Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review

Link to publication record in Ulster University Research Portal

Document Version
Publisher's PDF, also known as Version of record

REVIEW

JARID
Journal of Applied Research in Intellectual Disabilities   bild   WILEY

Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review

Stephanie Maguire | Jenny Davison | Marian McLaughlin | Victoria Simms |
Brendan Bunting

School of Psychology, University of Ulster, Coleraine, UK

Correspondence
Stephanie Maguire, School of Psychology, University of Ulster, Cromore Road, Coleraine, Co. Derry, BT52 1SA, UK.
Email: s.maguire1@ulster.ac.uk

Abstract

Background: Studies that have systematically reviewed the psychometric properties of health-related quality of life (HRQoL) and subjective wellbeing instruments for adolescents with intellectual disabilities narrowly focus on disease or health-specific conditions. This review aimed to critically appraise the psychometric properties of self-report instruments used to measure HRQoL and subjective wellbeing of adolescents with intellectual disabilities.

Method: A systematic search was undertaken in four databases. The quality of the included studies and their psychometric properties was assessed according to the COnsensus-based Standards for the selection of health Measurement Instruments Risk of Bias checklist.

Results: Seven studies reported psychometric properties of five different instruments. Only one instrument identified as having potential to be recommended for use but requires further validation research to assess its quality for this population.

Conclusions: There is insufficient evidence to support the recommendation of a self-report instrument to assess HRQoL and subjective wellbeing of adolescents with intellectual disabilities.

KEYWORDS
adolescents, health-related quality of life, intellectual disability, psychometric properties, self-report measures, subjective wellbeing

1 | INTRODUCTION

Intellectual disability (ID) is characterised by significant limitations in both intellectual functioning and adaptive behaviour as expressed in conceptual, social, and practical skills. This disability originates during the developmental period, which is defined operationally as before the individual attains age 22 (Schalock et al., 2021). The severity of ID is classified into four types, based on an intelligence quotient (IQ) test, namely mild, moderate, severe, and profound ID (Schalock et al., 2010). Adolescence is the phase of life between childhood and

adulthood, from ages 10 to 19. It is a unique stage of human development and an important time for laying the foundations of good health. Adolescents experience rapid physical, cognitive and psychosocial growth. This affects how they feel, think, make decisions and interact with the world around them (WHO, 2022). Health-related quality of life (HRQoL) and subjective wellbeing are important concepts for adolescents with ID given the health inequalities that exist within this population (Hamdani et al., 2018). In comparison with their neurotypical peers, adolescents with ID are at increased risk of developing health problems (e.g., co-occurring chronic disease and/or mental illness, obesity) (Frey et al., 2017), diminished physical health (Frey et al., 2008; WHO, 2018), lower socio-economic status and social exclusion (Allerton et al., 2011; Buckley et al., 2020; Emerson, 2021), premature and avoidable mortality (Hughes-McCormack et al., 2022; Smith et al., 2020). To address these health inequalities, it is important to measure the HRQoL and subjective wellbeing of these young people and identify those who are at risk of poorer outcomes so that effective interventions can be put in place as soon as possible.

Only recently has the study of HRQoL gained scientific interest and assumed paramount importance in the identification of health risks and populations' health status (Marques et al., 2019). HRQoL is a multi-dimensional construct that refers to the physical, psychological, and social domains of health, seen as distinct areas that are shaped by an individual's experiences, beliefs, expectations and perceptions (Guyatt et al., 1993; Solans et al., 2008; Stenman et al., 2010; Testa & Simonson, 1996; WHO, 1948). Its evaluation generally relies on the person's evaluation of wellbeing and/or functioning within the different domains comprising the overall construct (Solans et al., 2008; Testa & Simonson, 1996). The physical, psychological and social functioning aspect of HRQoL consists of behaviours that can be observed by other people (such as social relationships with family and friends), and the wellbeing part refers to internal, subjective perceptions such as vitality, pain, anxiety, depressive symptoms and general health perceptions (Hays & Fayers, 2021; Hays & Reeve, 2010). A related concept is that of subjective wellbeing (SWB). Despite some commonalities, HRQoL and wellbeing should be treated as separate concepts (Upton & Upton, 2015). HRQoL refers to the cognitive appraisal which a person makes about the impact their health has on their own lives, whilst SWB is 'people's cognitive and affective evaluations of their own lives' (Diener, 2000, p. 34). Wellbeing has long been considered key to the creation and maintenance of healthy, productive societies (Das et al., 2020; Durand, 2015). Wellbeing is a multifaceted construct (Delle Fave et al., 2011; Dodge et al., 2012; Forgeard et al., 2011) that encompasses objective (e.g., income, education, health) and subjective (e.g., inter-personal relationships, autonomy) aspects of a person's life (Bowling, 2011; Forgeard et al., 2011; Selwyn & Wood, 2015; Statham & Chase, 2010). This approach to measuring perceptions and life experiences has been characterised as SWB. There is no universally agreed definition of SWB, and the term is often used interchangeably with life satisfaction, happiness and quality of life (Selwyn & Wood, 2015; Statham & Chase, 2010). SWB falls within the hedonic perspective and can be understood as how people feel and function,

both on a personal and social level, and how they evaluate their lives as a whole (Diener, 2009; Michaelson et al., 2012). For example, how an individual feels about their life often stems from their inborn temperament and overall outlook, but the circumstances in which an individual lives, including access to basic resources, also play an important part in how happy and satisfied they feel.

Traditionally, HRQoL and subjective wellbeing of adolescents with ID have been assessed via proxy reports completed by parents, teachers, or carers (Ravens-Sieberer et al., 2005). However, limitations exist surrounding the robustness of proxy reflection of non-observable internal states (i.e., feelings), particularly in relation to people whose language limitations mean that they have not been able to tell even close proxies what they think (Emerson et al., 2013). Every individual has a unique perception of his/her health and wellbeing which is influenced by context, previous experiences and personal values (Noonan et al., 2016). This personal perspective can only be understood via individuals' self-reports. Therefore, adolescents' views should, where possible, be sought directly rather than being inferred from proxy-reports (Upton et al., 2008). Further, the United Nations Convention on the Rights of Persons with Disabilities highlights the need to ensure that children with disabilities 'have the right to express their views freely and are provided with appropriate assistance to realise that right' (Article 7). Although, measurement and identification of HRQoL and subjective wellbeing can be challenging within this heterogeneous group, research indicates that adolescents with ID can reliably report on these domains if instruments are appropriate to their age and cognitive functioning (Davison et al., 2022; Ingerski et al., 2010; Morrow et al., 2011). To date, studies have systematically reviewed the psychometric properties of self-report HRQoL and QoL instruments for children and young people with ID, however, they have a narrow focus on disease, chronic, or health-specific conditions (Davidson et al., 2017), examples include cerebral palsy (CP, Carlon et al., 2010), heart conditions (Jardine et al., 2014), neurodisability (Morris et al., 2015), asthma, cancer and epilepsy (Solans et al., 2008). Given this, an up-to-date systematic review is warranted to collate, summarise and appraise the psychometric properties of self-report instruments used to measure HRQoL and subjective wellbeing of adolescents with ID.

This systematic review will identify and synthesise evidence-based knowledge regarding what self-report instruments are being used in HRQoL and subjective wellbeing research amongst this population, and more importantly determine how reliable and valid these instruments are. This is urgently needed as the choice of high-quality instruments among clinicians and researchers is strongly determined by having robust psychometric properties including validity and reliability (Karanicolas et al., 2009; Kipfer & Pihet, 2020; Mason et al., 2018; Scholtes et al., 2011). The objectives of this review are to: (1) identify self-report instruments used to measure HRQoL and subjective wellbeing of adolescents with ID; and (2) formally evaluate the methodological quality and psychometric properties of these self-report instruments. Overall, this review will provide clinicians and researchers with a robust evidence-base to assist them when selecting instruments for this purpose.

2 | METHODS

This review followed the recommended guidelines for preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 (Page et al., 2021) and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) (Prinsen et al., 2018). The protocol was registered with PROSPERO, an international register for systematic reviews with health-related outcomes (PROSPERO 2021 CRD42021231697) and published open access in BioMed Central Systematic Reviews (Maguire et al., 2022). This review was conducted in four sequential steps:

1. *Step 1*: Systematic literature search formulating eligibility criteria (Step 1.1), searching the literature and selecting studies (Step 1.2);
2. *Step 2*: Evaluation of the methodological quality of included studies on psychometric properties of instruments using the COSMIN Risk of Bias checklist;
3. *Step 3*: Evaluation of the psychometric properties of instruments, rating the result of single studies against the criteria for good psychometric properties (Step 3.1), summarising all results of studies per instruments (Step 3.2), and grading the quality of evidence on psychometric properties (Step 3.3); and
4. *Step 4*: Recommendation of the most suitable instruments.

Each of these steps will be further described in the sections that follow.

2.1 | Step 1: Systematic literature search

Systematic literature search for this review was performed in two sub steps: formulating eligibility criteria (Step 1.1), searching literature and selecting studies (Step 1.2). These two steps are in accordance with the PRISMA 2020 guidelines (Page et al., 2021).

2.1.1 | Step 1.1: Eligibility criteria

Studies were included when: (1) instruments were designed for use by adolescents with ID to measure their HRQoL (physical health; psychological health and social health) and subjective wellbeing (life satisfaction; happiness and quality of life); (2) instruments were self-report; (3) the sample included adolescents aged between 11 and 16 years (if studies included adolescents with a broader or narrower age range than 11–16 years olds but encompass 11–16-year-olds, they were included) (4) the sample included adolescents with an ID diagnosis (mild, moderate, severe or profound); (5) they were written in English language and/or available in an English translation; (6) they were published in peer-reviewed journals and included either quantitative or mixed-methods research. Studies were excluded when: (1) reported instruments did not assess HRQoL and subjective wellbeing domains; (2) instruments were not self-report (i.e., proxy); (3) instruments were designed for adults or older people with ID; (4) it was not clear whether participants had an ID, or where they had

other conditions (i.e., autism, epilepsy, or physical disabilities) without specifically noting that they also had an ID; (5) they were not available in the English language; (6) they were review papers, editorials or case studies. Provided that HRQoL and SWB of adolescents with ID have traditionally been assessed via proxy reports completed by parents, teachers or carers (Ravens-Sieberer et al., 2005), the results were limited to publication available between 2000 and 2020 to focus on both old and new instruments utilised in the last 20 years. The following additional selection criteria was used for psychometric studies: (1) reported psychometric data of at least one of the following eight psychometric properties as defined by the COSMIN statement (Mokkink et al., 2010a, 2010b): content validity; structural validity; internal consistency; cross-cultural validity\measurement invariance; reliability; measurement error; criterion validity; and hypotheses testing for construct validity. Responsiveness was beyond the scope for this review because evaluation of responsiveness would require reviewing all studies that have used the identified instruments as an outcome measure and would require a different search strategy.

2.1.2 | Step 1.2: Literature search and study selection

During February 2021 systematic literature searches were performed in four electronic databases: Psychological Information Database (PsycINFO), Cumulative Index of Nursing and Allied Health Literature (CINAHL), Medical Literature Analysis Retrieval System Online (MEDLINE), and Education Resources Information Center (ERIC) to identify published articles reporting on the psychometric properties of self-report instruments used to measure HRQoL and subjective wellbeing instruments of adolescents with ID. Terms used in the literature search (with their synonyms and closely related words) included: 'health-related quality of life and subjective wellbeing', 'measures', 'psychometric properties', 'adolescence' and 'intellectual disability'. The search terms were applied to all databases and modified to meet the requirements of each database due to different field restrictions (see Supporting Information Appendix A for the full search strings for each database). The selection of studies was administered through a peer review process using the online software Covidence. The interrater agreement was assessed by calculating weighted κ (Cohen & Humphreys, 1968) and interpreted as very good (0.81–1.00), good (0.61–0.80), moderate (0.41–0.60), fair (0.21–0.40) and poor (0.00–0.20) (Altman, 1991). Next, reference lists of all included full texts were hand searched to identify additional eligible instruments and studies. Scholarly databases, resource catalogues and websites of government agencies were also searched to identify potential instruments. EThOS (e-thesis online service) was also used to search for e-thesis and dissertations.

2.2 | Step 2: Evaluation of methodological quality of studies

The methodological quality of the studies was assessed using the COSMIN Risk of Bias checklist, which is a standardised tool for

4 | **WILEY**_JARID_ bild Journal of Applied Research in Intellectual Disabilities   Published for the British Institute of Learning Disabilities

MAGUIRE ET AL.

evaluating study quality of psychometric studies (Yoon et al., 2021). The checklist contains 3–38 items for each psychometric property (Mokkink et al., 2018). The checklist items rate the quality of study design and the robustness of statistical analyses (Mokkink et al., 2018). Recently, the COSMIN group published updated guidelines for conducting systematic reviews on psychometric properties of patient-reported outcome instruments (Prinsen et al., 2018; Terwee et al., 2018). The COSMIN guidelines include the following practical tools: a taxonomy defining each psychometric property (Mokkink et al., 2010b), a checklist to assess methodological quality of psychometric studies (Mokkink et al., 2018), criteria to assess the result of each single study on a psychometric property, a rating system summarising all results of studies on each psychometric property and grading quality of all evidence used for the assessments of both the methodological and the psychometric quality (Prinsen et al., 2018; Terwee et al., 2018). The most significant advantage of the COSMIN guidelines over other methods is that they were designed to assess the quality of *all* domains of psychometric properties comprehensively, while other methods were designed for evaluating limited aspects of psychometric properties only. Although the COSMIN guidelines are comprehensive, precise and balanced, it is complex and requires in depth knowledge of psychometrics and quality rating criteria for conducting systematic reviews of the psychometric properties of an instrument (Christian et al., 2019; Dobbs et al., 2019).

Two members of the authorship team (J.D. and B.B.) independently applied the COSMIN Risk of Bias checklist (Mokkink et al., 2018) to evaluate the methodological quality of studies conducted on the psychometric properties (content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, and construct validity) of each instrument (see Supporting Information Appendix B for COSMIN definitions of psychometric properties). When rating the methodological quality of the included studies on psychometric properties, each checklist item was ranged on a four-point rating scale: 4 = very good, 3 = adequate, 2 = doubtful, or 1 = inadequate (Mokkink et al., 2018). A total rating for each psychometric property was obtained by calculating the ratio between 'the obtained total score minus the minimum score possible' and 'the maximum score possible minus the minimum score possible'. This approach was adapted instead of a worst score counts method (i.e., reporting total ratings obtained by taking the lowest rating among any of the checklist items) recommended by COSMIN guidelines (Mokkink et al., 2018), as determining the total ratings entirely based on the lowest rating scale tends to impede the detection of subtle differences in methodological quality between studies (Speyer et al., 2014). Therefore, the total score of methodological quality ratings per psychometric property was presented as a percentage of the ratings; inadequate = 0%–25%, doubtful = 25.1%–50.0%, adequate = 50.1%–75%, and very good = 75.1%–100%. J.D. and B.B. rated the methodological quality independently. Any discrepancies were resolved by consensus. The inter reviewer agreement between J.D. and B.B. was determined by calculating the weighted κ (Cohen & Humphreys, 1968).

After evaluating the methodological quality of the included studies, the following data were extracted from the included studies and instruments (Prinsen et al., 2018): (1) study characteristics (country, participant characteristics, recall period); (2) instrument characteristics (instrument names, construct measured, number of items, sub-scales, target population and rating options); and (3) psychometric properties assessed (see Table 1). S.M. extracted all relevant data from included studies, and J.D. checked the extracted data for accuracy and completeness.

2.3 | Step 3: Evaluation of psychometric properties of instruments

The psychometric properties of instruments were assessed for each of the eight psychometric properties in three consecutive steps: Step 3.1 rating the result of single studies, Step 3.2 summarising the results of all studies per instrument, and Step 3.3 grading the quality of evidence on psychometric properties. All ratings were conducted by J.D. and B.B. independently where after consensus ratings were determined by discussion between both authors.

2.3.1 | Step 3.1: Rating the result of single studies

Rating the results of single studies was completed for each psychometric property separately. The results of each psychometric property (structural validity; internal consistency; cross-cultural validity\measurement invariance, reliability; measurement error; criterion validity; and hypotheses testing for construct validity) in each individual study were rated as sufficient (above the quality criteria threshold: +); insufficient (below the quality criteria threshold: –); or indeterminate (less robust data that do not meet the quality criteria:?) using the predefined criteria for good psychometric properties (Prinsen et al., 2018; see Supporting Information Appendix C1). The results of content validity in each individual study were rated against the ten criteria for good content validity (Terwee et al., 2018; see Supporting Information Appendix C2).

2.3.2 | Step 3.2: Summarising the results of all studies per instrument

All results on each psychometric property were qualitatively summarised into overall ratings of the psychometric property per instrument (Prinsen et al., 2018; Terwee et al., 2018). An overall sufficient (+), insufficient (−), inconsistent (±) or indeterminate (?) rating was given for each psychometric property per instrument, with a 75% agreement rule used (Mokkink et al., 2018). That is, for an overall sufficient (+) or insufficient (−) rating on a psychometric property, 75% or more of the studies reporting the psychometric property must be sufficient (+) or insufficient (−); otherwise for an overall inconsistent

TABLE 1 Characteristics of the included instruments.

Instrument (reference)	Country	Construct	Sub-scales (number of items)	Number of items	Response options	Recall period	Sample size	Study population	Sample age	Psychometric properties assessed
WellSEQ (Boström et al., 2016)	Sweden	Subjective mental health and ill health, peer relations, school and family environment	Mental health (6) Mental ill health (13) Peer relations and conflict (11) School environment (6) Family relations (6)	42	3-point Likert: yes; sometimes; no.	The past week	113	Adolescents with mild or moderate ID.	12–16	Content validity Internal consistency Reliability Construct validity
CHEQOL-25 (Brabcova et al., 2014)	Czech Republic	QoL	Interpersonal/ Social Consequences (5) Worries and Concerns (4) Intrapersonal/ Emotional Issues (4) Disclosure and Normality (9)	25	NR	3–6 months	250	Children and adolescents suffering from epilepsy: 20.8% (n = 52) of sample with mild ID (IQ < 70), and 79.2% (n = 198) without ID.	8–15	Content validity Structural validity Internal consistency Reliability Construct validity
AUQUEI (Cui et al., 2008, 2010)	China	HRQoL	School activities Health Social life Leisure activities Family life	23	4-point Pictorial Likert: not happy at all; not happy; happy; very happy.	NA	168	Young people with ID: 31.5% with mild ID (IQ of 55–70), and 68.5% with moderate ID (IQ of 40–54).	8–18	Content validity Structural validity Internal consistency
CAPE (Longo et al., 2014)	Spain	Participation	Formal activities (15) Informal activities (40)	55	Five questions assess each item: 1. Diversity of activity 2. Intensity (frequency) 3. With whom 4. Where 5. Enjoyment	28 days	398	199 Young people with CP: 21.6% (n = 43) with mild ID; 16.6% (n = 33) with moderate ID; and 42.7% (n = 85) with severe ID and 19.1% (n = 38) had no ID. CAPE was completed proxy by father or mother. 199 young people without CP.	8–18	Reliability Measurement error Construct validity

(Continues)

TABLE 1 (Continued)

Instrument (reference)	Country	Construct	Sub-scales (number of items)	Number of items	Response options	Recall period	Sample size	Study population	Sample age	Psychometric properties assessed
PSI-VS-ID (Maïano et al., 2009, Maïano et al., 2011)	France France	Physical self-concept	Global self-worth (2) Physical self-worth (2) Physical condition (2) Sports competence (2) Physical attractiveness (2) Physical strength (2)	12	6-point Likert: that's just like me; not at all; very little; some; enough; a lot; entirely.		248–342	Adolescents and young adults: 131 with mild ID (IQ between 50 and 70); 81 with moderate ID (IQ between 35 and 49); and 11 with severe ID (IQ between 20 and 34).	12–20	Content validity Structural validity Internal consistency Cross-cultural validity Reliability Criterion validity Construct validity

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; CP, cerebral palsy; HRQoL, health-related quality of life; ID, intellectual disability; IDD, intellectual and developmental disability; IQ, intelligence quotient; NR, not reported; PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; QoL, quality of life; WellSEQ, Well-being in Special Education Questionnaire.

(±) rating, less than 75% of studies showed the same rating; and for overall indeterminate (?) rating, all studies must be indeterminate (?). For content validity, the overall ratings will be sufficient (+), insufficient (−) or inconsistent (±). An indeterminate overall rating (?) is not possible because the reviewer's rating is always available, which will be + or − or ±. If there are no content validity studies, or only content validity studies of inadequate quality, and the instrument development is of inadequate quality, the rating of the reviewers will determine the overall ratings. Indeterminate (?) ratings for development or content validity studies will be ignored (Terwee et al., 2018).

2.3.3 | Step 3.3: Grading the quality of evidence on psychometric properties

The quality of the evidence for overall ratings on each psychometric property (structural validity; internal consistency; cross-cultural validity\measurement invariance, reliability; measurement error; criterion validity; and hypotheses testing for construct validity) of an instrument was graded as high, moderate, low or very low using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Prinsen et al., 2018; see Supporting Information Appendix E). The GRADE approach considers the initial quality of evidence used for overall ratings to be high, but the evidence quality is subsequently downgraded by one or more levels (to moderate, low, or very low) if there are serious (one level down: −1), very serious (two levels down: −2), or extremely serious (three levels down: −3) concerns. The following four factors were considered in determining the ratings: (a) risk of bias (limitations in the methodological quality of studies: Step 2), (b) inconsistency of results (unexplained heterogeneity in results of studies: Step 3.2), (c) indirectness of evidence (evidence from different populations than the targeted population in the review), and (d) imprecision (a low total number of samples included in the studies) (Mokkink et al., 2018). For example, for downgrading one level (from high to moderate), only one factor is allowed to have a serious concern (−1); for two levels (from high to low), either only one factor with a very serious concern (−2) or two factors with serious concerns (−1) is allowed; for three levels (from high to very low), one factor with an extremely serious concern (−3), one factor with very serious concern (−2), and one factor with serious (−1) to extremely serious concerns (−3), or more than three factors with serious (−1) to extremely serious concerns (−3) is allowed. Quality of evidence was not graded when the overall rating was indeterminate (?) as this indicates lack of robust evidence (Prinsen et al., 2018). Further details on grading quality of evidence can be found in the COSMIN manual for systematic reviews of instruments (Mokkink et al., 2018) and the GRADE Handbook (Schünemann et al., 2022). The quality of evidence for content validity was rated using the modified GRADE approach (Terwee et al., 2018 see Supporting Information Appendix F); high = high level of confidence; moderate = moderate level of confidence; low = low level of confidence; very low = very low level of confidence; NE = not evaluated (instruments could not be retrieved).

2.4 | Step 4: Selection of instruments

The selection of instruments and recommendation of suitable instruments for future use was based on combining overall rating results of each psychometric property and grading results of evidence quality for each property (Prinsen et al., 2018). Each instrument was classified into three recommendation categories (Mokkink et al., 2018): (A) most suitable (i.e., instruments with high-quality evidence for sufficient content validity—in any aspects of relevance, comprehensiveness, and comprehensibility—and at least low-quality evidence for sufficient internal consistency); (B) promising but need further validation studies (i.e., instruments categorised not in A or C); and (C) not recommendable (i.e., instruments with high quality evidence for an insufficient psychometric property). The review did not consider interpretability (the degree to which clinical meaning can be assigned to an instrument's quantitative scores or change in scores) and feasibility (ease of use such a length, completion time, and access fee of an instrument) to recommend the most suitable HRQoL and subjective wellbeing instruments because neither interpretability nor feasibility are considered psychometric properties (Prinsen et al., 2018).

3 | RESULTS

3.1 | Step 1: Systematic literature search

In accordance with the PRISMA guidelines (Page et al., 2021), Figure 1 presents a flow chart detailing the studies identified during the literature searching and study selection (Step 2). The electronic literature searches yielded 5134 studies. The identified studies from each database were exported into Covidence and duplicates were removed, reducing the number to 3802. S.M. and J.D. independently scanned the 3802 article titles and abstracts in Covidence. Titles that mentioned HRQoL and subjective wellbeing measurement and indicated that the study sample included children or adolescents with ID were retained. Abstract summaries were also read in the screening phase to provide background information and objectives of the study, eligibility criteria, data sources and participant information. After the initial screening phase and conflicts were resolved, 3771 of the 3802 studies were irrelevant. The inter reviewer agreement for title and abstract screening between S.M. and J.D. was good (Altman, 1991): Weighted $\kappa = 0.690$ (95% CI [−0.741, 2.121]). The remaining 31 relevant studies were assessed for eligibility by S.M. and J.D. reading the full texts. Due to a range of reasons (see Figure 1), 24 articles were discarded, reducing the number of included articles to six. The remaining six articles were read in full, with five self-report HRQoL and subjective wellbeing instruments identified. Reference checking of the included six full-text articles identified an additional study that met all inclusion criteria. In total seven studies reported psychometric properties on five self-report instruments used to measure HRQoL and/or subjective wellbeing of adolescents with ID. The inter reviewer agreement for full-text review between the S.M. and J.D. was good (Altman, 1991): Weighted $\kappa = 0.734$ (95% CI [0.452, 1.016]).

3.2 | Characteristics of included studies and instruments

This review identified five self-report instruments of HRQoL and subjective wellbeing within an ID adolescent population across seven studies; (1) Well-being in Special Education Questionnaire (WellSEQ; Boström et al., 2016), (2) Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy (CHEQOL-25; Brabcova et al., 2014), (3) Autoquestionnaire Enfant Image (AUQUEI; Cui et al., 2008, 2010), (4) Children's Assessment of Participation & Enjoyment (CAPE; Longo et al., 2014), (5) ID version of the Short Form of the Physical Self-Inventory (PSI-VS-ID; Maïano et al., 2009, 2011). Overall, the sample size of the studies ranged from 113 to 398 participants, and the age of participants ranged from 8 to 20 years. Studies came from Sweden, Czech Republic, China, Spain and France. All but two instruments (CHEQOL-25 and CAPE) were purposefully designed and evaluated for children, adolescents, or young adults with ID. Longo et al.'s (2014) study included a sample of children and adolescents with and without CP, and Brabcova et al.'s (2014) study included a sample of children and adolescents with epilepsy. Table 1 provides an overview of the characteristics of the five instruments.

3.3 | Step 2: Methodological quality of included studies

The methodological quality of the seven included studies was assessed using the COSMIN Risk of Bias checklist (Mokkink et al., 2018). An overview of the methodological quality ratings for each study are displayed in Table 2. Among the psychometric properties, internal consistency (6/7) was most frequently assessed across all studies. Over half of the studies included psychometric data on structural validity (5/7), construct validity (5/7), reliability (4/7) and content validity (4/7). Two studies reported psychometric data on measurement error, and only one study reported psychometric data on cross-cultural validity and criterion validity. No information was retrieved on responsiveness in any study. The inter reviewer agreement for quality assessment of the included studies between J.D. and B.B. was good: Weighted $\kappa = 0.793$ (95% CI [0.730, 0.856]). An appraisal summary of the methodological quality for each psychometric property per instrument are presented below:

3.3.1 | Well-being in Special Education Questionnaire

Boström et al. (2016) developed and evaluated a self-report questionnaire (WellSEQ) on subjective mental health in a sample of 113 young people (aged 12–18 years) with mild or moderate intellectual developmental disorder (IDD) in Sweden. The methodological quality was rated as: (1) 'very good' for internal consistency because adequate sample size and methods were employed (i.e., internal consistency for
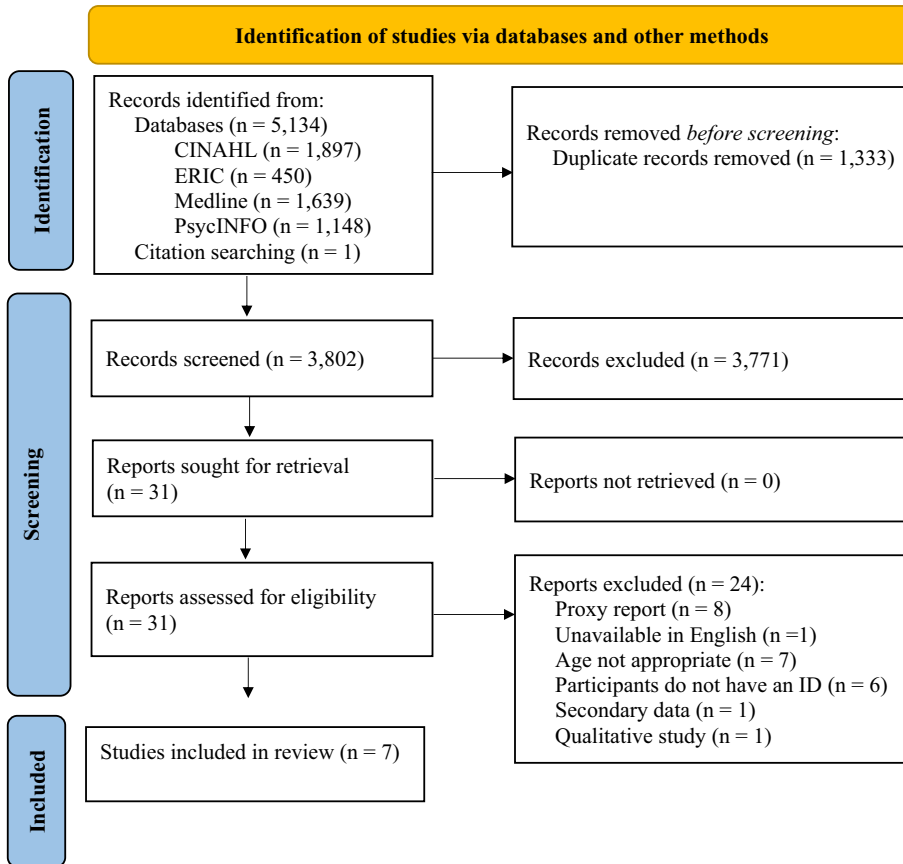
each subscales calculated); (2) 'adequate' for reliability within a 3-week time period; and, (3) 'doubtful' construct validity as the method used was not optimal.

3.3.2 | Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy

Brabcova et al. (2014) adapted and evaluated the psychometric properties (internal consistency, structural and construct validity) of a self-report QoL instrument (CHEQOL-25) with 250 children and adolescents (aged 8–15 years) with epilepsy (sample included 20.8% with a mild ID; IQ < 70) in the Czech Republic. The internal consistency was found to be 'very good'. The reliability was 'adequate' due to using a less preferred analysis method (i.e., EFA and Spearman's correlation), and construct validity was rated as 'doubtful' as a result of the method used not being optimal. It is important to note that only 20.8% (*n* = 52) of the sample in this study had ID (mild; IQ < 70). The data is not analysed or reported separately for those with or without ID.

3.3.3 | Autoquestionnaire enfant image

Cui et al. (2008) developed a self-report HRQoL instrument (AUQUEI) to assess subjective QoL in a population of Chinese children with

ID. Cui et al. (2010) adapted and investigated the psychometric properties of the AUQUEI in a sample of 168 school children (aged 8–18 years) with mild to moderate ID in China. In both papers, the internal consistency was found to be 'very good' and the structural validity was 'adequate' due to using a less preferred analysis method (i.e., EFA and Rasch analysis).

3.3.4 | Children's assessment of participation and enjoyment

The cross-cultural validity of the CAPE scale was evaluated by Longo et al. (2014) in a sample of 199 children and adolescents with CP (aged 8–18 years) who had either no (19.1%), mild (21.6%), moderate (16.6%) or severe (42.7%) ID. It is important to note that for the participants with severe ID, the CAPE was completed via proxy and the analysis is not reported separately for ID sub-groups. Reliability (rated as 'adequate') and construct validity (rated as 'doubtful') of the measure was explored across a 4-week period. Neither structural validity nor internal consistency of this measure is reported.

3.3.5 | The ID version of the short form of the physical self-inventory

The psychometric properties of the PSI-VS-ID were explored in two studies that met the criteria for the review (Maïano et al., 2009,

TABLE 2 Methodological quality assessment of studies on psychometric properties of the included instruments.

Study	Instrument	Psychometric property: Methodological quality per study							
		Content validity	Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity
Boström et al. (2016)	WellSEQ	AD (71.8%)	NR	DF (46.7%)	NA	AD (54.2%)	DF (44.4%)	NR	DF (33.3%)
Brabcova et al. (2014)	CHEQOL-25	VG (81.9%)	DF (30%)	AD (55.6%)	NA	AD (60%)	NR	NR	DF (42.9%)
Cui et al. (2008)	AUQUEI	VG (72.8%)	VG (88.9%)	VG (100%)	NA	NR	NR	NR	NR
Cui et al. (2010)	AUQUEI	NR	VG (88.9%)	VG (100%)	NA	NR	NR	NR	NR
Longo et al. (2014)	CAPE	NR	NR	NR	NA	AD (60%)	AD (73.3%)	NR	DF (42.9%)
Maïano et al. (2011)	PSI-VS-ID	NR	VG (100%)	AD (75%)	AD (75%)	NR	NR	NR	VG (80%)
Maïano et al. (2009)	PSI-VSF-ID	AD (62.3%)	VG (100%)	AD (75%)	NR	VG (100%)	NR	VG (100%)	VG (100%)

Note: The methodological quality was rated using the COSMIN Risk of Bias checklist (Mokkink et al., 2018): 4 = very good (VG), 3 = adequate (AD), 2 = doubtful (DF), and 1 = inadequate (IA). The overall methodological quality per study was presented as a percentage of the ratings (Cordier et al., 2015): inadequate (IA) = 0%–25%, doubtful (DF) = 25.1%–50.0%, adequate (AD) = 50.1%–75%, very good (VG) 75.1%–100%.

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; CP, cerebral palsy; NA, not applicable; NR, not reported (due to no psychometric data reported); PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.

2011). Maïano et al. (2009) tested the factor validity and reliability of PSI-VSF within a sample of 362 adolescents with mild to moderate ID in France. More recently, Maïano et al. (2011) investigated the robustness of the PSI-VSF psychometric properties in a new independent sample of 248 adolescents and young adults with ID in France. In both studies the structural validity, internal consistency and construct validity was found to be 'very good'. Maïano et al. (2009) evaluated the reliability and criterion validity which was rated as 'very good'. Maïano et al. (2011) assessed cross-cultural validity which was found to be 'very good'.

3.4 | Step 3: Psychometric properties and quality of evidence of the instruments

Tables 3 and 4 summarise ratings for each psychometric property of the included studies (Step 3.1). All data extracted from the seven included studies was evaluated against the criteria for good psychometric properties reported in each article (Prinsen et al., 2018; Terwee et al., 2018). A summary of the rating criteria is presented in detail in Supporting Information Appendices C.

Tables 5 and 6 present the overall ratings (Step 3.2) and the quality of evidence (Step 3.3) for each psychometric property per instrument. The results of all included studies on each psychometric property per instrument and their quality ratings are summarised in Supporting Information Appendices D. Grades for quality of evidence were reported in almost half of all overall ratings on psychometric quality for all five instruments, while all other quality of evidence was rated as NR due to no psychometric data reported or not evaluated. Only one instrument reported overall ratings for all six psychometric properties.

3.5 | Step 4: Recommendations for the most suitable instruments to measure HRQoL and subjective wellbeing

Table 7 provides recommendations for the use of self-report instruments to measure HRQoL and subjective wellbeing of adolescents with ID in the future. None of the instruments were rated as the most suitable (category A); one instrument (PSI-VS-ID) was considered the most promising but would require further validation studies; and four instruments (WELLSEQ, CHEQOL-25, AUQUEI, CAPE) were not recommendable.

4 | DISCUSSION

This systematic review summarised and critically evaluated the methodological quality and psychometric properties of self-report instruments used to measure HRQoL and subjective wellbeing of adolescents with ID using COSMIN methodology. Our intention was to provide an evidence-base that would help clinicians and

10 | WILEY—JARID _Journal of Applied Research in Intellectual Disabilities_ _Published for the British Institute of Learning Disabilities_ **bild**

MAGUIRE ET AL.

TABLE 3 Quality appraisal of the psychometric properties of each instrument.

Study	Instrument	Structural validity Rating	Internal consistency Rating	Reliability Rating	Measurement error Rating	Construct validity Rating	Cross-cultural validity Rating	Criterion validity Rating
Boström et al. (2016)	WellSEQ	NR	−	−	NR	+	?	NR
Brabcova et al. (2014)	CHEQOL-25	?	+	+	NR	−	?	NR
Cui et al. (2008)	AUQUEI	?	−	NR	NR	?	?	NR
Cui et al. (2010)	AUQUEI	−	+	NR	NR	?	?	NR
Longo et al. (2014)	CAPE	NR	NR	−	+	−	?	NR
Maïano et al. (2011)	PSI-VS-ID	+	+	NR	NR	+	+	NR
Maïano et al. (2009)	PSI-VSF-ID (Likert)	+	+	−	NR	+	?	+
	PSI-VSF-ID (Graphical)	−	+	+				

Note: Responsiveness was beyond the scope of this review. The psychometric properties was rated using the criteria for good psychometric properties (Prinsen et al., 2018); + = sufficient; ? = indeterminate; − = insufficient; ± = inconsistent (in case of rating one more results per psychometric property within a study, if <75% of ratings displayed the same scoring). Data and ratings on each psychometric property per study are available in Supporting Information S2.

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; CP, cerebral palsy; NA, not applicable; NR, not reported (due to no psychometric data reported); PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.

researchers when selecting instruments, based on the robust and comprehensive consensus-based COSMIN criteria. In total we identified five instruments (PSI-VS-ID, WELLSEQ, CHEQOL-25, AUQUEI, CAPE) which met the inclusion criteria for the review. The PSI-VS-ID (Maïano et al., 2009, 2011) appears to be the most psychometrically robust measure of subjective wellbeing for adolescents with ID, demonstrating good internal consistency, structural validity, construct validity, cross-cultural and criterion validity. However, the reliability across both studies was inconsistent with evidence based on a small sample size thus limiting the interpretation of this psychometric property, and the content validity was inconsistent given lack of reporting on all properties. Based on the results of this review, the evidence-base for self-report instruments used to measure HRQoL and subjective wellbeing of adolescents with ID is limited and as such, recommendations for use for each of the instruments are cautiously provided.

4.1 | Methodological quality of the included studies

For structural validity, two instruments (WellSEQ and CAPE) did not report any psychometric data and one instrument (CHEQOL-25) reported 'doubtful' study quality due to a less preferred factor analysis method being employed (EFA). As structural validity is to test a factor structure of existing instruments (Mokkink et al., 2018),

confirmatory factor analysis or item response theory analysis are the preferred methods in the COSMIN Risk of Bias checklist (Mokkink et al., 2018). One instrument (WellSEQ) reported on all three psychometric properties within the domain of reliability (Mokkink et al., 2010a, 2010b). Four instruments (WellSEQ, CHEQOL-25, CAPE, PSI-VS-ID) reported reliability, four instruments (WellSEQ, CHEQOL-25, AUQUEI, PSI-VS-ID) reported internal consistency, and two instruments (WellSEQ and CAPE) reported measurement error. Given the lack of reporting on all three psychometric properties makes it difficult to determine overall reliability for all instruments comprehensibly. One instrument (PSI-VS-ID) reported criterion validity between adapted (Likert/Graphical) and original (long) versions. Cross-cultural validity for different demographic groups was reported for one instrument (PSI-VS-ID), with a very good score for study quality achieved. Hypothesis testing for construct validity was reported for four instruments (WellSEQ, CHEQOL-25, CAPE, PSI-VS-ID) with ratings of 'very good' (PSI-VS-ID) and 'doubtful' (WellSEQ, CHEQOL-25, CAPE). Likewise, given the lack of reporting on all three content validity properties (i.e., relevance, comprehensiveness, and comprehensibility) makes it difficult to determine overall content validity for all instruments. Of the four studies reporting content validity studies, all overlapped with the development studies and conducted either item generation or cognitive interviewing. Two instruments (CHEQOL-25 and AUQUEI) reported 'very good' rating, and two instruments reported 'adequate' ratings (WELLSEQ and PSI-VS-ID). Few content validity studies asked parents, caregivers or professionals

TABLE 4 Quality of content validity per development and content validity study, and content of instrument.

Instrument	Reference	Relevance Development study	Relevance Content validity study	Relevance Content of instrument	Comprehensiveness Development study	Comprehensiveness Content validity study	Comprehensiveness Content of instrument	Comprehensibility Development study	Comprehensibility Content validity study	Comprehensibility Content of instrument
WellSEQ	Boström et al. (2016)	+	?	+	−	?	+	?	?	+
CHEQOL-25	Brabcova et al. (2014)	+	?	?	−	?	?	−	?	?
AUQUEI	Cui et al. (2010) Cui et al. (2008)	+	−	+	+	−	+	+	−	+
CAPE	Longo et al. (2014)	−	?	?	−	?	?	−	?	?
PSI-VSF-ID	Maïano et al. (2009) Maïano et al. (2011)	+	?	?	−	?	?	+	?	?

Note: The quality of content validity (relevance, comprehensiveness, and comprehensibility) per study and content of instrument was rated using the criteria for good content validity (Terwee et al., 2018); + = sufficient rating; ? = indeterminate rating; − = insufficient rating; ± = inconsistent rating. Rating for development and content validity studies was determined based on the data from development and content validity studies; rating for content of instrument was determined based on reviewers' subjective opinion on content of instrument itself (items and instructions).

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.

about relevance, comprehensiveness, and comprehensibility (i.e., how easy it is for respondents to understand instrument items) of the instruments as respondents lacked the required detail when reporting on the methodology (e.g., insufficient reporting on study design and results).

4.2 | Psychometric properties and quality of evidence of the instruments

The evidence on structural validity is a prerequisite for interpreting the evidence on internal consistency (Yoon et al., 2021). For example, if results on structural validity demonstrate that a scale has three factors, internal consistency of each of those three subscales is more relevant than that of the total scale. Of the four instruments reporting evidence on internal consistency, two instruments (CHEQOL-25 and PSI-VS-ID) displayed sufficient internal consistency, CHEQOL-25 with moderate evidence (due to 20.8% of the sample with ID) for sufficient validity and high Cronbach's alpha values, and PSI-VS-ID with high evidence (due to good study quality and consistent results) for sufficient structural validity and moderate Cronbach's alpha.

Two instruments (WellSEQ and CAPE) did not report any data on structural validity, one instrument (CHEQOL-25) reported indeterminate structural validity due to a less preferred factor analysis method being employed (EFA), and one instrument (AUQUEI) reported inconsistent results on the factor structure between studies. One instrument (PSI-VS-ID) reported high evidence for sufficient structural validity.

Of the four instruments reporting evidence on reliability (test–retest and interrater), two instruments (WellSEQ and CAPE) gained insufficient overall ratings due to reporting of other reliability statistics (WellSEQ) than the preferred reliability statistics (i.e., the intraclass correlation coefficient) in the COSMIN criteria for good psychometric properties (Prinsen et al., 2018). Although one instrument (CAPE) reported ICC, reliability was rated as insufficient due to the ICC below the criterion for good reliability. One instrument (CHEQOL-25) displayed sufficient reliability but with moderate evidence as a result of some evidence from a different population (i.e., adolescents without ID), and one instrument (PSI-VS-ID) reported inconsistent results on reliability between studies.

Evidence on the criterion validity of the PSI-VS-ID was sufficient because the correlation between the adapted (Likert/Graphical) and original (long) versions was over 0.70, which is the criterion for good criterion validity (Cronbach, 1951). In addition, evidence for cross-cultural validity was evaluated for the PSI-VS-ID with a sufficient rating overall.

Evidence on hypothesis testing for construct validity was evaluated for all instruments except AUQUEI. Two instruments (WellSEQ and PSI-VS-ID) reported sufficient hypothesis testing with moderate (WellSEQ) and high (PSI-VS-ID) evidence. Two instruments (CHEQOL-25 and CAPE) reported insufficient hypothesis testing with high (CHEQOL-25) and low (CAPE) evidence. For evaluation of hypothesis testing, two instruments (CHEQOL-25 and CAPE)

12 | **WILEY**_JARID_ bild
Journal of Applied Research in Intellectual Disabilities _Published for the British Institute of Learning Disabilities_

MAGUIRE ET AL.

TABLE 5 Overall quality of content validity and evidence quality per instrument.

	Relevance		Comprehensiveness		Comprehensibility	
Instrument	Overall quality of content validity[a]	Quality of evidence[b]	Overall quality of content validity[a]	Quality of evidence[b]	Overall quality of content validity[a]	Quality of evidence[b]
WellSEQ	+	Moderate	−	Moderate	+	Moderate
CHEQOL-25	+	Moderate	−	Low	−	Low
AUQUEI	−	Low	−	Low	+	Moderate
CAPE	?	NE	?	NE	?	NE
PSI-VSF-ID	+	Moderate	−	Low	+	Moderate

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; NE, not evaluated (instruments could not be retrieved); PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.
[a]The quality of content validity (relevance, comprehensiveness, and comprehensibility) per study and content of instrument was rated using the criteria for good content validity (Terwee et al., 2018): ? = indeterminate rating + = sufficient rating; − = insufficient rating; ± = inconsistent rating.
[b]The quality of evidence for content validity was rated using the modified GRADE approach (Terwee et al., 2018); high = high = level of confidence; moderate = moderate level of confidence; low = low level of confidence; very low = very low level of confidence.

presented only a _t_ value or _F_ value to confirm the statistical significance of the differences in scores between groups. As these two tests depend on sample size and do not account for direction or magnitude of difference, to address this weakness this review converted the _t_ value or _F_ value to an effect size estimate (i.e., Cohen's _d_) showing the direction and magnitude of differences between groups regardless of sample sizes (Thalheimer & Cook, 2002). An effect size of 0.5 or higher was used as criterion for sufficient hypothesis testing on group differences (Yoon et al., 2021).

Given that content validity is the first psychometric property to consider when selecting an instrument, the inadequate quality of evidence on content validity makes it difficult to select the best instrument (Terwee et al., 2018). All ratings on relevance, comprehensiveness, and comprehensibility based on the development and content validity studies were categorised as inconsistent. Due to these inconsistent study ratings, most overall ratings on relevance, comprehensiveness, and comprehensibility were determined based on reviewers' subjective opinion about the content of instrument itself only. The findings indicate lack of evidence on content validity or inappropriate methodological approaches used for instrument development and content validity studies (Terwee et al., 2018). Due to the inappropriate methodological approaches used when developing new instruments and assessing content validity of the instruments, evidence on the quality of relevance, comprehensiveness, and comprehensibility was low or moderate. Therefore, findings from this review indicate that evidence on the quality of content validity is very uncertain.

4.3 | Recommendation of the instruments

COSMIN based standards advise that in order to recommend an instrument, it should demonstrate any level of content validity and at least low-level evidence for sufficient internal consistency (Prinsen et al., 2018). None of the instruments included in this review met this criterion, therefore we are unable to recommend any of these

instruments for use. One instrument (PSI-VS-ID) was considered to have potential to be recommended for use but would require further validation studies to assess the quality of the psychometric properties and determine whether it could be recommendable (i.e., category A). As a criterion for category A, content validity should be further evaluated as a priority. While four instruments (WELLSEQ, CHEQOL-25, AUQUEI, and CAPE) should not be recommended at all (i.e., category C) due to high quality evidence for an insufficient measurement. Additional studies may change the assessment of the four instruments from not recommended (category C) to promising (category B).

Future research studies in this area should make use of the COSMIN based standards for designing and reporting validation research to ensure that the appropriate evidence-base is acquired for an instrument to be recommended for use. None of the articles in this review assessed the responsiveness of the instruments and future research should seek to ensure that these instruments are suitably responsive to assess HRQoL and subjective wellbeing of adolescents with ID. Furthermore, the design and evaluation of a new standardised instrument for use with this target group would be advantageous, enabling the conduct of high-quality research.

4.4 | Strengths and limitations

Strengths of this review include the use of a robust, thorough and consensus-based methodology and search filters for findings and reviewing the evidence for psychometric properties of instruments. As strongly recommended by the COSMIN guidelines, this review was conducted by a multidisciplinary team consisting of psychologists with relevant expertise and required knowledge to assess the quality of measurement instruments (Mokkink et al., 2018). Limitations include the inherent difficulty in developing self-report measures for individuals with ID; there is little research on self-reported HRQoL and subjective wellbeing experiences of adolescents with ID and further research into this would be beneficial. One article included 42.7% of

TABLE 6 Overall quality of psychometric properties and evidence quality per instrument.

Psychometric properties and quality of evidence per instrument

Instrument	Content validity		Structural validity		Internal consistency		Reliability		Measurement error		Hypothesis testing for construct validity		Cross-cultural validity		Criterion validity	
	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence	Overall Rating	Quality of Evidence
WellSEQ	±	Moderate	NR	NR	−	High	−	High	NR	NR	+	Moderate	?	NR	NR	NR
CHEQOL-25	±	Moderate	?	Low	+	Moderate	+	Moderate	NR	NR	−	High	?	NR	NR	NR
AUQUEI	±	Low	±	Moderate	±	Moderate	NR	NR	NR	NR	?	NR	?	NR	NR	NR
CAPE	−	NE	NR	NR	NR	NR	−	High	+		−	Low	?	NR	NR	NR
PSI-VS-ID	±	Low	+	High	+	High	±	Moderate	NR	NR	+	High	+	High	+	High

Note: The overall quality of psychometric properties (apart from the content validity) was rated using the criteria for good psychometric properties (Mokkink et al., 2018); + = sufficient rating; ? = indeterminate rating (due to less robust psychometric data); − = insufficient rating; ± = inconsistent rating; NR = not reported (due to no psychometric data). The overall quality of content validity (relevance, comprehensiveness and comprehensibility) was determined by qualitatively summarising all ratings on content validity per study of each instrument (Table 5) and reviewers' subjective ratings on content of instrument itself (Terwee et al., 2018); + = sufficient rating; − = insufficient rating; ± = inconsistent rating. The quality of evidence (confidence level for the overall quality rating of each psychometric property apart from content validity) was rated using a modified GRADE approach (Mokkink et al., 2018): high = high level of confidence; moderate = moderate level of confidence; low = low level of confidence; very low = very low level of confidence. The quality of evidence for content validity was rated using the modified GRADE approach (Terwee et al., 2018); high = high level of confidence; moderate = moderate level of confidence; low = low level of confidence; very low = very low level of confidence.

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; NE, not evaluated (instruments could not be retrieved); NR, not reported (due to no psychometric data reported); PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.

TABLE 7 Recommendations on suitable instruments for their future use adapted from Prinsen et al. (2018).

Category	Description on category (criteria)	Instruments
A: Most suitable	Instruments with high-quality evidence for sufficient content validity (any level) AND at least low-level quality evidence for sufficient internal consistency. Instruments can be recommended for use and results obtained with these instruments can be trusted.	None
B: Promising but need further validation study	Instruments have potential to be recommended for use, but they require further research to assess the quality of these (instrument categorised not in A or C).	PSI-VS-ID
C: Not recommendable	Instruments with high quality evidence for an insufficient measurement property and should not be recommended for use.	WELLSEQ CHEQOL-25 AUQUEI CAPE

Abbreviations: AUQUEI, Autoquestionnaire Enfant Image; CAPE, Children's Assessment of Participation & Enjoyment; CHEQOL-25, Modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy; PSI-VS-ID, the intellectual disability version of the short form of the physical self-inventory; WellSEQ, Well-being in Special Education Questionnaire.

data gathered via proxy (Longo et al., 2014) and only 20.8% of the sample had ID in another article (Brabcova et al., 2014), which may have confounded the results of the psychometric properties assessed. Information published in languages other than English were not included; therefore, some findings on psychometric properties of HRQoL and subjective wellbeing instruments published in other languages may have been overlooked. Finally, interpretability and feasibility were outside the scope of this article because they are not considered to be psychometric properties according to the COSMIN taxonomy, even though these two instrument characteristics should be considered when recommending the most suitable instruments (Mokkink et al., 2018; Prinsen et al., 2018).

5 | CONCLUSIONS

This review evaluated the psychometric properties of five self-report HRQoL and subjective wellbeing instruments using the COSMIN based standards. Evidence concerning psychometric property was limited and mostly of lower quality. Based on the available psychometric evidence, this review illustrates that currently there is no self-report instrument that can be recommended for use for the purpose of assessing HRQoL and subjective wellbeing of adolescents with ID. Only one instrument (PSI-VS-ID) was recommended as promising but requires further validation before any possible recommendations as most suitable instrument

may be made. Therefore, further validation work should focus on ensuring self-report instruments used to assess HRQoL and subjective wellbeing of adolescents with ID are sufficiently valid and reliable. Future research studies in this area are warranted to develop, evaluate and implement a new instrument designed to specifically assess HRQoL and subjective wellbeing. Studies should also make use of COSMIN based standards for designing and reporting validation research to ensure that the appropriate evidence-base is acquired for an instrument to be recommended.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Allerton, L. A., Welch, V., & Emerson, E. (2011). Health inequalities experienced by children and young people with intellectual disabilities: A review of literature from the United Kingdom. *Journal of Intellectual Disabilities*, 15(4), 269–278. https://doi.org/10.1177/1744629511430772

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

Boström, P., Åsberg, J., Thorson, M., & Broberg, M. (2016). Subjective mental health, peer relations, family, and school environment in adolescents with intellectual developmental disorder: A first report of a new questionnaire administered on tablet PCs. *Journal of Mental Health Research in Intellectual Disabilities*, 9(4), 207–231. https://doi.org/10.1080/19315864.2016.1186254

Bowling, A. (2011). Do older and younger people differ in their reported well-being? A national survey of adults in Britain. *Family Practice*, 28(2), 145–155. https://doi.org/10.1093/fampra/cmq082

Brabcova, D., Krsek, P., Kohout, J., & Zarubova, J. (2014). Psychometric properties of the modified Czech version of the children self-report Quality-of-life Measure for Children with Epilepsy (CHEQOL-25). *Epilepsy & Behavior*, 33, 31–38. https://doi.org/10.1016/j.yebeh.2014.02.013

Buckley, N., Glasson, E. J., Chen, W., Epstein, A., Leonard, H., Skoss, R., Jacoby, P., Blackmore, A. M., Srinivasjois, R., Bourke, J., Sanders, R. J., & Downs, J. (2020). Prevalence estimates of mental health problems in children and adolescents with intellectual disability: A systematic review and meta-analysis. *The Australian and New Zealand Journal of Psychiatry*, 54(10), 970–984. https://doi.org/10.1177/0004867420924101

Carlon, S., Shields, N., Yong, K., Gilmore, R., Sakzewski, L., & Boyd, R. (2010). A systematic review of the psychometric properties of quality-of-life measures for school aged children with cerebral palsy. *BMC Pediatrics*, 10, 81. https://doi.org/10.1186/1471-2431-10-81

Christian, B., Armstrong, R., Calache, H., Carpenter, L., Gibbs, L., & Gussy, M. (2019). A systematic review to assess the methodological quality of studies on measurement properties for caries risk assessment tools for young children. *International Journal of Paediatric Dentistry*, 29(2), 106–116. https://doi.org/10.1111/ipd.12446

Cohen, J., & Humphreys, L. H. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. https://doi.org/10.1037/h0026256

Cordier, R., Speyer, R., Chen, Y. W., Wilkes-Gillan, S., Brown, T., Bourke-Taylor, H., Doma, K., & Leicht, A. (2015). Evaluating the psychometric quality of social skills measures: A systematic review. *PloS one*, *10*(7), e0132299. https://doi.org/10.1371/journal.pone.0132299

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Cui, Y., Stapleton, F., Suttle, C., & Bundy, A. (2010). Health- and vision-related quality of life in intellectually disabled children. *Optometry and Vision Science*, *87*(1), 37–44. https://doi.org/10.1097/OPX.0b013e3181c1d533

Cui, Y., Stapleton, F., & Suttle, C. (2008). Developing an instrument to assess vision-related and subjective quality of life in children with intellectual disability: Data collection and preliminary analysis in a Chinese population. *Ophthalmic & Physiological Optics*, *28*(3), 238–246. https://doi.org/10.1111/j.1475-1313.2008.00564.x

Das, K. V., Jones-Harrell, C., Fan, Y., Ramaswami, A., Orlove, B., & Botchwey, N. (2020). Understanding subjective well-being: Perspectives from psychology and public health. *Public Health Reviews*, *41*(1), 25. https://doi.org/10.1186/s40985-020-00142-5

Davidson, G., Irvine, R., Corman, M., Kee, F., Kelly, B., Leavey, G., & McNamee, C. (2017). *Measuring the quality of life of people with disabilities and their families: Scoping study final report*. Department for Communities https://pureadmin.qub.ac.uk/ws/portalfiles/portal/164354073/Quality_of_Life_Scoping_study_final.PDF

Davison, J., Maguire, S., McLaughlin, M., & Simms, V. (2022). Involving children with intellectual disability in the adaptation of self-reported health and wellbeing measures: Participatory research and methodological considerations. *Journal of Intellectual Disability Research*, *66*(7), 628–641. https://doi.org/10.1111/jir.12936

Delle Fave, A., Brdar, I., Freire, T., Vella-Brodrick, D., & Wissing, M. P. (2011). The eudaimonic and hedonic components of happiness: Qualitative and quantitative findings. *Social Indicators Research*, *100*(2), 185–207. https://doi.org/10.1007/s11205-010-9632-5

Diener, E. (Ed.). (2009). *The science of well-being: The collected works of Ed Diener*. Springer Science + Business Media. https://doi.org/10.1007/978-90-481-2350-6

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34–43. https://doi.org/10.1037/0003-066X.55.1.34

Dobbs, T. D., Gibson, J. A. G., Hughes, S., Thind, A., Patel, B., Hutchings, H. A., & Whitaker, I. S. (2019). Patient-reported outcome measures for soft-tissue facial reconstruction: A systematic review and evaluation of the quality of their measurement properties. *Plastic and Reconstructive Surgery*, *143*(1), 255–268. https://doi.org/10.1097/prs.0000000000005112

Dodge, R., Daly, A. P., Huyton, J., & Sanders, L. D. (2012). The challenge of defining wellbeing. *International Journal of Wellbeing*, *2*(3), 222–235. https://doi.org/10.5502/ijw.v2i3.4

Durand, M. (2015). The OECD better life initiative: How's life? And the measurement of well-being. *Review of Income and Wealth*, *61*(1), 4–17. https://doi.org/10.1111/roiw.12156

Emerson, E. (2021). *Inequalities and inequities in the health of people with intellectual disabilities*. Oxford Research Encyclopedia of Global Public Health https://doi.org/10.1093/acrefore/9780190632366.013.326

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

Forgeard, M. J. C., Jayawickreme, E., Kern, M., & Seligman, M. E. P. (2011). Doing the right thing: Measuring wellbeing for public policy. *International Journal of Wellbeing*, *1*(1), 79–106. https://doi.org/10.5502/ijw.v1i1.15

Frey, G. C., Temple, V. A., & Stanish, H. I. (2017). Interventions to promote physical activity for youth with intellectual disabilities. *Salud Publica de Mexico*, *59*(4), 437–445. https://doi.org/10.21149/8203

Frey, G. C., Stanish, H. I., & Temple, V. A. (2008). Physical activity of youth with intellectual disability: Review and research agenda. *Adapted Physical Activity Quarterly*, *25*(2), 95–117. https://doi.org/10.1123/apaq.25.2.95

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, *118*(8), 622–629. https://doi.org/10.7326/0003-4819-118-8-199304150-00009

Hamdani, Y., Yee, T., Rowland, E., & McPherson, A. C. (2018). Examining wellness in children and youth with intellectual and developmental disabilities: A scoping review. *Current Developmental Disorders Reports*, *5*, 165–196. https://doi.org/10.1007/s40474-018-0146-1

Hays, R. D., & Fayers, P. M. (2021). Overlap of depressive symptoms with health-related quality-of-life measures. *PharmacoEconomics*, *39*(6), 627–630. https://doi.org/10.1007/s40273-020-00972-w

Hays, R. D., & Reeve, B. (2010). Measurement and modeling of health-related quality of life. In J. Killewo, H. K. Heggenhougen, & S. R. Quah (Eds.), *Epidemiology and demography in public health* (pp. 195–205). Academic Press.

Hughes-McCormack, L. A., Rydzewska, E., Cooper, S.-A., Fleming, M., Mackay, D., Dunn, K., Ward, L., Sosenko, F., Barlow, F., Miller, J., Symonds, J. D., Jani, B. D., Truesdale, M., Cairns, D., Pell, J., Henderson, A., & Melville, C. (2022). Rates, causes and predictors of all-cause and avoidable mortality in 163 686 children and young people with and without intellectual disabilities: A record linkage national cohort study. *BMJ Open*, *12*(9), 1–14. https://doi.org/10.1136/bmjopen-2022-061636

Ingerski, L. M., Modi, A. C., Hood, K. K., Pai, A. L., Zeller, M., Piazza-Waggoner, C., Driscoll, K. A., Rothenberg, M. E., Franciosi, J., & Hommel, K. A. (2010). Health-related quality of life across pediatric chronic conditions. *The Journal of Pediatrics*, *156*(4), 639–644. https://doi.org/10.1016/j.jpeds.2009.11.008

Jardine, J., Glinianaia, S. V., McConachie, H., Embleton, N. D., & Rankin, J. (2014). Self-reported quality of life of young children with conditions from early infancy: A systematic review. *Pediatrics*, *134*(4), 1129–1148. https://doi.org/10.1542/peds.2014-0352

Karanicolas, P. J., Bhandari, M., Kreder, H., Moroni, A., Richardson, M., Walter, S. D., Norman, G. R., Guyatt, G. H., & Collaboration for Outcome Assessment in Surgical Trials (COAST) Musculoskeletal Group. (2009). Evaluating agreement: Conducting a reliability study. *Journal of Bone and Joint Surgery*, *91*(3), 99–106. https://doi.org/10.2106/JBJS.H.01624

Kipfer, S., & Pihet, S. (2020). Reliability, validity and relevance of needs assessment instruments for informal dementia caregivers: A psychometric systematic review. *JBI Evidence Synthesis*, *18*(4), 704–742. https://doi.org/10.11124/JBISRIR-2017-003976

Longo, E., Badia, M., Orgaz, B., & Verdugo, M. A. (2014). Cross-cultural validation of the Children's Assessment of Participation and Enjoyment (CAPE) in Spain. *Child: Care, Health and Development*, *40*(2), 231–241. https://doi.org/10.1111/cch.12012

Maguire, S., Davison, J., McLaughlin, M., & Simms, V. (2022). Protocol for a systematic review exploring the psychometric properties of self-report health-related quality of life and subjective wellbeing measures used by adolescents with intellectual disabilities. *Systematic Reviews*, *11*(1), 81. https://doi.org/10.1186/s13643-022-01957-w

Maïano, C., Morin, A. J., Bégarie, J., & Ninot, G. (2011). The intellectual disability version of the very short form of the physical self-inventory (PSI-VS-ID): Cross-validation and measurement invariance across gender, weight, age and intellectual disability level. *Research in Developmental Disabilities*, *32*(5), 1652–1662. https://doi.org/10.1016/j.ridd.2011.02.019

Maïano, C., Bégarie, J., Morin, A. J., & Ninot, G. (2009). Assessment of physical self-concept in adolescents with intellectual disability: Content and factor validity of the very short form of the physical self-inventory. *Journal of Autism and Developmental Disorders*, *39*(5), 775–787. https://doi.org/10.1007/s10803-008-0686-z

Marques, A., Peralta, M., Santos, T., Martins, J., & Gaspar de Matos, M. (2019). Self-rated health and health-related quality of life are related with adolescents' healthy lifestyle. *Public Health*, *170*, 89–94. https://doi.org/10.1016/j.puhe.2019.02.022

Michaelson, J., Mahony, S., & Schifferes, J. (2012). *Measuring wellbeing: A guide for practitioners*. New Economics Foundation.

Mason, S. J., Catto, J., Downing, A., Bottomley, S. E., Glaser, A. W., & Wright, P. (2018). Evaluating patient-reported outcome measures (PROMs) for bladder cancer: A systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. *BJU International*, *122*(5), 760–773. https://doi.org/10.1111/bju.14368

Mokkink, L. B., de Vet, H., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010b.02.006

Morris, C., Janssens, A., Shilling, V., Allard, A., Fellows, A., Tomlinson, R., William, J., Thompson Coon, J., Rogers, M., Beresford, B., Green, C., Jenkinson, C., Tennant, A., & Logan, S. (2015). Meaningful health outcomes for paediatric neurodisability: Stakeholder prioritisation and appropriateness of patient reported outcome measures. *Health and Quality of Life Outcomes*, *13*(87), 1–9. https://doi.org/10.1186/s12955-015-0284-7

Morrow, A. M., Hayen, A., Quine, S., Scheinberg, A., & Craig, J. C. (2011). A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child: Care, Health and Development*, *38*(2), 186–195. https://doi.org/10.1111/j.1365-2214.2011.01240.x

Noonan, R. J., Boddy, L. M., Fairclough, S. J., & Knowles, Z. R. (2016). Write, draw, show, and tell: A child-centred dual methodology to explore perceptions of out-of-school physical activity. *BMC Public Health*, *16*(326), 1–19. https://doi.org/10.1186/s12889-016-3005-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*, 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., Auquier, P., Power, M., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J., & KIDSCREEN Group, E. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents.

Expert Review of Pharmacoeconomics & Outcomes Research, *5*, 353–364. https://doi.org/10.1586/14737167.5.3.353

Selwyn, J., & Wood, M. (2015). *Measuring well-being: A literature review*. University of Bristol Retrieved March 15, 2022, from https://research-information.bris.ac.uk/ws/portalfiles/portal/41278115/Measuring_Wellbeing_FINAL.pdf

Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury*, *42*(3), 236–240. https://doi.org/10.1016/j.injury.2010.11.042

Schünemann, H. J., Higgins, J. P. T., Vist, G. E., Glasziou, P., Akl, E. A., Skoetz, N., & Guyatt, G. H. (2022). Completing 'summary of findings' tables and grading the certainty of the evidence. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.3*. Cochrane Retrieved from www.training.cochrane.org/handbook

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Twenty questions and answers regarding the 12th edition of the AAIDD manual: Intellectual disability: definition, diagnosis, classification, and systems of supports*. American Association on Intellectual and Developmental Disabilities Retrieved October 28, 2022, from https://www.aaidd.org/docs/default-source/intellectualdisability/12th-ed-twenty-questions-faq.pdf?sfvrsn=a6403421_4

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H., Coulter, D. L., Craig, E. M., Shogren, K. A., Snell, M. E., Spreat, S., Tassé, M. J., Thompson, J. R., Verdugo-Alonso, M. A., Wehmeyer, M. L., & Yeager, M. H. (2010). *Intellectual disability: Definition, classification, and systems of supports*. American Association on Intellectual and Developmental Disabilities.

Smith, G., Fleming, M., Kinnear, D., Henderson, A., Pell, J. P., Melville, C., & Cooper, S. A. (2020). Rates and causes of mortality among children and young people with and without intellectual disabilities in Scotland: A record linkage cohort study of school children. *BMJ Open*, *10*, e034077. https://doi.org/10.1136/bmjopen-2019-034077

Solans, M., Pane, S., Estrada, M. D., Serra-Sutton, V., Berra, S., Herdman, M., Alonso, J., & Rajmil, L. (2008). Health-related quality of life measurement in children and adolescents: A systematic review of generic and disease-specific instruments. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *11*(4), 742–764. https://doi.org/10.1111/j.1524-4733.2007.00293.x

Speyer, R., Cordier, R., Kertscher, B., & Heijnen, B. J. (2014). Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: A systematic literature review. *BioMed Research International*, *2014*, 458678. https://doi.org/10.1155/2014/458678

Statham, J., & Chase, E. (2010). *Childhood wellbeing: A brief overview*. Childhood Wellbeing Research Centre.

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Boute, L. M., de Ve, H. C. W., & Mokkin, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0

Testa, M. A., & Simonson, D. C. (1996). Assessment of quality-of-life outcomes. *The New England Journal of Medicine*, *334*(13), 835–840. https://doi.org/10.1056/NEJM199603283341306

Thalheimer, W., & Cook, S. (2002). *How to calculate effect sizes from published research: A simplified methodology*. Work-Learning Research http://www.work-learning.com

Upton, P., Lawford, J., & Eiser, C. (2008). Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *17*, 895–913. https://doi.org/10.1007/s11136-008-9350-5

Upton, D., & Upton, P. (2015). Quality of life and well-being. In *Psychology of wounds and wound care in clinical practice*. Springer https://doi.org/10.1007/978-3-319-09653-7

World Health Organisation (WHO). (2022, October 22). *Adolescent health*. https://www.who.int/health-topics/adolescent-health

World Health Organization (WHO). (2018). *ACTIVE. A technical package for increasing physical activity*. Licence: CC BY-NC-SA 3.0 IGO WHO, Global Action Plan for physical activity 2018–2030. Resolution WHA71.6. Geneva. World Health Organization.

World Health Organization (WHO). (1948). *Constitution of the World Health Organization*. World Health Organization http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf

Yoon, S., Speyer, R., Cordier, R., Aunio, P., & Hakkarainen, A. (2021). A systematic review evaluating psychometric properties of parent or caregiver report instruments on child maltreatment: Part 2: Internal consistency, reliability, measurement error, structural validity, hypothesis testing, cross-cultural validity, and criterion validity. *Trauma, Violence & Abuse*, *22*(5), 1296–1315. https://doi.org/10.1177/1524838020915591

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.