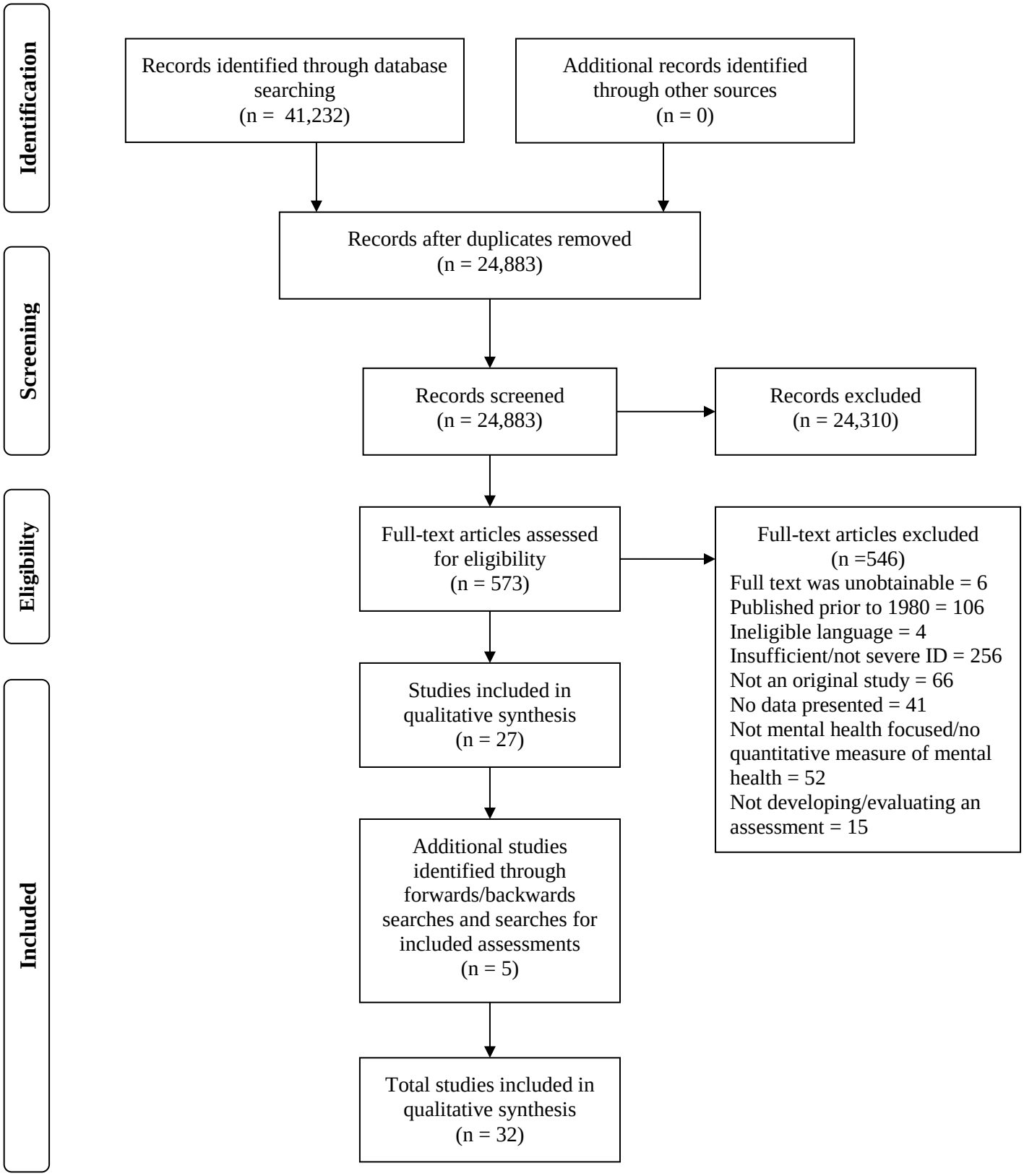


Figure 1. PRISMA flow diagram illustrating study selection