Psychometric properties of the Warwick-Edinburgh Mental Wellbeing Scale-Intellectual Disability version (WEMWBS-ID)

Maya Patel

DClinPsy thesis (Volume 1), 2022

University College London

UCL Doctorate in Clinical Psychology

Thesis Declaration Form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signatur[redacted]

Name: Maya Patel

Date: 17th June 2022

Overview

This thesis focussed on the psychometric properties of measures of mental wellbeing, for individuals with mild to moderate intellectual disabilities. The aim of the systematic review (Part 1) was to provide an update to previous evaluations of measures of common mental health problems and mental wellbeing, for adults with mild to moderate intellectual disabilities. The psychometric properties of nine measures were assessed using a quality assessment tool and a paucity of psychometrically robust measures of mental wellbeing for this population was observed.

The aim of the empirical paper (Part 2) was to assess the psychometric properties of a newly adapted version of the 14-item Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS; Tennant et al., 2007) and the Short 7-item WEMWBS (SWEMWBS), for individuals with mild to moderate intellectual disabilities (WEMWBS-ID/SWEMWBS-ID). This adapted version was developed by the primary research supervisor, experts by experience and colleagues in the intellectual disability field. The results suggest that the WEMWBS-ID and SWEMWBS-ID have promising psychometric properties. This empirical study was completed jointly with another trainee clinical psychologist, who examined predictors of self-esteem in individuals with intellectual disabilities.

The critical appraisal (Part 3) describes my reflections on the process of completing the systematic review and empirical study. A consideration of how my previous experiences influenced why I chose this project, the specific challenges faced whilst completing both papers and my reflections on quantitative measurement in the field of positive mental health are discussed.

Impact Statement

This study contributes to research in the field of positive mental health, or mental wellbeing, which is considered a valuable resource for individuals and communities. Measurement in this field is crucial, in order to gauge levels of mental wellbeing in the population, whilst allowing the effectiveness of interventions to improve mental wellbeing to be evaluated. The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS; Tennant et al., 2007) has been extensively validated in both adult and adolescent populations and is widely used in the UK and cross-culturally (e.g. Stewart-Brown et al., 2011). However, a specific measure of mental wellbeing for people with intellectual disabilities is still indicated.

The systematic review informs measurement choice for researchers and clinicians who may be interested in measuring mental health and mental wellbeing in individuals with mild to moderate intellectual disabilities. It indicates a paucity of psychometrically robust measures of mental wellbeing for this population, whilst also highlighting a need for continued research efforts into the quality of measures available for people with intellectual disabilities.

The empirical study, a preliminary evaluation of a newly adapted version of the WEMWBS and the Short 7-item WEMWBS (SWEMWBS) for individuals with mild to moderate intellectual disabilities (WEMWBS-ID/SWEMWBS-ID), suggests that these scales have promising psychometric properties. The WEMWBS-ID will enable clinicians working with individuals with intellectual disabilities to measure positive aspects of wellbeing and identify their strengths without relying solely on the report of carers or staff members.

Researchers in the intellectual disability field have expressed an interest in a psychometrically robust measure of wellbeing in this population, for example, to evaluate the impact of social prescribing. It is hoped that the WEMWBS-ID will continue to be piloted by researchers in the UK and internationally and that their data may contribute to a larger-scale evaluation of the measure, to further support its reliability and validity.

References

Stewart-Brown, S. L., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S., Tennant, R., Taggart, F. & Clarke, A. (2011). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A valid and reliable tool for measuring mental well-being in diverse populations and projects. *Journal of Epidemiology & Community Health, 65*, A38-A39. https://doi.org/10.1136/jech.2011.143586.86

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J. & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 1-13. https://doi.org/10.1186/1477-7525-5-63

Contents

List of Tables and Figures

Part 1: Literature Review

Part 2: Empirical Paper

Acknowledgements

I would like to thank and recognise the contribution of the following people who made this thesis possible:

- My supervisor, Professor Katrina Scior. I have had the pleasure of working with and being supervised by Katrina for a number of years and I continue to be thankful for her encouragement, warmth and wisdom.
- The participants who took part in the study. I am grateful to them for their time, enthusiasm and interest in the project. I wish them all the very best for the future.
- My research partner, Jun Yi Lee, for her humour, pragmatism and calmness throughout the process.
- Colleagues and collaborators including Lisa Richardson, for her time, advice and support with recruitment and Dr. Nikita Hayden, for her generous support and guidance with the analysis. Additionally, Professor Sarah Stewart-Brown, Professor Richard Hastings, Dr. Yona Lunsky, Dr. Abigail Goldsmith-Sumner, Michaela Osborne, Karuna Davies and the STORM research team for their contribution to the research.
- Friends on and outside of training, thank you for your unconditional kindness and motivation.

Part 1: Literature Review

Psychometric properties of measures used to assess common mental health problems and mental wellbeing in adults with mild to moderate intellectual disabilities:

A systematic review

Abstract

Introduction: Multiple measures of mental health problems and mental wellbeing for adults with intellectual disabilities are available but investigations into their reliability and validity are still in the early stages. The aim of this systematic review was to provide an update to previous evaluations of measures of common mental health problems and wellbeing in adults with mild to moderate intellectual disabilities.

Method: A systematic search was performed across three databases (MEDLINE, PsycINFO, SCOPUS). Papers were included if participants in the study were adults aged 18+ years old, at least 50% of the sample were reported to have a mild to moderate intellectual disability, the measure assessed common mental health problems or mental wellbeing, the article was published in a peer-reviewed journal in the English language and the measure was administered in the English language. Ten papers evaluating nine measures were reviewed and the psychometric properties of these measures discussed using the Characteristics of Assessment Instructions for Psychiatric Disorders in Persons with Intellectual Developmental Disorders (Zeilinger et al., 2013) as a framework.

Results: Four measures were deemed to have promising psychometric properties, as these measures had at least one rating of “good” across both dimensions of reliability and at least one dimension of validity. Additionally, these measures were developed through consultations with mental health professionals and/or people with intellectual disabilities, thus were deemed to have good content validity.

Conclusions: This review informs measurement choice for researchers and clinicians whilst highlighting a need for continued research efforts into the quality of measures available for people with intellectual disabilities. The results were limited by incomplete psychometric

evaluations of measures available, which made comparisons difficult. A paucity of psychometrically robust measures of mental wellbeing was observed.

Introduction

The term “learning disability” is the preferred term in the UK to refer to people who have “significantly reduced ability to understand new or complex information, to learn new skills” and a “reduced ability to cope independently which starts before adulthood with lasting effects on development” (Cluley, 2018; Department of Health, 2001 p. 14). Various countries use the term “intellectual disability” as their preferred label and this is used increasingly in British learning disability policy, practice and research, reflecting the changing international context (Cluley, 2018). The social construction of labels has implications for individuals, particularly if the labels are associated with stigma, as is the case with these terms (Green et al., 2005). Some self-advocates in England have expressed a preference for the term “learning difficulty”, rather than “disability” which may be considered socially limiting (Goodley, 2011). However, this term is used in UK educational settings to refer to specific learning difficulties such as dyslexia or dyspraxia and therefore may cause confusion if used interchangeably with the term learning disability. Whilst the potential disadvantages of using the term “intellectual disability” are acknowledged, this term will be used throughout the paper in line with research pertaining to this population. This decision was made to reflect the recent trend to use this term in a range of settings in the UK (Cluley, 2018).

People with intellectual disabilities may experience higher rates of mental health problems compared with the general population (Cooper et al., 2007; Dunn et al., 2020), although estimates of prevalence are difficult to determine and range from 10-39% (Emerson & Hatton, 2007; Pouls et al., 2021). However, research suggests that the number of people with intellectual disabilities accessing mental health services is disproportionately low, compared to the prevalence rates of mental health problems in this population (Cooper et al., 2007; Whittle et al., 2018). Whittle et al. (2018) reviewed the literature pertaining to barriers and facilitators to accessing mental health services for people with intellectual disabilities. They noted that ‘diagnostic overshadowing’ (Reiss et al., 1982), the misattribution of mental health symptoms to a person’s intellectual disability rather than being identified as related to mental health, may act as a barrier to accessing care. Therefore, the process of identifying mental health problems in people with intellectual disabilities is key to facilitating access to mental health services (Chinn & Abraham, 2016).

In addition to diagnostic overshadowing, researchers have described further challenges in assessing mental health problems and mental wellbeing in people with intellectual disabilities. Hartley and MacLean (2006) posited that Likert scales had better reliability and validity among people with borderline to mild intellectual disabilities, compared to those with a more severe intellectual disability. Emerson et al. (2013) suggested that difficulties for some people with intellectual disabilities in understanding the meaning of questions, recalling information and articulating responses may pose a challenge in acquiring self-report data. They highlighted that these difficulties may be minimised by designing scales specifically for people with intellectual disabilities using simplified wording and response formats. Mellor and Dagnan (2005) also suggested that people with mild to moderate intellectual disabilities may have difficulties in expressing their own emotions. However, Dagnan and Lindsay (2004) suggested that most people with intellectual disabilities who have good functional communication skills can accurately describe their mental state, when interview schedules are appropriately adapted (Hatton & Taylor, 2013). Finlay and Lyons (2001) reported that self-report measures for people with intellectual disabilities are often administered through face-to-face interviews, due to difficulties with reading and/or requiring support in understanding and responding to items. However, this may lead to under-reporting of thoughts, behaviours and experiences as interviews may be perceived as less private and more exposing than written questionnaires (O’Keeffe et al., 2019).

An alternative or addition to self-report assessments are measures designed to be used with an informant, usually a paid carer or family member. These may be useful when there are challenges with communication or comprehension of items in measures, but there are concerns regarding their validity (Emerson et al., 2013), particularly in gathering information on sensitive topics (O’Keeffe et al., 2019). Furthermore, Bertelli et al. (2019) suggested that a carer’s own concerns and prejudices may influence their report.

Although there is a wide range of measures of mental health problems for adults with intellectual disabilities available, investigations into their reliability and validity are still in the early stages (Hatton & Taylor, 2013). To date, two systematic reviews of measures of depressive symptoms in people with intellectual disabilities have been conducted (Hermans & Evenhuis, 2010; Perez-Achiaga et al., 2009). The earlier review concluded that the Reiss Screen for Maladaptive Behaviour (RSMB; Reiss, 1988) and the Psychiatric Assessment Schedule for Adults with Developmental Disabilities Checklist (PAS–ADD; Moss et al., 1993) demonstrated robust psychometric properties. However, Hermans and Evenhuis (2010) disagreed on the utility of the PAS-ADD Checklist and the RSMB for screening for depression, as sensitivity and specificity had not been measured, though agreed that the psychometric properties of the RSMB were promising. They concluded that the Glasgow Depression Scale for people with a Learning Disability (GDS-LD; Cuthill et al., 2003) was the most promising self-report instrument, while the Assessment of Dual Diagnosis (ADD; Matson & Bamburg, 1998), the RSMB and the Children’s Depression Inventory (CDI; Kovacs, 1985) were promising informant-report measures. However, they noted that none of these informant-report measures had yet been satisfactorily assessed with regards to their psychometric properties when used with this population. Furthermore, Hermans et al. (2011)

conducted a systematic review of measures of anxiety for people with intellectual disabilities. They concluded that the Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID; Mindham & Espie, 2003) was the most robust self-report instrument, whereas the Anxiety, Depression And Mood Scale (ADAMS; Esbensen et al. 2003) was the most promising informant-report instrument.

In the field of mental health, there is a growing interest in promoting ‘positive mental health’, or mental wellbeing. Several conceptualisations of mental wellbeing have been debated, though the consensus is that wellbeing encompasses ‘feeling well’ (hedonia) and ‘functioning well’ (eudaimonia), as opposed to a mere absence of symptoms of mental illness (Cooke et al., 2016; Deci & Ryan, 2008; Keyes, 2002; Stewart-Brown et al., 2015). There is growing evidence relating to the protective effect of mental wellbeing on mental and physical health (e.g. Keyes et al., 2010; Trompetter et al., 2017) Compared to the general population, there is less research pertaining to individuals with intellectual disabilities in this area (Raczka et al., 2020).

The term ‘mental wellbeing’ has often been used interchangeably with ‘Quality of Life’ (QoL) in the literature (Cooke et al., 2016), although it has been argued that they refer to different theoretical concepts (Skevington & Böhnke, 2018), with QoL referring to “an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” (WHOQOL Group, 1995, p. 1404). Two systematic reviews to date (Li et al., 2013; Townsend-White et al., 2012) have explored the measurement of QoL in adults with intellectual disabilities and these reviews included search terms related to “wellbeing”. In the first review, the authors concluded that the Choice Questionnaire (CQ; Stancliffe & Parmenter, 1999) and the Quality of Life Questionnaire (QOLQ; Schalock & Keith, 1993) were the most psychometrically robust measures, whilst the authors of the subsequent review concluded that six out of the 24 measures of QoL they evaluated were psychometrically sound, though they did not express a preference for a particular measure. Flynn et al. (2017) recently conducted a systematic review of measures of mental health problems and mental wellbeing in children and adults with severe or profound intellectual disabilities. The Aberrant Behaviour Checklist (ABC; Aman & Singh, 1985), the Diagnostic Assessment for the Severely Handicapped Scale-II (DASH-II; Matson, 1995) and the Mood, Interest and Pleasure Questionnaire (MIPQ; Ross & Oliver, 2002, 2003) were rated as having good methodological quality for use with individuals who had severe to profound intellectual disabilities. The authors noted that tools measuring mental wellbeing in this population were lacking.

The aim of this paper is to extend the results of the aforementioned systematic reviews and provide an update to previous psychometric evaluations of measures of mental health problems and wellbeing, in adults with intellectual disabilities. This will inform choice for clinicians and researchers interested in assessing mental health problems and mental wellbeing in this population.

As Flynn et al. (2017) recently evaluated measures of mental health problems and mental wellbeing for people with severe or profound intellectual disabilities, this paper will evaluate measures used for people with mild to moderate intellectual disabilities. In addition to mental wellbeing, the present review will focus on the measurement of anxiety disorders or depression, described by NICE (2011) as “common mental health problems”, because combined, they affect more people than other mental health problems.

The review sets out to answer the following questions:

1. Which measures have been used to assess common mental health problems and mental wellbeing in adults with mild to moderate intellectual disabilities?

2. What are the psychometric properties of these measurement tools?

Method

Design

The protocol for the present review was registered with Prospero (https://www.crd.york.ac.uk/prospero/; registration number: CRD42021270069).

The PICO (Population, Intervention, Comparator, Outcome) framework was considered to guide the search strategy. The population was adults (aged 18+) with a mild to moderate intellectual disability. The intervention was the psychometric evaluation of measures. With regards to the comparator, an evaluative and descriptive tool was used to allow comparison between the measures. The outcomes of interest were symptoms of anxiety disorders and depression, and mental wellbeing.

Search Strategy

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used to inform the methodology employed. Electronic searches of the databases MEDLINE, PsycINFO and SCOPUS were conducted on 13th September 2021. In previous relevant systematic reviews, studies published by the following dates were included: December 2008 (measures of depression; Hermans & Evenhuis, 2010); February 2010 (measures of anxiety; Hermans et al., 2011); May 2011 (measures of wellbeing; Li et al., 2013). The searches were therefore limited to papers published from January 2009 to September 2021, in order to minimise overlap with previous reviews whilst providing an update to the literature.

The full list of search terms used is summarised in Table 1. Search terms were identified based on previous similar reviews (Flynn et al., 2017; Hermans & Evenhuis, 2010) and related to four headings, truncated where appropriate and combined using the Boolean terms ‘OR’ and ‘AND’. These were as follows: (1) psychometric properties (e.g. validity, reliability, quality); (2) measurement (e.g. assessment, outcome, screening, questionnaire); (3) common mental health problems and mental wellbeing (e.g. anxiety, depression, mood, quality of life); (4) intellectual disabilities (e.g. learning disability, intellectual developmental disorder). As the focus of the present review was measures used with adults, an additional term, ‘NOT’, was used to exclude papers related to children and adolescents. Additional synonyms of these headings were also used and search terms accounted for both British English and American English spelling.

Table 1

List of terms used in systematic search.

Psychometric properties	Measures	Common mental health problems and wellbeing	Intellectual disabilities	NOT
valid* or properties or quality or methodolog* or reliab* or feasib* or psychometri* or sensitiv* or specificity	assess* or outcome* or index* or inventor* or item* or measur* or subscale* or scale* or screen* or tool* or survey* or self report* or test* or rating* or score* or questionnai*	Mental health or mood or depress* or anxiet* or anxiou* or phobia or panic or traum* or gad or worry or worrie* or sadness or post traumatic or ptsd or ocd or obsessive compuls* or body dysmorph* psychosocial or wellbeing or well being or quality of life or happiness or life satisfaction or personal growth or self accept* or self actuali* exp Mental Health/ or exp Happiness/ or exp Well Being/ or exp Emotional States/	exp intellectual development disorder/ (intellectual or development* or learning*) adj2 (disabilit* or impair* or difficult* or disorder* or handicap*) mental* retard* down* syndrome	Infant* or infancy or Newborn* or Baby* or Babies or Neonat* or Preterm* or Prematur* or Postmatur* or Child* or Schoolchild* or School age* or Preschool* or Kid or kids or Toddler* or Adolescen* or Teen* or Boy* or Girl* or Minors or Puberty or Pubescen* or Prepubescen* or Paediatric* or Pediatric* or Nursery or Kindergar* or Primary school* or Secondary school* or Elementary school* or High school* or Highschool*

Inclusion Criteria

Papers were included if:

(1) The participants in the study were adults aged 18+ years old. If a study included any participants who were aged 17 or below, the paper was included if the results for the participants aged above and below 18 years old were reported separately. A cut-off of 18 years old rather than 16 was set to ensure that the measures reviewed were appropriate for adults and also to reduce the overlap with systematic reviews of measures for children and adolescents with intellectual disabilities (e.g. Halvorsen et al., 2022).

(2) At least 50% of the sample were reported to have a mild to moderate intellectual disability. This was to ensure that there was a majority of people with mild to moderat intellectual disabilities in the study sample.

(3) The measure assessed common mental health problems or mental wellbeing. NICE (2011) identified depression and anxiety disorders including generalised anxiety disorder, panic disorder, specific phobias, Obsessive Compulsive Disorder (OCD), body dysmorphic disorder, Post-Traumatic Stress Disorder (PTSD), health anxiety or social anxiety as common mental health problems. Mental wellbeing was conceptualised as encompassing dimensions of hedonia (life satisfaction and positive affect) and eudaimonia (personal growth and self-acceptance). As previously mentioned, although it has been argued that they are theoretically different, the terms mental wellbeing and 'Quality of Life' have often been used interchangeably in the literature (Cooke et al., 2016). Therefore, to ensure that all of the relevant papers were included, studies which referred to measures of either mental wellbeing or QoL were included. Measures of health-related QoL were not included due to the narrow focus on physical health-based constructs, which does not capture broader aspects of QoL or mental wellbeing.

(4) The main aim of the study was to evaluate the psychometric properties of a measure.

(5) The article was published in a peer-reviewed journal in the English language.

(6) The measure was administered in the English language. This was so that the review may inform measurement choice for fellow English-speaking researchers and clinicians.

Screening Process

See Figure 1 for the PRISMA flow diagram which summarises the systematic review screening process. The initial search yielded 3936 papers which reduced to 2434 following the removal of duplicates. The titles were initially screened and where indicated, abstracts were reviewed. Additionally, citation searches and an inspection of reference lists were undertaken to ensure that no further eligible studies were missed. The full text of potentially eligible articles (*n* = 121) was reviewed against the inclusion criteria. This process was audited by the research supervisor. One disagreement arose during this audit, as one paper met all of the inclusion criteria except for one (majority of sample reported to have a mild to moderate intellectual disability). In order for a paper to be included in the review, the minimum percentage of the sample required to have a mild to moderate intellectual disability was initially set at 70%. This was based on the methodology employed by Flynn et al. (2017), who only included studies with at least 70% of the sample reportedly having a severe or profound intellectual disability. Following discussions with the research supervisor, it was agreed that lowering this threshold to at least 50% would still ensure that the majority of a sample comprised individuals with mild to moderate intellectual disabilities, whilst reducing the likelihood of relevant papers being excluded in the present review. A consensus was reached and this paper was subsequently included in the final selection. As a result, ten eligible articles, which evaluated nine measures, were included in the review.

As the search was limited to papers published from January 2009 onwards, in order to provide a comprehensive overview of the psychometric properties of a measure, previously published articles which also reported on the quality of one of the nine included measures were reviewed for data extraction, if these studies also met the inclusion criteria.

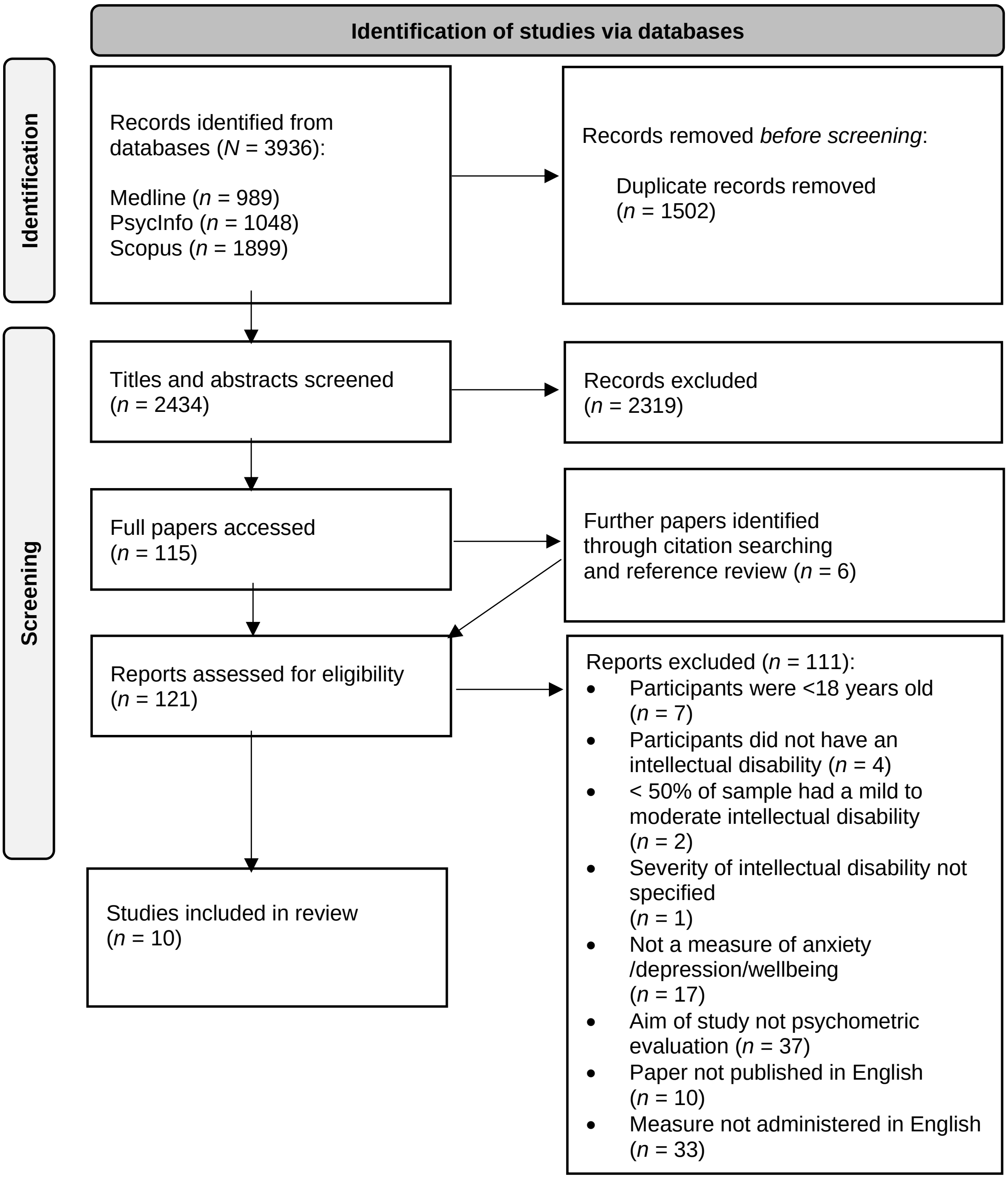


Figure 1. PRISMA flow diagram of study selection for review

Quality Appraisal

The Characteristics of Assessment Instructions for Psychiatric Disorders in Persons with Intellectual Developmental Disorders (CAPs-IDD; Zeilinger et al., 2013) is a comprehensive framework for evaluating and describing measures of psychiatric disorders in people with intellectual disabilities. The CAPs-IDD does not produce a total score, but it was used in the present review to summarise the psychometric properties of the included measures. This was guided by the methodology employed by Flynn et al. (2017), who conducted a similar review for the severe to profound intellectual disability population, also using the CAPs-IDD.

There are two parts to the CAPs-IDD. Part 1 relates to the conceptual and measurement model of an instrument, describing basic information about the measure, how the measure was developed and measurement characteristics. Part 2 pertains to psychometric properties and summarises information about the validity (criterion, content, construct and face), reliability (internal consistency, test-retest reliability and measurement error), objectivity of application, objectivity of interpretation and feasibility of a measure. The review used Part 2 of the CAPs-IDD framework (pertaining to psychometric properties) to discuss the findings of the quality appraisal. For the purpose of this review, measurement error was not discussed as no information pertaining to this was identified in the reviewed papers. Furthermore, face validity was not reviewed separately, as it overlapped with content validity in the CAPs-IDD framework, which encompassed both the relevance and comprehensiveness of items in the measures.

In order to determine the psychometric quality of included measures, they were subsequently rated on the four-point scale used by Flynn et al. (2017) (++ excellent; + good; − fair; - poor). Further information can be found in Table 2. With regards to interpreting results of factor analyses, a Root Mean Square Error of Approximation (RMSEA) of ≤ .06 and a Comparative Fit Index (CFI) of ≥ 0.95 were rated as a good fit (Hu & Bentler, 1999). It should be noted that there is variability in the literature with regards to cut-offs for ‘good’ and ‘poor’ ratings and therefore the criteria in Table 2 serve as a guide for interpretation.

A second reviewer independently extracted data from the included papers and rated the quality of the measures. Differences in ratings were discussed until a consensus was reached.

Table 2

Criteria used to interpret the results of psychometric evaluations.

Measure	Range	Rating	Source
Correlation coefficient	.7 – 1	++	Hinkle et al. (2002)
	.5 – .69	+	
	.3 – .49	–	
	< .29	-	
Cronbach’s alpha	0.9 – 1	++	Cicchetti (1994)
	0.8 – 0.89	+	
	0.7 – 0.79	–	
	0.0 – 0.69	-	
Intra-class correlations	.7 – 1	++	Cicchetti and Sparrow (1981)
	.5 – .69	+	
	.3 – .49	–	
	< .29	-	

Results

The screening process yielded ten papers, which reviewed nine measures. The measures included in the review are listed in Table 3. Five papers presented initial studies of the psychometric properties of a measure (Brooks et al., 2013; Chaplin et al., 2013; McGillivray et al., 2009; Raczka et al., 2020; Wigham et al., 2011) and four papers detailed further analyses of measures which had been previously validated in other studies (Briscoe et al., 2019; Devine et al., 2010; Rojahn et al., 2011; Wigham et al., 2021). The paper by Hall et al. (2014) included results from an initial study of a measure in addition to results from further analyses of another measure, which had been previously validated in different paper.

Table 3

List of measures included in the present review.

Measure	Acronym	Original authors
Anxiety, Depression and Mood Scale	ADAMS	Esbensen et al. (2003)
Assessment of Dual Diagnosis	ADD	Matson & Bamburg (1998)
Clinical Outcomes in Routine Evaluation-Learning Disabilities	CORE-LD	Brooks et al. (2013)
Impact of Events Scale-Intellectual Disabilities	IES-IDs	Hall et al. (2014)
Lancaster and Northgate Trauma Scales	LANTS	Wigham et al. (2011)
Mini-Maslow Assessment of Needs Scale-Learning Disabilities	Mini-MANS-LD	Raczka et al. (2020); Skirrow & Perry (2009)
Mini Psychiatric Assessment Schedule for Adults with a Developmental Disability	Mini PAS-ADD	Moss (2002); Prosser et al. (1998)
Personal Wellbeing Index-Intellectual Disability	PWI-ID	Cummins et al. (2003); McGillivray et al., (2009)
Self Assessment and Intervention (self-report section)	SAINT	Chaplin et al. (2012)

Further details about the measures and data on their reliability and validity are presented in Tables 4 and 5. Five measures included items which pertained to a broader spectrum of disorders or mental health difficulties (ADAMS, ADD, CORE-LD, Mini PAS-ADD, SAINT), two measured PTSD (IES-IDs, LANTS), and two measured QoL (Mini-MANS-LD, PWI-ID). Five of the measures reviewed were self-report (CORE-LD, IES-IDs, Mini-MANS-LD, PWI-ID, SAINT) and three were designed to be used with an informant (ADAMS, ADD, Mini-PAS-ADD). The LANTS included both a self-report and an informant-report scale. Three measures were recommended to detect changes over time and/or in response to an intervention (e.g., as a routine outcome measure; CORE-LD, Mini-MANS-LD, PWI-ID).

The CAPs-IDD tables which provide a comprehensive description of the conceptual and measurement model and psychometric properties of the nine measures are presented in Appendix A (Tables A1-A9).

Table 4

Description of measures included in the present review.

Measure	Concept measured	Recommendation for use	Experts involved in test development	Respondent	Measure content
ADAMS	Broader spectrum of disorders: manic/hyperactive behaviour, depressed mood, social avoidance, general anxiety and obsessive/compulsive behaviour	Screening	Mental health professionals	Third person: Caregiver (e.g. direct care staff, family carer, teacher)	28 items 4-point rating scale that combines frequency and severity ratings; from 0 (behaviour has not occurred or is not a problem) to 3 (behaviour occurs a lot, or is a severe problem)
ADD	Broader spectrum of disorders: mania, depression, anxiety, post-traumatic stress disorder, substance abuse, somatoform disorder, dementia, conduct disorder, pervasive developmental disorder, schizophrenia, personality disorders, eating disorders and sexual disorders	Screening	Mental health professionals	Third person: Caregiver (e.g. direct care staff, family carer, teacher)	79 items Three 3-point rating scales: (1) frequency, ranging from 0 (not at all) to 2 (more than 10 times), (2) duration, ranging from 0 (less than 1 month) to 2 (over 12 months), and (3) severity, ranging from 0 (no disruptions or damage) to 2 (caused property damage or injury)
CORE-LD	Broader spectrum of disorders: psychosocial functioning, emotional difficulties and wellbeing	Screening, research, evaluation of interventions	Mental health professionals and persons with intellectual disabilities	Person with intellectual disability	14 items 3-point rating scale with visual iconic representations. Beakers depict frequency (an empty beaker representing ‘not at all’, half-full beaker for ‘sometimes’ and a full beaker for ‘a lot’)

Measure	Concept measured	Recommendation for use	Experts involved in test development	Respondent	Measure content
IES-IDs	Specific Disorder: Anxiety disorder (PTSD)	Screening	Mental health professionals and direct care staff	Person with intellectual disability	22 items 3-point rating scale augmented with a visual scale. Respondents are asked whether they have experienced the symptom (yes/no) and then asked, ‘how much has that upset or scared you?’ (‘a little bit’, ‘in the middle’ or ‘a lot’).
LANTS	Specific Disorder: Anxiety disorder (PTSD) The informant version measures the outward presentation of a trauma effect, i.e. observable behaviours, and the self-report version measures subjective affective, biological and emotional state.	Screening, research	Mental health professionals, persons with intellectual disabilities, direct-care staff and family carers	Person with intellectual disability and caregiver (e.g. direct care staff, family carer, teacher)	Self-report scale: 29 items 4-point visual rating scale (‘no’, ‘a little’, ‘sometimes’, or ‘a lot’) indicating the frequency of subjective states experienced. Informant scale: 43 items Three subscales: behavioural changes, frequency and severity. Each question was rated for frequency on a 6-point scale (‘none’, ‘monthly’, ‘weekly’, ‘several times a week’, ‘daily’, or ‘several times a day’) and severity on a 3-point scale (‘mild’, ‘moderate’ or ‘severe’).

Measure	Concept measured	Recommendation for use	Experts involved in test development	Respondent	Measure content
Mini-MANS-LD	Other: Quality of life	Screening and evaluation of interventions	Mental health professionals and persons with ID	Person with intellectual disability	9 items 5-point Likert-type scale, using a pictorial and verbal scale. Items reflect Maslow's (1943) hierarchy of needs: physiological, safety, social, esteem and self-actualisation.
Mini PAS-ADD	Broader spectrum of disorders: depression, anxiety, hypo-mania, obsessive compulsive disorder, psychosis, unspecified disorder (including dementia) and autistic spectrum disorder	Screening	Mental health professionals	Third person: Health professional or caregiver	66 items Provides a framework for staff to collect information from informants. All items rated on a 4-point scale depending on their presence and/or level of severity over the previous 4 weeks ('symptoms not present', 'mild symptoms', 'moderate symptoms', or 'severe symptoms'.

Measure	Concept measured	Recommendation for use	Experts involved in test development	Respondent	Measure content
PWI-ID	Other: Quality of life	Screening and evaluation of interventions	Other: Scholars	Person with intellectual disability	7 items Items reflect satisfaction with standard of living, health, life achievement, personal relationships, personal safety, community-connectedness, future security. The scale incorporates a pre-testing protocol to determine which response scale to use. If the standard 11-point response scale was too complex, a visual scale (drawings of faces) with a choice of either 5, 3 or 2 points was used instead.
SAINT	Broader spectrum of disorders: feelings, emotions, and symptoms of poor mental health	Other: To assist people with intellectual disabilities in recognising and reporting symptoms of mental distress	Mental health professionals and persons with intellectual disabilities	Person with intellectual disability	10 items Self-report statements accompanied by pictures. The tool uses a dichotomous yes/no format.

Table 5

Data on the reliability and validity of measures included in the present review.

Measure	**Reliability**		**Validity**		
	Internal consistency	Test-retest	Criterion	Content	Construct
ADAMS	Total score: α = 0.90 (++) Mean subscale: α = 0.83 (+) Subscales: α = 0.78 – 0.85 (– to +)	ND	ND	ND	**Confirmatory factor analysis** (-): CFI = .89 RMSEA = .10 **Convergent** (++): ADAMS depression and ADD depressed mood subscales ($r_s = .77$, $p < .000$) ADAMS general anxiety and ADD anxiety subscales ($r_s = .75$, $p < .000$) ADAMS Manic/Hyperactive and ABC Hyperactivity subscales ($r_s = .75$, $p < .000$)
ADD	Total score: α = 0.91 (++) Mean subscale: α = 0.59 (-) Subscales: α = 0.18 – 0.84 (- to +)	ND	ND	ND	**Confirmatory factor analysis** (-): Model would not converge after over 60 iterations **Convergent** (+ to ++): ADD depressed mood and ADAMS depression subscales ($r_s = .77$, $p < .000$) ADD anxiety and ADAMS general anxiety subscales ($r_s = .75$, $p < .000$) ADD Depression subscale and ABC Lethargy subscale ($r_s = .63$, $p < .000$)

Measure	**Reliability**		**Validity**		
	Internal consistency	Test-retest	Criterion	Content	Construct
CORE-LD	$\alpha = 0.80$ (+)	1 week $n = 50$ $\rho = .64$ (+)	ND	Methodology supports content validity (items adapted from an established measure, focus group contributed to measure development)	**Convergent** (+): CORE-OM ($r_s = .68$, $p < .001$)
IES-IDs	Total score: $\alpha = 0.90 – 0.91$ (++) Subscales: $\alpha = 0.61 – 0.88$ (– to +)	2 weeks $n = 40$ Total score: ICC = .86 (++) Subscales: ICC = .65 – .85 (+ to ++)	ND	Methodology supports content validity (items adapted from an established measure, mental health professionals were consulted during measure development)	**Convergent** (- to ++): GDS self-report scale ($r = .60$, $p < .001$) GDS informant scale ($p > .05$) GAS ($r = .40$, $p < .01$) LANTS self-report scale ($r = .76$, $p < .001$) LANTS informant subscales ($r = .28 – .37$, $p < .05$) Trauma frequency ($r = .35$, $p < .05$)

Measure	**Reliability**		**Validity**		
	Internal consistency	Test-retest	Criterion	Content	Construct
LANTS	Self-report scale: $\alpha = 0.84 – 0.89$ (+) Informant subscales: $\alpha = 0.80 – 0.92$ (+ to ++)	Wigham et al. (2011): 5-6 weeks *Self-report scale:* $n = 48$ $r_s = .72, p < .01$ (++) *Informant subscales:* $n = 33$ $r_s = .57 – .59, p < .01$ (+) Hall et al. (2014): 2 weeks *Self-report scale:* $n = 40$ ICC = .92 (++) *Informant subscales:* $n = 32$ ICC = .80 – .84 (++)	ND	Methodology supports content validity (items derived by consulting a number of sources e.g. literature reviews and focus groups)	**Factor analysis of self-report scale** (–): Four factor structure which accounted for 53.46% of the variance. Seven items removed due to particularly high skewness and/or kurtosis. Kaiser-Meyer-Olkin value was acceptable (0.75). **Convergent**: *Self-report scale* (– to ++): IES (general population measure) subscales ($r_s = .58 – .62, p < .01$) IES-IDs ($r = .76, p < .001$) BLESID ($r_s = .45, p < .01$) BSI ($r_s = .62, p < .01$) *Informant subscales* (- to +): IES-IDs ($r = .28 – .37, p < .05$) PEDS ($r_s = .35 – .50, p < .01$) PAS-ADD ($r_s = .26 – .62, p < .01$) BLSEID ($r_s = .54 – .64, p < .01$) *Convergence between LANTS self-report and informant subscales* (-): Behavioural changes ($r_s = .20, p < .005$) Frequency ($r_s = .13, p > .05$) Severity ($r_s = .16, p > .05$)

Measure	**Reliability**		**Validity**		
	Internal consistency	Test-retest	Criterion	Content	Construct
Mini-MANS-LD	α = 0.74 (–)	ND	ND	Methodology supports content validity (items adapted from an established measure, experts by experience were consulted)	**Convergent** (+): PWI-ID (*r* = .67, *p* < 0.001)
Mini PAS-ADD	ND	ND	Sensitivity: 100% (informants and psychiatrist correctly identified the same five individuals who had a potential mental health problem) Specificity: 77% (informants and psychiatrist correctly identified 17 people who did not a have mental health problem; psychiatrist did not identify a mental health problem for five people,	Methodology supports content validity (items adapted from an established measure)	ND

Measure	Reliability		Validity		
	Internal consistency	Test-retest	Criterion	Content	Construct
			whereas the informants using the Mini PAS-ADD scored them above the threshold)		
PWI-ID	α = 0.76 (–)	1-2 weeks $n = 31$ $r = .67$, $p < .05$ (+) ICC = .57 (+)	ND	ND	**Convergent** (- to –): Life as a whole: $r = .27 – .44$, $p < .05$ **Factor analysis**: Two factor structure which accounted for 57.97% of the variance. However, as one of the items loaded onto both factors and the minimum number of items for a factor was not met, the analysis resolved to a single factor.
SAINT	α = 0.83 (+)	1 week $n = 20$ $r = 0.90$, $p < .01$ (++)	ND	Methodology supports content validity (Delphi methods and focus groups were employed to inform the contents of the measure)	**Convergent** (+ to ++): GDS: $r = 0.71$; $p < .01$ GAS: $r = 0.58$; $p < .01$

Note: ++ Excellent; + Good; – Fair; - Poor; ND No Data; CFI Comparative Fit Index; ICC Intraclass Correlation Coefficient; RMSEA Root Mean Square Error of Approximation

Measures: BLESID Bangor Life Events Schedule for Intellectual Disabilities (Hulbert-Williams et al., 2011); BSI Brief Symptom Inventory (Derogatis, 1993); CORE-OM Clinical Outcomes in Routine Evaluation - Outcome Measure (Evans et al., 2002); GAS Glasgow Anxiety Scale; GDS Glasgow Depression Scale; IES Impact of Events Scale (Horowitz et al., 1979); PAS-ADD Psychiatric Assessment Schedule for Adults with Developmental Disability; PEDS Paediatric Emotional Distress Scale (Saylor et al., 1999)

Sample Characteristics

A summary of the sample characteristics is presented in Table 6. Most of the included studies were conducted in the UK, in addition to one study which was conducted in Australia and another in the USA. Sample sizes ranged from 33 to 324. The percentage of male participants in each sample ranged from 40.7% to 85.9%. Seven studies reported the mean age of the participants and this ranged between 33.0 to 45.6 years old. The age range for participants was not reported in all of the papers, although they specified that only adults aged 18+ were recruited. Participants were recruited from a range of sources, including clinical services for people with intellectual disabilities, residential services and day centres. Participants' ethnicity was reported in four studies, in which at least 71.1% of the samples were White. With regards to the severity of intellectual disability, one study included participants with severe and profound intellectual disabilities (who made up < 50% of the sample) and another included participants with borderline intellectual disabilities (6.3% of the sample). The remaining studies only recruited participants with either a mild or moderate intellectual disability, although the breakdown of severity of intellectual disability within the sample was not always reported.

Table 6

Summary of sample characteristics included in the present review.

Authors	Measure(s)	Recruitment source	Country	Sample size (% Male)	Mean age in years (range)	Ethnicity	Severity of intellectual disability
Rojahn et al., (2011)	ADAMS & ADD	Referrals to a psychiatric outpatient clinic and behavioural support programme	USA	263 (66.2%)	37.8 (range not stated)	71.1% White American, 12.5% African American, 7.2% Asian/Pacific Islander, 4.2% Hispanic American, 1.1% American Indian, 3.4% Other	24% mild, 34.6% moderate, 23.3% severe, 17.5% profound
Brooks et al. (2013)	CORE-LD	Intellectual disability services (NHS or local authority), voluntary sector organisations and supported living	UK	324 (40.7%)	Mean age not stated (18–80)	83.6% White British, 7.1% Asian/Asian British, 7.1% Black/Black British, 2.2% Other	Recorded for 261 participants: 81% mild, 19% moderate
Briscoe et al. (2019)	CORE-LD	Two inpatient hospitals	UK	41 (48.8%)	36.5 (20–64)	Not stated	All participants had mild or moderate intellectual disabilities
Hall et al. (2014)	IES-IDs & LANTS	NHS teams, residential services and day centres	UK	40 (42.5%)	37.0 (range not stated)	Not stated	All participants had mild intellectual disabilities

Authors	Measure(s)	Recruitment source	Country	Sample size (% Male)	Mean age in years (range)	Ethnicity	Severity of intellectual disability
Wigham et al. (2011)	LANTS	NHS services, day centres, social services and independent service providers	UK	99 (85.9%)	Female: 45.6 (21–60) Male: 40.2 (19–75) Not stated for 19 participants	99.0% White British, 1.0% Other	All participants had mild or moderate intellectual disabilities
Wigham et al. (2021)	LANTS	NHS inpatient services and community service providers	UK	98 (85.7%)	41.0 (range not stated)	100% White British	All participants had mild or moderate intellectual disabilities
Raczka et al. (2020)	Mini-MANS-LD	Intellectual disability services	UK	33 (42.4%)	39.1 (22–69)	78.8% White British, 12.1% Black/Black British, 6.1% Asian/Asian British, 3.0% Other	87.9% mild, 12.1% moderate
Devine et al. (2010)	Mini PAS-ADD	Community-based adult services, day-care services	UK	96 (50.0%)	Mean age not stated (18–65)	Not stated	6.3% borderline, 38.5% mild, 55.2% moderate

Authors	Measure(s)	Recruitment source	Country	Sample size (% Male)	Mean age in years (range)	Ethnicity	Severity of intellectual disability
McGillivray et al. (2009)	PWI-ID	Government and non-government services or programs for people with intellectual disabilities	Australia	114 (54%)	33.0 (18–60) Not stated for 2 participants	Not stated	71.9% mild, 28.1% moderate
Chaplin et al. (2012)	SAINT	Supported group homes, NHS services, third sector and local authority providers	UK	54 (57.4%)	39.4 (18–77)	Not stated	92.6% mild, 7.4% moderate

Reliability

Internal Consistency

Internal consistency is the extent to which items in a questionnaire are correlated and therefore measure the same concept (Terwee et al., 2007). Internal consistency was assessed for eight measures and was generally high. Three measures had excellent total score internal consistencies (ADAMS, ADD, IES-IDs), three measures were rated as having "good" internal consistency (CORE-LD, LANTS, SAINT) and two were rated as "fair" (Mini-MANS-LD, PWI-ID). However, subscale internal consistencies across the measures were generally lower, as detailed in Table 5.

Test-retest

Test-retest reliability refers to the degree to which repeated administrations of a measure provide similar responses (Terwee et al., 2007). The time period between administrations is often one or two weeks, to prevent recall whilst ensuring that clinical change has not occurred (Terwee et al., 2007). Good test-retest reliability suggests that the construct being measured does not change over time (Collins, 2007) and therefore is not state-dependent. Test-retest reliability was reported for five measures (CORE-LD, IES-IDs, LANTS, SAINT, PWI-ID). The time period between administrations of the measure ranged from one to six weeks. The coefficients ranged between good (CORE-LD, LANTS informant scale, PWI-ID) and excellent (IES-IDs, LANTS self-report scale, SAINT).

Validity

Criterion Validity

Criterion validity refers to the extent to which scores on an instrument relate to a 'gold standard' measure (Terwee et al., 2007). Hatton et al. (2013) noted that an issue with testing validity stringently in this field is the lack of ‘gold standard’ measures of mental health problems for people with intellectual disabilities. Clinical opinion is still preferred by many researchers as the gold standard method (Perez-Achiaga et al., 2009). In the present review, criterion validity was evaluated only for one measure (Mini PAS-ADD), by examining the sensitivity and specificity the measure compared to an assessment by a psychiatrist. Sensitivity analysis was found to be perfect, however specificity analysis was lower.

Content Validity

Content validity refers to the extent to which concepts are represented by the items in the measure (Terwee et al., 2007). Content validity is deemed to be good if a clear description of the concept being measured and item selection is provided, in addition to the target population and experts being involved in the measure development process (Terwee et al., 2007). In the present review, six measures addressed at least one aspect of content validity (CORE-LD, IES-IDs, LANTS, Mini-MANS-LD, Mini-PAS-ADD, SAINT). The items in four of these (CORE-LD, IES-IDs, Mini-MANS-LD, Mini-PAS-ADD) were derived from established measures of mental health or wellbeing. Four measures were reported to be developed through consultation with both people with intellectual disabilities and mental health experts (CORE-LD, LANTS, Mini-MANS-LD, SAINT). Mental health professionals were consulted during the development of the IES-IDs.

Construct Validity

Construct validity refers to the extent to which measures of theoretically related constructs converge, whilst theoretically unrelated constructs do not. It also refers to the structural validity of a measure and whether different dimensions within the measure correlate.

Convergent validity was examined in all measures except for the Mini-PAS-ADD. The CORE-LD and Mini-MANS-LD were correlated with only one other measure which may limit assessment of convergent validity, while the other measures were correlated with more than one measure. There was a broad range in the strength of significant correlations with other measures and five measures had a minimum rating of ‘good’ (ADAMS, ADD, CORE-LD, Mini-MANS-LD, SAINT). Six measures demonstrated excellent convergent validity with at least one other measure (ADAMS, ADD, IES-IDs, LANTS, SAINT). It appeared that the relationship between self-report and informant-report measures was poorer, as the correlation between the IES-IDs and the GDS informant scale was not significant, whilst the correlations between the IES-IDs and the LANTS informant subscales ranged from poor to fair. The correlation between the LANTS informant subscales and PAS-ADD ranged from poor to good. Furthermore, the convergence between the LANTS self-report and informant-report subscales was also poor, as the magnitude of the correlation with the behavioural changes subscale was low, whilst the correlations with the frequency and severity subscales were not significant.

There were insufficient investigations into the factorial structures of included measures. A factor analysis was attempted for four of the measures (ADAMS, ADD, LANTS self-report scale, PWI-ID). This was rated as “poor” for the ADAMS, whilst the model would not converge for the ADD. The anticipated factor structures were confirmed for the PWI-ID and LANTS, though seven items were removed from the analysis of the LANTS due to particularly high skewness and/or kurtosis. There are different recommendations in the literature for the number of participants required for factor analytic techniques, for example, Guadagnoli and Velicer (1988) suggested n = 100–200, whereas a minimum of 1:5 item: case ratio was recommended by Floyd and Widaman (1995). With regards to the adequacy of sample sizes in the included studies which examined factor structure, this was considered acceptable for the ADAMS, ADD (both *n* = 263) and PWI-ID (*n* = 114). The sample size for the LANTS self-report study was slightly below the recommendation (*n* = 98).

Objectivity of Application, Interpretation, Norming and Fairness

With regards to the objectivity of application and interpretation, some instructions for administration of the CORE-LD and IES-IDs were reported in the published papers although instructions for coding were not. A short manual was available for administration and coding for the Mini-MANS-LD. A more comprehensive manual is available for the Mini-PAS-ADD and PWI-ID. Guidelines for administration and coding were not found through a web search for the remaining measures. With regards to the SAINT, the authors reported that it was not intended to be coded as it has not been designed as a diagnostic tool and its purpose was to help individuals with intellectual disabilities to recognise symptoms of mental distress.

In the included studies, little to no information was reported with regards to normative or comparative data from the general population. McGillivray et al. (2009) compared PWI-ID ratings with PWI ratings from general population samples and found that the total scores did not differ significantly. For included measures which were adapted from measures designed for the general population (for example, the CORE-LD and IES-IDs), normative data may be found in the published papers assessing the psychometric properties of the general population measures, for example the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2002) and The Impact of Events Scale-Revised (IES-R; Weiss & Marmar, 1997).

In terms of representativeness and generalisability, convenience samples were used in all of the included studies, which may have limited fairness concerning culture, gender and age. Data on ethnicity and age was not reported consistently in the studies. Further information may be found in the “Sample Characteristics” section of this review.

Feasibility

No information on the percentage of missing values was reported in the included papers. No information regarding the ease of administration, burden of completing the measure or acceptability was reported for the ADAMS or ADD. Briscoe et al. (2019) reported that participants indicated that they found completing the CORE-LD easier than the CORE-OM. In terms of the IES-IDs, professionals in an Adult Community Learning Disability Team were consulted to modify the language of the IES-R in order to ensure acceptability. An interviewer script was developed to increase the ease of administration so that the IES-IDs could be administered as a semi-structured interview. The LANTS was developed via consultation with a clinical sample, carers, advocates and clinicians, to ensure acceptability and inclusiveness. The authors reported that administration took between 10 – 20 minutes to complete. The Mini-MANS-LD was also developed following consultation with a group of experts by experience. Accessibility was enhanced by using pictures and colour-coded faces as prompts. The authors reported that it was rated by participants as "easy to use" and acceptable to people with intellectual disabilities and that administration took on average less than 12 minutes to complete. Information on the acceptability was not reported for the Mini PAS-ADD, but the authors reported that interviewers were provided with training in how to administer the measure and that finding appropriate time to administer the tool was identified as a difficulty by informants. The PWI-ID included a pre-testing protocol to enhance ease of administration, by identifying the level of complexity respondents were able to use the scale. Participants were given the choice of an 11-, 5-, 3-, and 2-point scale. Administration took on average between 10 – 20 minutes. Finally, the developers of the SAINT consulted with professional experts and service user experts to enhance the acceptability and feasibility of the measure.

Discussion

Summary of Results

The main aim of this review was to provide an update to previous evaluations of the psychometric properties of measures developed to assess common mental health problems and mental wellbeing in adults with mild to moderate intellectual disabilities. The psychometric properties of nine measures were examined across the ten papers considered. Although internal consistency was examined in eight of the nine measures, test-retest reliability was only assessed for five measures. Furthermore, criterion validity was only assessed for one measure. Six measures addressed at least one aspect of content validity. Convergent validity was examined for eight measures, although two measures were correlated with only one other measure. Factor analyses were attempted for four of the measures.

The results from the present review suggest that self-report and informant-report scales were poorly correlated. It has been argued that the degree of convergence between self-report and informant-report scales may not reflect validity (Stancliffe, 1995) as informants cannot directly access the subjective experiences of individuals with intellectual disabilities (Hartley & MacLean, 2006). Although informant scales offer valuable information, it has been suggested that a mental health assessment of an individual with intellectual disabilities should also include self-report questionnaires, which may add unique information about affective and cognitive symptoms which may not be apparent to caregivers (Mileviciute & Hartley, 2015).

Based on the results of the present review, the CORE-LD, IES-IDs, LANTS and SAINT were deemed to have promising psychometric properties, as these measures had at least one rating of ‘good’ across both dimensions of reliability and at least one dimension of validity. Additionally, these measures were developed through consultation with mental health professionals and/or people with intellectual disabilities, thus were deemed to have good content validity. Although Hermans and Evenhuis (2010) previously suggested the ADD was a promising informant-report measure, the present review indicates a lack of evidence on the quality of this measure for the mild to moderate adult intellectual disability population.

Several limitations and strengths of the studies which validated these four measures should be noted. With regards to the CORE-LD, Brooks et al. (2013) noted that the sample size meant that it was not possible to establish a cut-off score or to investigate whether the measure was more appropriate for some groups of people with intellectual disabilities and not others. Briscoe et al. (2019) reported that the strength of the correlation coefficient between the CORE-LD and CORE-OM was lower compared to the correlation between other related measures, for example the GAS-ID and the Beck Anxiety Inventory (Beck et al., 1988a; Mindham & Espie, 2003) or the GDS-LD and the Beck Depression Inventory (Beck et al., 1996; Cuthill et al., 2003). Furthermore, Briscoe et al. (2019) commented on the sample characteristics in their study which comprised of forensic inpatients with intellectual disabilities and other comorbidities. This sample may be unrepresentative of individuals with intellectual disabilities in the general population as it is likely that these individuals experienced a higher level of distress compared to a community sample. A strength of the CORE-LD was the emphasis on inclusivity and collaboration in the development of the measure. Brooks et al. (2013) reported receiving feedback from the individuals with intellectual disabilities who were involved developing the measure, such as, “I have enjoyed every minute of this research” and “I felt valued” (p. 328). This is incredibly important given the barriers to participation in research that individuals with intellectual disabilities face (Lennox et al., 2005).

In terms of the IES-IDs, Hall et al. (2014) reported that a limitation of their study was the small sample size, which meant that the factor structure of the measure could not be examined. However, the authors reported that a study strength was that the sample had experienced at least one traumatic event in their lives and were at risk of experiencing PTSD. This allowed an investigation into the whether there was a relationship between trauma frequency and symptomatology as measured by the IES-IDs, so that convergent validity could be assessed. The authors also contrasted the IES-IDs with the LANTS and noted that conceptually, the IES-IDs specifically assessed PTSD symptomology in response to a specific trauma, whereas the LANTS assessed more general trauma-related psychopathology, in addition to symptoms of anxiety and depression which are comorbid with PTSD.

Regarding the LANTS, Wigham et al. (2011, 2021) considered the inclusion of participants from both inpatient and community settings to be a study strength as this suggests that the LANTS may be utilised in both settings. Furthermore, the LANTS was developed via consultation with individuals with intellectual disabilities, carers and clinicians which supports their content validity. However, the samples recruited were 99-100% White British, which may limit the applicability for other ethnic groups. Wigham et al. (2011) also highlighted that the self-report version of the LANTS was only significantly correlated with one of the informant LANTS subscales and that the strength of this correlation was low. They posited that this may be because the two scales measured different aspects of trauma; the self-report version measured internal states whereas the informant scale measured observable behaviours. They suggested that construct validity was not compromised as both scales correlated significantly with the number of adverse life events experienced.

Chaplin et al. (2013) reported that a limitation of the SAINT was that test-retest reliability was assessed on a small proportion (37%) of the participants (*n* = 20) and that retest data were collected on the same day. A strength of the study was that the convergent validity was examined using the GDS-LD and GAS-ID, which were reported in previous systematic reviews to have promising psychometric properties. Furthermore, the SAINT was developed through consultation with experts and service users. Chaplin et al. (2013) reported that an advantage of the SAINT was that it measured psychological distress more generally, rather than specific symptoms of depression and anxiety, which may present similarly in people with intellectual disabilities. It is also unique, as the SAINT is not only a measure of distress, but also covers specific coping strategies to reduce distress.

Limitations of the Present Review

The inclusion criteria were selected to ensure that measures of common mental health problems and mental wellbeing for adults with intellectual disabilities were identified. The first criterion was that participants in the included studies were aged 18 years or over. This was because the aim of the review was to identify measures which may be used with adults with intellectual disabilities, for example in adult learning disability services or research pertaining to the adult intellectual disability population. Studies which included participants who were aged below 18 years old were therefore excluded, if the results for the participants aged above and below 18 years old were not reported separately. This may have limited the number of papers included.

The restriction of studies to full English publication is another limitation of the present review. Furthermore, only studies which administered measures in the English language were included and therefore 33 articles which assessed the quality of measures administered in other languages were excluded. The decision was made to only include measures which were validated in the English language, to inform measurement choice for fellow English-speaking researchers and clinicians. When selecting measures for people with intellectual disabilities in English-speaking countries, one may choose to translate measures which were validated in other languages, into English. However, it is recognised that cross-cultural adaptation of measures may be problematic for several reasons, such as the two languages having non-equivalent words, or items having very different meanings based on the specific cultural context (Epstein et al., 2015).

Suggested guidelines for cross-cultural adaptation of measures, which are commonly followed in cross-cultural research, are as follows: multiple independent initial translations; synthesis/reconciliation of these translations into a single translation; back translation and an assessment of the equivalence of the original and back-translated versions; expert committee review to reach a consensus and pretesting (Beaton et al., 2000; Epstein et al., 2015; Guillemin et al., 1993).

One of the 33 excluded papers in the present review (Hermans et al., 2012) was a psychometric evaluation of the Dutch translation of the ADAMS. In this study, three members of the research team completed an initial translation of the ADAMS into Dutch and a final consensus was reached. Subsequently, a native English speaker back-translated the Dutch version into English. With regard to this back-translation, the authors reported, “it appeared that the phrasing of his translation was different for several questions and the response format, but the purport of both versions corresponded for all questions and responses” (Hermans et al., 2012, p. 438). Although it was important to compare the back-translation with the original version and to examine any discrepancies (Kuliś et al., 2017), it was unclear how the translation differed from English version and how the authors assessed equivalence in the meaning between the two versions. This example highlights difficulties which may be faced when translating measures into other languages. Future researchers may wish to complete a review including reports on measures administered in languages other than English, to inform clinicians and researchers interested in selecting an appropriate measure in another language.

A further limitation was that the quality appraisal was somewhat limited by the lack of comprehensive evaluations of the psychometric properties of the included measures. There was very little information available on criterion and structural validity. Furthermore, little to no normative data from the general population were reported and there was a lack of information on the time taken to complete the measures or how they were scored. It was therefore difficult to make comparisons between measures as not all aspects of reliability and validity were assessed for each measure. This highlights the need for continued research efforts into the quality of measures available for people with intellectual disabilities.

Finally, it is recognised that the systematic search in the present review started from the month following the last month covered by previous systematic reviews. As databases may only be updated every few months and the searches did not overlap, it is possible that some papers published around this time point may have been missed, which is a further limitation.

Implications for Research and Practice

Researchers and clinicians may use the findings of this review to make informed decisions when choosing a mental health or wellbeing measure for adults with mild to moderate intellectual disabilities. Mileviciute and Hartley (2015) reported that self-report questionnaires may capture internalised experiences of people with intellectual disabilities, which may not be apparent to carers. Therefore, although informant questionnaires offer valuable information, assessments should also include self-reported information. The CORE-LD, IES-IDs, LANTS and SAINT all include self-report scales. The CORE-LD may be used by clinicians and researchers interested in measuring the wellbeing, psychosocial functioning and emotional difficulties experienced by adults with mild to moderate intellectual disabilities. The SAINT self-report section forms part of a guided self-help tool for people with intellectual disabilities and may be used to assist people with intellectual disabilities in recognising and reporting symptoms which indicate mental distress. Finally, the IES-IDs and LANTS may be used to screen for symptoms of PTSD. The LANTS also assesses for comorbid symptoms of anxiety and depression and, additionally, allows informants to provide information based on their observations of individuals with intellectual disabilities.

Although two measures of QoL were identified in the present review, a lack of psychometrically sound measures of mental wellbeing, encompassing dimensions of hedonia and eudaimonia, for adults with mild to moderate intellectual disabilities, was observed. The CORE-LD only included one positively worded item, “Have you felt happy with the things you have done?” and so this tool may not be sufficient for those interested in measuring positive aspects of mental wellbeing. The results from this review therefore have implications for research as they highlight a need to develop and validate measures of positive mental health, or mental wellbeing, in adults with mild to moderate intellectual disabilities.

Conclusions

This review evaluated the psychometric properties of nine measures of common mental health problems and mental wellbeing in adults with mild to moderate intellectual disabilities, administered in English. Four of these (CORE-LD, IES-IDs, LANTS and SAINT) were deemed to have promising psychometric properties. The results were limited by incomplete psychometric evaluations of measures which made it difficult to compare measures. A paucity of psychometrically robust measures of mental wellbeing was observed. This review informs measurement choice for researchers and clinicians whilst highlighting a need for continued research efforts into the quality of measures available for people with intellectual disabilities.

References

Aman, M. G., Singh, N. N., & Turbott, S. H. (1987). 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., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal on Mental Deficiency, 89*, 485-491. https://doi.org/10.1037/t10453-000

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, fourth edition text revision (DSM-1V-TR).* Washington, DC.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191. https://doi.org/10.1097/00007632-200012150-00014

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988a). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893. https://doi.org/10.1037/0022-006X.56.6.893

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988b). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*(1), 77-100. https://doi.org/10.1016/0272-7358(88)90050-5

Beck, A.T., Steer, R.A. and Brown, G.K. (1996). *Beck Depression Inventory* - 2nd Edition Manual. The Psychological Corporation, San Antonio, TX. https://doi.org/10.1037/t00742-000

Bertelli, M. O., Bianco, A., Rossi, A., Mancini, M., La Malfa, G., & Brown, I. (2019). Impact of severe intellectual disability on proxy instrumental assessment of quality of life.

Journal of Intellectual & Developmental Disability, 44(3), 272-281. https://doi.org/10.3109/13668250.2017.1350835

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Briscoe, H., Ashworth, S., & Shelton, L. (2019). Establishing the concurrent validity of the Clinical Outcome in Routine Evaluation-Learning Disabilities scale. *Advances in Mental Health and Intellectual Disabilities, 13*(6), pp. 237-244. https://doi.org/10.1108/AMHID-02-2019-0005

Brooks, M., Davies, S., & Twigg, E. (2013). A measure for feelings-using inclusive research to develop a tool for evaluating psychological therapy (Clinical Outcomes in Routine Evaluation-Learning Disability). *British Journal of Learning Disabilities, 41*(4), 320-329. https://doi.org/10.1111/bld.12020

Chaplin, E., Chester, R., Tsakanikos, E., McCarthy, J., Craig, T., & Bouras, N. (2013). Reliability and validity of the SAINT: A guided self-help tool for people with intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities, 6*(3), 245-253. https://doi.org/10.1080/19315864.2012.670367

Chinn, D., & Abraham, E. (2016). Using "candidacy" as a framework for understanding access to mainstream psychological treatment for people with intellectual disabilities and common mental health problems within the English Improving Access to Psychological Therapies service. *Journal of Intellectual Disability Research, 60*(6), 71-582. https://doi.org/10.1111/jir.12274

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284. https://doi.org/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, 86*, 127-137.

Cluley, V. (2018). From "Learning disability to intellectual disability"-Perceptions of the increasing use of the term "intellectual disability" in learning disability policy, research and practice. *British Journal of Learning Disabilities, 46*(1), 24-32. https://doi.org/10.1111/bld.12209

Collins, L. M. (2007). *Research Design and Methods in Encyclopedia of Gerontology* (2nd ed.). Pennsylvania State University, USA. pp. 433-442. https://doi.org/10.1016/B0-12-370870-2/00162-1

Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist, 44*(5), 730-757. https://doi.org/10.1177/0011000016633507

Cooper, S.-A., Smiley, E., Morrison, J., Williamson, A., & Allan, L. (2007). Mental ill-health in adults with intellectual disabilities: Prevalence and associated factors. *The British Journal of Psychiatry, 190*(1), 27-35.https://doi.org/10.1192/bjp.bp.106.022483

Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale-Revised. *Behaviour Research and Therapy, 41*, 1489-1496. https://doi.org/10.1016/j.brat.2003.07.010

Cummins, R. A. (1997). *Comprehensive Quality of Life Scale - Adult, Manual* (5th ed.). School of Psychology, Deakin University, Melbourne, pp. 1-51.

Cummins, R.A., Eckersley, R., Pallant, J., van Vugt, J. & Misajon, R. (2003). Development of a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social Indicators Research, 64*, 159-190. https://doi.org/10.1023/A:1024704320683

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. Individual and carer supplement versions. *British Journal of Psychiatry, 182*, 347-353. https://doi.org/10.1192/bjp.182.4.347

Dagnan, D., & Lindsay, W. R. (2004). Cognitive therapy with people with learning disabilities. *International Handbook of Research and Evaluation in Intellectual Disabilities*, 517-530.

Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and wellbeing: An introduction. *Journal of Happiness Studies, 9*(1), 1-11. https://doi.org/10.1007/s10902-006-9018-1

Department of Health (2001). *Valuing people*. London: HMSO.

Derogatis, L. R. (1993). *Brief Symptom Inventory: Administration and scoring procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.

Devine, M., Taggart, L., & McLornian, P. (2010). Screening for mental health problems in adults with learning disabilities using the Mini PAS-ADD Interview. *British Journal of Learning Disabilities, 38*(4), 252-258. https://doi.org/10.1111/j.1468-3156.2009.00597.x

Dunn, K., Rydzewska, E., Fleming, M., & Cooper, S. A. (2020). Prevalence of mental health conditions, sensory impairments and physical disability in people with co-occurring intellectual disabilities and autism compared with other people: A cross-sectional total population study in Scotland. *BMJ Open, 10*(4), e035280. https://doi.org/10.1136/bmjopen-2019-035280

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. https://doi.org/10.1352/1934-9556-51.5.333

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology, 68*(4), 435-441. https://doi.org/10.1016/j.jclinepi.2014.11.021

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*, 617-629. https://doi.org/10.1023/B:JADD.0000005999.27178.55

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry, 180*(1), 51-60. https://doi.org/10.1192/bjp.180.1.51

Finlay, W. M., & Lyons, E. (2001). Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. *Psychological Assessment, 13*(3), 319-335. https://doi.org/10.1037/1040-3590.13.3.319

Floyd, F. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299. https://doi.org/10.1037/1040-3590.7.3.286

Flynn, S., Vereenooghe, L., Hastings, R. P., Adams, D., Cooper, S. A., Gore, N., Hatton, C., Hood, K., Jahoda, A., Langdon, P., 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, 57*, 32-44. https://doi.org/10.1016/j.cpr.2017.08.006

Goodley, D. (2011). *Self-advocacy in the lives of people with learning disabilities.* Maidenhead: Open University Press.

Green, S., Davis, C., Karshmer, E., Marsh, P., & Straight, B. (2005). Living stigma: The impact of labeling, stereotyping, separation, status loss, and discrimination in the lives of individuals with disabilities and their families. *Sociological Inquiry, 75*(2), 197–215. https://doi.org/10.1111/soin.2005.75.issue-2Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265-275. https://doi.org/10.1037/0033-2909.103.2.265

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*(12), 1417-1432. https://doi.org/10.1016/0895-4356(93)90142-N

Hagerty, M. R., Cummins, R., Ferriss, A. L., Land, K., Michalos, A. C., Peterson, M., ... & Vogel, J. (2001). Quality of life indexes for national policy: Review and agenda for research. *Bulletin of Sociological Methodology, 71*(1), 58-78. https://doi.org/10.1177/075910630107100104

Hall, J. C., Jobson, L., & Langdon, P. E. (2014). Measuring symptoms of post-traumatic stress disorder in people with intellectual disabilities: The development and psychometric properties of the Impact of Event Scale-Intellectual Disabilities (IES-s). *British Journal of Clinical Psychology, 53*(3), 315-332. https://doi.org/10.1111/bjc.12048

Halvorsen, M. B., Helverschou, S. B., Axelsdottir, B., Brøndbo, P. H., & Martinussen, M. (2022). General Measurement Tools for Assessing Mental Health Problems Among Children and Adolescents with an Intellectual Disability: A Systematic Review. *Journal of Autism and Developmental Disorders*, 1-73. https://doi.org/10.1007/s10803-021-05419-5

Hartley, S. L., & MacLean Jr, W. E. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research, 50*(11), 813-827. https://doi.org/10.1111/j.1365-2788.2006.00844.x

Hatton, C., & Taylor, J. L. (2013). The assessment of mental health problems in adults with intellectual disabilities. In J. L. Taylor, W. R. Lindsay, R. P. Hastings & C. Hatton (Eds.), *Psychological therapies for adults with intellectual disabilities* (pp. 31-54). Chichester: Wiley & Sons. https://doi.org/10.1002/9781118329252.ch3

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. https://doi.org/10.1016/j.ridd.2010.04.023

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. https://doi.org/10.1016/j.ridd.2011.09.018

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. https://doi.org/10.1016/j.ridd.2011.01.034

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

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*(3), 209-218. https://doi.org/10.1097/00006842-197905000-00004

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. https://doi.org/10.1080/10705519909540118

Hulbert-Williams, L., Hastings, R. P., Crowe, R., & Pemberton, J. (2011). Self-reported life events, social support and psychological problems in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 24*(5), 427-436. https://doi.org/10.1111/j.1468-3148.2011.00624.x

International Wellbeing Group. (2006). *Personal Wellbeing Index*. Australian Centre on Quality of Life: Deakin University.

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. https://doi.org/10.1111/j.1365-2788.2012.01544.x

Keyes, C. L. M. (2002). The mental health continuum: from languishing to flourishing in life. *Journal of Health and Social Behavior, 43*(2), 207-22. https://doi.org/10.2307/3090197

Keyes, C. L., Dhingra, S. S., & Simoes, E. J. (2010). Change in level of positive mental health as a predictor of future risk of mental illness. *American Journal of Public Health, 100*(12), 2366-2371. https://doi.org/10.2105/AJPH.2010.192245

Kovacs, M. (1985). The children's depression, inventory (CDI). *Psychopharmacological Bulletin, 21*, 995-998.

Kuliś, D., Whittaker, C., Greimel, E., Bottomley, A., Koller, M., & EORTC Quality of Life Group. (2017). Reviewing back translation reports of questionnaires: the EORTC conceptual framework and experience. *Expert review of Pharmacoeconomics & Outcomes Research, 17*(6), 523-530. https://doi.org/10.1080/14737167.2017.1384316

Lamers, S. M., Westerhof, G. J., Glas, C. A., & Bohlmeijer, E. T. (2015). The bidirectional relation between positive mental health and psychopathology in a longitudinal representative panel study. *The Journal of Positive Psychology, 10*(6), 553-560. https://doi.org/10.1080/17439760.2015.1015156

Lau A. L. D., Cummins R. A. & McPherson W. (2005) An investigation into the cross-cultural equivalence of the personal wellbeing index. *Social Indicator Research, 75,* 403-430. https://doi.org/10.1007/s11205-004-0561-z

Lennox, N., Taylor, M., Rey-Conde, T., Bain, C., Purdie, D. M., & Boyle, F. (2005). Beating the barriers: recruitment of people with intellectual disability to participate in research. *Journal of Intellectual Disability Research, 49*(4), 296-305. https://doi.org/10.1111/j.1365-2788.2005.00618.x

Li, C., Tsoi, E. W., Zhang, A. L., Chen, S., & Wang, C. J. (2013). Psychometric properties of self-reported quality of life measures for people with intellectual disabilities: A systematic review. *Journal of Developmental and Physical Disabilities, 25*(2), 253-270. https://doi.org/10.1007/s10882-012-9297-x

Maslow, A. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-396. https://doi.org/10.1037/h0054346

Matson, J. L. (1995). *The Matson evaluation of social skills for individuals with severe retardation*. Baton Rouge, LA: Scientific Publishers, Inc.

Matson, J. L., & Bamburg, J. W. (1998). Reliability of the assessment of dual diagnosis (ADD). *Research in Developmental Disabilities, 19*, 89-95. https://doi.org/10.1016/S0891-4222(97)00031-0

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

Matson, J. L., Kazdin, A. E., & Senatore, V. (1984). Psychometric properties of the psychopathology instrument for mentally retarded adults. Applied *Research in Mental Retardation, 5*(1), 81-89. https://doi.org/10.1016/S0270-3092(84)80021-1

McGillivray, J. A., Lau, A. L. D., Cummins, R. A., & Davey, G. (2009). The utility of the personal wellbeing index intellectual disability scale in an Australian sample. *Journal of Applied Research in Intellectual Disabilities, 22*(3), 276-286. https://doi.org/10.1111/j.1468-3148.2008.00460.x

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. https://doi.org/10.1177/1744629505056696

Mileviciute, I., & Hartley, S. L. (2015). Self-reported versus informant-reported depressive symptoms in adults with mild intellectual disability. *Journal of Intellectual Disability Research, 59*(2), 158-169. https://doi.org/10.1111/jir.12075

Mindham, J., & Espie, C. A. (2003). Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID): Development and psychometric properties of a new measure for use with people with mild intellectual disability. *Journal of Intellectual Disability Research, 47*, 22-30. https://doi.org/10.1046/j.1365-2788.2003.00457.x

Moss, S. (2002). *The Mini PAS-ADD interview pack*. Pavilion Pub.

Moss, S., Patel, P., Prosser, H., Goldberg, D., Simpson, N., Rowe, S., & Lucchino, R. (1993). Psychiatric morbidity in older people with moderate and severe learning disability: I: Development and reliability of the patient interview (PAS-ADD). *The British Journal of Psychiatry, 163*(4), 471-480. https://doi.org/10.1192/bjp.163.4.471

National Institute for Health and Care Excellence (NICE) (2011). *Common mental health disorders: identification and pathways to care.* https://www.nice.org.uk/guidance/cg123

Novaco, R. W. (1994). Anger as a risk factor for violence among the mentally disordered. In J. Monahan & H. J. Steadman (Eds.), *Violence mental disorder: Developments in risk assessment* (pp. 21-59). Chicago: University of Chicago Press.

O'Keeffe, L., Guerin, S., McEvoy, J., Lockhart, K., & Dodd, P. (2019). The process of developing self-report measures in intellectual disability: A case study of a complicated grief scale. *British Journal of Learning Disabilities, 47*(2), 134-144. https://doi.org/10.1111/bld.12261

Overall, J. E., & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological Reports, 10*(3), 799-812. https://doi.org/10.2466/pr0.1962.10.3.799

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. https://doi.org/10.1177/1744629509104487

Pouls, K. P., Koks-Leensen, M. C., Mastebroek, M., Leusink, G., & Assendelft, W. (2021). Adults with intellectual disabilities and mental health disorders in primary care: A scoping review. *British Journal of General Practice.* https://doi.org/10.3399/BJGP.2021.0164

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. https://doi.org/10.1046/j.1365-2788.1998.00146.x

Raczka, R., Theodore, K., & Williams, J. (2020). An initial validation of a new quality of life measure for adults with intellectual disability: The Mini-MANS-LD. *Journal of Intellectual Disabilities*, *24*(2), 177-193. https://doi.org/10.1177/1744629518787895

Raczka, R., Williams, J., & Theodore, K. (2014). Pilot evaluation of routine outcome measures for use with adults with learning disabilities. *The Bulletin of the Faculty for People with Intellectual Disabilities, 12*(1), 60-65.

Reiss S., Levitan G. & Szysko (1982) Emotional disturbance and mental retardation: diagnostic overshadowing. *American Journal of Mental Deficiency, 86,* 567-574.

Reiss, S. (1988). *Reiss screen for maladaptive behavior: Test manual.* IDS Publishing Corporation.

Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2001). The Behaviour Problems Inventory: An instrument for the assessment of self-injury, stereotyped behavior, and aggression/destruction in individuals with developmental disabilities. *Journal of Autism & Developmental Disorders, 31*(6), 577-588. https://doi.org/10.1023/A:1013299028321

Rojahn, J., Rowe, E. W., Kasdan, S., Moore, L., & van Ingen, D. J. (2011). Psychometric properties of the aberrant behavior checklist, the anxiety, depression and mood scale, the assessment of dual diagnosis and the social performance survey schedule in adults with intellectual disabilities. *Research in Developmental Disabilities, 32*(6), 2309-2320. https://doi.org/10.1016/j.ridd.2011.07.035

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. https://doi.org/10.1046/j.1365-2788.2002.00397.x

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. https://doi.org/10.1348/014466503762842039

Saylor, C. F, Swenson, C. C., Reynolds, S. S., & Taylor, M. (1999). The Pediatric Emotional Distress Scale: A brief screening measure for young children exposed to traumatic events. *Journal of Clinical Child Psychology, 28*(1), 70-81. https://doi.org/10.1207/s15374424jccp2801_6

Schalock, R. L., & Keith, K. D. (1993). *Quality of life questionnaire manual*. Worthington, Ohio: IDS Publishing Corporation. https://doi.org/10.1037/t10624-000

Schmidt S, Muhlan H and Power M (2006) The EUROHIS-QOL 8-item index: Psychometric results of a cross-cultural field study. *European Journal of Public Health, 16*, 420-428. https://doi.org/10.1093/eurpub/cki155

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. *BMJ, 351*, h432. https://doi.org/10.1136/bmj.h4326

Skevington, S. M., & Böhnke, J. R. (2018). How is subjective well-being related to quality of life? Do we need two concepts and both measures?. *Social Science & Medicine, 206*, 22-30. https://doi.org/10.1016/j.socscimed.2018.04.005

Skirrow, P., & Perry, E. (2009). *The Maslow assessment of needs scales.* Liverpool: Mersey care NHS Trust.

Smith, J. A. (2003). *Qualitative psychology: A practical guide to research methods*. Sage Publications, Inc.

Stancliffe R. J. (1995). Assessing opportunities for choice-making: a comparison of self- and staff reports. *American Journal on Mental Retardation, 99*, 418-29.

Stancliffe, R. J., & Parmenter, T. R. (1999). The choice questionnaire: a scale to assess choices exercised by adults with intellectual disability. *Journal of Intellectual & Developmental Disability, 24*, 107-132. https://doi.org/10.1080/13668259900033911

Stewart-Brown, S., Samaraweera, P. C., Taggart, F., Kandala, N. B., & Stranges, S. (2015). Socioeconomic gradients and mental health: implications for public health. *The British Journal of Psychiatry, 206*(6), 461-465. https://doi.org/10.1192/bjp.bp.114.147280

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J. & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 1-13. https://doi.org/10.1186/1477-7525-5-63

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Townsend-White, C., Pham, A. N. T., & Vassos, M. V. (2012). A systematic review of quality of life measures for people with intellectual disabilities and challenging behaviours. *Journal of Intellectual Disability Research, 56*(3), 270-284. https://doi.org/10.1111/j.1365-2788.2011.01427.x

Trompetter, H. R., de Kleine, E., & Bohlmeijer, E. T. (2017). Why does positive mental health buffer against psychopathology? An exploratory study on self-compassion as a resilience mechanism and adaptive emotion regulation strategy. *Cognitive Therapy and Research, 41*(3), 459. https://doi.org/10.1007/s10608-016-9774-0

Weiss, D. S., & Marmar, C. R. (1997). The Impact of Event Scale-Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD: A handbook for practitioners* (pp. 399-411). New York, NY: Guilford Press.

Whittle, E. L., Fisher, K. R., Reppermund, S., Lenroot, R., & Trollor, J. (2018). Barriers and enablers to accessing mental health services for people with intellectual disability: a scoping review. *Journal of Mental Health Research in Intellectual Disabilities, 11*(1), 69-102. https://doi.org/10.1080/19315864.2017.1408724

WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine, 41*(10), 1403-1409. https://doi.org/10.1016/0277-9536(95)00112-K

Wigham, S., Hatton, C., & Taylor, J. L. (2011). The Lancaster and Northgate Trauma Scales (LANTS): The development and psychometric properties of a measure of trauma for people with mild to moderate intellectual disabilities. *Research in Developmental Disabilities, 32*(6), 2651-2659. https://doi.org/10.1016/j.ridd.2011.06.008

Wigham, S., Hatton, C., & Taylor, J. L. (2021). Short report and initial evaluation of the factor structure of the Lancaster and Northgate Trauma Scales (LANTS). *Research in Developmental Disabilities*, *112*, 103914. https://doi.org/10.1016/j.ridd.2021.103914

Willie N, Badia X, Bonsel G, et al. (2010) Development of the EQ-5D-Y: A child friendly version of the EQ-5D. *Quality of Life Research, 19*(6), 875-886. https://doi.org/10.1007/s11136-010-9648-y

World Health Organization (WHO). (2018). *International classification of diseases for mortality and morbidity statistics* (11th Revision).

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*(8), 737-746. https://doi.org/10.1111/jir.12003

Part 2: Empirical Paper

Psychometric properties of the Warwick-Edinburgh Mental Wellbeing Scale-Intellectual Disability version (WEMWBS-ID)

Abstract

Aims: Mental wellbeing, encompassing dimensions of hedonia (feeling good) and eudaimonia (functioning well), is considered a valuable resource for individuals and communities. The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS; Tennant et al., 2007), a 14-item positively worded measure of mental wellbeing, has been extensively psychometrically validated within the UK and cross-culturally. However, it is yet to be validated for use with individuals with intellectual disabilities, a priority given the paucity of measures of mental wellbeing for this population. The aim of this study was to assess the psychometric properties of a newly adapted version of the WEMWBS and the Short 7-item WEMWBS (SWEMWBS) for individuals with mild to moderate intellectual disabilities (WEMWBS-ID/SWEMWBS-ID). This adapted version was developed by the primary research supervisor, experts by experience and colleagues in the intellectual disability field.

Method: Individuals aged 16+ with mild to moderate intellectual disabilities were recruited using volunteer sampling. Data from three studies conducted in the UK were collated to evaluate the psychometric properties of the WEMWBS-ID (*n* = 96) and additional data from a study conducted in Canada (*n* = 27) was used in addition to the UK data to evaluate the SWEMWBS-ID (*n* = 123). The WEMWBS-ID was administered by researchers via an online meeting using the screenshare function. A subsample (*n* = 22) completed the scale twice for test-retest reliability. Furthermore, as a relationship between wellbeing and self-esteem has been demonstrated in previous research, 95 of the UK participants also completed an adapted 6-item version of the adapted Rosenberg Self-Esteem Scale (RSES) to assess convergent validity.

Results: The WEMWBS-ID demonstrated good internal consistency (α = 0.86), excellent test-retest reliability (ICC = .88) and good convergent validity (*r* = .68) with RSES scores. Confirmatory factor analysis supported the hypothesised one-factor structure and the measure demonstrated an adequate model fit. The SWEMWBS-ID showed poor internal consistency (α = 0.67), good test-retest reliability (ICC = .67) and good convergent validity (*r* = .61). The results from the confirmatory factor analysis indicated a good model fit. Examination of the response distribution highlighted a tendency for participants to select the options "sometimes" and "always" (from a choice of "never", "sometimes", "often" and "always"). A linear transformation of scores suggested that the present WEMWBS-ID scores were slightly lower compared to UK general population data, whilst SWEMWBS-ID scores were slightly higher.

Conclusions: The results from this initial evaluation of the WEMWBS-ID and SWEMWBS-ID suggest that the scales have promising psychometric properties, when administered by a researcher to individuals with mild to moderate intellectual disabilities. The generalisability of the findings is limited by the sample size and sampling strategy and a further exploration of the scales with larger samples is warranted. Suggestions for future research are discussed.

Introduction

Terminology

The term “learning disability” is the preferred term in the UK to refer to people who have “significantly reduced ability to understand new or complex information, to learn new skills” and a “reduced ability to cope independently which starts before adulthood with lasting effects on development” (Cluley, 2018; Department of Health, 2001 p. 14). Various countries use the term “intellectual disability” as their preferred label and this is used increasingly in British learning disability policy, practice and research, reflecting the changing international context (Cluley, 2018). The social construction of labels has implications for individuals, particularly if the labels are associated with stigma, as is the case with these terms (Green et al., 2005). Some self-advocates in England have expressed a preference for the term “learning difficulty”, rather than “disability” which may be considered socially limiting (Goodley, 2011). However, this term is used in UK educational settings to refer to specific learning difficulties such as dyslexia or dyspraxia and therefore may cause confusion if used interchangeably with the term learning disability. Whilst the potential disadvantages of using the term “intellectual disability” are acknowledged, this term will be used throughout the paper in line with research pertaining to this population. This decision was made to reflect the recent trend to use this term in a range of settings in the UK (Cluley, 2018).

Background

There has been growing interest in the promotion of positive mental health, or mental wellbeing, in public health, as it is considered a valuable resource for individuals and communities (Faculty of Public Health and Mental Health Foundation, 2016; Stewart-Brown et al., 2015). Although mental wellbeing has been conceptualised and defined in different ways, there is a growing consensus that it encompasses dimensions of hedonia (feeling good, including happiness and positive affect) and eudaimonia (functioning well, including personal growth and having positive relations with others) (Diener et al., 1999; Galderisi et al., 2015; Keyes et al., 2002; Ryan & Deci, 2001). Furthermore, it is recognised that mental wellbeing is not merely the absence of mental illness but rather positive mental functioning and therefore the two may represent two separate continua, rather than extreme ends of a single continuum (Huppert, 2009; Keyes, 2005; Stewart-Brown et al., 2015). Although it has been argued that wellbeing research to date has been largely based in "WEIRD" (Western, Educated, Industrialised, Rich, and Democratic) populations (Henrich et al., 2010), positive psychology research is found to have a growing global presence and researchers are seeking to better understand and measure wellbeing globally (Kim et al., 2018; Lambert et al., 2020).

Evidence suggests that people with intellectual disabilities experience higher rates of major mental disorders compared to the general population (Dunn et al., 2020; Evans et al., 2012), not simply due to their intellectual disability or related medical conditions but to multiple factors, including stigma and prejudice (Ali et al., 2012). Stigmatisation can have a negative impact on wellbeing, self-esteem and mood (Paterson et al., 2012). There is growing evidence relating to the protective effect of mental wellbeing on mental and physical health (e.g. Keyes et al., 2010; Siahpush et al., 2008) and an increasing focus in public health on moving away from illness, towards the promotion of resilience, social purpose and autonomy (Faculty of Public Health and Mental Health Foundation, 2016). Measurement is therefore crucial, in order to gauge levels of mental wellbeing in the population, whilst allowing the effectiveness of interventions to improve mental wellbeing to be evaluated. A range of scales are available to measure mental wellbeing in the general population (Taggart & Stewart-Brown, 2019), such as the World Health Organisation-Five Wellbeing Index.

Challenges Relating to Measurement

It is now widely accepted that actively involving people with intellectual disabilities in research is crucial for empowerment and needs to be increased (Cleaver et al., 2010; Sigstad, 2014). Self-report measures allow people with intellectual disabilities to share their own experiences and provide a valuable insight into their subjective perspectives (Hartley & MacLean, 2006; O'Keeffe et al., 2019). It has been established that many people with intellectual disabilities are able to provide reliable and valid information using self-report measures (Kooijmans et al., 2022). However, researchers have also documented several challenges relating to the use of self-report measures in this population, which require the individual to understand the questions, response alternatives and produce a response which is not influenced by factors such as response option order or demand characteristics (Mileviciute & Hartley, 2015). Furthermore, self-report measures require individuals to access short- and long-term memory to recall past behaviour and experiences when selecting a response, which may be challenging for some people with intellectual disabilities (Hartley & MacLean, 2006). Response bias is an additional challenge, as it has been documented that people with intellectual disabilities tend to choose the most positive option in Likert-type scales, though it has been demonstrated that Likert-type scales are suitable for self-report measures for individuals with borderline to mild intellectual disabilities (Hartley & MacLean, 2006). Using pictorial representations, such as pictures of containers filled to increasing levels, may enhance the reliability of Likert-type scales (Marshall & Willoughby-Booth, 2007; O'Keeffe et al., 2019).

Measures designed to be used with an informant such as a family member or support worker may be an alternative or additional tool to self-reporting. They may be useful when a measure requires a level of understanding beyond that of the individual with intellectual disability, or when there are communication difficulties (O'Keeffe et al., 2019). However, several studies have expressed caution about the validity of informant measures, the degree to which they actually reflect the experience of the individual concerned and whether responses are influenced by the informant’s own prejudices (Bertelli et al., 2017; Claes et al., 2012; McGillivray et al., 2009; Schalock et al., 2002). Therefore, an overreliance on informant reports should be avoided and self-reports should be used where possible (O’Keeffe et al., 2019).

Attention must be given to the design and language used when developing self-report measures for individuals with intellectual disabilities, because of the difficulties with comprehension and responding which may be encountered (O’Keeffe et al., 2019). This is particularly true for questions about abstract concepts such as emotions, which may be more difficult for people with intellectual disabilities to understand, compared to more concrete concepts such as physical health (O’Keeffe et al., 2019; Ruddick & Oliver, 2005). There are various recent examples in the literature of self-report measures which have been developed or adapted in collaboration with people with intellectual disabilities, to ensure their acceptability and accessibility. The Mini-Maslow Assessment of Needs Scale-Learning Disabilities (Mini-MANS-LD; Raczka et al., 2014) measures quality of life for people with intellectual disabilities. It was developed in collaboration between psychologists, speech and language therapists and people with intellectual disabilities to ensure that the wording was clearly understood by respondents. Additionally, the development of scripts for administrators of the measure ensured that respondents could be supported to give their responses in a standardised way. The development of the self-report version of the Complicated Grief Questionnaire for People with Intellectual Disabilities (Guerin et al., 2009; O’Keeffe et al., 2019) also involved consultation with adults with intellectual disabilities in focus groups, to understand the type of language and phrases which may be used to describe feelings in this population. The measure was subsequently piloted and interviewers paid attention to participants’ perceived ability to understand the concepts and language used in the measure. These examples highlight the importance of including individuals with intellectual disabilities in the measurement development process, in order to overcome some of the challenges which may be faced when using self-report scales with this population.

Existing Measures of Mental Wellbeing for People with Intellectual Disabilities

Often in research, wellbeing has been assessed using a limited set of items relating to life satisfaction or happiness. There is general agreement that wellbeing is a multidimensional construct and therefore an informative measure should include items which measure both hedonic and eudaimonic aspects of wellbeing (Huppert, 2009; Ruggeri et al., 2020; VanderWeele et al., 2020).

The term ‘mental wellbeing’ has often been used interchangeably with ‘quality of life’ in the literature (Cooke et al., 2016), although it has been argued that they refer to different theoretical concepts and support separate measurement fields (Skevington & Böhnke, 2018). Quality of life refers to “an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” (WHOQOL Group, 1995, p. 1404). Previous systematic reviews have identified many measures of quality of life which may be used for people with intellectual disabilities (e.g. Li et al., 2013; Townsend-White et al., 2012). However, these measures may not be suitable for those interested in measuring mental wellbeing or positive mental health.

The Clinical Outcomes in Routine Evaluation–Learning Disability (CORE-LD; Brooks et al., 2013) measures various factors such as wellbeing, mental health and interpersonal relationships for people with intellectual disabilities. However, it only includes one positively worded item, “Have you felt happy with the things you have done?” and therefore may not be sufficient for those interested in measuring aspects of positive mental health. The Psychological Therapies Outcome Scale– Intellectual Disabilities (PTOS-ID; Vlissides et al., 2017) is a recently developed brief psychological therapies outcome measure. It includes items loading onto three factors: anger and mood, positive wellbeing and anxiety. During the development of the PTOS-ID, it was noted that it was difficult to initially develop an item pool for mental wellbeing as there were no measures currently used for people with intellectual disabilities which assess this. The CORE-LD and PTOS-ID offer a positive advance in assessing positive aspects of mental wellbeing in this population, though a specific measure of mental wellbeing for people with intellectual disabilities is still indicated.

The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) (Tennant et al., 2007)

The WEMWBS is a 14-item positively worded measure of mental wellbeing developed in the UK. The conceptual underpinning of the measure is based on the hedonic (e.g. “I’ve been feeling cheerful”) and eudaimonic (e.g. “I’ve been dealing with problems well”) dimensions of wellbeing. It was designed to enable monitoring of positive mental health and evaluation of public health interventions to promote mental wellbeing. The WEMWBS is included in the Health Survey for England and in national surveys in Scotland and Wales and has been extensively validated with adults and adolescents in the UK and cross-culturally, demonstrating robust psychometric properties (e.g. Fung et al., 2019; McKay & Andretta, 2017; Trousselard et al., 2016). Furthermore, Crawford et al. (2011) supported the acceptability of the measure, as they reported that the WEMWBS was commended by service users for asking questions about “good mental health”, rather than long lists of questions about mental ill health, which many found upsetting.

In 2009, a seven-item version of the measure, the Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS, Stewart-Brown et al., 2009) was developed using Rasch modelling, enabling interval scale measurement. The items in the SWEMWBS mostly represent aspects of eudaimonic wellbeing, with fewer items covering hedonic wellbeing. The SWEMWBS has also been validated in general and clinical populations (e.g. Bass et al., 2016; Koushede et al., 2019; Melendez-Torres et al., 2019; Vaingankar et al., 2017), demonstrating sensitivity to change as a clinical measure (Shah et al., 2018, 2021).

In summary, the WEMWBS has demonstrated promising psychometric properties in general and clinical adult and young people populations. However, it is yet to be validated for use with individuals with intellectual disabilities. The original 14-item version was piloted in an outpatient clinic with individuals with intellectual disabilities by Vlissides et al. (2017), who reported that the wording of some of the questions was too difficult for them. This suggested a need to develop and evaluate an adapted version of the WEMWBS specifically for individuals with intellectual disabilities.

Psychometric Evaluation

Best practices for scale development and validation have been set out by Boateng et al. (2018). They outlined three steps pertaining to psychometric evaluation. The first step involves testing the hypothesis that there is a relationship between items and their underlying latent construct. This may be assessed using a confirmatory factor analysis. Secondly, tests of reliability establish whether responses are consistent when repeated. This may be assessed through calculating the internal consistency and test-retest reliability. Finally, tests of validity ensure that the scale measures the intended latent construct. If there is a ‘gold standard’ scale available, concurrent validity may be assessed by correlating scores between the two scales. If not, convergent validity, which contributes to the overall construct validity of a measure, may examined by measuring the relationship between the scale and other measures of similar or related constructs. With regards to cross-cultural research, translating measures which have been psychometrically evaluated in other cultures may be problematic for several reasons, such as the two languages having non-equivalent words, or items having very different meanings based on the specific cultural context (Epstein et al., 2015). The World Health Organization sets out recommendations for the translation and adaptation of scales, which includes forward-translation, expert panel back-translation, pre-testing and cognitive interviewing (WHOQOL, 1995).

Aims and Hypotheses

The aim of this study was to assess the psychometric properties of a newly adapted version of the WEMWBS and SWEMWBS for individuals with mild to moderate intellectual disabilities aged 16 or older (WEMWBS-ID/SWEMWBS-ID). The adapted version was developed by the primary research supervisor, experts by experience and colleagues in the intellectual disability field as part of an NIHR funded study **(**NIHR PHR project 17/149/03 - The STanding up fOR Myself [STORM] psychosocial group intervention for young people and adults with intellectual disabilities: Feasibility study)**.**

The investigation was split into two parts: Study 1 was the psychometric evaluation of the 14-item WEMWBS-ID; Study 2 was the psychometric evaluation of the seven items which constitute the SWEMWBS-ID. Criteria for evaluating reliability and validity were adopted from the three steps pertaining to psychometric evaluation, outlined by Boateng et al. (2018).

Reliability was assessed by calculating Cronbach’s alpha to determine internal consistency and intra-class coefficients (ICC) were calculated to examine test-retest reliability. The standard error of measurement was also calculated. In the validation of the WEMWBS by Tennant et al. (2007), Cronbach’s alpha ranged between .89 – .91 across samples, indicating good internal consistency. The ICC in their study was .83 ($p < 0.01$), indicating high retest reliability (Cicchetti & Sparrow, 1981). Therefore, it was hypothesised that the WEMWBS-ID and SWEMWBS-ID would also demonstrate good internal consistency and test-retest reliability.

Construct validity was also assessed. Self-esteem, which has been defined as an individual's general sense of his or her value or worth (Rosenberg, 1979), is one of the most widely examined predictors of wellbeing (e.g. Kong et al., 2013; Padhy et al., 2011). Previous researchers (e.g. Paterson et al., 2012) have used a version of the Rosenberg Self-Esteem Scale (RSES, Rosenberg, 1965, 1982), adapted for people with intellectual disabilities (Dagnan & Sandhu, 1999) to measure self-esteem. Ringdal et al. (2018) found that scores on the RSES were positively correlated (r = .70) with scores on the WEMWBS, among Norwegian adolescents. Therefore, it was hypothesised that there would be a positive correlation between scores on the WEMWBS-ID/SWEMWBS-ID and the adapted RSES, demonstrating convergent validity. Criterion validity was not evaluated due to a paucity of gold standard instruments in this area. With regards to structural validity, previous validation studies confirmed a one-factor structure of the WEMWBS and SWEMWBS. Therefore, it was hypothesised that a confirmatory factor analysis would support that the WEMWBS-ID and SWEMWBS-ID correspond to a one-factor structure.

Furthermore, response distribution and floor and ceiling effects were assessed for both scales. Although response bias has been documented in the intellectual disability literature in relation to the use of self-report measures, the WEMWBS-ID was developed in collaboration with individuals with intellectual disabilities to ensure accessibility and comprehensibility of the response scale. Therefore, this analysis was exploratory in nature and a hypothesis relating to the degree of response bias was not made. Finally, the scores on the adapted measures were transformed, to allow comparison with general population normative data.

Method

Design and Ethics

A cross-sectional study was conducted. The study involved individuals with intellectual disabilities completing a battery of measures via an online meeting using the screenshare function. Measures were administered to participants by a researcher. For test-retest reliability, a subsample completed the measure again one to two weeks later.

Data collection was carried out as an extension of the STORM study. Ethical approval was obtained from the University College London Research Ethics Committee (Project ID Number: 0241/005). An ethics amendment was approved on 26th May 2021 (see Appendix B), which enabled additional participants with intellectual disabilities to be recruited for the purposes of psychometric evaluation of the WEMWBS-ID. These participants were not offered the STORM intervention but completed measures at one or two time points. This group of participants will be referred to as "Sample 1" in this paper.

Analyses of data collected from Samples 1 to 3 (which were integrated into the STORM study) were covered by ethical approval for the STORM study. This was because examining the psychometric properties of the WEMWBS-ID was an important aspect of the study. Regarding Sample 4, a secondary analysis of anonymised data was conducted and therefore did not require additional ethical approval. Participants in Sample 4 were informed that their anonymised data could be subject to further analysis.

This study was conducted jointly with a fellow trainee, who assessed predictors of self-esteem in individuals with intellectual disabilities. See Appendix C for an explanation of the contributions by each trainee to this project.

Participants

Data were collated from four different samples, as outlined in Table 1 along with the demographic information. Data from the different samples were collected at different timepoints during the COVID-19 pandemic and therefore data collection was conducted online. Data from Sample 1 were collected as part of the present study with a fellow trainee. Data from Sample 2 were collected through the STORM study by separate group of researchers (Scior et al., 2022). The STORM intervention involved attending five group-based sessions consisting of activities designed to help people with intellectual disabilities challenge stigma in their everyday lives. Data from Sample 3 were collected by a former DClinPsy trainee (Goldsmith-Sumner, 2021) who developed a tool to assess how people with intellectual disabilities respond to stigma. Data from Sample 4 were collected by a group of intellectual disability researchers in Canada led by Dr Yona Lunsky, University of Toronto, who piloted the WEMWBS-ID as part of their study evaluating a mental health intervention for individuals during the COVID-19 pandemic (St. John et al., 2022). This dataset was fully anonymised before it was shared and therefore the General Data Protection Regulation did not apply.

Statistical Power

During the planning phase of the study, baseline data from Samples 2 and 3 (n = 52) were being collected and access to data from Sample 4 had been agreed. Various recommendations for sample sizes when using factor analytic techniques have been issued. Some researchers recommend a sample size of 100-200 participants (Guadagnoli & Velicer, 1988; MacCallum et al., 1999) whilst other recommendations (e.g. Kline, 2000; Terwee et al., 2007) range from four to ten participants per variable, which would equate to n = 56–140 for the WEMWBS. With respect to the correlational analysis, the power analysis was informed by Ringdal et al. (2018), who reported a correlation of r = .70 (large) between scores on the RSES and WEMWBS. A G*Power 3.1 analysis based on an effect size of .70 or .50 (two-tailed, specifying alpha = 5% and desired power = 80%) produced a minimum total sample size of 13 or 29, respectively.

In order to obtain a total sample size which was larger than the lowest recommended number of participants required for a factor analysis, whilst considering resource limitations, the aim was to recruit 50 additional participants, of whom half would complete the measure twice to allow assessment of test-retest reliability. All participants were offered the opportunity to complete the measures a second time, until 50% of participants had done so. Due to time constraints, at the end of the recruitment process, 44 participants were recruited in Sample 1, with 22 participants completing the measures twice.

Table 1

Outline of participants included in the present study.

	Study 1 14-item WEMWBS-ID *n* = 96 (Samples 1 to 3)	**Study 2** 7-item SWEMWBS-ID *n* = 123 (Samples 1 to 4)	Sample 1 *n* = 44	Sample 2 *n* = 22	Sample 3 *n* = 30	Sample 4 *n* = 27
Gender *n* (%)						
Female	45 (46.88)	60 (48.78)	19 (43.18)	15 (68.18)	11 (36.67)	15 (55.56)
Male	51 (53.13)	62 (50.51)	25 (56.82)	7 (31.82)	19 (63.33)	11 (40.74)
Other	0	1 (0.81)	0	0	0	1 (3.70)
Mean Age *years* (SD)	39 (12.17)	39 (12.40)	41 (12.87)	34 (10.00)	40 (11.86)	41 (13.24)
Range	18-74	18-74	18-70	21-59	24-74	22-67
Ethnicity *n* (%)						
White British/Other	81 (84.38)	No additional data from Sample 4	39 (88.64)	16 (72.73)	26 (86.67)	No data
Asian British/Other	6 (6.25)		2 (4.55)	1 (4.55)	3 (10.00)	
Black British/African/ Caribbean/Other	5 (5.21)		1 (22.73)	4 (18.18)	0	
Other	4 (4.17)		2 (4.55)	1 (4.55)	1 (3.33)	
Schooling *n* (%)						
Mainstream school only	31 (32.29)	No additional data from Sample 4	12 (27.27)	9 (40.91)	10 (33.33)	No data
Special needs school only	45 (46.88)		24 (54.55)	8 (36.36)	13 (43.33)	
Both special needs and mainstream schools	16 (16.67)		7 (15.90)	2 (9.09)	7 (23.33)	
Unsure	4 (4.17)		1 (22.73)	3 (13.64)	0	

	Study 1 14-item WEMWBS-ID *n* = 96 (Samples 1 to 3)	**Study 2** 7-item SWEMWBS-ID *n* = 123 (Samples 1 to 4)	Sample 1 *n* = 44	Sample 2 *n* = 22	Sample 3 *n* = 30	Sample 4 *n* = 27
Living arrangements *n* (%)						
On my own	25 (26.04)	32 (26.02)	13 (29.55)	4 (18.18)	8 (26.67)	7 (25.93)
With parents/family	35 (36.46)	42 (34.15)	14 (31.82)	11 (50.00)	10 (33.33)	7 (25.93)
With a partner/spouse	8 (8.33)	17 (13.82)	5 (11.36)	0	3 (10.00)	9 (33.33)
In supported living	23 (23.96)	23 (18.70)	10 (22.73)	5 (22.73)	8 (26.67)	0
Other	5 (5.21)	9 (7.32)	2 (4.55)	2 (9.09)	1 (3.33)	4 (14.81)
Member of a Self-Advocacy Group *n (%)*						
Yes	67 (69.79)	No additional	29 (65.91)	15 (68.18)	23 (76.67)	100%
No	26 (27.08)	data from	15 (34.09)	4 (18.18)	7 (23.33)	0
Unsure	3 (3.13)	Sample 4	0	3 (13.63)	0	0

Recruitment

Recruitment for Sample 1

Using volunteer sampling, participants were recruited through third sector organisations for people with intellectual disabilities in the UK. An Easy Read information sheet and a carers information sheet (Appendices D and E) were sent to 76 organisations which were identified through internet searches. Facilitators from these organisations acted as mediators in the recruitment process. The researchers also attended virtual group meetings for people with intellectual disabilities at various organisations to promote the study. Advertisements were shared on social media websites such as Twitter and Facebook. Interested individuals contacted the researchers directly to learn more about the study, or a family member or carer expressed an interest on their behalf. Participants were given a £10 gift voucher in recognition of the time taken to participate. Those who completed the measures a second time were given an additional £5 voucher.

Recruitment for Samples 2 to 4

Sample 2 was recruited through three community organisations across England and Wales and one Educational provider. Sample 3 was recruited using the same strategy employed to recruit Sample 1. With regards to Sample 4, individuals with intellectual disabilities in Canada were recruited through various national and provincial self-advocacy organisations.

Inclusion and Exclusion Criteria

Participants were included in Samples 1 to 3 if they lived in the UK, were aged 16 years or over and had an intellectual disability (by an administrative definition, in terms of receipt of specialist services for people with intellectual disabilities). They were also required to have the cognitive skills to be able to complete the measures (with support if necessary) and have sufficient expressive and receptive verbal communication skills in English to provide consent. These abilities were likely to equate to having a mild to moderate intellectual disability. Additionally, participants were required to have access to the internet and a computer, smartphone or tablet to be able to join a video call.

Participants were excluded if they were unable to communicate using English, were deemed to not have capacity to consent, or if they did not have an intellectual disability (there were instances of individuals with diagnoses of autism, dyslexia or ADHD who expressed an interest in the study, but they did not additionally have a known intellectual disability and were therefore excluded).

With regards to Sample 4, participants with mild to moderate intellectual disabilities, who were over the age of 18, could provide informed consent, had sufficient verbal skills and access to the internet, were included. Therefore, all participants in Samples 1-4 were deemed to have mild to moderate intellectual disabilities, on the basis of their cognitive and communicative functioning, rather than a formal assessment.

Measures

Warwick-Edinburgh Mental Wellbeing Scale-Intellectual Disability (WEMWBS-ID)

The WEMWBS-ID was designed to be either self-administered (for individuals who were most likely to have the cognitive and communicative abilities required to understand and respond to the items and use the response scale with minimal assistance) or administered by a researcher or supporter presenting each item in a neutral way (for individuals who may need support reading and/or fully comprehending the items and/or the response scale). In the present study, the WEMWBS-ID was researcher administered to all four samples to ensure that all participants were offered appropriate support and also to ensure standardisation across the studies. An administration guide was provided to all researchers involved in data collection (Appendix F). Items 1, 2, 3, 6, 7, 9 and 11 of the WEMWBS-ID constitute the SWEMWBS-ID. Higher total scores indicate higher levels of mental wellbeing. The WEMWBS-ID and SWEMWBS-ID can be found in Appendices F and G, respectively.

Adapting the WEMWBS. The WEMWBS had been adapted for individuals with intellectual disabilities prior to the present study, by the primary research supervisor and colleagues. First, item wordings were revised by a group of clinicians and researchers expert in the field of intellectual disability. Proposed revisions, alternative rewordings for individual items and different options for the response scale and its visual representation were then reviewed in detail by two research advisory panels of people with intellectual disabilities and their recommendations were integrated into the final version piloted. Changes to the measure included altering the item stem, e.g. “I’ve been feeling” to simple past tense, i.e. “I felt”. Specific concepts were explained through more familiar words e.g. “optimistic about the future” was changed to “hopeful about the future”. The original reference period of two weeks was considered too complex for individuals who frequently struggle with concepts of time and reduced to one week, ensuring retention of the focus on current wellbeing. The Likert scale was changed from a five-point scale (none of the time; rarely; some of the time; often; all of the time) to a four-point scale (never; sometimes; often; always). Each item was scored from zero to three and therefore total scores ranged from 0-42. A visual aid to the scale in the form of a diagram of blocks in ascending size with the scale wording were provided and two practice items (“I watched sports on TV” and “I ate rotting food”) were added to help with familiarisation and to assess understanding of the response scale. A response of ‘never’ to the latter item was expected, but if a participant selected another option and was able to explain why, this suggested that they were able to understand and reliably respond to the items. If the participant was not able to explain why they chose a given response, or it appeared that they did not understand the items and/or response scale, data collection was discontinued, as this suggested that they may not have been able to complete the measures reliably. Further information regarding administration of the two practice questions may be found in the "Guide for administration of the WEMWBS-ID" section in Appendix F.

Adapted RSES

The RSES (Rosenberg, 1965) is a widely used measure of self-esteem for the general population, originally consisting of ten items on a four-point Likert scale ranging from 'strongly agree' to 'strongly disagree'. The scale was later reduced to six items by Rosenberg (1982). The six item RSES was adapted by Dagnan and Sandhu (1999) for use with individuals with intellectual disabilities. This adapted RSES comprised six items (four positively worded and two negatively worded) in simplified language and a five-point Likert scale ranging from 'never true' to 'always true'. Examples of items are, "I feel that I have a lot of good qualities" and, "At times I think I am no good at all". Visual cues were presented alongside the response categories to indicate the increasing magnitude of the response. The adapted RSES has demonstrated reasonable concurrent validity, good internal reliability and a factor structure similar to that predicted by Rosenberg's model of self-esteem (Dagnan & Sandhu, 1999).

In the present study, the response scale was simplified from five to four points to correspond with the response scale used for the WEMWBS and other measures, and in line with consultation with people with intellectual disabilities. Responses were scored from zero to three, therefore the possible range of scores was 0-18. Negatively worded items were reverse-scored and higher total scores indicated higher levels of self-esteem. The adapted RSES was administered to Samples 1-3, also using an interview style format.

Procedure

Procedure for Sample 1

Following recruitment, researchers met with those who expressed an interest in participating (and their carers if this was preferred) via a video call, to talk through the information sheet and answer any questions. Capacity to consent and communication skills were gauged during these meetings by the researcher. Participants were given at least 24 hours to decide whether they wished to take part in the project. Interested participants met with the researcher again at a later date via video call to provide informed consent. The consent form (see Appendix H) was shared via screenshare and each item was discussed. Participants who were happy to participate were asked to state their name, the date and that they agreed participate to participate in the project. This was recorded and stored securely on a UCL research drive, separately from all other data collected. The measures were subsequently completed on Qualtrics via screen-sharing so that participants were able to see the items and response options. Items were read aloud to facilitate comprehension. Participants were given the option to have a supporter present during the video call, but it was emphasised that the participants' own views were of interest. A small number of participants expressed a preference to have a supporter present.

At the initial administration of the measures, participants answered demographic questions, followed by the two practice items, the WEMWBS-ID and adapted RSES. Participants subsequently completed additional measures: a four-item adapted version of the Perceived Stigma in People with Intellectual Disabilities scale (Reactions to Discrimination subscale) (Ali et al., 2008), a four-item adapted version of the Sense of Power Scale (Anderson et al., 2012), a single item relating to self-efficacy in rejecting prejudice and the five-item EuroQol-Youth EQ-5D-Y (Wille et al., 2010). The results from these additional measures will not be discussed in the present study.

Participants were subsequently asked if they wished to complete the measures again one to two weeks later and if they agreed, a second meeting was arranged. At the second meeting, the participants completed the scales again, in the same order.

The average time taken to complete the consent process and baseline measures with Sample 1 was 30 minutes. The average time taken to complete the measures for test-retest was 14 minutes.

Procedures for Samples 2 to 4

Although the procedures for gathering data from Samples 2 to 4 differed from the procedure for Sample 1, data from all four samples were collected via a video call with a researcher. All measures were researcher administered via screen-sharing, so that participants were able to see the items and response options and items were read aloud by a researcher.

Informed consent from Sample 2 was obtained by trained researchers following the same protocol. Participants in Sample 2 completed the same measures as Sample 1, at baseline and again after completion of the STORM intervention, around 3 to 4 months from baseline. Participants in Sample 2 also completed the Client Service Receipt Inventory (Beecham & Knapp, 2001).

Informed consent from Sample 3 was also obtained following the same protocol as Sample 1. Participants completed the WEMWBS-ID, adapted RSES and the Responding to Intellectual Disability Stigma tool (Goldsmith-Sumner, 2021).

A researcher obtained informed consent from Sample 4 who completed the WEMWBS-ID and additional measures virtually at three time points (baseline, post-intervention and eight-week follow-up).

Analysis

Analyses were conducted using SPSS and SPSS AMOS versions 28. Although Samples 2 and 4 had completed the measures at multiple timepoints, for the purposes of the present study, only the baseline data were analysed.

Study 1: Psychometric Properties of the 14-item WEMWBS-ID

Data from Samples 1 to 3 were analysed in Study 1 (*n* = 96). All participants in these three samples had completed the 14-item WEMWBS-ID and 95 participants had completed the adapted RSES (one participant in Sample 2 did not fully complete this measure).

Reliability was assessed by calculating Cronbach’s alpha to determine internal consistency. Internal consistency measures the extent to which items in a scale correlate, thus measure the same concept (Terwee et al., 2007). A low Cronbach’s alpha indicates a lack of correlation between the items, whereas a very high Cronbach’s alpha indicates high correlations between items, which may suggest redundancy of one or more items (Terwee et al., 2007). With regards to test-retest reliability, it has been recommended that the period of time between administrations of a scale should be long enough to prevent recall, but short enough to ensure that change has not occurred. Therefore, a period of 1 or 2 weeks is often appropriate (Terwee et al., 2007). Intra-class coefficients (ICC) were calculated to examine test-retest reliability for the subsample of Sample 1 who completed the WEMWBS-ID again 1 to 2 weeks later (*n* = 22). A correlational analysis between total scores on the WEMWBS-ID and adapted RSES was conducted in order to assess convergent validity. Table 2 summarises the criteria used to interpret the results of these analyses. It should be noted that there is variability in the literature with regards to interpretation. For example, George and Mallery (2003) suggest that Cronbach’s alpha of >0.7 should be interpreted as ‘acceptable’.

Table 2

Criteria used to interpret the results of reliability and convergent validity tests.

Measure	Range	Rating	Source
Correlation coefficient	.7 – 1	Excellent	Hinkle et al. (2002)
	.5 – .69	Good	
	.3 – .49	Fair	
	< .29	Poor	
Cronbach's alpha	0.9 – 1	Excellent	Cicchetti (1994)
	0.8 – 0.89	Good	
	0.7 – 0.79	Fair	
	0.0 – 0.69	Poor	
Intra-class correlations	.7 – 1	Excellent	Cicchetti and Sparrow (1981)
	.5 – .69	Good	
	.3 – .49	Fair	
	< .29	Poor	

The standard error of measurement (SEM) was calculated to evaluate the variability of error within the WEMWBS-ID. A person's observed score on any measure equals their true score in addition to the SEM (Leong & Huang, 2010). The SEM was calculated as follows: SEM= σ (√1− α), where σ is the standard deviation and α is its Cronbach's alpha. 95% confidence intervals were then calculated using, 95% CI = Score ± (1.96*SEM).

A confirmatory factor analysis was conducted to assess structural validity. Model of fit indices calculated included: (i) chi-square and *p* value (chi-square value closer to zero and *p* > .05 indicate good fit), (ii) standardised Root Mean Square Error of Approximation (RMSEA; ≤ .06 indicates an good fit [Hu & Bentler, 1999]), (iii) Comparative Fit Index (CFI; ≥ 0.95 indicates good fit [Hu & Bentler, 1999]), and (iv) Tucker-Lewis Index (TLI; ≥ .90 indicates a good fit [Bentler & Bonett, 1980]). Factor loadings < .40 were considered weak and those ≥ .60 were considered strong (Garson, 2010).

Response distribution and floor and ceiling effects were also examined. Floor and ceiling effects were considered absent if less than 15% of participants scored the minimum or maximum score possible (Terwee et al., 2007).

In order to compare the scores reported in the present study with data from the general population for the original WEMWBS, a linear transformation of the scores from the four- to five-point scale was conducted. Table 3 summarises the original and transformed WEMWBS-ID scores. The mean total transformed score was calculated for the WEMWBS-ID.

Table 3

Transformed WEMWBS-ID and SWEMWBS-ID scores.

Original Score	Reduced-Scale Score	Transformed Score
0	0	1
1	0.33	2.33
2	0.66	3.66
3	1	5

Study 2: Psychometric Properties of the 7-item SWEMWBS-ID

Responses to the seven items which constitute the SWEMWBS-ID were analysed separately in Study 2, which used data from Samples 1 to 4 (n = 123). Unfortunately, due to human error, participants in Sample 4 completed a version of the WEMWBS-ID which included an incorrectly worded item (item 13 was worded, “I was interested in things”, instead of, “I was interested in new things”). As this item did not form part of the SWEMWBS, data from Sample 4 were only included in the analysis of the SWEMWBS-ID.

Internal consistency (Cronbach's alpha), test-retest reliability (ICC), convergent validity (correlational analysis) and structural validity (confirmatory factor analysis) of the SWEMWBS-ID were assessed, in addition to response distribution and floor and ceiling effects. The SEM was calculated and the transformed mean total score is also reported for the SWEMWBS-ID to allow comparison with reported general population norms.

Results

Study 1: WEMWBS-ID

Combined data from 96 participants (51 males) with a mean age of 39 (*SD* = 12.17) and a range of 18-74 years old were analysed. There were no missing data for the WEMWBS-ID.

Reliability Analysis

Cronbach's alpha was good ($\alpha = 0.86$), with no item indicated for potential removal. Item-total correlations ranged between .178 - .706 (lowest for item 11 ["I felt able to make my own decisions"]; highest for item 8 ["I felt good about myself"]).

With regards to test-retest reliability, the average number of days between the initial completion of the WEMWBS-ID and repeat administration was eight days. Test-retest reliability was excellent (ICC = .88, 95% CI = .72, .95).

Standard Error of Measurement (SEM)

The SEM was 2.87 (SEM = 7.66 ($\sqrt{}1 - 0.86$)). The SEM can determine confidence levels of scores on the WEMWBS-ID. The observed total WEMWBS-ID scores ranged from 10 to 42 (M = 27.39; SD = 7.66; 95% CI [21.76, 33.02]).

Convergent Validity

The mean average score for the adapted RSES was 12.67 (*SD* = 3.42). The scores were negatively skewed (-.500). Therefore, a Spearman's rank correlation (one-tailed) was computed to assess the relationship between total scores on the WEMWBS-ID and the adapted RSES. There was a significant positive correlation between scores on the two measures, $r(93) = .678$, $p < .001$, indicating good convergent validity and, as expected, a positive relationship between self-esteem and mental wellbeing, as assessed by these measures.

Structural Validity

The results from the confirmatory factor analysis are presented in Tables 4 and 5. The indices indicated an adequate model fit. Factor loadings were all statistically significant and varied from .28-.69. Factor loadings for items 1, 6 and 11 were weak, whilst those for items 3, 5, 7, 8 and 10 were strong. The average value explained by each item was $R^2 = 0.32$ (range = .076 - .618).

Table 4

Results from the confirmatory factor analysis of the WEMWBS-ID.

Index of fit	Value
Chi-square	117.45
Probability level	.002
Degrees of freedom	77
RMSEA	.066
CFI	.895
TLI	.857

Table 5

Confirmatory factor analysis factor loadings for the WEMWBS-ID.

Item	1	2	3	4	5	6	7
Factor loading	0.36	0.43	0.72	0.46	0.69	0.38	0.60
Item	8	9	10	11	12	13	14
Factor loading	0.79	0.52	0.74	0.28	0.58	0.59	0.57

Response Distribution and Floor and Ceiling Effects

Possible total scores on the WEMWBS-ID range from 0 to 42, with a higher score indicating higher levels of mental wellbeing. The observed total WEMWBS-ID scores ranged from 10 to 42 (*M* = 27.39; *SD* = 7.66). No participants scored the minimum score and one participant (1.04%) scored the maximum of 42, suggesting that floor and ceiling effects were absent. The distribution of the scores (Figure 1) showed the skewness of the data to be approximately symmetric (0.084) with a kurtosis value of -0.880. Visual examination of the histogram approximated to a normal distribution.

The distribution of responses per item is outlined in Table 6. All four response options were used by at least one person for all 14 items. The two most positive response options, ‘often’ and ‘always’, were chosen for 59.52% of the total responses. The data suggested that ‘always’ was the most frequently selected response option, followed by ‘sometimes’.

Following the linear transformation, the mean total WEMWBS-ID score was 50.48 (*SD* = 10.23). This is slightly lower than the mean score of 51.0 in UK general population samples who completed the original WEMWBS (Tennant et al., 2007).

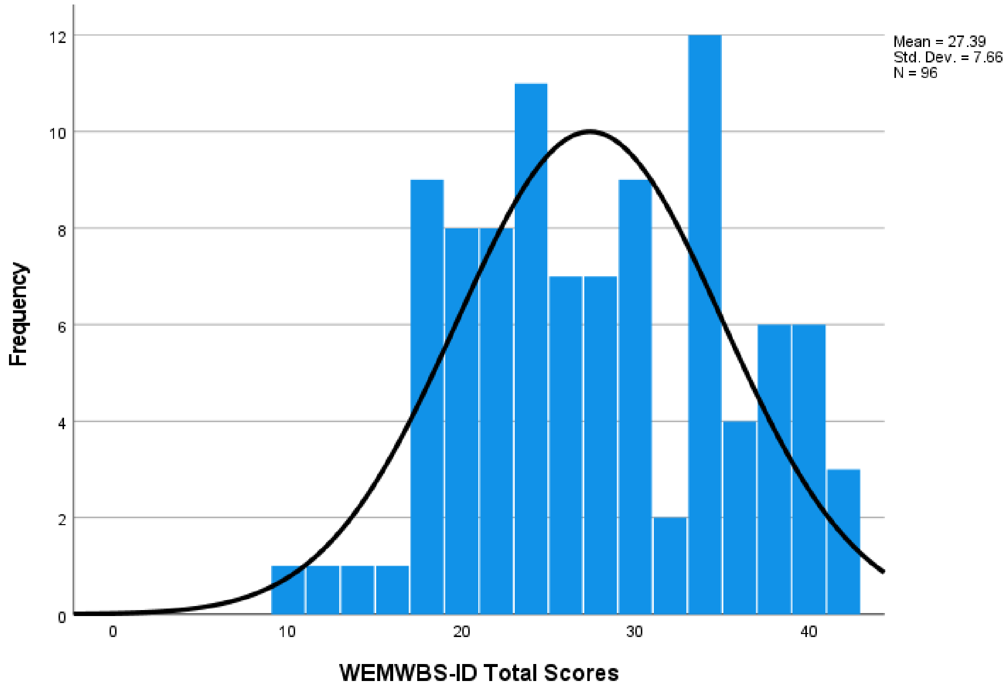


Figure 1. Distribution of WEMWBS-ID total scores.

Table 6

Frequency of responses per item (*n* = 96).

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	**Total (%)**
Response (*n*)															
Never	5	5	2	2	3	9	3	6	2	3	1	2	8	17	5.06
Sometimes	46	31	36	28	49	54	48	29	32	34	24	16	22	27	35.42
Often	19	25	18	22	14	14	24	22	18	21	22	15	20	21	20.46
Always	26	35	40	44	30	19	21	39	44	38	49	63	46	31	39.06

Study 2: SWEMWBS-ID

To test the SWEMWBS-ID, combined data from 123 participants (62 males) with a mean age of 39 (*SD* = 12.40) and a range of 18 to 74 years old were analysed. There were no missing data for the SWEMWBS-ID.

Reliability Analysis

The Cronbach's alpha was rated as poor ($\alpha = 0.67$), with no improvement indicated from removing any item. Item-total correlations ranged between .251-.517 (lowest for item 11; highest for item 7 ["I thought clearly"]). Test-retest reliability was good, ICC = .67, 95% CI [.58, .76].

Standard Error of Measurement

The SEM was 2.13 (SEM = 3.70 ($\sqrt{}1- 0.67$)). The observed total SWEMWBS-ID scores ranged from 4 to 21 (M = 12.87; SD = 3.70; 95% CI [8.70, 17.04]).

Convergent Validity

A Spearman's rank correlation (one-tailed) was computed to assess the relationship between total scores on the SWEMWBS-ID and the adapted RSES for the 95 participants who completed both measures. There was a significant positive correlation between scores on the two measures ($r(93) = .614$, $p < .001$), which was rated as 'good'.

Structural Validity

The results from the confirmatory factor analysis are presented in Tables 7 and 8. The indices indicated a good model fit. Factor loadings were all statistically significant and varied from .31-.67. The factor loading for item 11 was considered weak, whilst the factor loading for item 7 was considered strong. The average value explained by each item was $R^2 = 0.24$ (range = .097 - .443).

Table 7

Results from the confirmatory factor analysis of the SWEMWBS-ID.

Index of fit	Value
Chi-square	18.93
Probability level	.168
Degrees of freedom	14
RMSEA	.054
CFI	.947
TLI	.920

Table 8

Confirmatory factor analysis factor loadings for the SWEMWBS-ID.

Item	1	2	3	6	7	9	11
Factor loading	0.41	0.45	0.56	0.48	0.67	0.48	0.31

Response Distribution and Floor and Ceiling Effects

Possible total scores on the SWEMWBS-ID range from 0 to 21. The observed total SWEMWBS-ID scores ranged from 4 to 21 (*M* = 12.87; *SD* = 3.70). No participants scored the minimum score and three participants (2.44%) scored the maximum of 21, suggesting that floor and ceiling effects were absent. The distribution of the scores (Figure 2) showed the skewness of the data to be approximately symmetric (0.113) with a kurtosis value of -0.280. Visual examination of the histogram approximated to a normal distribution.

The distribution of responses per item is outlined in Table 9. All four response options were used by at least one person for all seven items. The two most positive response options, ‘often’ and ‘always’, were chosen for 54.93% of the total responses. The data suggested that ‘sometimes’ was the most frequently selected response option, followed by ‘always’.

Following the linear transformation, the mean total SWEMWBS-ID score was 24.14 (*SD* = 4.93). This is slightly higher than the mean score of 23.5 in UK general population samples using the original scale (Ng Fat et al., 2016). It is important to note that 27 participants in this sample lived in Canada and therefore a comparison with data from the UK general population should be made with caution.

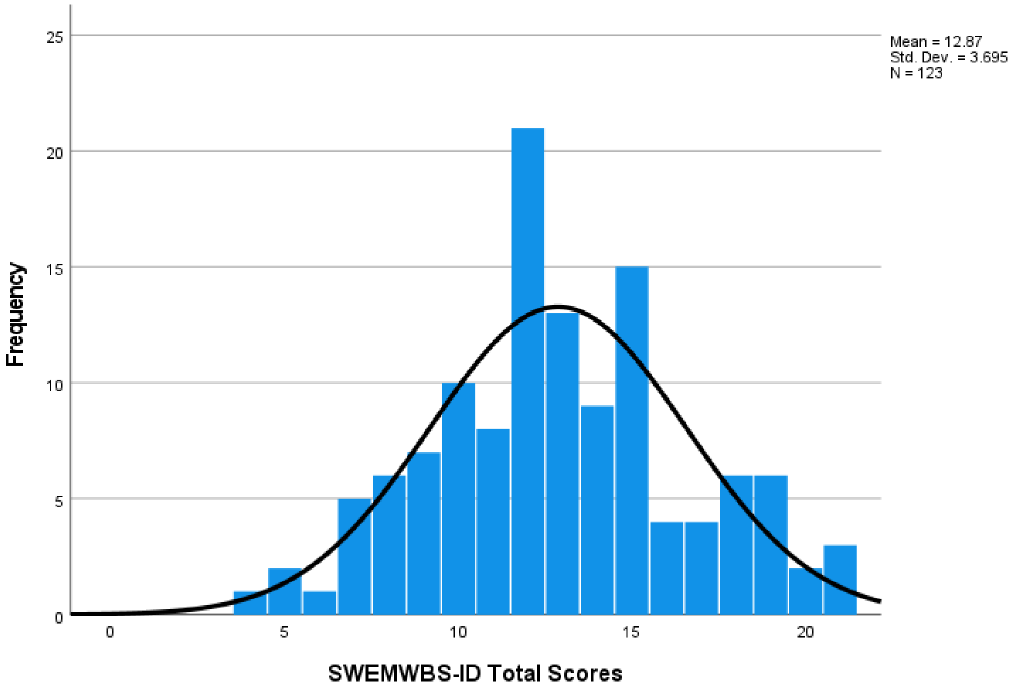


Figure 2. Distribution of SWEMWBS-ID total scores.

Table 9

Frequency of responses per item (*n* =123).

Item	1	2	3	6	7	9	11	**Total (%)**
Response (*n*)								
Never	7	7	2	12	4	2	1	4.07
Sometimes	54	39	50	68	64	44	34	41.00
Often	31	34	25	19	30	24	26	21.95
Always	31	43	46	24	25	53	62	32.98

Discussion

The aim of this study was to assess the psychometric properties of a newly adapted version of the WEMWBS and SWEMWBS for individuals with intellectual disabilities (WEMWBS-ID/SWEMWBS-ID). It was hypothesised that both scales would demonstrate good internal consistency, test-retest reliability and convergent validity. Additionally, it was hypothesised that a confirmatory factor analysis would support that the WEMWBS-ID and SWEMWBS-ID correspond to a one-factor structure.

Summary of Main Findings

The results of the present study suggest that the 14-item WEMWBS-ID demonstrated good internal consistency and excellent test-retest reliability. There was a significant positive correlation with the adapted RSES, demonstrating good convergent validity. Confirmatory factor analysis supported the hypothesised one-factor structure and the measure demonstrated an adequate model fit. These results were consistent with the hypotheses. Floor and ceiling effects did not appear to be present. With regards to response distribution, the majority

(59.52%) of participants selected one of the two most positive response options, but the option ‘sometimes’ was among the two most frequently chosen responses.

The SWEMWBS-ID demonstrated poor internal consistency, good test-retest reliability and good convergent validity. The results from the confirmatory factor analysis indicated a good model fit. The response option ‘sometimes’ was most popular, followed by ‘always’.

The internal consistency of the 7-item measure (α = 0.67) was lower than expected, compared to the original SWEMWBS (α = 0.84; Ng Fat et al., 2017). In order to try to understand this, individual items were explored further. Item 11 (“I felt able to make my own decisions”) had the lowest inter-item correlation for both scales. Furthermore, the factor loadings for item 11 in the confirmatory factor analysis of both scales were considered weak. On visual examination of the frequency of responses per item, with regards to the WEMWBS-ID, 51.04% of participants chose the option ‘always’ in response to this item, which was greater than the average proportion of ‘always’ responses overall (39.06%). This distribution was also reflected in the SWEMWBS-ID (50.41% of participants selected ‘always’ in response to item 11 compared to 32.98% of responses overall). It has been reported that the rights of individuals with intellectual disabilities to autonomous decision making have often been restricted (Werner, 2012, 2015) and that they are often perceived as being incapable of handling decisions. Therefore, one may expect individuals with intellectual disabilities to respond less positively to this item. However, it was noted that 69.79% of participants in Study 1 were members of self-advocacy groups. In recent systematic reviews (Fenn & Scior, 2019; Tilley et al., 2020) it was reported that self-advocacy group membership was associated with empowerment and increases in confidence in people with intellectual disabilities. Therefore, the participants in the present study may have been more likely to report that they were always able to make their own decisions.

Although removing item 11 from the scale in the future may improve the overall reliability and validity of the measure, it seems important to retain an item which provides valuable information about empowerment and autonomy in a group of individuals who are at increased risk of experiencing restrictions to their rights to decision making.

With regards to response distribution, participants had a tendency to select the response options ‘sometimes’ and ‘always’, with fewer participants choosing the option ‘often’. This suggests that there was not a response bias towards the more positive response choices. The response distribution for the original WEMWBS completed by the general population was presented graphically by Tennant et al. (2007). On visual examination, ‘sometimes’ and ‘often’ were the most frequently chosen options, with ‘often’ being the most popular response overall. This raises questions about the way that the term ‘often’ was being interpreted by participants with intellectual disabilities in the present study. It may be that this word is conceptually more abstract compared to ‘sometimes’ and ‘always’ and may require an individual to estimate how frequently something occurs beyond “it happens but not all of the time”, (i.e. ‘sometimes’) or “it happens all of the time” (i.e. ‘always’). In the paper by Tennant et al. (2007), ‘never’ represented less than 10% of chosen responses. ‘Never’ was also the least frequently chosen option by participants in the present study.

The results from the present study suggest that the transformed scores on the WEMWBS-ID were slightly lower than that reported in general population samples using the original scale (mean difference of -0.52). The SWEMWBS-ID scores were slightly higher than those found in general population samples, although it is important to note that this scale included item 11, which had a disproportionately higher number of ‘always’ responses. As a proportion of this sample lived in Canada, this finding should be interpreted with caution when being compared to data from the UK general population. The differences in the administration of the measure should also be acknowledged when making comparisons with data from the general population who completed the original WEMWBS (the adapted version was researcher-administered, as opposed to being completed independently, which may have introduced bias). It is also important to consider the context in which this study took place when interpreting these results. It has been reported that the COVID-19 pandemic had a disproportionate impact on individuals with intellectual disabilities, who were more likely to be infected and who suffered higher rates of hospitalisation and mortality (Lunsky et al., 2022). Understandably, the pandemic negatively impacted on the wellbeing of these individuals (e.g. Lake et al., 2021). Therefore, it may be expected that levels of wellbeing in the samples analysed in the present study were lower compared to those reported in studies which were conducted prior to the pandemic. However, it was noted that the levels of wellbeing reported were only slightly lower compared to the general population. It is possible that being a member of a self-advocacy group also had a protective effect, as research suggests that increases in self-esteem and subjective wellbeing are associated with self-advocacy group membership in people with intellectual disabilities (Fenn & Scior, 2019; Tilley et al., 2020).

Study Strengths and Limitations

Although the sample sizes in the present study were adequate for an initial exploration of the psychometric properties of the measure, they were relatively small which somewhat limited the results. The recommendations for sample sizes for factor analyses vary considerably and although the item: case ratios for Studies 1 and 2 were approximately 1:7 and 1:18 respectively. the sample sizes were considerably smaller compared to other psychometric evaluations of the WEMWBS and SWEMWBS.

Although the WEMWBS-ID may be self-administered, it was administered during an interview by a researcher for all four samples. This was to ensure that all participants were offered appropriate support to comprehend the items and response scale and also to ensure that the measure was administered in a standardised way across the different studies. This method required greater resources compared to administering the measure via a survey, which limited the sample sizes. Furthermore, it is possible that this format may have introduced forms of bias, such as socially desirable responding or acquiescence (Kooijmans et al., 2022). However, a strength of this approach was that during the interviews, the researchers were able to assess whether participants understood the items and response options well enough to complete the measure, which would not have been possible if the scale was completed independently. A further strength of this approach is that by administering the scale during an interview, a 100% completion rate for the WEMWBS-ID was achieved and there were no missing data.

With regards to sample characteristics, due to the recruitment strategy, the majority of participants in Study 1 were self-advocacy group members. Therefore, it is likely that the sample was less representative of individuals with intellectual disabilities who, for example, were more socially isolated and therefore had lower levels of overall wellbeing. Furthermore, as the participants' level of intellectual disability was not assessed, it is not possible to draw conclusions about the suitability of the WEMWBS-ID for subgroups of the population (i.e. whether it is more suitable for those with mild compared to those with moderate intellectual disabilities). Due to the cognitive and communicative abilities required to participate in the studies, it was likely that the majority of the participants had a mild intellectual disability. Additionally, participants were implicitly required to have sufficient visual, auditory and attentional skills. This was in addition to being required to have access to a device which was connected to the internet. These factors are likely to have inadvertently excluded a number of individuals with intellectual disabilities, thus limiting the representativeness of the samples. However, in terms of representativeness of the samples, the study has many strengths. Firstly, there were approximately even numbers of male and female participants recruited (50.41% vs 48.78% in across all four samples). With regards to the ethnicity of participants, in Study 1, 84.38% of participants reported that they were White British/Other, which is close to the proportion of people identifying as White in the UK (84.8%; Office for National Statistics, 2019). With regards to the recruitment strategy, a further strength of the study was that meeting with participants virtually allowed individuals from across the country to participate without additional resources required. Additionally, those who would not have otherwise been able to meet with a researcher in person, due to government restrictions during the pandemic, or pre-existing health conditions, were able to participate using this method.

It is also important to note that the linear transformation conducted in the present study only offers an approximation of scores if participants had used the five-point scale to enable comparisons, rather than an exact figure. This was included to put the results into context when interpreted. There are limitations associated with translating the four-point scale to a five-point scale for comparison, for example, making the assumption that the intervals between the points are equivalent. The wording in the five-point scale also differed to the four-point scale (for example, the option “some of the time” was used instead of “sometimes”) and therefore the transformed scores should be interpreted as equivalent to the original scores on the five-point scale.

Suggestions for Future Research

Further exploration of the psychometric properties of the WEMWBS-ID and SWEMWBS-ID with a larger sample size is required to provide further support for the reliability and validity of the scales. Future research should go beyond self-advocacy groups to ensure that samples are more representative of the intellectual disability population. As discussed previously, there was a disproportionately high number of ‘always’ responses for item 11 (“I felt able to make my own decisions”) which is likely to have affected overall internal consistency. This may have been due to the high number of participants recruited from self-advocacy groups. Therefore, recruiting a more representative sample may improve the overall psychometric quality of the scales.

Future research involving self-administration of the scale may allow data collection on a greater scale. Larger sample sizes would also enable further investigation into other psychometric properties, such as responsiveness or sensitivity to change. Responsiveness of the original SWEMWBS has been evaluated and therefore it would be useful to also establish whether the adapted version for individuals with intellectual disabilities can also detect clinically important changes in wellbeing following an intervention. It has been recommended that for analyses of internal responsiveness using paired *t*-tests, sample sizes of at least 30 participants per study are required (Husted et al., 2000). In the present study, Samples 2 and 4 had completed the WEMWBS-ID post-intervention, but the sample sizes were too small to investigate internal responsiveness ($n = 21$ and 24, respectively). Additionally, a reference measure of wellbeing had not been administered to either sample and therefore explorations into external responsiveness (the extent to which changes in a measure relate to corresponding changes in a reference measure) were not possible. Therefore, future researchers using the WEMWBS-ID to measure change following an intervention may also wish to administer an additional measure of a related construct, such as quality of life. The SEM of both scales have been calculated in the present study to aid future researchers in ascertaining the smallest detectable change (the number of points on a scale for a change in scores to have been considered a result of an intervention rather than chance), using the following equation: $1.96 * \sqrt{2} * SEM$. Furthermore, divergent validity (the extent to which believed unrelated constructs are, in fact, unrelated) was not assessed in the present study and therefore future researchers may assess this in order to provide further support for the scales' validity.

Gathering information such as participants' IQ scores or severity of intellectual disability may allow further exploration of within-group differences. It may be of interest to assess whether the scales are more reliable or valid with for example, individuals with mild or moderate intellectual disabilities.

Future researchers may also wish to further explore the utility of the four-point response scale of the WEMWBS-ID to ascertain whether the response option 'often' is acceptable and produces reliable responses. A three-point Likert scale ('never', 'sometimes', 'always') could potentially be piloted as this may be more suitable, however this has implications for detecting change due to the reduced variability in scores.

Conclusions and Clinical Implications

To date, there is a paucity of psychometrically validated self-report measures of positive mental health or mental wellbeing for individuals with intellectual disabilities. The results from this preliminary evaluation of the WEMWBS-ID and SWEMWBS-ID suggest that the scales had promising psychometric properties, when administered by a researcher to individuals with mild to moderate intellectual disabilities. The WEMWBS-ID appeared to have good reliability and adequate validity, whilst the 7-item version of the scale demonstrated adequate reliability and good validity. A further exploration of the scales in larger sample sizes is warranted.

There are numerous stakeholders interested in measuring wellbeing, including clinicians, researchers and policy makers in public health. Individuals presenting to mental health services are often administered a range of measures of mental health problems or distress. The WEMWBS-ID may be a helpful additional measure, enabling clinicians to measure positive aspects of wellbeing and identify a person’s strengths and resources, which may inform their intervention. Furthermore, this measure allows clinicians to gain an understanding of the individuals’ own experiences from their perspective, without relying solely on the report of carers or staff members. Once responsiveness of the SWEMWBS-ID has been established, the scale may be used in clinical settings to monitor wellbeing in clinical populations over time. With regards to implications for research, the scale may also be used in clinical trials and intervention studies aimed at improving mental wellbeing in this population. In public health practice, it is recognised that measurement enables understanding of population health status and trends over time (Faculty of Public Health and Mental Health Foundation, 2016). The WEMWBS-ID will enable the wellbeing of individuals with mild to moderate intellectual disabilities to be better understood and measured at a population level.

References

Ali, A., Hassiotis, A., Strydom, A., & King, M. (2012). Self stigma in people with intellectual disabilities and courtesy stigma in family carers: A systematic review. *Research in Developmental Disabilities, 33*(6), 2122-2140. https://doi.org/10.1016/j.ridd.2012.06.013

Ali, A., Strydom, A., Hassiotis, A., Williams, R., & King, M. (2008). A measure of perceived stigma in people with intellectual disability. *The British Journal of Psychiatry, 193*(5), 410-415. https://doi.org/10.1192/bjp.bp.107.045823

Anderson, C., John, O. P., & Keltner, D. (2012). The personal sense of power. *Journal of Personality, 80*(2), 313-344. https://doi.org/10.1111/j.1467-6494.2011.00734.x

Bass, M., Dawkin, M., Muncer, S., Vigurs, S., & Bostock, J. (2016). Validation of Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in a population of people using secondary care Mental health services. *Journal of Mental health, 25*(4), 323-329.

Beecham, J. & Knapp, M. (2001) Costing psychiatric interventions, in G. Thornicroft (ed.) *Measuring Mental Health Needs* (2nd edition), 200-224.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Brooks, M., Davies, S., & Twigg, E. (2013). A measure for feelings-using inclusive research to develop a tool for evaluating psychological therapy (Clinical Outcomes in Routine Evaluation-Learning Disability). *British Journal of Learning Disabilities, 41*(4), 320-329. https://doi.org/10.1111/bld.12020

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284. https://doi.org/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, 86,* 127-137.

Claes, C., Vandevelde, S., Van Hove, G., van Loon, J., Verschelden, G., & Schalock, R. (2012). Relationship between self-report and proxy ratings on assessed personal quality of life-related outcomes. *Journal of Policy and Practice in Intellectual Disabilities*, *9*(3), 159-165. https://doi.org/10.1111/j.1741-1130.2012.00353.x

Cleaver, S., Ouellette-Kuntz, H., & Sakar, A. (2010). Participation in intellectual disability research: a review of 20 years of studies. *Journal of Intellectual Disability Research, 54*(3), 187-193. https://doi.org/10.1111/j.1365-2788.2010.01256.x

Cluley, V. (2018). From "Learning disability to intellectual disability"-Perceptions of the increasing use of the term "intellectual disability" in learning disability policy, research and practice. *British Journal of Learning Disabilities, 46*(1), 24-32. https://doi.org/10.1111/bld.12209

Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist, 44*(5), 730-757. https://doi.org/10.1177/0011000016633507

Crawford, M. J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., ... & Rose, D. (2011). Selecting outcome measures in mental health: the views of service users. *Journal of Mental Health, 20*(4), 336-346. https://doi.org/10.3109/09638237.2011.577114

Dagnan, D., & Sandhu, S. (1999). Social comparison, self-esteem and depression in people with intellectual disability. *Journal of Intellectual Disability Research, 43*(5), 372-379. https://doi.org/10.1046/j.1365-2788.1999.043005372.x

Department of Health (2001). *Valuing people*. London: HMSO.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*(2), 276. https://doi.org/10.1037/0033-2909.125.2.276

Dunn, K., Rydzewska, E., Fleming, M., & Cooper, S. A. (2020). Prevalence of mental health conditions, sensory impairments and physical disability in people with co-occurring intellectual disabilities and autism compared with other people: A cross-sectional total population study in Scotland. *BMJ Open, 10*(4), e035280. https://doi.org/10.1136/bmjopen-2019-035280

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology, 68*(4), 435-441. https://doi.org/10.1016/j.jclinepi.2014.11.021

Evans, E., Howlett, S., Kremser, T., Simpson, J., Kayess, R., & Trollor, J. (2012). Service development for intellectual disability mental health: a human rights approach. *Journal of Intellectual Disability Research, 56*(11), 1098-1109. https://doi.org/10.1111/j.1365-2788.2012.01636.x

Faculty of Public Health and Mental Health Foundation. (2016). *Better mental health for all: A public health approach to mental health improvement*. London.

Fenn, K., & Scior, K. (2019). The psychological and social impact of self-advocacy group membership on people with intellectual disabilities: A literature review. *Journal of Applied Research in Intellectual Disabilities, 32*, 1349–1358. https://doi.org/10.1111/jar.12638

Fung, S. F. (2019). Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese university students. *Health and Quality of Life Outcomes, 17*(1), 1-9. https://doi.org/10.1186/s12955-019-1113-1

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2015). Toward a new definition of mental health. *World Psychiatry, 14*(2), 231. https://doi.org/10.1002/wps.20231

Garson, D. (2010). Statnotes: *Topics in Multivariate Analysis: Factor Analysis.* Retrieved from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference (4th edition). Boston: Allyn & Bacon.

Goldsmith-Sumner, A. (2021). Development of a tool to assess how people with intellectual disabilities respond to stigma [Doctoral thesis, University College London]. https://discovery.ucl.ac.uk/id/eprint/10135570/8/Goldsmith-Sumner_10135570_thesis_sigs_removed.pdf

Goodley, D. (2011). *Self-advocacy in the lives of people with learning disabilities.* Maidenhead: Open University Press.

Green, S., Davis, C., Karshmer, E., Marsh, P., & Straight, B. (2005). Living stigma: The impact of labeling, stereotyping, separation, status loss, and discrimination in the lives of individuals with disabilities and their families. *Sociological Inquiry*, *75*(2), 197–215. https://doi.org/10.1111/soin.2005.75.issue-2

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265-275. https://doi.org/10.1037/0033-2909.103.2.265

Guerin, S., Dodd, P., Tyrell, J., McEvoy, J., Buckley, S., & Hillery, J. (2009). An initial assessment of the psychometric properties of the Complicated Grief Questionnaire for People with Intellectual Disabilities (CGQ-ID). *Research in Developmental Disabilities*, *30*(6), 1258-1267. https://doi.org/10.1016/j.ridd.2009.05.002

Hartley, S. L., & MacLean Jr, W. E. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research*, *50*(11), 813-827. https://doi.org/10.1111/j.1365-2788.2006.00844.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

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

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. https://doi.org/10.1080/10705519909540118

Huppert, F. A. (2009). Psychological well-being: Evidence regarding its causes and consequences. *Applied Psychology: Health and Well-Being, 1*(2), 137-164. https://doi.org/10.1111/j.1758-0854.2009.01008.x

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology, 53*(5), 459-468. https://doi.org/10.1016/S0895-4356(99)00206-1

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73*(3), 539. https://doi.org/10.1037/0022-006X.73.3.539

Keyes, C. L., Dhingra, S. S., & Simoes, E. J. (2010). Change in level of positive mental health as a predictor of future risk of mental illness. *American Journal of Public Health, 100*(12), 2366-2371. https://doi.org/10.2105/AJPH.2010.192245

Keyes, C. L., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: the empirical encounter of two traditions. *Journal of Personality and Social Psychology, 82*(6), 1007. https://doi.org/10.1037/0022-3514.82.6.1007

Kim, H., Doiron, K., Warren, M., & Donaldson, S. (2018). The international landscape of positive psychology research: A systematic review. *International Journal of Wellbeing, 8*(1). https://doi.org/10.5502/ijw.v8i1.651

Kline, P. (2013). *Handbook of psychological testing.* Routledge: London. https://doi.org/10.4324/9781315812274

Kong, F., Zhao, J., & You, X. (2013). Self-esteem as mediator and moderator of the relationship between social support and subjective well-being among Chinese university students. *Social Indicators Research, 112*(1), 151-161. https://doi.org/10.1007/s11205-012-0044-6

Kooijmans, R., Mercera, G., Langdon, P. E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: A systematic review. *Clinical Psychology: Science and Practice.* https://doi.org/10.1037/cps0000058

Koushede, V., Lasgaard, M., Hinrichsen, C., Meilstrup, C., Nielsen, L., Rayce, S. B., ... & Santini, Z. I. (2019). Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings. *Psychiatry Research, 271*, 502-509. https://doi.org/10.1016/j.psychres.2018.12.003

Lake, J. K., Jachyra, P., Volpe, T., Lunsky, Y., Magnacca, C., Marcinkiewicz, A., & Hamdani, Y. (2021). The wellbeing and mental health care experiences of adults with intellectual and developmental disabilities during COVID-19. *Journal of Mental Health Research in Intellectual Disabilities, 14*(3), 285-300. https://doi.org/10.1080/19315864.2021.1892890

Lambert, L., Lomas, T., van de Weijer, M. P., Passmore, H. A., Joshanloo, M., Harter, J., ... & Diener, E. (2020). Towards a greater global understanding of wellbeing: A proposal for a more inclusive measure. *International Journal of Wellbeing, 10*(2). https://doi.org/10.5502/ijw.v10i2.1037

Leong, F. T. L., & Huang, J. L. (2010). Standard error of measurement. In Salkind. N. J. (Ed.), *Encyclopaedia of research design.* Sage Reference.

Li, C., Tsoi, E. W., Zhang, A. L., Chen, S., & Wang, C. K. (2013). Psychometric properties of self-reported quality of life measures for people with intellectual disabilities: A systematic review. *Journal of Developmental and Physical Disabilities, 25*(2), 253-270. https://doi.org/10.1007/s10882-012-9297-x

Lunsky, Y., Durbin, A., Balogh, R., Lin, E., Palma, L., & Plumptre, L. (2022). COVID-19 positivity rates, hospitalizations and mortality of adults with and without intellectual and developmental disabilities in Ontario, Canada. *Disability and Health Journal, 15*(1), 1011174. https://doi.org/10.1016/j.dhjo.2021.101174

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-89. https://doi.org/10.1037/1082-989X.4.1.84

Marshall, K., & Willoughby-Booth, S. (2007). Modifying the clinical outcomes in routine evaluation measure for use with people who have a learning disability. *British Journal of Learning Disabilities*, *35*(2), 107-112. https://doi.org/10.1111/j.1468-3156.2006.00422.x

McGillivray, J. A., Lau, A. L. D., Cummins, R. A., & Davey, G. (2009). The utility of the personal wellbeing index intellectual disability scale in an Australian sample. *Journal of Applied Research in Intellectual Disabilities*, *22*(3), 276-286. https://doi.org/10.1111/j.1468-3148.2008.00460.x

McKay, M. T., & Andretta, J. R. (2017). Evidence for the psychometric validity, internal consistency and measurement invariance of Warwick Edinburgh Mental Well-being Scale scores in Scottish and Irish adolescents. *Psychiatry Research*, *255*, 382-386. https://doi.org/10.1016/j.psychres.2017.06.071

Melendez-Torres, G. J., Hewitt, G., Hallingberg, B., Anthony, R., Collishaw, S., Hall, J., ... & Moore, G. (2019). Measurement invariance properties and external construct validity of the short Warwick-Edinburgh mental wellbeing scale in a large national sample of secondary school students in Wales. *Health and Quality of Life Outcomes, 17*(1), 1-9. https://doi.org/10.1186/s12955-019-1204-z

Mileviciute, I., & Hartley, S. L. (2015). Self-reported versus informant-reported depressive symptoms in adults with mild intellectual disability. *Journal of Intellectual Disability Research, 59*(2), 158-169. https://doi.org/10.1111/jir.12075

Ng Fat, L., Mindell, J., Boniface, S., & Stewart-Brown, S. (2016). Evaluating and establishing national norms for the short Warwick-Edinburgh mental well-being scale (SWEMWBS) using the health survey for England. *Quality of Life Research, 26*(5), 1129-44.

Office for National Statistics (2019). *Population estimates by ethnic group and religion, England and Wales: 2019.* https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/populationestimatesbyethnicgroupandreligionenglandandwales/2019

O'Keeffe, L., Guerin, S., McEvoy, J., Lockhart, K., & Dodd, P. (2019). The process of developing self-report measures in intellectual disability: A case study of a complicated grief scale. *British Journal of Learning Disabilities, 47*(2), 134-144. https://doi.org/10.1111/bld.12261

Padhy, M., Rana, S., & Mishra, M. (2011). Self esteem and subjective wellbeing: Correlates of academic achievement of students. *Research Journal of Social Science & Management, 1*(7), 148-156.

Paterson, L., McKenzie, K., & Lindsay, B. (2012). Stigma, social comparison and self-esteem in adults with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 25*(2), 166-176. https://doi.org/10.1111/j.1468-3148.2011.00651.x

Raczka, R., Williams, J., & Theodore, K. (2014). Pilot evaluation of routine outcome measures for use with adults with learning disabilities. *The Bulletin of the Faculty for People with Intellectual Disabilities, 12*(1), 60-65.

Ringdal, R., Bradley Eilertsen, M. E., Bjørnsen, H. N., Espnes, G. A., & Moksnes, U. K. (2018). Validation of two versions of the Warwick-Edinburgh mental well-being scale among Norwegian adolescents. *Scandinavian Journal of Public Health, 46*(7), 718-725. https://doi.org/10.1177/1403494817735391

Rosenberg M. (1982). Self conceptions: Configurations of content. In M. Rosenberg & H. B. Kaplan (Ed.), *Social Psychology of the Self Concept*, pp. 110-123. Wheeling, IL: Harlan Davidson Inc.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press, Princeton, NJ. https://doi.org/10.1515/9781400876136

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. (2020). Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes, 18*(1), 1-16. https://doi.org/10.1186/s12955-020-01423-y

Ruddick, L., & Oliver, C. (2005). The development of a health status measure for self-report by people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 18*(2), 143-150.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*(1), 141-166. https://doi.org/10.1146/annurev.psych.52.1.141

Schalock, R. L., Brown, I., Brown, R., Cummins, R. A., Felce, D., Matikka, L., ... & Parmenter, T. (2002). Conceptualization, measurement, and application of quality of life for persons with intellectual disabilities: Report of an International Panel of Experts. *Mental Retardation, 40*(6), 457-470. https://doi.org/10.1352/0047-6765(2002)040<0457:CMAAOQ>2.0.CO;2

Scior, K., Richardson, L., Randell, E., Osborne, M., Bird, H., Ali, A., Bonin, E., Brown, A., Brown, C., Burke, C., Bush, L., Crabtree, J., Davies, K., Davies, P., Gillespie, D., Jahoda, A., Johnson, S., Hastings, R., Kerr, L., … & Zhang, K. (2022). *The STanding up fOR Myself (STORM) psychosocial group intervention for young people and adults with intellectual disabilities: Adaptation for digital delivery and pilot.* Manuscript submitted for publication.

Shah, N., Cader, M., Andrews, B., McCabe, R., & Stewart-Brown, S. L. (2021). Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health and Quality of Life Outcomes, 19*(1), 1-9. https://doi.org/10.1186/s12955-021-01882-x

Shah, N., Cader, M., Andrews, B., McCabe, R., & Stewart-Brown, S. L. (2021). Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health and Quality of Life Outcomes, 19(*1), 1-9. https://doi.org/10.1186/s12955-021-01882-x

Shah, N., Cader, M., Andrews, W. P., Wijesekera, D., & Stewart-Brown, S. L. (2018). Responsiveness of the short Warwick Edinburgh mental well-being scale

(SWEMWBS): evaluation a clinical sample. *Health and Quality of Life Outcomes, 16*(1), 1-7. https://doi.org/10.1186/s12955-018-1060-2

Siahpush, M., Spittal, M., & Singh, G. K. (2008). Happiness and life satisfaction prospectively predict self-rated health, physical health, and the presence of limiting, long-term health conditions. *American Journal of Health Promotion, 23*(1), 18-26. https://doi.org/10.4278/ajhp.061023137

Sigstad, H. M. H. (2014). Characteristic interviews, different strategies: Methodological challenges in qualitative interviewing among respondents with mild intellectual disabilities. *Journal of Intellectual Disabilities, 18*(2), 188-202. https://doi.org/10.1177/1744629514523159

Skevington, S. M., & Böhnke, J. R. (2018). How is subjective well-being related to quality of life? Do we need two concepts and both measures?. *Social Science & Medicine, 206*, 22-30. https://doi.org/10.1016/j.socscimed.2018.04.005

St. John, L., Volpe, T., Jiwa, M. I., Durbin, A., Safar, Y., Formuli, F., Thakur, A., Lake, J., & Lunsky, Y. (2022). 'More together than apart': The evaluation of a virtual course to improve mental health and well-being of adults with intellectual disabilities during the COVID-19 pandemic. *Journal of Applied Research in Intellectual Disabilities.* Advance online publication. https://doi.org/10.1111/jar.13024

Stewart-Brown, S., Samaraweera, P. C., Taggart, F., Kandala, N. B., & Stranges, S. (2015). Socioeconomic gradients and mental health: implications for public health. *The British Journal of Psychiatry, 206*(6), 461-465. https://doi.org/10.1192/bjp.bp.114.147280

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale

(WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes, 7*(1), 1-8.

Taggart F, Stewart-Brown S (2019) A review of questionnaires designed to measure mental wellbeing. https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/research/validation/frances_taggart_research.pdf. Accessed 15 February 2022.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J. & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 1-13. https://doi.org/10.1186/1477-7525-5-63

Terwee, C. B., Bot, S. D., de Boer, M. R., van Der Windt, D. A., Knol, D. L., Dekker, J., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Tilley, E., Strnadová, I., Danker, J., Walmsley, J., & Loblinzk, J. (2020). The impact of self-advocacy organizations on the subjective well-being of people with intellectual disabilities: A systematic review of the literature. *Journal of Applied Research in Intellectual Disabilities, 33*(6), 1151-1165. https://doi.org/10.1111/jar.12752

Townsend-White, C., Pham, A. N. T., & Vassos, M. V. (2012). A systematic review of quality of life measures for people with intellectual disabilities and challenging behaviours. *Journal of Intellectual Disability Research, 56*(3), 270-284. https://doi.org/10.1111/j.1365-2788.2011.01427.x

Trousselard, M., Steiler, D., Dutheil, F., Claverie, D., Canini, F., Fenouillet, F., ... & Franck, N. (2016). Validation of the Warwick-Edinburgh mental well-being scale

(WEMWBS) in French psychiatric and general populations. *Psychiatry Research, 245*, 282-290. https://doi.org/10.1016/j.psychres.2016.08.050

Vaingankar, J., Edimansyah, A., Chong, S., Sambasivam, R., Seow, E., Jeyagurunathan, A., et al. (2017). Psychometric properties of the short Warwick Edinburgh mental well-being scale (SWEMWBS) in service users with schizophrenia, depression and anxiety spectrum disorders. *Health and Quality of Life Outcomes, 15*, 153. https://doi.org/10.1186/s12955-017-0728-3

VanderWeele, T. J., Trudel-Fitzgerald, C., Allin, P., Farrelly, C., Fletcher, G., Frederick, D. E., ... & Kubzansky, L. D. (2020). Current recommendations on the selection of measures for well-being. *Preventive Medicine*, *133*, 106004. https://doi.org/10.1016/j.ypmed.2020.106004

Vlissides, N., Beail, N., Jackson, T., Williams, K., & Golding, L. (2017). Development and psychometric properties of the Psychological Therapies Outcome Scale-Intellectual Disabilities (PTOS-ID). *Journal of Intellectual Disability Research*, *61*(6), 549-559. https://doi.org/10.1111/jir.12361

Werner, S. (2012). Individuals with intellectual disabilities: A review of the literature on decision-making since the Convention on the Rights of People with Disabilities (CRPD). *Public Health Reviews*, *34*, 1-27. https://doi.org/10.1007/BF03391682

Werner, S. (2015). Public stigma and the perception of rights: Differences between intellectual and physical disabilities. *Research in Developmental Disabilities*, *38*, 262-271. https://doi.org/10.1016/j.ridd.2014.12.030

WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine*, *41*(10), 1403-1409. https://doi.org/10.1016/0277-9536(95)00112-K

Wille, N., Badia, X., Bonsel, G., Burström, K., Cavrini, G., Devlin, N., ... & Ravens-Sieberer, U. (2010). Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. *Quality of Life Research*, *19*(6), 875-886. https://doi.org/10.1007/s11136-010-9648-y

Part 3: Critical Appraisal

Introduction

This critical appraisal will focus on my reflections on the process of completing the systematic review and empirical paper. The appraisal begins with a consideration of how my previous experiences influenced why I chose this project. I will then outline the specific challenges I faced whilst completing both parts of the thesis. Finally, I will offer my reflections on quantitative measurement in the field of positive mental health.

Interest in this Project

My first experience of conducting research in the intellectual disability field was when I completed my undergraduate dissertation, supervised by my thesis supervisor Katrina. The focus of this research was on the relationship between implicit and explicit attitudes towards individuals with intellectual disabilities. Subsequently, I was involved in another study which reviewed attitudes to intellectual disability around the globe (Scior et al., 2015). I was struck by the high levels of stigma, prejudice and discrimination that individuals with intellectual disabilities still face worldwide. I also recognised that often the voices of these individuals were not captured within intellectual disability research and that studies frequently involved interviewing their carers or relatives instead.

Prior to clinical psychology training, I worked for three years as a Psychological Wellbeing Practitioner (PWP) in an Improving Access to Psychological Therapies (IAPT) service and also at a paediatric hospital, as part of a research project. I was responsible for routinely collecting data from self-report outcome measures, using questionnaires such as the GAD-7 (Spitzer et al., 2006) and the PHQ-9 (Kroenke et al., 2001). These measures served various purposes, for example enabling the evaluation of IAPT services nationally, monitoring clients' progress and the effectiveness of interventions, in addition to monitoring service performance outcomes such as 'access' and 'recovery' rates. On reflection, I had accepted that the use of routine outcome measures was a key part of the role of a PWP, without much consideration of how these measures had been developed, validated for use, or their limitations. When I started training, I learned more about strengths-based and values-led therapeutic models such as Acceptance and Commitment Therapy (Hayes & Strosahl, 2004), which focus less on the reduction of symptoms of mental health and more on helping clients to identify and engage with their own strengths and resources. I reflected on how this differed to my own experiences of working in mental health services prior to training and questioned why I had not come across a measure of positive mental health or mental wellbeing.

I was aware that that there was a growing interest in positive psychology and the promotion of positive mental health in the field. Having focussed on symptoms of mental ill-health for most of my career, I was keen to be involved in research about wellbeing and learn more about this area. I also wished to promote the active participation of individuals with intellectual disabilities in research, rather than relying on informant reports.

Challenges: Systematic Review

As somebody who was new to the field of wellbeing, the main challenge I faced when conducting the systematic review was understanding the terminology, the way in which mental wellbeing was conceptualised in the literature and the differences between mental wellbeing and related concepts, such as quality of life, health-related quality of life, subjective wellbeing and others. I often used the terms wellbeing and quality of life interchangeably and had not considered that there were differences between these terms. I endeavoured to gain more of an understanding of this and was surprised by the difficulties I experienced in finding published papers which clearly and consistently described the differences between mental wellbeing and other related constructs. Although the focus of my thesis was on mental wellbeing, I made the decision to include measures of quality of life in the review. It became apparent that researchers in the field were also using the terms interchangeably and therefore I did not wish to exclude articles which may have been relevant. The results of the review highlighted that there were a range of existing measures of mental health and quality of life for individuals with intellectual disabilities. However, there was a paucity of scales which were specifically focussed on measuring the eudaimonic and hedonic dimensions of mental wellbeing.

A further challenge related to the quality assessment tool used to evaluate the papers. The Characteristics of Assessment Instructions for Psychiatric Disorders in Persons with Intellectual Developmental Disorders (CAPs-IDD; Zeilinger et al., 2013), which is a framework for evaluating and describing measures of psychiatric disorders in people with intellectual disabilities, was chosen for the systematic review. Although the CAPs-IDD did not produce a total score pertaining to the quality of each measure, a strength of the tool was that it provides a comprehensive summary of the psychometric properties of measures. Furthermore, it was specific to measures for people with intellectual disabilities and therefore included items which examined e.g., the level of intellectual disability, whether individuals with intellectual disabilities were included in test development and whether the measure was acceptable to them. This tool therefore seemed to be the most relevant and suitable for the systematic review. However, as the CAPs-IDD was very thorough, using it turned out to be fairly time consuming, due to the high level of detail required to comprehensively summarise each measure. Also, the lack of a total score meant that a comparison of the quality of different measures was not easily done. Therefore, a further quality assessment component (the four-point scale used by Flynn et al., 2017) was included in the systematic review to enable such a comparison. I was not able to identify many published systematic reviews which had utilised the CAPs-IDD framework and therefore future researchers in the intellectual disability field may wish to consider using it to further explore its utility.

Challenges: Empirical Paper

Reflecting on the empirical study, I faced numerous challenges during the process. Firstly, during the initial planning of this project, researchers examining the impact of COVID-19 on individuals with intellectual disabilities had piloted the Warwick-Edinburgh Mental Wellbeing Scale-Intellectual Disability (WEMWBS-ID) and had agreed to share their data with us to contribute to the psychometric evaluation of the measure. Their research was conducted on a large scale and they had recruited around 600 participants. Unfortunately, due to human error, one item included in both the full and short versions of the scale had been worded incorrectly and therefore, after careful consideration, the decision was made to exclude these data from the analyses in the empirical paper altogether. Coincidentally, a different group of researchers in Canada who had also agreed to share their data with us had also accidently misworded an item. Fortunately, this item was not included in the 7-item short version of the measure and so we were able to include these data in the final analyses. However, the final sample sizes were consequently far smaller than we had initially hoped for. This experience highlighted one of the challenges which may be faced when collaborating with different teams of researchers, though there are of course many advantages in doing so.

During the initial planning phase, we also faced the dilemma of whether to collect data from Sample 1 in person, or virtually. We agreed on virtual data collection for a number of reasons. As data were collected during the COVID-19 pandemic, we were mindful that many potential participants may have been shielding or self-isolating. There was also uncertainty around whether further government restrictions limiting travel and face to face contact might prevent us from meeting with participants in person. Furthermore, we knew that many individuals with intellectual disabilities had comorbid long-term health conditions, which made them more vulnerable to experiencing severe symptoms associated with

COVID-19. Virtual data collection also enabled us to collect data from participants who were living across the country, rather than limiting us to Greater London. There were, however, disadvantages associated with limiting data collection to virtual interviews. For example, it was likely that the requirement for potential participants to have access to a smartphone, tablet or computer which was connected to the internet, inadvertently excluded many individuals with intellectual disabilities who may have otherwise been interested and eligible to participate. On balance, it was felt that the advantages of virtual data collection outweighed the disadvantages and therefore the decision was made to collect data virtually.

It is also worth considering the impact of the global pandemic on the results of research pertaining to mental wellbeing. The impact of the pandemic on the mental health of those with and without intellectual disabilities have been widely reported (Lake et al., 2021). During the interviews, many of the participants who took part in this project talked about difficulties such as feeling isolated and unhappy about social groups and activities taking place online rather than in person. It would be interesting to compare the results from the empirical study to scores on the WEMWBS-ID in a few years' time, as this may potentially help researchers understand the impact of the pandemic on mental wellbeing in this population and measure post-pandemic levels of wellbeing.

Another challenge faced during the data collection process pertained to the initial screening process. The poster which was used to advertise the study specified that we were interested in speaking to people who had a "learning disability", the most commonly used term in the UK outside of a research context. The use of an administrative definition of an intellectual disability (in terms of receipt of specialist services for people with intellectual disabilities), rather than requesting evidence of a formal diagnosis or IQ scores, posed a challenge. During the initial screening process, there were many instances of individuals expressing an interest in participating and then during the initial screening process reporting conditions such as autism, dyslexia or ADHD. Often this was labelled by potential participants as a “learning disability”, due to the effect of these conditions on their ability to learn when they were at school, or to conduct daily tasks. However, they were not in receipt of specialist services for individuals with intellectual disabilities, nor were they involved with charities or third sector organisations for people with intellectual disabilities. These individuals were therefore excluded from the project as there was no evidence that they had a comorbid intellectual disability. This highlights a potential challenge in intellectual disability research in the UK as often terms such as “learning difficulties”, or other neurodevelopmental conditions such as autism or ADHD, are used interchangeably with “learning disabilities” in the general population. On occasions when not completely certain whether a potential participant had an intellectual disability or not, we made the decision to exclude these individuals, in order to not compromise the validity of our results. Future researchers may wish to request information pertaining to an individual’s formal intellectual disability diagnosis, however this may inadvertently put people off from participating or exclude individuals who for whatever reason may not be able to provide this information.

Challenges in recruiting individuals with disabilities have been widely documented (e.g. Banas et al., 2019), due to barriers including the inability of researchers to directly contact potential participants and difficulties obtaining consent (Cleaver et al., 2010). The feedback generally from participants was that they enjoyed completing the questionnaires, administered via a one to one interview. I was also asked on a number of occasions whether I knew of other research projects for individuals with intellectual disabilities which they might also be able to participate in. I got the sense that these individuals were keen to contribute to research and were seeking opportunities for their voices to be heard by others. This to me highlighted the importance of directly involving individuals with intellectual disabilities in research, rather than relying on informant reports. As the majority of participants in the empirical study were recruited through contacting self-advocacy organisations, there remains a need to consider how individuals who may be less socially engaged could be recruited into studies so that their views are also represented in research.

Reflections on Quantitative Measurement

Despite criticisms around the use of outcome measures and questionnaires in services such as IAPT, as a clinician I recognise that there are advantages of quantitative measurement and that it may be clinically useful in many ways, for example, enabling the measurement of change in specific symptoms following an intervention in order to evaluate its effectiveness. In research, measures or scales may also be useful in capturing data from a large number of individuals fairly quickly, compared to, for example, in-depth interviews about individuals' wellbeing.

I believe that it is important to note that although to a large extent the concept of mental wellbeing may be universal, it is likely that there are differences in the way in which individuals nationally but also cross-culturally define or experience wellbeing. Take item 5 on the WEMWBS-ID, "I had lots of energy", as an example. Personally, having "lots of energy" does not feel central to my own sense of wellbeing, compared to other items, such as "I felt good about myself". Furthermore, there may be differences in how individuals with intellectual disabilities interpret the items. For example, item 11, "I felt able to make my own decisions". It has been documented that individuals with intellectual disabilities often experience restrictions to their decision making (Werner, 2012, 2015). At times, I considered the insight that these individuals may have into this. For example, I recall a few participants reporting that they were able to "choose what to eat", or "choose which movie to watch" and therefore chose the option 'Often' or 'Always' in response to this item. However, I wondered whether they felt able to make their own decisions about other matters, such as where they lived, whether they had a job or children. This suggests that responses on measures such as the WEMWBS-ID should be interpreted cautiously and be used to guide further discussions to better understand an individual’s wellbeing.

I believe that quantitative measurement has an important place in the field. As a clinician, I can see how the information captured by the WEMWBS-ID could be very useful in a clinical setting, allowing me to facilitate further discussions about an individual’s strengths and resources or meaningful goals for therapy. For example, I may wish to further explore why an individual does or does not feel useful, hopeful about the future or connected to others, as this may guide therapeutic interventions. Further investigations into the measure’s sensitivity to change may also allow clinicians to use the SWEMWBS-ID to assess changes in an individual’s wellbeing following an intervention.

Conclusions

To conclude, the systematic review highlighted a paucity of self-report measures of mental wellbeing for individuals with intellectual disabilities and the empirical paper attempted to offer a solution by psychometrically evaluating the WEMWBS-ID. Although a range of challenges were faced throughout the process, there were many strengths, such as the active engagement of individuals with intellectual disabilities in the research. It is hoped that this initial report on the WEMWBS-ID will encourage utilisation of the measure clinically and further research into its validity in a representative sample.

References

Banas, J. R., Magasi, S., The, K., & Victorson, D. E. (2019). Recruiting and retaining people with disabilities for qualitative health research: Challenges and solutions. *Qualitative Health Research*, *29*(7), 1056-1064. https://doi.org/10.1177/1049732319833361

Cleaver, S., Ouellette-Kuntz, H., & Sakar, A. (2010). Participation in intellectual disability research: A review of 20 years of studies. *Journal of Intellectual Disability Research, 54*(3), 187-193. https://doi.org/10.1111/j.1365-2788.2010.01256.x

Kroenke, K., Spitzer, R. L., & Williams, J, B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Hayes S. C., & Strosahl K. D. (2004). *A practical guide to acceptance and commitment therapy* (eds.). Springer.

Lake, J. K., Jachyra, P., Volpe, T., Lunsky, Y., Magnacca, C., Marcinkiewicz, A., & Hamdani, Y. (2021). The wellbeing and mental health care experiences of adults with intellectual and developmental disabilities during COVID-19. *Journal of Mental Health Research in Intellectual Disabilities, 14*(3), 285-300. https://doi.org/10.1080/19315864.2021.1892890

Scior, K., Hamid, A., Hastings, R., Werner, S., Belton, C., Laniyan, A., Patel, M., & Kett, M. (2015). *Intellectual Disabilities: Raising Awareness and Combating Stigma- a Global Review*. London: University College London.

Spitzer, R, L., Kroenke, K., Williams, J. B., and Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*, 1092-7. https://doi.org/10.1001/archinte.166.10.1092

Werner, S. (2012). Individuals with intellectual disabilities: A review of the literature on decision-making since the Convention on the Rights of People with Disabilities (CRPD). *Public Health Reviews, 34*, 1-27. https://doi.org/10.1007/BF03391682

Werner, S. (2015). Public stigma and the perception of rights: Differences between intellectual and physical disabilities. *Research in Developmental Disabilities, 38,* 262-271. https://doi.org/10.1016/j.ridd.2014.12.030

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*(8), 737-746. https://doi.org/10.1111/jir.12003

Appendix A: Characteristics of Assessment Instructions for Psychiatric Disorders in Persons with Intellectual Developmental Disorders (CAPs-IDD) based summaries of the measures reviewed in the present paper

Table A1

CAPs-IDD based summary of the ADAMS.

Measure: The Anxiety, Depression and Mood Scale (ADAMS; Esbensen et al., 2003)		
Part 1: Conceptual and Measurement Model		
B: Basic Information		**Comments**
B1: Concept to be measured:	Broader spectrum of disorders	The ADAMS is an informant-based rating scale that assesses behaviour-based affective symptoms of individuals with intellectual disabilities (ID). The ADAMS was developed a screening measure for anxiety and depression. In the original study by Esbensen et al. (2003), the ADAMS was validated with people aged 10 years and older with participants with borderline, mild, moderate, severe and profound ID (<50% of the sample had mild-moderate ID). In the study by Rojahn et al. (2011), 263 (174 male, 89 female) adults with ID were recruited. Of these, 24% had mild, 34.6% moderate, 23.3% severe, and 17.5% had profound ID (two participant’s level of ID was unavailable). The measure was completed by people most familiar with the participants (parents and guardians, group home staff, teachers or other day placement staff). In the study by Rojahn et al. (2011), 263 (174 male, 89 female) adults with ID were recruited. Of these, 24% had mild, 34.6% moderate, 23.3% severe, and 17.5% had profound ID (two participant’s level of ID was unavailable). In terms of ethnicity, 71.1% of the sample were Caucasian, 12.5% were African American, 7.2% Asian/Pacific islander, 4.2% Hispanic American, 1.1% American Indian, 3.4% Other.
B2: Level of IDD:	Mild, Moderate, Severe, Profound	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Childhood, Adolescence, Adulthood	
B5: Primary purpose/recommendation for use:	Screening	
B6: Available modes of administration:	Setting: Single setting. Application: Paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): No information	
B7: Respondent requirements:	Person with IDD: Person with IDD is not a respondent. Third person: Caregiver (e.g. direct care staff, family carer, teacher).	

B8: Competence Level needed for administration:	None	

T: Test Development		**Comments**
T1: Main underlying theory for generating items:	Empirical	Esbensen et al. (2003) developed the scale in collaboration with psychologists and psychiatrists. They initially generated a list of 131 items from DSM-IV criteria and various assessment instruments including the Diagnostic Assessment for the Severely Handicapped-II (DASH; Matson et al., 1991) and the Psychopathology Instrument for Mentally Retarded Adults (PIMRA; Matson et al., 1984). Through several rounds of consultations and psychometric analyses, the list was revised, items were reworded, and redundant items were eliminated, resulting in a 28-item scale. These final items tapped aspects of anxiety, fear, compulsiveness, appetite, communication, concentration, depression, energy level, mood, physical state, sleep disturbance, and social interaction.
T2: Experts involved in test development:	Mental health professionals	
T3: Based on classification models:	DSM	

C: Measurement		**Comments**
C1: Item content:	Questions as third party report: Problem behaviour, Emotional, e.g. feelings	The ADAMS consists of 28 items which are scored on a 4-point rating scale that combines frequency and severity ratings; they range from 0 (behaviour has not occurred or is not a problem) to 3 (behaviour occurs a lot, or is a severe problem). Raters should consider behaviours which had occurred during the previous six months. The items are distributed to five subscales are labelled Manic/Hyperactive Behaviour (5 items), Depressed Mood (7 items), Social Avoidance (7 items), General Anxiety (7 items), and Obsessive/Compulsive Behaviour (3 items); one of the items contributes to two subscales.
C2: Item coding:	Response format: Polytomous (includes Likert Scale). Facet/Quality assessed: Presence (e.g. of a problem, symptom), Frequency, Severity. Perspectice/locus of response for specific items (symptom, behaviour):	

	Presence, Problem for the person with IDD	
C3: Timeframe:	Time frame given: 6 months	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	The Root Mean Square Error of Approximation (RMSEA) and the comparative fit index (CFI) were used to evaluate goodness of fit between the measurement model and the data. The RMSEA value for the ADAMS was .10 (90% confidence interval: .095 – .11), suggesting a poor fit between the model and data. A confirmatory factor analysis yielded a CFI of .89, also indicating a poor fit between the measurement model and the data. Evidence for convergent validity was assumed if subscales that were expected to assess the same construct e.g. the ADAMS Depressed Mood subscale and the Assessment of Dual Diagnosis (ADD; Matson & Bamburg, 1998) Depression subscale, or a clinically related construct e.g. the ADAMS Depressed Mood subscale and the ADD Anxiety subscale. Positive correlations were demonstrated between the ADAMS depression subscale and ADD Depression subscale ($r_s = .77$, $p < .000$; excellent) and the ADD Anxiety subscale and ADAMS general anxiety subscale ($r_s = .75$, $p < .000$; excellent). Examples of discriminant validity were the non-existing correlations between the five ADAMS subscales and the ADD's Sexual Disorders subscale, which was a clinically unrelated construct. The ADAMS Manic/Hyperactive subscale showed significant positive correlations ($\rho = .35$ - $.75$; fair to excellent) with the Aberrant behaviour Checklist (ABC; Aman et al., 1985) Irritability, Stereotypy, Hyperactivity, and Inappropriate Speech subscales, which were expected to correlate and hence were indicative of good convergent validity. However, against expectations, the ABC Lethargy subscale
V2: Content validity:	No information	
V3: Construct validity:	Confirmatory factor analysis, Convergent validity	
V4: Face validity:	Face validity rated by authors	

		correlated significantly with the ADAMS Manic/Hyperactive subscale (ρ = .45; fair), which was questionable.

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Rojahn et al. (2011) reported that the Cronbach's alpha was 0.9 (excellent). The mean subscale alpha coefficient for the ADAMS was 0.83 (subscales ranged from α = 0.78 - 0.85; fair to good). Although Esbensen at al. (2003) analysed test–retest and interrater reliability in their original paper, the results have not been reported as their study did not meet the eligibility requirements for the present review (participants were <18 years old and <50% of the sample had mild to moderate ID). Rojahn et al. (2011) did not examine reliability in their study.
R2: Reliability:	No information	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	No information	
O2: Coding:	No information	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	
N2: Cut-offs:	No information	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	In terms of value, Esbensen et al. (2003) explained that the ADAMS expanded on screening instruments which were available at the time, as it was the first instrument
F2: Ease of administration/burden:	No information	

F3: Value:	For health professionals, for person with IDD, for researchers	to successfully obtain empirically derived factors of both anxiety and depression with one sample of subjects.
F4: Acceptability:	No information	
F5: Availability:	No information	

Table A2

CAPs-IDD based summary of the ADD.

Measure: The Assessment of Dual Diagnosis (ADD; Matson & Bamburg, 1998)		
Part 1: Conceptual and Measurement Model		
B: Basic Information		**Comments**
B1: Concept to be measured:	Broader spectrum of disorders	The ADD is a screening instrument that was designed for adults with mild and moderate ID to reflect common mental disorders (although it was also administered to people with severe and profound in the study by Rojahn et al., 2011). It contains 79 items. The subscales are organised along the lines of the DSM-IV disorders, as follows: Mania (6 items), Depression (8 items), Anxiety (11 items), Post Traumatic Stress Disorder (6 items), Substance Abuse (4 items), Somatoform Disorder (6 items), Dementia (5 items), Conduct Disorder (8 items), Pervasive Developmental Disorder (10 items), Schizophrenia (9 items), Personality Disorders (9 items), Eating Disorders (6 items), and Sexual Disorders (4 items). Ten of the items contribute to two different subscales, and one item contributes to three. In the study by Rojahn et al. (2011), 263 (174 male, 89 female) adults with ID were recruited. Of these, 24% had mild, 34.6% moderate, 23.3% severe, and 17.5% had profound ID (two participant's level of ID was unavailable). The measure was completed by people most familiar with the participants (parents and guardians, group home staff, teachers or other day placement staff).
B2: Level of IDD:	Mild, Moderate, Severe, Profound	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Screening	
B6: Available modes of administration:	Setting: Single setting. Application: Paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): No information	
B7: Respondent requirements:	Person with IDD: Person with IDD is not a respondent. Third person: Caregiver (e.g. direct care staff, family carer, teacher).	In the study by Rojahn et al. (2011), 263 (174 male, 89 female) adults with ID were recruited. Of these, 24% had mild, 34.6% moderate, 23.3% severe, and 17.5% had profound ID (two participant's level of ID was unavailable). In terms of ethnicity, 71.1% of the sample were Caucasian, 12.5% were African American, 7.2% Asian/Pacific islander, 4.2% Hispanic American, 1.1% American Indian, 3.4% Other.

B8: Competence Level needed for administration:	None	

T: Test Development		Comments
T1: Main underlying theory for generating items:	Empirical	The items were compiled from symptoms of particular disorders listed in the DSM-IV.
T2: Experts involved in test development:	Mental health professionals	
T3: Based on classification models:	DSM	

C: Measurement		Comments
C1: Item content:	Questions as third party report: Problem behaviour, emotional e.g. feelings	The ADD contains 79 items that are scored on three 3-point rating scales: (1) frequency, ranging from 0 (not at all) to 2 (more than 10 times), (2) duration, ranging from 0 (less than 1 month) to 2 (over 12 months), and (3) severity, ranging from 0 (no disruptions or damage) to 2 (caused property damage or injury).
C2: Item coding:	Response format: Polytomous (includes Likert scale). Facet/quality assessed: Presence (of a problem, symptom), Frequency, Severity, Chronicity. Perspective/locus of response for specific items (symptom, behaviour): Presence (does the person show the symptom/behaviour at all), Problem for person with IDD (is the	

	symptom/behaviour a problem for person with IDD), Problem for third persons/surroundings (is the symptom/behaviour a problem for third persons/surroundings)	
C3: Timeframe:	Time frame given: 1 month	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	In the study by Rojahn et al. (2011), evidence for convergent validity was assumed if subscales that were expected to assess the same construct e.g. the ADAMS Depressed Mood subscale and the ADD Depression subscale, or a clinically related construct e.g. the ADAMS Depressed Mood subscale and the ADD Anxiety subscale. Positive correlations were demonstrated between the ADAMS depression subscale and ADD Depression subscale ($r_s = .77$, $p < .000$; excellent) and the ADD Anxiety subscale and ADAMS general anxiety subscale ($r_s = .75$, $p < .000$; excellent). The correlation between the Lethargy subscale of the ABC (Aberrant Behaviour Checklist) and the ADD Depression subscale was good ($r_s = .63$, $p < .000$). Examples of discriminant validity were the non-existing correlation between the five ADAMS subscales and the ADD's Sexual Disorders subscale. Confirmatory Factor Analysis was attempted with the ADD, however after over 60 iterations, the model would not converge. Rojahn et al. (2011) noted that given the number of items and scales on the ADD, the model was estimating more parameters than the data set had participants and therefore future analyses of the measurement model with the ADD will likely require a larger dataset.
V2: Content validity:	No information	
V3: Construct validity:	Confirmatory factor analysis, Convergent validity	
V4: Face validity:	Face validity rated by authors	

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Rojahn et al. (2011) reported that the Cronbach's alpha was for the total ADD score was .91 (excellent). However, the mean subscale alpha coefficient was 0.59 (ranging from 0.18 - 0.84; poor to good). The authors posited that this may because the ADD was originally developed for individuals with mild and moderate ID and in the study by Rojahn et al. (2011), participants ranged from having a mild to profound ID. However, when they recalculated the Cronbach's alpha with the 154 participants with mild to moderate ID, no obvious improvements were observed (the mean alpha coefficient changed from 0.59 to 0.61). With regards to reliability, although Matson & Bamburg (1998) analysed test–retest reliability in their original paper, the results have not been reported as their study did not meet the eligibility requirements for the present review (sample included participants who were <18 years old). Rojahn et al. (2011) did not examine test-retest reliability in their study.
R2: Reliability:	Test-retest	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	No information	
O2: Coding:	No information	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	
N2: Cut-offs:	No information	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	

F2: Ease of administration/burden:	No information	In terms of value, Matson & Bamburg (1998) reported that as the ADD screens a wide range of disorders, it may aid in the assessment of psychopathology in individuals with mild to moderate ID, allowing informed treatment choice. They also reported that the ADD may aid in advancing research in this area and in the delivery of optimal services for individuals with mild to moderate ID.
F3: Value:	For health professionals, for person with IDD, for researchers	
F4: Acceptability:	No information	
F5: Availability:	No information	

Table A3

CAPs-IDD based summary of the CORE-LD.

Measure: The Clinical Outcomes in Routine Evaluation – Learning Disabilities (CORE-LD) (Brooks et al., 2013)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
B1: Concept to be measured:	Broader spectrum of disorders	The CORE-LD is an adapted version of the Clinical Outcomes in Routine Evaluation - Outcome Measure (CORE-OM; Evans et al., 2002), which measures various factors such as wellbeing, mental health and interpersonal relationships, for people with ID. The CORE-LD enables the monitoring of mental health and evaluation of the results of psychological therapies. Brook et al. (2013) developed the CORE-LD as they identified an absence of items within CORE-OM which related specifically to the lived experience of people with ID and their wellbeing. The measure was validated with 272 participants in the study by Brooks et al. (2013) and the level of ID was reported for 96% of the sample (81% had a mild LD and 19% had a moderate LD). The participants ranged between 18 - 80 years old. In a further study by Briscoe et al. (2019), the measure was validated with 41 participants with a mild to moderate ID, with an age range of 20-64. The CORE-LD is designed to be completed collaboratively with a therapist, though people with ID are given their own copy of the questionnaire and are encouraged to mark their own response. In the study by Briscoe et al. (2019), if participants could not read or write, items were read aloud verbatim and their responses recorded. Participants were encouraged to respond to each item according to the wording of available responses (not at all–all the time) but in some cases when completing the CORE-LD participants would point to their pictorial response (empty beaker–full beaker). The sample in the study by Brooks et al. (2013) was described as follows: 272 clinical clients [162 (60%) female; 110 male] took part in the first phase of CORE-LD data collection, along with a further 52 nonclinical clients [30 (58%) female; 22 male] who contributed test–retest data (2 of whom did not contribute time 2 data).
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood, Elderly	
B5: Primary purpose/recommendation for use:	Screening, Research, Evaluation of interventions	
B6: Available modes of administration:	Setting: Single Setting. Application: Paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): Available.	
B7: Respondent requirements:	Person with IDD: Some verbal skills required. Third person: Third person is not a respondent	
B8: Competence Level needed for administration:	Health professional: Other	

		The majority of clients [37 (71%) nonclinical; 234 (86%) clinical] were White. In the nonclinical population, 8 (15%) of those remaining were Asian/Asian British; 5 (10%) Black/Black British and 2 (4%) Chinese/Other. In the clinical population, 15 (6%) were Asian/Asian British; 18 (7%) Black/Black British; 3 (1%) Mixed and 2 (1%) Chinese/Other. Ages in the sample ranged between 18 and 80 years. Clients were drawn from tertiary sector community services (*n* = 17); a high-security NHS setting (*n* = 14) and other NHS settings (*n* = 241). Level of learning disability was recorded for 261 of the 272 clinical participants, with 212 (81%) 'mild' and 49 (19%) 'moderate'. Briscoe et al. (2019) recruited 41 participants with an age range of 20-64 from two UK inpatient hospitals (a medium secure unit with female participants [*n* = 21] and male participants [*n* = 9], and a low secure unit with male participants](*n* =11]). They noted that that preliminary analyses demonstrated that the female sample had statistically significantly higher CORE-LD scores than the male sample, which warranted further exploration in future research.

T: Test Development		**Comments**
T1: Main underlying theory for generating items:	Empirical: factor analysis of pool of items	Brooks et al. (2013), who are Practitioner Researchers (a Dramatherapist and Specialist Counsellor) facilitated a Collaborative Research Group (CoRG) which included five people with ID, where they discussed the impact of having an LD on feelings. These sessions were recorded and the data was analysed using interpretative phenomenological analysis (Smith, 2003) and thematic analysis (Braun & Clarke, 2006). They noted that many statements described emotional difficulties and feelings which were not in CORE-OM. These were grouped into five themes (Feeling put down; Feeling left out; Feeling confused; Feeling angry inside; Feeling vulnerable).
T2: Experts involved in test development:	Mental health professionals, Persons with IDD (e.g. focus group, pre-testing)	
T3: Based on classification models:	None	

C: Measurement		**Comments**
C1: Item content:	Questions as self-report - Emotional, e.g. feelings	The CORE-LD has 14 items, which includes questions about feeling lonely, confused and unhappy. The CoRG felt that the original 5-point rating scale of

C2: Item coding:	Response format: Polytomous. Facet/quality assessed: Presence (e.g. problem, symptom), Frequency. Perspective/locus of response for specific items (symptom, behaviour): Presence	CORE-OM was confusing and suggested a simplified 3-point scale with visual iconic representations. Beakers were chosen to depict frequency, an empty beaker representing ‘not at all’, half-full beaker for ‘sometimes’ and a full beaker for ‘a lot’. Individuals are asked to answer the questions based on how they had been feeling over the last week.
C3: Timeframe:	1 week	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes (e.g. as a ROM - routine outcome measurement) - Yes	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	Brooks et al. (2013) completed a preliminary psychometric analysis on a 17-item version of the CORE-LD data looked at: item omission rates; test–retest stability of items; discrimination between clinical and non-clinical clients; the contribution individual items made to overall reliability and low fall in item score over treatment. These analyses revealed four problematic items though it was not apparent that these should be dropped. The CoRG researchers developed a qualitative questionnaire on accessibility and usability to gain feedback from therapists who had used the measure. They were also asked to comment on items which they deemed especially problematic or helpful. This feedback, in conjunction with results of psychometric analysis, led to a decision to retain the item, "Are you pleased with the things you have done?" (as it was the only positive item in the measure) and omit the remaining three problem items. This left the final, 14-item CORE-LD. Psychometric analyses were repeated on the 14-item CORE-LD. The
V2: Content validity:	Concerning relevance of items, Concerning comprehensiveness of items	
V3: Construct validity:	Convergent validity	
V4: Face validity:	Face validity rated by author(s), Face validity rated by person with IDD	

		results of analyses on validity were not reported in the paper. Briscoe et al. (2019) established the concurrent validity of the CORE-LD by correlating it with the CORE-OM. The data indicated the CORE-LD was significantly positively correlated to the CORE-OM ($r_s = .68$, $p < .001$).

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Brooks et al. (2013) reported the results of analyses of reliability. The CORE-LD was administered to 50 people living in the community who were not receiving therapy, in addition to the clinical sample used in the main study. In order to measure test-retest reliability, the nonclinical sample were asked to complete the CORE-LD twice with a week in between. Test–retest stability was good, with $\rho = .64$. There was no significant test–retest mean shift and no total score mean shift ($P = .85$). Furthermore, Cronbach's alpha was good ($\alpha = 0.80$) for both the clinical and nonclinical samples.
R2: Reliability:	Test-retest	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	Some instructions available (e.g. published paper)	The published paper gives brief instructions on the structure and administration. High scores indicate more distress, suggesting poor outcome of therapy.
O2: Coding:	No information	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	Evans et al. (2002) administered the CORE-OM to a non-clinical sample and a clinical sample from the general population during the initial validation of the measure. Further information may be found in the published paper.
N2: Cut-offs:	None	
N3: Fairness:	No information	

F: Feasibility		**Comments**

F1: Missing values:	No information	The authors concluded that the CORE-LD is a robust and valid measure, which appears to be highly acceptable to clients and practitioners. Briscoe et al. (2019) noted that participant anecdotal feedback indicated they found completing the CORE-LD easier than doing the CORE-OM.
F2: Ease of administration/burden:	For respondent: Evidence that instrument places no undue physical or emotional strain on respondent. For administration: Easy to understand instructions, user-friendly design/layout	
F3: Value:	For health professionals (e.g. psychologist, psychiatrist), For person with IDD, For researchers	
F4: Acceptability:	For health professionals (e.g. psychologist, psychiatrist), For person with IDD, For researchers	
F5: Availability:	Available for free, Published as test(manual)	

Table A4

CAPs-IDD based summary of the IES-IDs.

Measure: The Impact of Events Scale-Intellectual Disabilities (IES-IDs; Hall et al., 2014)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
Specific Disorder: Anxiety disorder	Specific Disorder: Anxiety disorder	The Impact of Events Scale-Revised (IES-R; Weiss & Marmar, 1997) measures PTSD symptomatology and is one of the most widely used measures of traumatic stress as it has well established psychometric properties. The Impact of Events Scale-Intellectual Disabilities (IES-IDs; Hall et al., 2014) was developed by modifying the IES-R. The IES-R was chosen as was designed as a screening self-report questionnaire which corresponds directly to the DSM-IV-TR (APA, 2000) symptoms of PTSD (avoidance, intrusion, and hyperarousal symptoms). Secondly, it is relatively short and easy to complete. A score of 33 on the IES-R has been suggested to be an appropriate clinical cut-off. In the study by Hall et al. (2014), 40 adults with mild ID (aged 18+, 57.5% women) were recruited. 36 carers were also recruited and asked to complete the informant versions of several measures. An interviewer script was developed so that the IES-IDs could be administered as a semi-structured interview.
Mild	Mild	
None	None	
Adulthood	Adulthood	
Screening	Screening	
Other: Semi structured interview. Augmentative and alternative communication (e.g. visual aids, symbols): Available	Other: Semi structured interview. Augmentative and alternative communication (e.g. visual aids, symbols): Available	
Person with IDD: Some verbal skills required.	Person with IDD: Some verbal skills required.	Forty adults with IDs (Mean age = 36.95, SD = 14.84, 57.5% women, Mean IQ = 60.68, SD = 6.13) were recruited from National Health Service (NHS) teams, residential services and day centres across the counties of Norfolk, Cambridgeshire, and Suffolk in England. They were aged 18 or above and spoke English and had mild ID as evidenced by a Full Scale IQ that ranged from 50 to 70. Participants were excluded if they did not have a history of at least one traumatic event.
Third person: Third person is not a respondent.	Third person: Third person is not a respondent.	
Health professional: Psychologist, Other	Health professional: Psychologist, Other	

T: Test Development		**Comments**

T1: Main underlying theory for generating items:	Empirical	To modify the IES-R, and create the IES-IDs, professionals in an Adult Community Learning Disability Team were consulted (two clinical psychologists, a speech and language therapist, and an assistant psychologist). The language of the IES-R, format and item organisation were changed to ensure that items were comprehensible and appropriate for individuals with IDs. Guidance was followed to ensure that the question structure was simple and avoided the use of technical vocabulary, and the text was made larger to increase the accessibility of the measure.
T2: Experts involved in test development:	Mental health professionals, Direct-care staff	
T3: Based on classification models:	DSM	

C: Measurement		**Comments**
C1: Item content:	Questions as self-report: Emotional, e.g. feelings, Problem behaviour.	The measure has 22 items. The rating scale was simplified from 5 points to 3. Other changes included that respondents were asked to respond 'yes/no' based on whether they have experienced each symptom, initially, before rating the distress experienced. If the symptom had not been experienced, the item was coded a zero. This approach was adopted to simplify the assessment and reduce confusion. If the respondent had experienced the symptom, they were then asked ‘how much has that upset or scared you?’ Possible responses and associated scores were then: ‘a little bit’ (score = 1), ‘in the middle’ (score = 2) or ‘a lot’ (score = 3). This was augmented with a visual scale to improve the reliability and validity of the Likert scales. If the respondent had selected more than one traumatic event, they were then asked to pick the one event that had upset them the most and then questions were asked in reference to this selected event. Following this, further questions were asked to assess whether the event happened recently, less than a year ago, more than a year ago, or when they were a child.
C2: Item coding:	Response format: Polytomous (includes Likert scale), Facet/quality assessed: Presence (e.g. of a skill, problem, symptom), Severity. Perspective/locus of response for specific items (symptom, behaviour): Presence, Problem for person with IDD.	
C3: Timeframe:	No time frame given (e.g. how the person is feeling/behaving now)	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	There were good to excellent positive correlations between the Intrusion, Hyperarousal, and Avoidance subscales of the IES-IDs (*r* =.67 - .74). All three subscales also significantly correlated with the IES-IDs Total Score (*r* =.87 - .9; excellent). With regards to convergent validity, the IES-IDs Total Score significantly correlated with the self-report Glasgow Depression Scale (GDS; Cuthill et al., 2003) (*r* =.60; good) but not the informant version of the GDS. The authors suggested that this may be the because covert symptoms of depression are more difficult for carers to rate. Furthermore, the IES-IDs Total Score and subscales all significantly correlated with the Glasgow Anxiety Scale (GAS; Mindham & Espie, 2003) (*r* =.40 = fair). The IES-IDs Intrusion, Hyperarousal and Avoidance subscales (*r* =.63 - .71; good to excellent) and Total Score (*r* =.76; excellent) significantly correlated with the self-report version of the LANTs. The IES-IDs Total Score also correlated significantly with the Behaviour Changes, Frequency, and Severity Scales (*r* =.28 - .37 = poor to fair) from the informant version of the LANTS. Hall et al. (2014) hypothesised that people who have experienced a higher frequency of trauma would score higher on the IES-IDs and their results indicated that there was a positive association between trauma frequency and symptomatology as measured by the IES-IDs (*r* =.35 = fair).
V2: Content validity:	Concerning relevance of items, concerning comprehensiveness of items	
V3: Construct validity:	Convergent validity	
V4: Face validity:	Face validity rated by author(s) and health professionals.	

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	The Cronbach's alpha for the IES-IDs Total ranged between α = 0.90 - 0.91 from T1 to T2, demonstrating excellent internal consistency. Subscales ranged from α = 0.61 – 0.88 (fair to good). Test–retest reliability was determined by calculating the intraclass correlation coefficient. Participants completed measures on two occasions, separated by 2 weeks. The authors reported that the test–retest reliability of the IES-IDs Total Score (ICC = .86) was excellent. Overall reliability for the subscales were as follows: Intrusion (ICC = .85; excellent), Avoidance (ICC = .65; good) and Hyperarousal (ICC = .82; excellent).
R2: Reliability:	Test-retest	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	Some guidelines available (e.g. published paper)	Procedure detailed in published paper.
O2: Coding:	No information	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	The IES-IDs was adapted from the IES-R. A score of 33 on the IES-R has been suggested to be an appropriate clinical cut-off (Creamer et al., 2003). A cut-off score for the IES-IDs was not suggested. Creamer et al. (2003) assessed the psychometric properties of the IES-R using a community-based sample and a clinical sample. Further information may be found in the published paper.
N2: Cut-offs:	No information	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	In terms of the ease of administration/burden, an interviewer script was developed so that the IES-IDs could be administered as a semi-structured interview and the language of the IES-R was changed to ensure that items were comprehensible and appropriate for individuals with IDs. Guidance was followed to ensure that the question structure was simple and avoided the use of technical vocabulary and the text was made larger to increase the accessibility of the measure. Professionals in an Adult Community Learning Disability Team were consulted to modify the IES-R in order to ensure acceptability. With regards to the value of the measure, the authors reported that the IES-IDs was developed response to a lack of well-developed instruments to assess PTSD symptomatology with people who have IDs. The items and Likert Scale for the IES-IDs can be found in Table 1 in the published paper. The IES-IDs is also available from the authors.
F2: Ease of administration/burden:	For respondent: Reading and/or comprehensive level. For administration: Easy to understand instructions, user-friendly design/layout	
F3: Value:	For health professionals, for person with IDD	
F4: Acceptability:	For health professionals	
F5: Availability:	Cost: No information. Finding instrument: Whole instrument is published in scientific journal or book (all items are printed)	

Table A5

CAPs-IDD based summary of the LANTS.

Measure: The Lancaster and Northgate Trauma Scales (LANTS) (Wigham et al., 2011)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
B1: Concept to be measured:	Specific Disorder: Anxiety disorder	The LANTS were developed to measure symptoms of trauma in adults with ID. The LANTS comprises two scales: a self-report version, for people with mild/moderate ID and an informant LANTS for carers or clinicians. The informant version measures the outward presentation of a trauma effect, i.e., observable behaviours, and the self-report version measures subjective affective, biological and emotional state. Respondents are not asked to identify a specific traumatic incident. The self-report measure includes a visual scale, short sentences, and large text were used to increase the accessibility. In the original study by Wigham et al. (2011), data were collected from service users over a maximum of 3 x 30 min interview sessions, with a carer present if requested. The informant data was primarily collected via interviews or surveys. Three pairs of screening questions were included at the start of the measure to check response validity and comprehension of the response scale, for example respondents were asked to name their favourite food, and then to rate how much they liked their favourite food on the response scale. To reduce the chances of acquiescence, half the questions in the service user measure were reverse worded. The psychometric evaluation conducted by Wigham et al. (2011) involved 99 adults (85 men and 14 women) with mild to moderate ID receiving services from NHS, day centre, social services and independent service providers in northeast England. Forty-six lived in community settings and 53 in inpatient settings. Of the inpatients 7 were from mental health wards and the rest from forensic wards, with 19 from
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Screening, Research	
B6: Available modes of administration:	Setting: Single setting. Application: Paper-pencil, Other. Augmentative and alternative communication (e.g. visual aids, symbols): Available	
B7: Respondent requirements:	Person with IDD: Some verbal skills required. Third person: Health professional, Caregiver.	
B8: Competence Level needed for administration:	None	

		low or enhanced low secure, 10 from medium secure and 17 from open or rehabilitation wards. Of the community participants 29 were recruited from day centres, 8 from an independent residential service provider, 2 from an NHS supported residential service, 4 from a community residential forensic service, and 3 lived independently. The mean age of female service users was 45.6 years (SD = 12.2; range = 21–60). The mean age of male service users was 40.2 years (SD = 13.9; range = 19–75). The ages of 19 participants were not collected. All participants barring one were white. The informant participant group comprised 88 paid carers (44 men and 44 women), who had known the service user for a minimum of a year.

T: Test Development		**Comments**
T1: Main underlying theory for generating items:	Empirical	Wigham et al. (2011) created a pool of 48-items pertaining to the possible effects of a traumatic life event on a person with ID, by consulting a number of sources (a systematic literature review of the empirical evidence, the general population trauma literature and the views of service users, carers, advocates and staff). Diagnostic criteria were also included in the item pool e.g. PTSD triad of symptoms (APA, 2000) including re-experiencing, avoidance, and arousal. The item pool was presented as a survey with 48 questions asking if the effects listed could be seen in a person with ID after an adverse life event and how different areas of their daily living would be affected. Data were generated using the 48-item pool during the surveys, interviews and focus groups. Service user participants in were 7 adults with mild to moderate ID (1 woman and 6 men; age range of 22–56 years). Sixteen staff members and clinicians, two family carers and two advocates were also recruited (15 women and 6 men). All of the study participants were white British. The data were subsequently analysed via content analysis. The themes and categories were used to construct questions for a self-report and an informant LANTS measure. During a pilot study, five informant measures were completed by staff and five service users completed self-report measures during interviews, which took between 10 and 20 min to complete. Respondents were asked for feedback on the measures
T2: Experts involved in test development:	Mental health professionals, Persons with IDD (e.g. focus group, pre-testing), Direct-care staff, Family carers	
T3: Based on classification models:	DSM, Other	

		including their comprehensibility and inclusiveness. Based on this feedback, the measures were revised and after the pilot the final version of the self-report LANTS included 34 questions, and the informant LANTS included 47 questions.

C: Measurement		**Comments**
C1: Item content:	Questions as self-report: Emotional, e.g. feelings. Questions as third party report: Problem behaviour, Emotional e.g. feelings	The LANTS comprises two scales: a self-report version, for people with mild/moderate ID, with 29-items rated on a 4-point visual Likert scale. There is also an informant LANTS for carers or clinicians comprising 43-items rated for frequency, severity and changes from usual. The self-report LANTS included questions about the frequency of subjective states experienced over the last week, for example ‘Worries have been going round and round in my head’, to be rated on a visual 4-point adjectival response scale (‘no’; ‘a little’; ‘sometimes’; ‘a lot’). The informant questions pertained to the observable effects of trauma during the last month, for example ‘Fearful – expressing an expectation that something bad will happen’. The informant measure comprised ‘behavioural changes’, ‘frequency’ and ‘severity’ subscales. Each question was rated for frequency on a 6-point scale (‘none’, ‘monthly’, ‘weekly’, ‘several times a week’, ‘daily’, and ‘several times a day’) and severity on a 3-point scale (‘mild’, ‘moderate’ and ‘severe’). The response option for the ‘behavioural changes’ subscale prompted respondents to indicate whether the behaviour was the ‘same as usual’ for the person.
C2: Item coding:	Response format: Polytomous (includes Likert scale). Facet/quality assessed: Presence (e.g. of a skill, problem, symptom), Frequency, Severity. Perspective/locus of response for specific items (symptom, behaviour): Presence	
C3: Timeframe:	Time frame given: 1 week, 1 month	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	

V2: Content validity:	Concerning relevance of items, Concerning comprehensiveness of items	For the psychometric evaluation, Wigham et al. (2011) recruited a new sample of 99 adults (85 men and 14 women) with mild to moderate ID. Spearman correlations were calculated for the LANTS self-report and informant scales and established measures of trauma. The correlations between the self-report LANTS and the Impact of Events Scale (IES; Horowitz et al., 1979) self-report trauma subscales were of a good magnitude ($r_s = .58 - .62$, $p < .01$). Similarly, the correlations between the LANTS informant subscales and the Paediatric Emotional Distress Scale (PEDS; Saylor et al., 1999) informant trauma scales were all significant and of an acceptable magnitude ($r_s = .35 - .50$, $p < .01$). The LANTS self-report scale correlation with the self-report Brief Symptom Inventory (BSI; Derogatis, 1993) measure of psychopathology symptoms was good ($r_s = .62$, $p < .01$). The informant LANTS subscales showed significant correlations ($r_s = .26 - .62$, $p < .01$) with conceptually linked informant-rated measures of psychopathology, problem behaviour and anger disposition (PAS-ADD Checklist, Behaviour Problems Inventory [Rojahn et al., 2001] and Ward Anger Rating Scale Part B [Novaco, 1994]), but low and insignificant correlations with the self-rated BSI measure. Furthermore, relationships between the LANTS and number of adverse life events a measured by the Bangor Life Events Schedule for Intellectual Disabilities (BLESID; Hulbert-Williams et al., 2011) were analysed. The correlation between the LANTS and BLESID self-report scales was fair ($r_s = .45$, $p < .01$). Similarly, the correlations between the LANTS informant subscales and the BLESID informant measure were good ($r_s = .54 - .64$, $p < .01$). The LANTS measures were also examined for their convergence with each other. Significant convergence was only found between the self-report LANTS and the behavioural changes informant LANTS subscale (.204), and the magnitude of this correlation was low. The authors suggested that this may have been due to differences in administration of the measures, or the idea that the two versions focus on different aspects of trauma (internal states vs. observable behaviours). Since the study by Wigham et al. (2011), there were changes to the clinical conceptualisation of trauma, reflected in the criteria for complex PTSD in the Eleventh Revision of the International Classification of Diseases (ICD-11; World Health Organisation, 2018). Therefore, Wigham et al. (2021) carried out further psychometric analysis of the factor
V3: Construct validity:	Hypotheses testing: Convergent validity. Other: Exploratory factor analysis	
V4: Face validity:	Face validity rated by author(s), health professionals, person with IDD, caregiver	

		structure of the self-report measure and assessed compatibility with ICD-11 criteria for complex PTSD. They recruited 98 adults with mild to moderate ID. It was anticipated that the self-report LANTS would have a 4-factor structure (PTSD, affect dysregulation, interpersonal problems, and negative self-concept) and this was confirmed. The factor structure suggests four potential LANTS subscales which accounted for 53.46 % of the variance. The items were organised into the following groups: PTSD (eight items), negative self-concept (six items), affect dysregulation (six items) and interpersonal difficulties (nine). Seven items (pertaining to anger and self-harm) were removed due to particularly high skewness and/or kurtosis prior to running the factor analysis. The Kaiser-Meyer-Olkin (KMO) value was acceptable (0.75). The internal consistency (Cronbach's alpha) of the 22 LANTS items was fair (α = 0.79).

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Wigham et al. (2011) reported the internal consistency of the measures as follows: LANTS self-report scale (α = 0.84; good); LANTS informant behavioural changes subscale (α = 0.82; good), frequency subscale (α = 0.80 = good); and severity subscale (α = 0.84; good). In a further study, Hall et al. (2014) also administered the LANTS in their study to 40 adults with ID and reported that the Cronbach's alpha for the self-report measure was α = 0.89 (excellent) and ranged between α = 0.86 – 0.92 for the informant reports (*n* = 36), across T1 and T2. With regards to test-retest reliability, Wigham et al. (2011) administered the LANTS for retest with 48 self-report and 33 informant participants, on average 5–6 weeks later for service users and 6 weeks later for informants. Significant correlations were demonstrated between initial and repeat administrations of the self-report LANTS (r_s = .72, $p < .01$; excellent) and for the informant subscales (r_s = .57 - .59; good). Hall et al. (2014) also assessed test–retest reliability for the LANTS self-report (ICC = .92) and informant versions (ICC = .80 - .84) and reported that this was also excellent.
R2: Reliability:	Test-retest	
R3: Measurement error:	No information	

O: Objectivity of Application		Comments
O1: Application:	No information	
O2: Coding:	No information	

N: Objectivity of Interpretation, Norming and Fairness		Comments
N1: Norms:	No information	
N2: Cut-offs:	No information	
N3: Fairness:	No information	

F: Feasibility		Comments
F1: Missing values:	Other	Wigham et al. (2011) developed the LANTS via consultation with a clinical sample, carers, advocates and clinicians, supports their content validity to ensure acceptability and inclusiveness. Administration of the self-report measure took between 10 and 20 min to complete. Respondents were asked three screening items at the beginning of the questionnaire to ensure that participants understood the response options, before proceeding with the measure. Missing values were substituted with the mean score for that item across all respondents on the other measures if no more than 5% of responses on the measure were missing. In terms of value, Wigham et al. (2011) reported that the aim of the measure is to facilitate case recognition in clinical settings so that effective treatments may be facilitated and also to further trauma research in ID populations.
F2: Ease of administration/burden:	For respondent (person with ID or third person): Time needed for completion, Reading and/or comprehensive level. For administration: User-friendly design/layout	
F3: Value:	For health professionals, for person with IDD, for researchers	
F4: Acceptability:	For health professionals, for person with IDD, for researchers	
F5: Availability:	No information	

Table A6

CAPs-IDD based summary of the Mini-MANS-LD.

Measure: The Mini Maslow Assessment of Needs Scale - Learning Disabilities (Mini-MANS-LD; Raczka et al., 2020; Skirrow & Perry, 2009)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
B1: Concept to be measured:	Other	The Maslow Assessment of Needs Scale - Learning Disabilities (MANS-LD; Skirrow & Perry, 2009) is a measure quality of life for people with ID. The Mini-MANS-LD (Raczka et al., 2020) was adapted from the MANS-LD following a pilot development study (Raczka et al., 2014), with one item derived from the Adapted World Health Organisation Quality of Life Measure (WHOQOL-8; Schmidt et al., 2006). The MANS-LD may be used to measure outcomes of psychological interventions. In the study by Raczka et al. (2020), a convenience sample of 33 adults with ID were recruited. Participants were accessing support from services for people with ID. Fourteen (42.44%) were male and participants the age range was 22-69 years. A clinical opinion of level of intellectual disability was elicited from the clinical team and 87.89% were considered to have mild ID whilst the remainder were reported to have a moderate ID. The measures were administrated by either qualified clinical psychologists (*n* = 3), a trainee clinical psychologist (*n* = 1) or assistant psychologists (*n* = 3) who worked within the services, all of whom were trained in administration of the measures. A convenience sample of 33 adults with ID who were accessing support from services for people with ID were recruited between June and November 2013. Fourteen (42.44%) were male. The mean age of participants was 39.11 years (*SD* = 14.63) and the range was from 22 years to 69 years. In terms of ethnicity: 77.78% were White, 12.12% Black, 6.06% Asian and 3.03% Other.
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Screening, Evaluation of interventions	
B6: Available modes of administration:	Setting: Single setting. Application: Paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): Available	
B7: Respondent requirements:	Person with IDD: Some	

	verbal skills required. Third person: Third person is not a respondent	
B8: Competence Level needed for administration:	Health professional: Psychologist, Other.	

T: Test Development		**Comments**
T1: Main underlying theory for generating items:	Other	The main underlying theory for generating items was Maslow’s (1943) seminal hierarchy of needs. Items reflect the five motivational needs: basic physiological, safety, social (love and belonging), esteem (feeling of accomplishment) and self-actualisation (achieving one’s full potential). Raczka et al. (2014) conducted a pilot study in collaboration with speech and language therapist and a small group of 18 experts by experience (people with mild to moderate ID) to review a range of routine outcome measures. The piloting, revealed that both the MANS-LD and WHOQOL-8 were rated as too lengthy and repetitive to be used as regularly administered outcome measures. The EQ-5D-Y (Willie et al., 2010) which measures health-related QoL was rated as more acceptable. This led to the development of the Mini-MANS-LD, which encompassed the most important items of the MANS-LD together with one introductory item adapted from the WHOQOL-8. It was designed to be used in conjunction with the EQ-5D-Y, to capture all the important aspects of self-reported quality of life of people with ID.
T2: Experts involved in test development:	Mental health professionals, Persons with IDD (e.g. focus group, pre-testing), Other	
T3: Based on classification models:	None	

C: Measurement		**Comments**
C1: Item content:	Questions as self-report: Emotional, e.g. feelings	The Mini-MANS-LD is a nine-item assessment instrument to be completed with support as required. It uses a five-point Likert-type scale, using a pictorial as well as verbal scale. All except one of the participants were able to understand the five-point scale. The nine items included in the Mini-MANS-LD cover factors including self-reported satisfaction with environment, safety, social

C2: Item coding:	Response format: Polytomous (includes Likert scale). Facet/quality assessed: Frequency, Other. Perspective/locus of response for specific items (symptom, behaviour): Other	relationships, esteem and self-actualisation. There is also an item pertaining to overall life satisfaction. A script for administrators was also developed, to enhance the overall standardisation of the measure. Individual item scores were summed to give an overall score. With regards to responsiveness, Raczka et al. (2020) planned to repeat administration approximately 3 months after baseline, to explore the sensitivity of the measure to detect changes in outcomes over time. This was achieved for 25 participants. Preliminary exploratory analysis suggested that the measure did not detect a significant change in scores over a 3-month period, however, the Mini-MANS-LD showed a difference in the median scores between the two-time points approaching significance. There was also a significant correlation between the number of days between first and second administration and the difference in Mini-MANS-LD scores which may suggest that a significant change in Mini-MANS-LD scores may have been seen with a greater length of time between T1 and T2.
C3: Timeframe:	No time frame given: None specified	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: Yes	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	With regards to the relevance and comprehensiveness of items, the eight items adapted from the original MANS-LD were included in the Mini-MANS-LD as they broadly represent the theoretical constructs underpinning Maslow's (1943) hierarchy of needs. The item derived from the WHOQOL-8 ('overall, [how] do you feel your life is...?) also taps into Maslow's concept of self-actualization. The domain not felt to be covered by the Mini-MANS-LD, physical well-being, was felt to be
V2: Content validity:	Concerning relevance of items,	

	concerning comprehensiveness of items	adequately addressed by EQ-5D-Y which focuses on health-related quality of life. Congruent validity was analysed by correlating the Mini-MANS-LD with the PWI-ID. The Mini-MANS-LD was significantly correlated with the PWI-ID ($r(33) = .67$, $p < 0.001$), suggesting it has good congruent validity with an already validated measure. In terms of face validity, the pilot study conducted by Raczka et al. (2014) allowed speech and language therapists and people with ID to review and adapt the measures.
V3: Construct validity:	Convergent validity	
V4: Face validity:	Face validity rated by author(s), health professionals, person with IDD	

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Cronbach’s alpha for the Mini-MANS-LD was $\alpha = 0.74$, indicating fair internal consistency. None of the individual items increased the overall alpha if deleted and the item-total correlations ranged from $\alpha = 0.29$ to 0.58 (poor to good).
R2: Reliability:	No information	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	Instructions: Short manual available	User guide available: https://www.bps.org.uk/sites/www.bps.org.uk/files/Member%20Networks/Faculties/Intellectual%20Disabilities/Mini%20MANS-LD%20User%20Guide.pdf
O2: Coding:	Guidelines for coding: Short manual available	The authors reported that the Mini MANS-LD should be scored at the individual item level and a total score, summing the individual scores. This enables the Mini-MANS-LD to identify areas for interventions and then to investigate both changes in overall scores over time and also in individual items and relate any shifts back to the interventions.

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	
N2: Cut-offs:	No information	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	With regards to ease of administration/burden, if a participant was assessed as having a tendency to acquiesce, the measures were not administered. The measure was developed following consultation with a small group of experts by experience. Accessible prompts was developed for each measure to use with participants, using Photosymbols and colour-coded ‘smiley faces’ to enhance accessibility. Additionally, scripts were developed for administrators to use to ensure consistent clarification of potentially problematic concepts and the five-point scale. In terms of acceptability, The Mini-MANS-LD was rated by administrators using a 10-point Likert-type scale. It was rated easy to use (rated on average 9/10) and acceptable to people with ID (rated 9.2/10). Designed to be used with the EQ-5D-Y, the two measures together took, on average, less than 12 min to administer. The authors suggested that that they could therefore be appropriate measures to be used fairly routinely within clinical health and social care services for people with ID. With regards to availability, the Mini-MANS-LD items are reported in the published paper however pictorial scale is not.
F2: Ease of administration/burden:	For respondent: reading and/or comprehensive level, when/under what circumstances the instrument is not suitable for respondent. For administration: Time needed for completion, training needed, easy to understand instructions, user-friendly design/layout.	
F3: Value:	For health professionals,	

	for person with IDD
F4: Acceptability:	For health professionals, for person with IDD
F5: Availability:	Cost: No information. Finding instrument: Other

Table A7

CAPs-IDD based summary of the Mini PAS-ADD.

Measure: The Mini Psychiatric Assessment Schedule for Adults with a Developmental Disability (Mini PAS-ADD) (Moss, 2002; Prosser et al., 1998)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
B1: Concept to be measured:	Broader spectrum of disorders	The Mini PAS-ADD provides a framework for front-line staff to collect relevant information on psychiatric symptomatology from an informant. The Mini PAS-ADD is aimed at case identification, rather than full ICD-10 diagnostic evaluation. It consists of a 66-item questionnaire, with accompanying probes, designed to detect seven psychiatric disorders including depression, anxiety, hypo-mania, obsessive compulsive disorder, psychosis, unspecified disorder (including dementia) and autistic spectrum disorder. Devine et al. (2010) recruited informants of 96 participants. There were 48 (50%) males and 48 (50%) females. Twenty-four people (25%) fell within the age range 18–29 years, 20 (20.8%) within the range 30–39 years, 18 (18.8%) within the range 40–49 years, 22 (22.9%) within the range 50–59 years and 12 (12.5%) were aged between 60 and 65 years. In terms of their level of ID, 6 (6.3%), people were identified as having a borderline learning disability, 37 (38.5%) a mild learning disability and 53 (55.2%) a moderate learning disability. In the study by Devine et al. (2010), the category of Autism was not considered as individuals with a diagnosis of Autism would have already been identified and diagnosed previously. The Mini PAS-ADD was used to collect symptom information from a range of front-line staff and carers who knew the individual with ID well (for at least 1 year), via a semi-structured interview. Community LD Nurses (CLDN) and social workers acted as 'interviewers'. The interviewers were provided with training in how to administer the Mini PAS-ADD Interview by the first and second author.
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Screening	
B6: Available modes of administration:	Other: Semi-structured interview	
B7: Respondent requirements:	Person with IDD: Person with IDD is not a respondent. Third person: Health professional, Caregiver	
B8: Competence Level needed for administration:	Health professional: Nurse, Other. Other professional: Direct care staff, other	
T: Test Development		**Comments**

T1: Main underlying theory for generating items:	No information	Prosser et al. (1998) reported that the Mini PAS-ADD is derived from the PAS-ADD semi-structured clinical interview, which is based on ICD-10 diagnostic algorithms. Unlike the PAS-ADD, Mini PAS-ADD is not an interview, but a framework for gathering data on psychiatric symptomatology which is available without the need for interviewing. A glossary within the Mini PAS-ADD schedule provides a specific framework to follow during the interview, thereby making much more precise judgements around ratings of severity. All items are rated depending on their presence and/or level of severity over the previous 4 weeks. 4-point scale: 0, symptoms not present; 1, mild symptoms; 2, moderate symptoms; or 3, severe symptoms.
T2: Experts involved in test development:	Mental health professionals	
T3: Based on classification models:	ICD	

C: Measurement		**Comments**
C1: Item content:	Questions as third party report: Emotional, e.g. feelings, Other	The Mini PAS-ADD generates a series of sub-scores for the seven domains.
C2: Item coding:	Response format: Polytomous. Facet/quality assessed: Presence (e.g. of a problem, symptom), Severity. Perspective/locus of response for specific items: Presence, Problem for person with IDD.	
C3: Timeframe:	1 month	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**

V1: Criterion validity:	Sensitivity/Specificity	Devine et al. (2010) conducted sensitivity and specificity analyses to examine the agreement between the informants’ ratings on the Mini PAS-ADD and the outcome of a psychiatrist’s interview following her clinical assessment. A random sample of 27 (28.1%) were identified. Sensitivity analysis was found to be perfect (1 or 100%) as both the informants using the Mini PAS-ADD and psychiatrist both correctly identified the same five individuals who had a potential mental health problem. However, specificity analysis was found to be lower at .77 (77%), as both the community informants and psychiatrist correctly identified 17 people who did not a have mental health problem. For five individuals, the psychiatrist did not identify a mental health problem, whereas the informants using the Mini PAS-ADD scored them above the threshold. The authors also noted that out of 18 adults with a known diagnosis of a psychiatric disorder, the Mini PAS-ADD only detected the presence of a mental health problem in 11 cases. Therefore, the Mini PAS-ADD Interview failed to detect the presence of mental problem for seven people who were known to have a psychiatric diagnosis. The authors suggested that it was possible that these individuals may have been in remission or not presenting symptoms at the time of interview. There have been no further studies which have examined the psychometric properties of the recent version of the Mini PAS-ADD (Moss 2002) in adults with mild to moderate ID in English; although Janssen & Maes (2013) validated the measure in Dutch.
V2: Content validity:	No information	
V3: Construct validity:	No information	
V4: Face validity:	No information	

R: Reliability		**Comments**
R1: Internal consistency:	No information	
R2: Reliability:	No information	
R3: Measurement error:	No information	

O: Objectivity of Application		**Comments**
O1: Application:	Comprehensive manual available	

O2: Coding:	Comprehensive manual available	Threshold scores are provided for each of the above seven diagnostic areas. The symptom scales are provided with threshold or referral scores, and it is necessary to have at least one core symptom of a disorder to reach the referral score.

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	No information	Moss (2012) explained that the Mini PAS-ADD produces a single threshold score for each diagnostic area, which can be compared with standardised threshold scores that have been predetermined by the authors of the instrument (Moss 2002).
N2: Cut-offs:	No explanation (but cut-offs are available)	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	The Mini PAS-ADD can be used by unqualified observers, such as family members or staff, who have known the person for at least 6 months. The interviewers in the study by Devine et al. (2010) were provided with training in how to administer the Mini PAS-ADD by the first and second author. The interviews either took place in the person's day-centre, work placement, college or family home. The authors identified that "difficulties were reported by the community informants in identifying appropriate time to undertake the screening process with a suitable informant, as this study was undertaken alongside their current workload/working hours." The measure can be purchased at: https://www.pavpub.com/moss-pas
F2: Ease of administration/burden:	For respondent: No information. For administration: Time needed for completion, Time needed for scoring, Training needed	
F3: Value:	For health professionals, For person with IDD, For caregiver	
F4: Acceptability:	No information	
F5: Availability:	Cost: Not free. Finding instrument: Published as test (manual)	

Table A8

CAPs-IDD based summary of the PWI-ID.

Measure: The Personal Wellbeing Index – Intellectual Disability (PWI-ID) (Cummins et al., 2003)

Part 1: Conceptual and Measurement Model

B: Basic Information		**Comments**
B1: Concept to be measured:	Other	The PWI is a measure of quality of life and life satisfaction. It is an improved version an earlier widely used scale, the Comprehensive Quality of Life Scale (Com-Qol) (Cummins, 1997) which was rated as having satisfactory to excellent validity, reliability, sensitivity, and utility (Hagerty et al. 2001). The PWI-ID was designed for people who have ID or other forms of cognitive impairment, with simple and concretely worded questions and the option of substituting numerical response scales with a reduced choice format illustrated as a series of outline faces (from very happy to sad) to enhance comprehension. Reduced-choice formats (5-, 3-, and 2-point) are available for those who are not able to use the standard 11-point scale. Furthermore, the PWI-ID includes a pre-testing protocol to determine whether respondents with ID are able to use it. This ensures that individuals whose response to scales is unlikely to be valid, such as those with severe or profound ID, are excluded from the sample. The pre-testing protocol allows screening of potential respondents for acquiescent responding, and a test for response scale competence. The PWI-ID may be used as an outcome measure for interventions and service delivery evaluations. It is designed to be administered on an individual basis to the person with an intellectual disability and not by a caregiver or someone who ‘knows the person well’. The protocol specifies that caregivers must not be present during testing and that the interviewees should be seen alone or, if absolutely necessary, with one supportive friend of their choice. In the study by McGillivray et al. (2009), convenience sampling was used to recruit 114 people with mild (82) or moderate (32) level ID in Victoria, Australia. They were recruited from government and non-government organizations and agencies
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Screening, Evaluation of interventions	
B6: Available modes of administration:	Setting: Single setting. Application: Paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): Available	
B7: Respondent requirements:	Person with IDD: Some verbal skills required. Third person: Third person is not a respondent	
B8: Competence Level needed for administration:	None	

		that provide services or programs for people with ID, such as sheltered workshops and supported-employment settings. Of the 114 participants, 62 (54%) were male and 52 (46%) were female. Half of the sample was in the age category 18–30 years, and the mean age was 33 years. There was no significant gender difference in PWI-ID and domains scores. There was no age difference, except for the ‘life achievement’ domain, in which the younger group (18–30) had a significantly higher rating (76.70) than the older group (51–65; 49.50). The PWI-ID scores of respondent groups with mild and moderate ID were also compared. There was generally no appreciable difference in scores between the two groups, although two domain scores (standard of living and health) were significantly higher the mild ID group.

T: Test Development		**Comments**
T1: Main underlying theory for generating items:	Empirical	The PWI was developed by an international community of scholars (International Wellbeing Group, 2006). The scale reflects contemporary understanding of the QoL construct and the items tap into aspects of "satisfaction with life as a whole.”
T2: Experts involved in test development:	Other	
T3: Based on classification models:	None	

C: Measurement		**Comments**
C1: Item content:	Questions as self report: Emotional, e.g. feelings.	There are seven items. The measure asks how happy people are with seven life domains: standard of living, health, life achievement, personal relationships, personal safety, community-connectedness, future security. Questions on ‘satisfaction’ from the original PWI are substituted by the term ‘happiness’ in the PWI-ID. The ID version also uses more simple and concrete wordings. An additional question asks how happy or sad the respondent is with life as a whole is included. The PWI-ID scale incorporates a pre-testing protocol to determine whether, and to what level of complexity, respondents are able to use the scale. For those who are unable to comprehend the standard 11-point response scale, a set of reduced-choice format scales (5-, 3-, and 2-point) was provided. These use
C2: Item coding:	Response format: Other. Facet/quality assessed: Other. Perspective/locus of response for specific items (symptom, behaviour): Other	
C3: Timeframe:	No time frame given: None specified	

C4: Floor/ceiling effects:	Other	drawings of faces (from very happy to very sad) in the place of numbered scale choice points. In the study by McGillivray et al. (2009), 72 people were able to use the 11-point scale, 20 used the 5-point scale, 19 used the 3-point scale, and 3 used the 2-point scale. In the analysis, a total of 37 data sets from individual respondents were eliminated prior to analysis because they showed maximum scores for all seven items.
C5: Responsiveness:	Recommended to detect changes (e.g. as a ROM - routine outcome measurement): Yes	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	McGillivray et al. (2009) conducted a factor analysis and two factors explained 57.97% of the variance. However, one of the items loaded onto both factors. As three items is considered the minimum number for a factor, the analysis resolved to a single factor, in line with the previous literature. Subjective wellbeing was traditionally measured by asking a single question regarding how people rate their satisfaction with "life as a whole". The PWI-ID deconstructs satisfaction with life as a whole into a number of life domains. All domains of the PWI-ID correlated significantly with ‘life as a whole’ and ranged between .27 and .44 (poor to fair). To determine the unique contribution of the domain to ‘life as whole’, the former were regressed against the latter. ‘Personal relationships’ and ‘Safety’ constituted the largest unique contribution to the prediction of ‘life as a whole’, although all domains were statistically significant contributors (adjusted $R_2 = .27$). This is a much lower level of explained variance than is typically found in general population samples. The domain inter-correlations ranged between 0.16 and 0.54 (poor to good). The highest correlations were between community connectedness and personal relationships (.54); standard of living and life achievement (.49); and personal relationships and future security (.45).
V2: Content validity:	No information	
V3: Construct validity:	Factor analysis, convergent validity	
V4: Face validity:	Face validity rated by author(s)	

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	The Cronbach's alpha coefficient was α = 0.76, which demonstrated fair internal reliability, and is comparable to previous research of the PWI in the Australian
R2: Reliability:	Test-retest	

R3: Measurement error:	No information	general population (α = 0.80; Lau et al. 2005). To examine test–retest reliability of the scale, it was re-administered to 31 randomly selected participants (representative in demographic and socioeconomic characteristics, of the 114 group) at 1–2 weeks following its initial administration. The analysis revealed an intraclass correlation coefficient of .57 (good), and test–retest coefficient of .58 (good).

O: Objectivity of Application		**Comments**
O1: Application:	Instructions: Comprehensive manual available	http://www.acqol.com.au/uploads/pwi-id/pwi-id-english.pdf
O2: Coding:	Guidelines for coding: Comprehensive manual available	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**
N1: Norms:	Available	McGillivray et al. (2009) reported that the PWI is forms a part of the Australian Unity Wellbeing Index, which has been used since 2001 in regular surveys of the Australian general population, to measure how satisfied people are with their lives. McGillivray et al. (2009) compared PWI-ID ratings with PWI ratings from general population samples. Further information is available in the published paper.
N2: Cut-offs:	None	
N3: Fairness:	No information	

F: Feasibility		**Comments**
F1: Missing values:	No information	The pre-testing and the full scale administration take from 10 to 20 minutes to complete, depending on the extent of pre-testing required though it should be emphasised that there is no time limit. Participants were initially screened to identify and exclude participants who could not cope with the test demands due to
F2: Ease of administration/burden:	For respondent: Time needed for completion, when/under what	

	circumstances the instrument is not suitable for respondent. For administration: Time needed for completion, Norms available, Easy to understand instructions, User-friendly design/layout	level of disability. Each interview lasted approximately 45 min. The authors reported that the use of the PWI-ID may assist in meeting the needs of people with ID and inform the planning and delivery of resources and services. They concluded that the advantages of the PWI-ID were that it was easy and convenient to use, includes a pre-testing protocol, has simple and concretely worded questions and reduced-choice response scales.
F3: Value:	For health professionals, for person with IDD	
F4: Acceptability:	For health professionals, for researchers	
F5: Availability:	Cost: Available for free. Finding instrument: Published as test (manual)	

Table A9

CAPs-IDD based summary of the SAINT.

Measure: The Self Assessment and Intervention (SAINT; Chaplin et al., 2012)		
Part 1: Conceptual and Measurement Model		
B: Basic Information		**Comments**
B1: Concept to be measured:	Broader spectrum of disorders	The SAINT is a guided self-help approach for people with ID. The SAINT's self-report section consists of 10 self-report statements and is tool designed to assist people (with or without support) to report feelings, emotions, and symptoms of mental health problems. Each statement is accompanied by a picture in the booklet. It has not been designed as a diagnostic tool and therefore does not use a cut-off score for screening depression or anxiety. Instead, the aim is to help the person with ID to recognise and report symptoms which indicate mental distress. The SAINT also provides 10 coping strategies that the individual may use to help manage the symptoms identified. The authors reported that utilising the self-report tool regularly can also provide a chronology of feelings experienced by the person and interventions the person has used. The SAINT was developed originally through consultation with professional experts and service user experts. It can be used either independently or with support. The sample consisted of 54 participants (M = 31, F = 23). Fifty participants had mild ID and four had moderate ID. Participants were between 18 and 77 years old with a mean of 39.4 years of age (*SD* = 13.57). Participants were recruited from community (*n* = 31, 57.4%) and inpatient settings (*n* = 23, 42.6%).
B2: Level of IDD:	Mild, Moderate	
B3: Aetiology of target group:	None	
B4: Age of Target group:	Adulthood	
B5: Primary purpose/recommendation for use:	Other	
B6: Available modes of administration:	Setting: Single setting. Application: paper-pencil. Augmentative and alternative communication (e.g. visual aids, symbols): Available	
B7: Respondent requirements:	Person with IDD: No information. Third person: Third person is not a respondent	
B8: Competence Level needed for administration:	None	
T: Test Development		**Comments**

T1: Main underlying theory for generating items:	No information	Delphi methods and focus groups were employed to gather opinions from two expert groups: professionals or clinical experts (ranged *N* = 15-33); and service users (*n* = 9), to inform the contents of the SAINT. Focus groups were asked to address the questions, "how do people feel when they are becoming distressed or mentally unwell? When do we know when we are at risk from becoming unwell?" Results from each round were shared between the two groups to develop a consensus. A list of 52 items from existing measures of psychopathology (e.g. Beck depression inventory (Beck et al., 1988b), brief psychiatric rating scale (Overall and Gorham, 1962), GDS-LD (Cuthill et al., 2003)), were used to inform the list of self-report statements. Participants were also asked for comments and to suggest other items for consideration that had not already been listed.
T2: Experts involved in test development:	Mental health professionals, Persons with IDD (E.g. focus group, pre-testing)	
T3: Based on classification models:	Other	

C: Measurement		**Comments**
C1: Item content:	Questions as self-report - Emotional, e.g. feelings	The SAINT has a feelings list of 10 items with an additional four examples, which has a dichotomous yes/no format.
C2: Item coding:	Response format: Dichotomous (yes/no). Facet/quality assessed: Presence (of a problem/symptom). Perspective/locus of response for specific items (symptom, behaviour): Presence, Problem for person with IDD.	
C3: Timeframe:	No time frame given: Now (e.g. how the person is feeling/behaving now)	
C4: Floor/ceiling effects:	No information	
C5: Responsiveness:	Recommended to detect changes: No information	

Part 2: Psychometric Properties

V: Validity		**Comments**
V1: Criterion validity:	No information	With regards to convergent validity, the total SAINT scores were correlated with GDS-LD total and GAS-ID worries subscale scores. The SAINT showed significant correlation at the .01 level (two-tailed) with the GDS-LD (*r* = 0.71; *p* < .01; strong correlation), GAS-ID worries (*r* = 0.58; *p* < .01; moderate correlation). In terms of face validity, the SAINT was piloted among two nurses and one research assistant all working in specialist mental health services for people with ID.
V2: Content validity:	Concerning relevance of items, Concerning comprehensiveness of items	
V3: Construct validity:	Hypotheses testing: Convergent validity	
V4: Face validity:	Face validity rated by health professionals	

R: Reliability		**Comments**
R1: Internal consistency:	Cronbach's alpha	Cronbach’s alpha (α = 0.83) suggested good internal consistency for the SAINT questionnaire. The inter-item correlation for internal consistency of the SAINT for any item deleted ranged from .79 to .83 (excellent). Test-retest reliability was analysed for 20 participants (37% of the sample). There was statistically significant correlation between T1 and T2 (*r* = 0.90, *p* < .01; strong correlation). The period between test and retest was one week.
R2: Reliability:	Test-retest	
R3: Measurement error:	No information	

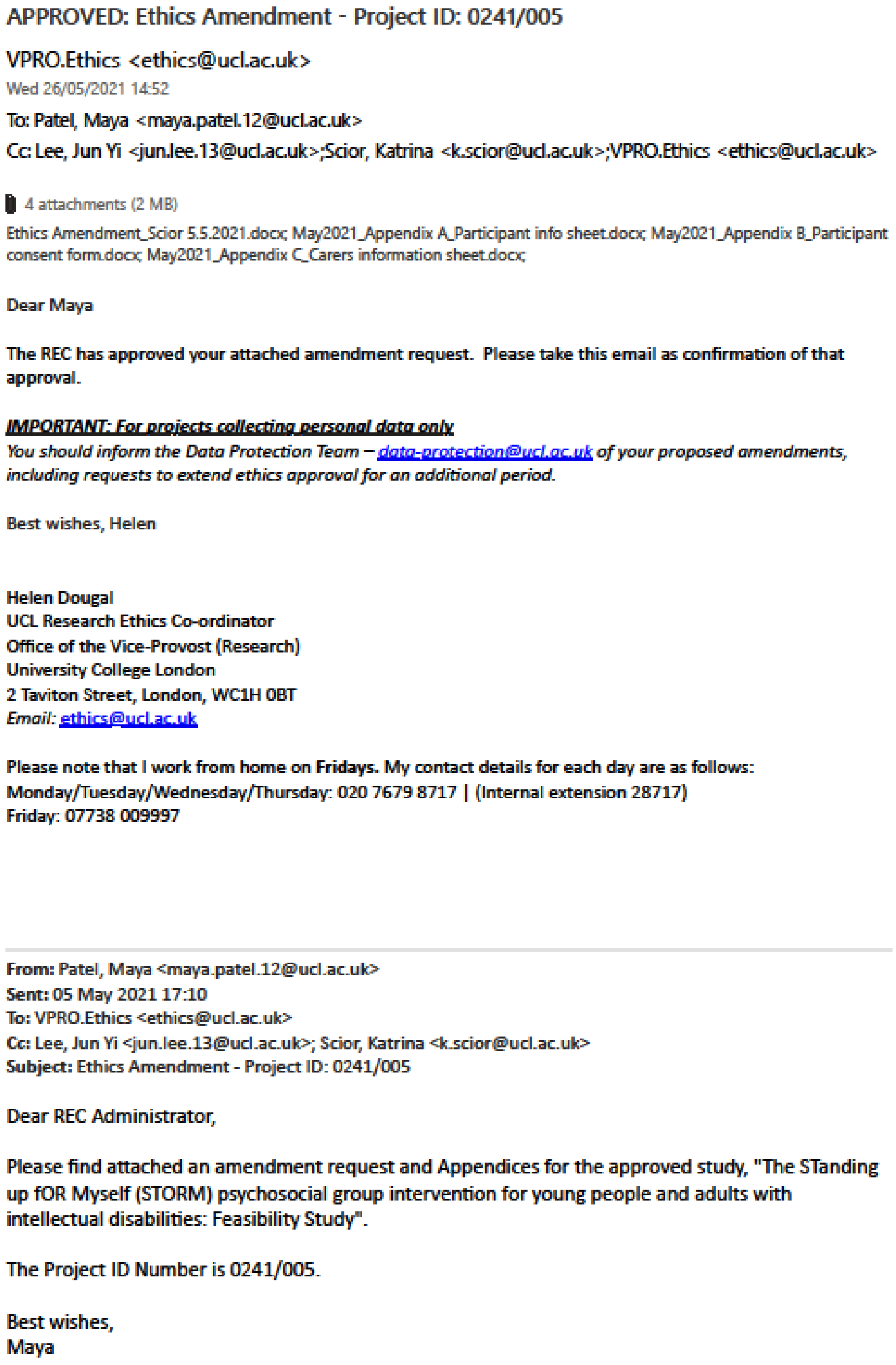
O: Objectivity of Application		**Comments**
O1: Application:	No information	The SAINT has not been designed as a diagnostic tool and therefore there are no cut-off scores.
O2: Coding:	None	

N: Objectivity of Interpretation, Norming and Fairness		**Comments**

N1: Norms:	No information
N2: Cut-offs:	None
N3: Fairness:	No information

F: Feasibility		**Comments**
F1: Missing values:	No information	The focus group was asked for feedback comments regarding the ease of use. The authors reported that using the tool on a regular basis will enable people with ID to become increasingly familiar with the tool and signs and symptoms associated with mental distress.
F2: Ease of administration/burden:	For respondent: Reading and/or comprehension level. Evidence that instrument places no undue physical or emotional strain on respondent. For administration: User-friendly design/layout	
F3: Value:	For person with IDD	
F4: Acceptability:	For person with IDD	
F5: Availability:	No information	

Appendix B: Ethical Approval



APPROVED: Ethics Amendment - Project ID: 0241/005

VPRO.Ethics <ethics@ucl.ac.uk>
Wed 26/05/2021 14:52
To: Patel, Maya <maya.patel.12@ucl.ac.uk>
Cc: Lee, Jun Yi <jun.lee.13@ucl.ac.uk>;Scior, Katrina <k.scior@ucl.ac.uk>;VPRO.Ethics <ethics@ucl.ac.uk>

4 attachments (2 MB)
Ethics Amendment_Scior 5.5.2021.docx; May2021_Appendix A_Participant info sheet.docx; May2021_Appendix B_Participant consent form.docx; May2021_Appendix C_Carers information sheet.docx;

Dear Maya

The REC has approved your attached amendment request. Please take this email as confirmation of that approval.

IMPORTANT: For projects collecting personal data only
You should inform the Data Protection Team – data-protection@ucl.ac.uk of your proposed amendments, including requests to extend ethics approval for an additional period.

Best wishes, Helen

Helen Dougal
UCL Research Ethics Co-ordinator
Office of the Vice-Provost (Research)
University College London
2 Taviton Street, London, WC1H 0BT
Email: ethics@ucl.ac.uk

Please note that I work from home on **Fridays.** My contact details for each day are as follows:
Monday/Tuesday/Wednesday/Thursday: 020 7679 8717 | (Internal extension 28717)
Friday: 07738 009997

From: Patel, Maya <maya.patel.12@ucl.ac.uk>
Sent: 05 May 2021 17:10
To: VPRO.Ethics <ethics@ucl.ac.uk>
Cc: Lee, Jun Yi <jun.lee.13@ucl.ac.uk>; Scior, Katrina <k.scior@ucl.ac.uk>
Subject: Ethics Amendment - Project ID: 0241/005

Dear REC Administrator,

Please find attached an amendment request and Appendices for the approved study, "The STanding up fOR Myself (STORM) psychosocial group intervention for young people and adults with intellectual disabilities: Feasibility Study".

The Project ID Number is 0241/005.

Best wishes,
Maya

Appendix C: Contributions by Each Trainee to this Project

This project was completed jointly with another trainee, Jun Yi Lee. Jun Yi's thesis focussed on predictors of self-esteem in individuals with intellectual disabilities.

We jointly submitted an ethics amendment request, data protection and risk assessment forms. We jointly amended the STORM Qualtrics survey so that it was suitable for our project and developed a research database.

We jointly created a poster for distribution on social media and managed a Twitter page to promote the study.

We jointly contacted 57 charities and third sector organisations for individuals with intellectual disabilities to recruit participants and subsequently attended virtual group meetings to promote the project. With regards to data collection, I met with 25 participants (15 twice for test-retest) and Jun Yi met with the remaining 19 participants.

Jun Yi also double rated the quality assessment in the systematic review.

Our analyses and the writing up of our theses were completed independently.

Appendix D: Easy Read Participant Information Sheet

Information about our research

You can ask someone you know to help you read this letter.

Research means finding out about things.

We are the research team from University College London:

Katrina

Jun Yi

Maya

Lisa

You can contact Maya or Jun Yi to talk about this research:

Maya.patel.12@ucl.ac.uk or jun.lee.13@ucl.ac.uk

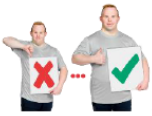
FUNDED BY NIHR | National Institute for Health Research

This study is funded by the National Institute for Health Research (NIHR) Public Health Research programme (NIHR PHR Project 17/149/03).

Why we are doing this research

We want to find out how people with learning disabilities think about themselves and how they feel.

We have made a new questionnaire to find this out.



We want some people to help us try out the new questionnaire by answering the questions. This will help us know if the questionnaire is working.

We would like you to complete some online questionnaires about you and how you feel.

We will meet you using video calls.

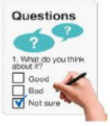
A video call means you will see us on a digital computer or phone screen. We will not be together in the same room.

We are asking you to take part

This letter gives you information about the research and what we will ask you to do if you want to take part.

You can talk about it with other people like your family and friends if you want.

What will happen if you take part



You will answer some questions online about you and how you feel.

We will be asking some of you to answer the questions two times.

There will be a one or two week break in between.

You will answer the questions on a video call with a researcher.

We can send you Easy Read information on how to use the video call.

What we do with your information

Your answers to the questions are confidential.

That means we don't share your answers with anyone who is not working on the research.

If you tell us something that makes us worry about you or someone else's safety, we may need to tell someone. This is to keep you and others safe.

Your answers to the questions will be kept in a locked place on a computer.

The computer will use passwords to keep what you talked about safe.

Choosing to take part in research

You can choose if you want to take part in the research.

It's up to you.

You can say **no**.

You can say no later on if you change your mind.

If you say **yes** to the research, we will ask you to give your consent on a video call. We will record this part of the call.

This means that you agree to take part in the research and that you are happy to use your information.

Changing your decision

You can stop taking part in the research at any time.

You don't have to tell us why.

Good things about taking part

We hope you find the research interesting and learn new things.

What we find out may help you and other people with learning disabilities.

Your answers will help us make sure the questionnaire is working.

You will be given a £10 thank you voucher for answering questions.
If we meet with you two times, you will be given an extra £5 voucher.

Things that might be difficult

Some questions might be difficult or make you feel sad.

We will make sure that there is someone to talk to about anything you found difficult or sad.

How to make a complaint

If there is a problem, you can talk to someone from the research team.

We will try to help you.

Contact details for the research team are on page 1.

We will tell you when we think the problem has been fixed.

Thank you for reading about the project.

Appendix E: Carers information sheet

Study Information for Carers

UCL Research Ethics Committee Approval ID Number: 0241/005

Dear Family/Supporter of ___________________________,

As you support ____________________, we are writing to let you know that they have been invited to take part in a research study that the University College London (UCL) are running. The study is about understanding how people with learning disabilities view themselves and how they feel.

Please find information about the study enclosed with this letter.

If you would like to contact the research team, our details are at the bottom of this letter.

Kind Regards,

Katrina Scior, Lisa Richardson, Michaela Osborne, Maya Patel and Jun Yi Lee

Contact details:

Should you have any queries or require any further information, please do not hesitate to contact Maya Patel or Jun Yi Lee by email: maya.patel.12@ucl.ac.uk or jun.lee.13@ucl.ac.uk.

Maya and Jun Yi work Wednesdays and Fridays on this project.

Should you need to get hold of someone urgently at other times please contact Katrina Scior (Study Lead): Email- k.scior@ucl.ac.uk, Tel- 0207 6791897

This study is funded by the National Institute for Health Research (NIHR) Public Health Research programme (NIHR PHR Project 17/149/03).

FUNDED BY

NIHR | National Institute for Health Research

Study Information

About the research study

- We are a team of researchers from University College London (UCL).
- We are working with researchers, clinicians and self-advocates with learning disabilities from different parts of the UK.
- We are researching how people with learning disabilities view themselves and how they feel.
- The current study is funded by the National Institute of Health Research.

Why has the person you support been invited to take part?

- We are interested in speaking to people with intellectual disabilities aged 16 and above.

What will the study involve for the person you support?

- The person you support will receive information via an accessible information sheet and an explanation from the research team during a video call. Each participant will be provided with an accessible consent form to agree to verbally if they understand and agree to take part. They do not have to take part in this research study and will be informed that it is OK to say ‘no’.
- If the person you support agrees to take part in the study, they will be asked to complete a set of questionnaires. They will be guided by a member of the research team during a video call. The questions will be about themselves and how they feel, including how empowered they feel and how having a learning disability affects them. You or another supporter can be present if the person you support requests that. As we are also interested in seeing how consistent our measures are over time, we will be asking some of our participants to meet with us on a second occasion, one to two weeks later.

Thank you for your support, please do not hesitate to contact a member of the research team or the group facilitator if you have any queries or concerns regarding this.

What are the possible disadvantages and risks of taking part?

Taking part in questionnaires-
Participants will be asked to complete questionnaires that ask questions about their mental wellbeing and how they feel about themselves. They will do this independently or jointly with a researcher who will support them if they find any questions difficult to answer and look out for any signs of discomfort or upset. If the person is unduly upset the researcher will inform their carers/supporters and arrange for further support. Participants may wish to speak to their carers/supporters and/or friends afterwards. You are welcome to contact us if you are concerned about the person in any way. We can help sign-post to support.

What are the benefits of taking part?

- Participants may enjoy completing the questionnaires jointly with a researcher and may learn new things about themselves.
- We hope to be able to better understand and measure self-esteem and levels of wellbeing in people with mild to moderate learning disabilities.
- To say thank you for participants' time they will receive £10 in retail vouchers. Participants will receive an additional £5 voucher if we asked them to meet with us on a second occasion.

What will happen to the questionnaires?

The anonymised responses to the online questionnaires will be available to the UCL research team.

No one outside the UCL research team will have access to any personal or identifying information about participants. The information will be stored safely on a password protected computer. The questionnaire responses will be saved securely on a password protected computer without any identifiable information about participants. The UK Data Protection Act 2018 will be adhered to at all times.

What will happen to the results of the research study?

This study will be written into reports, which could be published. It will not be possible to identify any of the individuals who take part in the study from the reports, as all the information will be anonymised, with information from many individuals grouped together.

Local Data Protection Privacy Notice

The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data.

This 'local' privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our 'general' privacy notice for participants in health and care research studies.

Questionnaires and interviews (data collection) will be completed over web-based meetings. The categories of personal data used will be the name of the organisation through which the person participates in this research, name of person, age and gender as well as contact details (to facilitate arranging meetings for the purposes of data collection without necessarily having to rely on the group facilitators). The lawful basis that will be used to process this personal data are: 'Public task' for personal data and' Research purposes' for special category data. The data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data provided we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how personal data is being processed, or if you would like to contact us about your and/or the rights of the participants you are supporting, please contact a member of the research team in the first instance (contact details on page 1). In addition the UCL data protection team can also be contacted at the following e mail dataprotection@ucl.ac.uk should you have any further concerns.

What if something goes wrong?

If you have any concerns or complaints about the way the research is being managed you can contact Katrina Scior in the first instance, our contact details are on page 1.

You can also contact the Chair of the UCL Research Ethics Committee with complaints – ethics@ucl.ac.uk

Thank you for reading this information sheet.

Supporting people should they become upset

The topics discussed can affect people differently. We recommend to remain alert not only to verbal disclosures, but also changes to individuals' usual demeanour and body language, for example, if someone who is usually very outspoken is avoiding eye contact and appearing withdrawn. In these instances we would recommend that you approach people after the session to see if they want to talk with you about anything on a one-to-one basis.

What to do if someone is distressed

If someone is in distress, try the following:

- Allow them time to talk and say how they feel
- Acknowledge their feelings and that this may be distressing for them and others
- Reassure them that it is OK to be upset because of the nature of what they heard or talked about
- Ask them if there is anything you could do to help
- Support them to do something they find relaxing when they are ready
- Consider with the person whether they need more support, you can contact the G.P in the first instance to discuss what support is available locally, e.g. talking therapy

Emergency and crisis contacts-

If someone mentions an **on-going problem, abuse or crime** this should be reported by follow local safeguarding procedures and/or to the Police. Contact details can be found on the website of your Local Authority under Adult Safeguarding.

On the next page are some emergency contact details should the person you care for and support need them.

Support for carers/supporters-

If you feel that you would like support as a carer/supporter of someone with a learning disability, please see the resources below:

Mencap – FamilyHub: An online community for parents and family carers to connect with others, to share triumphs and challenges, and a place for support and tips https://www.mencap.org.uk/familyhub

National Family Carer Network: a charity that aims to promote the voice and rights of family carers supporting a person with a learning disability http://www.familycarers.org.uk/ info@familycarers.org.uk

Crisis Emergency Contacts

If you are in **immediate crisis**,
for example if you think:

- you may act on suicidal feelings or
- you have seriously hurt yourself or
- are thinking of seriously hurting yourself:

Go to any hospital **Accident & Emergency department** and ask for help, or

If you need to, you can **call 999** and ask for an ambulance

Support

Contact **Samaritans** for free 24/7 telephone support: 116 123 (UK & ROI)

If you don’t feel you need to go to A&E, other services can offer support:

Your **GP surgery** should be able to offer urgent appointments with the next available GP

SANE runs a mental health helpline from 4:30pm to 1 0:30pm daily (0300 304 7000)

Appendix F: WEMWBS-ID

- Removed due to copyright -

Appendix G: SWEMWBS-ID

- Removed due to copyright -

Appendix H: Consent Form

V3

Research Consent Form

UCL Research Ethics Committee Approval ID Number: 0241/005

Please tick the box if your answer is 'Yes'.

1. Have you read the information letter or has it been read to you? ☐

2. Do you understand what the project is about? ☐

3. Do you understand the good things about taking part? ☐

4. Do you understand what might be difficult about taking part? ☐

5. Have you asked all the questions you want? ☐

6. Were your questions answered in a way you understand? ☐

7. Have you had time to think about if you want to take part? ☐

8. Do you understand that we will meet with you either one or two times on a video call? ☐

9. Do you understand that all of the information you give the research team will be kept confidential? ☐

10. Do you understand that it is OK to stop at any time? ☐

11. Are you happy to take part? ☐

If you want to take part, please sign below:

Name: ____

Signature: ____

Date: ____

If you don’t want to take part, you don’t have to sign.

The researcher will complete this section:

Researcher’s initials:_________

Signature: ___________________________Date: ________