

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/162509>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

**The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review.**

PRE-PUBLICATION VERSION (ACCEPTED MANUSCRIPT) OF THE FOLLOWING ARTICLE:

Kooijmans, R., Mercera, G., Langdon, P.E., & Moonen, X. (2021). The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clinical Psychology: Science and Practice*.

This article may not exactly replicate the authoritative document published in the APA journal. It is not the copy of record.

**Abstract**

*Background*

Persons with intellectual disabilities (ID) may have difficulties providing reliable and valid accounts of their personal experiences through self-report measures. The aim of the current study was to systematically review the peer-reviewed research literature on the adaptations needed to develop 'ID-inclusive' self-report measures.

*Method*

A search of PsycINFO, PubMed, Web of Science and Google Scholar identified 49 studies that met inclusion criteria. A GRADE-CERQual assessment was performed to determine the level of confidence in the review findings.

*Results*

161 recommendations for the development of 'ID-inclusive' self-report measures were extracted from 49 included studies. Recommendations were presented in a GRADE-CERQual Summary of Findings table, according to a 5-stage model of instrument development.

### *Conclusions*

This review offers much-needed practical guidance for clinicians and researchers on how to develop 'ID-inclusive' self-report measures. Recommendations for future research about self-report instrument development for use with people with ID are presented.

*Keywords: intellectual disabilities, self-report, questionnaires, interviews, psychometrics*

*Public Health Statement: it is important to include the personal views of persons with intellectual disabilities in clinical practice. However, guidance on how to attune self-report measures to the needs of persons with ID is lacking. We conducted a systematic review to compile evidence-based recommendations on how to develop 'ID-inclusive' self-report measures.*

*Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.*

## **Introduction**

An intellectual disability (ID) is associated with a range of challenges including problems with reasoning, verbal expression, reading, abstract thinking and judgment (American Psychiatric Association, 2013; Schalock et al., 2010). These challenges can interfere with inclusive participation in society and make it more difficult for people with ID to voice their opinions, feelings and thoughts. In clinical work and research, assessing the perspectives of people with ID on a variety of topics is vital and routinely undertaken. For example, within diagnostic procedures, support needs assessment, routine outcome

monitoring, and studies on the efficacy of interventions, either self-report or proxy-rated measures are used as the source of information.

### **Self-Reports versus Proxy Information**

To gain insight into the personal experiences of people with ID, self-reported information is generally preferred over proxy ratings (Emerson et al., 2013; Schalock et al., 2002; Scott & Havercamp, 2018). Aside from moral-ethical considerations that stress the importance of involving people with ID directly in decision making and research (Freedman, 2001; Huus et al., 2015), there are further methodological objections to the use of proxy measures. Evidence suggests that information given by proxies about opinions, feelings and thoughts of people with ID themselves may be less accurate and less sensitive, relative to self-report (Hulbert-Williams et al., 2011; Lewis & Morrissey, 2010; Scott & Havercamp, 2018). Agreement about opinions, feelings and thoughts between proxies and persons with ID is especially poor for personal or sensitive topics (Mileviciute & Hartley, 2015; White-Koning et al., 2005).

### **Suitability of Self-Report Measures for People with ID**

Whereas there is virtually no debate for researchers and clinicians as to whether the opinions, feelings and thoughts of people with ID themselves should be a primary source of information, there are concerns about gaining access to information in a valid and reliable way when using self-report questionnaires or interviews as many people with ID will struggle to understand and respond to questions (Nicolaidis et al., 2020). In order to address the personal experiences of people with ID using self-report measures, developers and researchers have turned to one of three options: (a) use measures that can be administered to persons with ID in an unaltered form, (b) use measures that can be administered to persons with ID after making adaptations to the original, and (c) use measures specifically designed to be administered to persons with ID.

Some authors have argued that from an ethical and theoretical perspective, it would be appropriate to explore the application of existing unaltered assessment measures before modifying existing instruments, or developing new measures for special target groups (Kellett et al., 1999; Wieland et al., 2012). Mainstream instruments usually have a long development history, with much attention being paid to the theoretical underpinnings of concepts and the operationalisation of the constructs under study. There is often no reason to assume that these concepts relate to people with ID in a fundamentally different way than for the general population. Several measures have been identified that can be administered to persons with ID without making changes to the content or associated procedures (e.g. the Brief Symptom Inventory; Wieland et al., 2012).

If measures are unsuitable for use in their original format with persons with ID, it is common practice to adapt assessments and to test the modified versions for usability, reliability and validity (Stancliffe et al., 2017). Many measures have been adapted, for example the revised version of the How I Think Questionnaire (Daniel et al., 2018), Impact of Events Scale (Hall et al., 2014), the Self- and Other-Deception Questionnaire (Langdon et al., 2010), and the Zung Self-Rating Anxiety Scale (SAS-ID; Lindsay & Michie, 1988).

Instead of adapting existing measures, some researchers have constructed measures specifically designed to account for the needs of people with ID. Examples include the Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID; Mindham & Espie, 2003) and the Lancaster and Northgate Trauma Scales (LANTS; Wigham et al., 2011). Notably, researchers in the field of Quality of Life (QoL) research, like Schalock and Verdugo, have designed psychometrically robust measures of life satisfaction (Schalock et al., 2008).

### **Tailoring Self-Report Measures to the Needs of People with ID**

Notwithstanding the progress that has been made in the past years, in some areas there are still few measures available that can be used successfully with people with ID. Vlot-van Anrooij et al. (2018)

for instance, stated that “suitable and valid scales to collect self-reports on health and health-related behaviour among people with ID remain to be scarce”. Similarly, a paucity of psychometrically sound self-report measures is noted in the field of treatment efficacy (Vlissides et al., 2016) and diagnostics (Scott & Havercamp, 2018). Thus, ongoing efforts from developers of measures are needed to provide researchers and clinicians with ID-inclusive measures.

However, developers that attempt to meet this challenge, rarely justify the reasons for specific adaptations and not all published evidence is incorporated in the construction procedures. Although authors report some adaptations from ‘mainstream’ instruments and procedures, and most adaptations are substantiated by at least some form of evidence, they usually do not account for *all* possible challenges associated with the collection of self-reported experiences of people with ID. This - at least in part - seems to be due to a lack of a complete overview into these factors. The information available to build ‘ID-inclusive’ measures seems to be scattered among the research literature. Some efforts have been made to compile and review the available evidence. These reviews generally address specific topics, for instance ‘acquiescence’ (Finlay & Lyons, 2002) or the use of Likert-type scales (Hartley & MacLean, 2006). One very notable attempt to comprehensively discuss a wide range of issues about the use of self-reports is the review by Finlay and Lyons (2001).

The guidance offered by these studies are presented as applicable to ‘persons with ID’, which by definition includes a very broad range of cognitive and adaptive functioning. In most studies that explore under which conditions reliable and valid self-reports can be obtained from people with ID, many potential participants are excluded on the basis of their level of disability. Usually, a certain level of verbal adequacy or comprehension is required to participate, barring many persons with levels of functioning lower than mild intellectual disability (MID) from participation (Hartley & MacLean, 2006). This reduces the applicability of many recommendations to people with borderline intellectual functioning or MID - a limited proportion of the total ‘people with ID’ population. And even within this

subgroup there is much variation in terms of cognitive, verbal and adaptive functioning, necessitating a nuanced view of the recommendations offered by the research.

### **The Need for Guidance**

In the absence of comprehensive practical and evidence-based guidance for developers, the instruments that are developed may be poorly attuned to the needs of people with ID. As the validity of clinical and research outcomes largely depends on the validity of the measures used, this may pose a serious threat to the credibility and validity of the research in this field. While the group of people with ID is ‘...too heterogeneous in terms of personal history and linguistic and cognitive abilities for any single questionnaire to be valid for the whole population’ (Finlay & Lyons, 2001), and ‘it would overlook the heterogeneity of the population to propose that gaining self-report from everyone is possible’ (Emerson et al., 2013), it would be markedly valuable to develop self-report measures for people with ID that are as inclusive as possible.

### **Objective**

Research about the important factors that need to be kept in mind when creating or adapting self-report measures for use with people with ID is scant. The objective of the current study is to systematically map and synthesize the research literature about evidence-based suggestions for adaptations necessary for creating inclusive measures and administration procedures for people with ID. The aim is to provide professionals with evidence-based guidelines for developing, adapting or using measures suitable to use by people with ID. Since the vast majority of studies on this topic excluded persons with more severe levels of ID, it is expected that most results pertain to persons with borderline intellectual functioning and mild to moderate ID. However, those with severe or even profound levels of ID were not excluded from our review, and issues associated with this group will be accentuated within the results where possible.

An associated goal of this study is to define possible areas of interest that have not been extensively researched yet. Wherever appropriate and possible, the current review will distinguish between recommendations for different subpopulations (e.g. with respect to levels of ID or specific age groups).

## **Methods**

### **Protocol and Registration**

To report the outcomes of the current review, the 2020 Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for conducting and reporting systematic reviews (Page et al., 2020) were followed. The PRISMA-S extension for reporting literature searches in systematic reviews (Rethlefsen et al., 2021) was used to report the search strategy. Following these guidelines, a full description of the strategy used for systematically searching the literature and the protocol for study screening and selection was registered in PROSPERO (registration number CRD42019138765). The protocol for screening and selection of included studies, including search terms and strings for all databases can be accessed at:

[https://www.crd.york.ac.uk/PROSPEROFILES/138765\\_STRATEGY\\_20210203.pdf](https://www.crd.york.ac.uk/PROSPEROFILES/138765_STRATEGY_20210203.pdf)

### **Eligibility Criteria**

Studies were eligible for inclusion if they reported on studies including persons with ID, outcomes included recommendations for the construction or adaptation of measures designed for people with ID and were published in peer-reviewed journal papers written in English. Both primary research and secondary sources (e.g. from literature reviews) were included. The publication period was restricted to studies published after 1995 and records had to be available electronically for practical purposes.

### **Information Sources and Search**

The Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews and the International Register of Prospective Systematic Reviews (PROSPERO) were searched to identify if any reviews on the topic of self-report measure construction and adaptation had recently been planned or carried out. No results were found.

A systematic search of the PsycInfo, PubMed and Web of Science databases was performed by the first author and an information specialist at the University of Amsterdam. These databases were selected to maximize the reach across disciplines. Additionally, the first 200 records of a Google Scholar search were scanned.

### **Study Selection**

After merging results across databases and deduplication, articles were screened for relevance on the basis of titles and abstracts, using the Rayyan software tool (Ouzzani et al., 2016). The remaining articles were screened full-text. Additional records were retrieved through forward and backward citation searching. All steps in the process of study selection were guided by the study screening and selection protocol and carried out by two reviewers (RK and GM) independently. After each step, interrater agreement was assessed. All discrepancies between raters were resolved on the basis of discussion until consensus was reached.

### **Quality Appraisal**

The included designs were expected to be very diverse in nature, ranging from small-scale qualitative designs to literature reviews. The Mixed Methods Appraisal Tool (Hong et al., 2018) was used to appraise the methodological quality of five categories of studies: qualitative research, randomized controlled trials, quantitative non-randomized studies, quantitative descriptive studies and mixed methods studies. The MMAT is one of few formal tools to concurrently evaluate the quality of studies with varying designs in a review. It is found to be a reliable and efficient tool (Pace et al., 2012). As the MMAT is not suitable for use with literature reviews, the Joanna Briggs Institute (JBI) Checklist for Text

and Opinion Papers was used for non-systematic reviews, and the JBI Checklist for Systematic Reviews was used to appraise included systematic reviews (Aromatis & Munn, 2017).

For each type of study, the appropriate tool was selected and the appraisal was carried out by the first two authors (RK and GM). The outcomes of the appraisal were integrated in the GRADE-CERQual assessment of the strength of the evidence (see below).

### **Synthesis of Results**

The primary outcome variables consisted of suggestions and recommendations put forward by authors. These are qualitative statements, based on quantitative and qualitative data from both primary and secondary sources. To map and structure the outcomes, the steps outlined in the 'Best fit framework synthesis' approach by Carroll et al. (2013) were followed. In this approach the researcher first sets out to identify pre-existing models or frameworks that may underlie the outcomes under study. These models are then integrated using thematic analysis to form an a priori framework to code the results from the outcome studies against. In subsequent steps the results are axially coded against the framework and new codes are created by performing thematic analysis on any evidence that cannot be coded against the framework. This results in a revised framework composed of new and a priori themes supported by the evidence.

Several authors present models to structure the recommendations for future adaptations. Three attempts to summarize practical suggestions were integrated into the a priori coding framework: the framework used for the presentation of results from the review by Finlay and Lyons (2001), Tourangeau's model of survey responses, as adapted by Jen-Yi et al (2015), and the overview of possible adaptations of self-report measures for people with ID presented by Bell and colleagues (Bell et al., 2018). The following model resulted from the synthesis of these three models and was used to guide the initial deductive coding process.

**Table 1***A Priori Coding Framework*

<u>Overarching factor</u>	<u>Subfactors / themes</u>	<u>Examples include</u>
Content factors	Language	Wording, phrasing of questions and answer formats
	Response formats	Closed- and open formats, number of response alternatives
	Design	Use of supportive visualisation, <u>lay-out, 'survey flow'</u> .
Procedural factors	Assessment procedure	Use of pre-tests, ability screening, standardisation vs flexibility
	Context of the assessment	Interviewer-interviewee relation factors (including social desirability bias), interviewer skills
	Construction and psychometric evaluation	Item generation, piloting, procedures for establishing <u>validity and reliability</u>

**Assessing the strength of evidence with GRADE-CERQual**

GRADE-CERQual (Confidence in Evidence from Reviews of Qualitative research; Lewin et al., 2018) was used to determine how much certainty can be placed in each recommendation. This approach provides guidance for assessing how much confidence to place in results from systematic reviews of qualitative findings. The confidence that can be put in each of the recommendations is rated as high, moderate, low, or very low, based on the assessment of four components that contribute to the robustness of each review finding. These components are (a) methodological limitations of studies contributing to the recommendation, (b) coherence of findings, (c) adequacy (richness) of data, and (d) relevance of the contributing studies to the context of the review question. The assessment of these components collectively contribute to an overall assessment of whether the individual recommendation provides a reasonable representation of the research interest (Lewin et al., 2018). A key product of the assessment is a transparent summary of findings. In line with the guidance by Munthe-Kaas et al. (2018), the results from the quality appraisal procedure outlined above were used for the methodological limitations component of the GRADE-CERQual assessment.

The GRADE-CERQual assessment was performed by the first two authors (RK and GM) and results were discussed among the contributing authors. The elaborate assessment results for each recommendation can be found in the GRADE Evidence Profile (Supplemental Material). The corresponding Summary of Findings table can be found in Table 2.

## **Results**

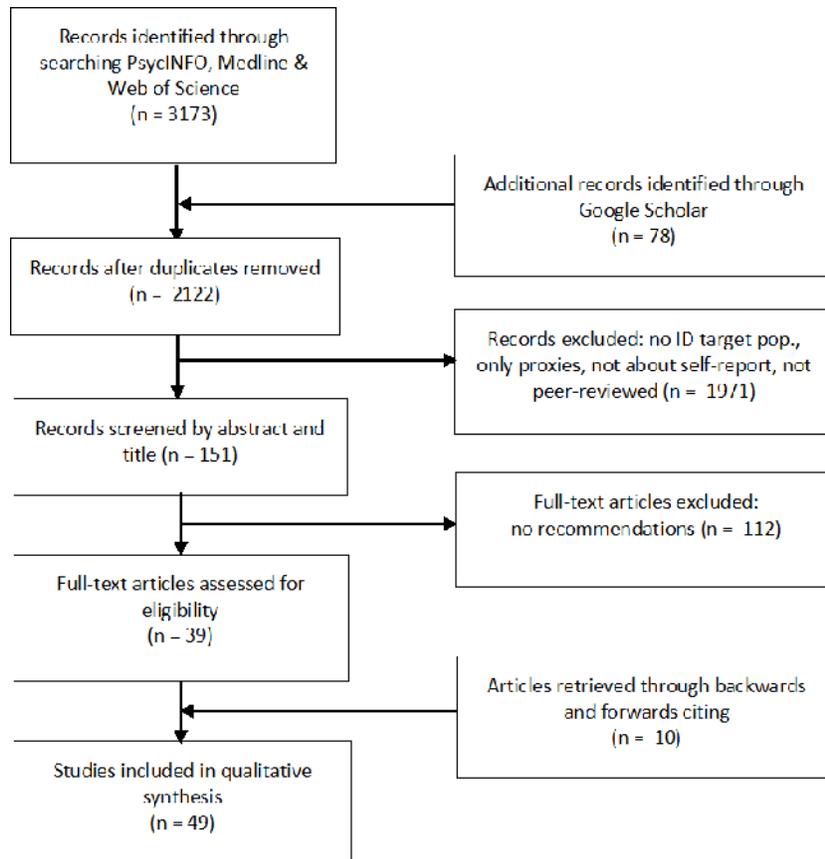
### **Study Selection**

The selected databases were searched in February 2020 and updated in February 2021. A total of 3173 records were found. After deduplication 2122 articles were scanned for eligibility by reading the titles and abstracts. The first 100 records were reviewed by the first two authors (RK and GM) simultaneously. According to the Landis and Koch (1977) guidelines interrater agreement was observed to be 'near perfect',  $k = 0.89$ ,  $p < 0.01$ . Disagreements on study screening were resolved based on consensus and discussion. The remaining records were screened by either the first or second author, resulting in a total of 152 articles to be appraised full-text.

All full-text of remaining articles were read by both reviewers (RK and GM) independently. Interrater agreement at this stage was observed to be excellent,  $k = 0.95$ ,  $p < 0.01$ . Again, conflicting results were resolved through discussion. Thirty-nine articles were retained for inclusion in the final dataset. Citation tracking and manual searching of reference lists of all articles that were included in the full-text appraisal yielded an additional 10 articles to be included, totalling the number of included studies to 49.

Figure 1

PRISMA flowchart



## Study Characteristics

The main characteristics of the included studies can be found in Appendix A. The studies showed a great variety in subject and design, and included quantitative experimental and observational designs, qualitative designs and mixed-method studies, as well as (systematic) literature reviews. The aim of most of the included studies was to investigate which characteristics of instruments were best suited for people with ID in general, but recommendations were generally made for specific levels of cognitive functioning within the broader 'ID range'. Persons with levels of functioning below the moderate disability level were usually excluded. Only a handful of studies explicitly included persons with 'severe' or even profound levels of disability. For many studies, the level of ID of participants was not specified

or described in very general terms. The resulting summary of recommendations can therefore not be applied for 'persons with ID' in general, but should be assessed relative to the level of functioning of the population under study.

### **Results of the Quality Appraisal**

Quality appraisal scores for all studies are reported in Appendix A. In general, the methodological quality of included studies was adequate, albeit not flawless. No studies were excluded on the basis of quality appraisal outcomes. For all types of studies, the target population was often not clearly defined. For quantitative studies, points were deducted for unclear sampling procedures or use of convenience samples without regard to generalisation issues. Further, in some of the studies small samples of participants were used. In most studies, the study sample consisted of persons with borderline intellectual functioning and mild ID, often because persons with moderate or severe levels of ID failed to meet inclusion criteria as a result of (the sometimes presumed, sometimes tested) limited verbal abilities. If the chosen eligibility criteria meant that a large proportion of the target population was excluded, this was seen as a risk for 'nonresponse bias'.

For many studies that used qualitative methods to make inferences, the process of data synthesis was not (clearly) described. The conclusions drawn from the data on some occasions seemed to be highly dependent on the researcher's interpretation, which is of course an artefact of the qualitative nature of the design and was generally accounted for in the reported limitations of each study. Sample sizes in some of the qualitative studies were small, verging on questionable. In some literature reviews, it remained unclear how the literature was searched (sources, search terms), and on some occasions, claims were made that were not substantiated by the literature.

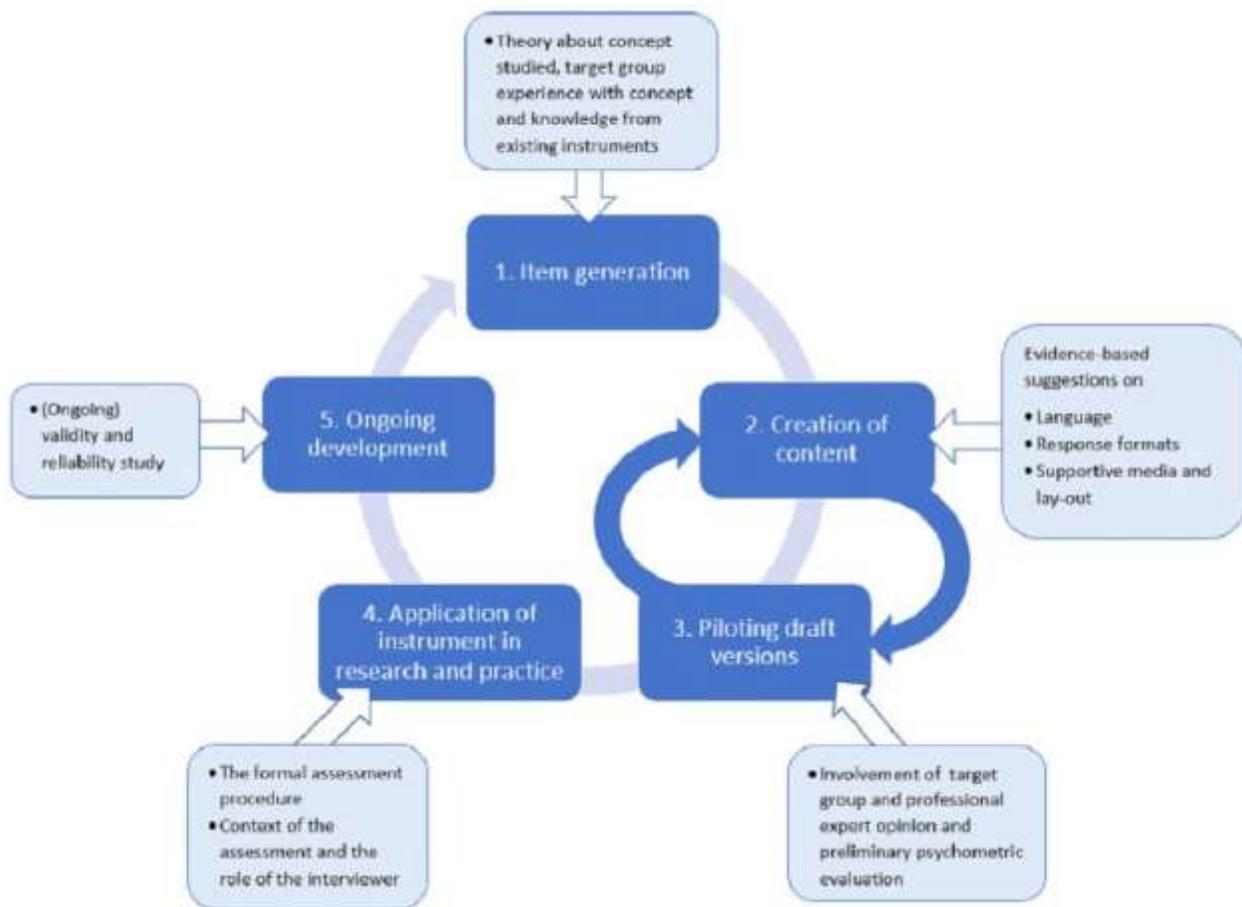
### **Synthesis of Results: Revising the A Priori Framework**

A total of 161 suggestions or recommendations were extracted from the 49 included studies. Many suggestions were mentioned by more than one author. Similar suggestions were collated. 74

unique suggestions were forwarded, that addressed a total of 25 general issues. While it was possible to code all the suggestions using the factors within the a priori framework (Table 1), some factors did not seem to mirror the sequence in which the development or adaptation process took place. Therefore, to provide optimal guidance for developers, we rearranged the factors to match the stages in this process during our thematic analysis. The revised framework with the resulting sequence of stages is presented in Figure 2.

**Figure 2**

*Sequence of Stages in ID Instrument Development*



### Recommendations from Researchers by Development Stage

In the Summary of Findings (Table 2), all 74 unique suggestions are tabulated under the corresponding stages of instrument development (Figure 2). For each recommendation, the GRADE-

CERQual level of confidence (high – moderate – low – very low) is presented. A more elaborate summary that includes the detailed GRADE-CERQual component assessment can be found in the Evidence Profile (Supplemental material). For only a handful of recommendations the level of confidence was high. This was the case for some relatively well-researched topics such as the optimal number of response alternatives and some well-established good practices in research such as extensive psychometric evaluation of any newly constructed measure. By far most recommendations received a ‘moderate confidence’ evidence level because substantial empirical research was lacking, recommendations were not very specific, or few studies contributed to the finding. Many of these recommendations originated in practical experience in research or clinical practice and expert opinion. Low confidence recommendations were mostly the result of relatively low-quality research, contradictory findings, and results that solely reflected the researcher’s opinion.

A summary of the findings per topic or development phase is provided below. The recommendations apply to persons with BIF to moderate ID. Where relevant, distinctions are made between recommendations for different levels of ID. If recommendations extend across *all* levels of functioning, including more severe levels of ID, this is made explicit.

### ***Stage 1: Item Generation***

In the first stage of instrument development, the concept under study is explored. Several researchers stressed the importance of involving persons with ID in this process, to discuss how the concept translates to their everyday life experiences. Developers should not assume that the concepts operationalised in the original measure hold the same meaning and value for persons with ID. ***Stage***

### ***2: Creation of content***

Many suggestions for the creation of ‘ID-inclusive’ questions and responses were put forward by researchers. These apply to language aspects, choosing appropriate response formats and the use of media to support the meaning of questions and responses.

**Language.** One of the most frequently discussed topics was the use of simple wording and grammar for questions and responses. It is of note that recommendations such as ‘simplify complex language’ (Bell et al., 2018) may not offer developers sufficiently concrete guidance. And what level of simplification is needed varies greatly for different levels of cognitive impairment. A particularly concrete and useful suggestion was to use established guidelines for the ‘translation’ of plain language to more accessible language, such as the British Easy Read guidelines (UK Department of Health, 2010), the Dutch ‘Taal voor Allemaal’ (‘Language for All’) guidelines (Taal voor Allemaal, 2021) or the German ‘Leichte Sprache’ (‘Easy Language’) guidelines (Bredel & Maass, 2016).

**Response Formats.** The question of which response formats are optimal for persons with ID was addressed frequently, and this is one of few topics that has been researched extensively using quantitative study designs. Research on this topic explored the impact of using different response options on comprehension, answering patterns, psychometric properties, and bias. However, for many issues there was no clear solution; findings were quite contradictory and dependent on many variables, such as subpopulation characteristics, the topic under study, and how much emphasis is put on the threat of bias to the validity of results that is associated with some response formats (Finlay & Lyons, 2001).

Yes/no-type questions are understood by the largest proportion of persons with ID and may be used even with people with severe levels of disability (Ikeda et al., 2016; Ramirez, 2005). But the appropriateness of simple yes/no answers for self-report questionnaires involving people with ID requires consideration. Sigelman et al.’s (1981) conclusions that yes/no statements should generally be avoided because they promote acquiescent responses - have been echoed by subsequent scholars and developers ever since, attributing acquiescence to submissiveness on the part of ID participants. However, research trying to replicate both the higher prevalence of acquiescence in persons with ID compared to typically developing persons and the finding that submissive acquiescence to yes/no

formats increases as the level of intellectual functioning decreases, showed mixed results (Finlay & Lyons, 2002) or were refuted (Matikka & Vesela, 1997; Ramirez, 2013; Rapley & Antaki, 1996). A suggestion by some researchers was to follow-up yes/no questions with open-ended questions for explanation or examples when the assessor suspects acquiescent responding tendencies (Finlay & Lyons, 2001; Stenfert-Kroese et al., 1998), but this requires respondents to be able to verbally express themselves (Boland, 2018) and may therefore not be a suitable strategy for persons at the lower end of intellectual functioning.

The following recommendations for the use of response options were relatively well-established and backed-up by empirical evidence:

- No more than 3 response options should be used in Likert scales for people with mild to moderate ID and no more than 5 options for persons with borderline intellectual functioning to mild ID (Bell et al., 2018; Cummins, 1997; Dagnan & Ruddick, 1995; Fang et al., 2011; Hartley & MacLean, 2006; Power et al., 2010).
- Adding a 'don't know' option is advisable as this prevents participants with ID from choosing a random response when they do not understand the question (Bell et al., 2018; Finlay & Lyons, 2001; Finlay & Lyons, 2002; Ramirez & Lukenbill, 2008).

**Supportive Visualisation and Lay-Out.** Another frequently considered adaptation was the use of visual supports to enhance comprehension of questions and responses, with little in the way of definitive conclusions. Whereas the general consensus seemed to be that supporting written content with pictures is helpful (Ikeda et al., 2016; O'Keeffe et al., 2019; Reid et al., 2009; Stenfert-Kroese et al., 1998), the visualisations used as support in self-report measures were hugely diverse in form, shape and meaning. Research on this topic often relied on small sample sizes (e.g. Dagnan, 1995; de Knecht et al., 2017) and there were some issues with generalising findings from specific subpopulations (children with ID, persons with Down Syndrome) to the broader ID population (e.g. Reid et al., 2009).

Aside from the potential benefits of adding pictures for enhancing comprehension, some authors raised awareness that pictures may be confusing or not helpful if not recognised by participants (e.g. Barker et al., 2020; Finlay & Lyons, 2001; Payne, 2004). Unless the individuals' understanding of the meanings of these supports is assessed, they may decrease the reliability and validity of answers rather than ensure better quality data (Cuskelly et al., 2013). This topic remains a largely under-researched area of investigation and not many concrete suggestions can be given.

### ***Stage 3: Piloting draft versions***

Researchers stressed that the targeted group of persons with ID should be included in the process of developing and piloting instruments, as their involvement can be very helpful for revision and refinement of questionnaire items (Emerson et al., 2013; O'Keeffe et al., 2019; Williams et al., 2007).

### ***Stage 4: Application of the Instrument in Research in Practice***

Several recommendations were made that address the processes and procedures of using instruments in practice. These recommendations relate to either descriptions of the formal assessment procedure, or the role of the interviewer or the person assisting the assessment procedure.

**The Formal Procedure for Assessment.** Suggestions under this topic related to the formal procedures for carrying out the assessment. A need to balance between standardised procedures and the need for flexibly adapting to the person under study was observed by many researchers. Whereas the assessment of individuals without ID can be performed with a high level of standardisation, interviewing persons with varying degrees of ID requires a much greater deal of flexibility to accommodate for individual variations in cognitive functioning and language abilities. Researchers offered several suggestions to flexibly adapt, while ensuring standardisation within acceptable limits. Notably, to avoid heterogeneity in the formulation of questions, the use of standardized scripts or prompts was recommended in the case a question needs to be reformulated.

A general consensus to use pre-tests was found. Depending on the outcome and intended use or goal for the measure, the results of pre-tests can be used to: (a) exclude participants from the study that are expected to return invalid results because of problems with comprehension; (b) detect biased responding patterns (e.g. acquiescence, social desirability) and establish validity of the results at the individual and population level; and (c) offer participants an opportunity to practice with the response formats.

**Role of the interviewer in clinical and research practice.** Self-report questionnaires are frequently administered in a structured interview format. To guide the assistance and minimise the impact of interviewer-interviewee dynamics on results, suggestions were forwarded regarding interviewer skills and optimising the assessment setting.

***Stage 5: Ongoing development***

After implementation of the measure in practice, researchers advocate that ongoing psychometric evaluation is carried out to evaluate the quality and structure of the measure (Finlay & Lyons, 2001; Lindsay, 2002; Stancliffe et al., 2014). Developers should not assume that the adapted version holds the same structure for varying subpopulations and the psychometric properties of the adapted version should be re-evaluated as if it were a new measure (Blasingame et al., 2011; Zabalía, 2013).

**Table 2**

*Summary of Findings: Recommendations for the Construction or Adaptation of Self-report Instruments for People with Intellectual Disabilities.*

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
<b>1. Item generation</b>			
Participation of people with ID	Use focus groups to gather insights in cognitive processes, personal experiences and idiom used by ID persons in relation to the object of study.	Glenn 2003, Ramirez 2007, Schalock 2002, Vlot van Anroij 2018, O'Keeffe 2019, Finlay 2001	High confidence Mostly based on practical experience and expert opinion.
	Assess the cognitive and language skills of participants involved in the construction process to determine suitability for target populations.	Emerson 2013	Moderate confidence Based on expert opinion.
<b>2.1 Creation of content: format and language</b>			
Vocabulary	Keep the vocabulary simple to prevent acquiescence and non-responding arising from not understanding the question.	Bell 2018, Finlay 2001, Finlay 2002, Gjertsen 2019, Jen-Yi 2015, Scott 2018, Sigstad 2018, White Koning 2005	Moderate confidence Mostly based on clinical expertise. What constitutes simple vocabulary is not specified.
	Use literal meaning of words, do not use metaphors and proverbs.	Ikeda 2014	Low confidence Finding seems common sense, but this specific study does not contribute to the evidence base for the recommendation.
	Replace abstract concepts with more concrete concepts where possible.	Bell 2018, Finlay 2001, Ikeda 2014, Scott 2018, Sigstad 2018	Moderate confidence. Finding aligns with expert opinions, but not substantiated empirically.
	Avoid adding 'no' and 'not' to positive phrasings. Use negative form of words.	Finlay 2001	High confidence It is known from linguistics that this reduces complexity.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Sentence structure	Keep the sentence structure as clear and simple as possible.	Bell 2018, Finlay 2001, Sigstad 2018, White Koning 2015	Moderate confidence  What constitutes simple grammar is not specified.
	Use only positively phrased questions as negatively formulated items can be confounding.	Bell 2018, Payne 2004	Moderate confidence  Some substantiation from primary empirical research and secondary sources.
	Use short sentences to minimise the strain on working and short-term memory.	Bell 2018, Gjertsen 2018	Moderate confidence  How short sentences should be is not made specific More empirical research needed.
	Do not use questions containing modifiers or more than one clause, direct comparisons, or contractions.	Bell 2018, Finlay 2001, Heal 1995, White Koning 2005	Moderate confidence  Linguistics research confirms that this reduces linguistic complexity.
	Use active formulations (as opposed to passive phrasings) as they are structurally less complex.	Finlay 2001, White Koning 2005	Moderate confidence  More empirical research needed.
	Avoid double negatives as this can be confusing, especially in combination with positive and negative response options.	Finlay 2001	Moderate confidence  More empirical research needed.
	Use established guidelines, such as Easy-Read standards, to reword items.	Kent 2018	Moderate confidence  Replication needed.
	Use established evaluative statistics to check for readability (e.g. Flesch Reading Ease, Gunning-Fog Index).	Keeling 2017	Moderate confidence  Replication needed for broader ID population.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Time frames	Use anchor events and situations to support the respondent's view of the time frame to help support the retrieval of events from (long-term) studies. memory.	Bell 2018, Finlay 2001, Jen-Yi 2015, Scott 2018	High confidence Coherent evidence base form diverse
	Don't ask to retrieve detailed information over longer periods of time.	Vlot van Anroij 2018	Moderate confidence Replication needed for broader ID population.
	Ask respondents to reflect on what they <i>currently</i> feel and think. Stay in the here and now.	Williams 2007	Moderate confidence Replication needed for broader ID population.
Content: other/miscellaneous	Refrain from presumptions in questions, e.g. "You do like the food at your home, don't you?".	Bowles 2014	Moderate confidence Based on only one small-sample study.
	Consider rephrasing questions to accommodate for specific cultural issues.	Jen-Yi 2015	Low confidence Although common sense, the recommendation solely reflects professional opinion of researcher.
<b>2.2 Creation of content: Response format</b>			
Dichotomous answer options	Yes/no type questions are understood by the largest proportion of persons with ID.	Ikeda 2016, Ramirez 2005, Stancliffe 2015	High confidence Decisive results from quantitative studies.
	Be aware of acquiescent responding on yes/no response formats. Consider using an either/or response format instead of yes/no.	Heal 1995	Low confidence Results from more current studies show conflicting results with this finding. Contemporary update necessary.
	Follow-up yes/no answers with open questions.	Finlay 2001, Heal 1995, Stenfert-Kroese 1998	Moderate confidence Plausible for persons with adequate verbal abilities.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
More than two response options	Tailor the number of response options to the need for nuance in answers and the ability of the target population. Rule of thumb: - It is generally preferable to use 3-point Likert scales. - Up to 5-point Likert scales can be used for most persons with borderline intellectual functioning to mild ID.	Hartley 2006, Bell 2018, Cummins 1997, Fang 2011, Power 2010, Dagnan 1995	High confidence  Compelling evidence form a broad variety of research designs. Adequate differentiation of findings between subpopulations.
	Either/or- and yes/no questions can be broken down into two stages to obtain nuanced answers. Affirmative answers on the first question can be followed-up with sometimes/always, a little/a lot, etc.	Finlay 2001, Ramirez 2008, Cuthill 2003	High confidence  Sensible suggestion that is backed-up by some empirical evidence.
	Use a single set of one- or two-word descriptors (Often, sometimes, never) with a series of questions instead of elaborate self-descriptive statements that vary for every question.	Hartley 2006	Moderate confidence  Sensible suggestion, but limited empirical evidence base.
	Check validity of answers on multiple choice questions by asking for examples or further (scripted) probing questions.	Finlay 2001, Hartley 2006	Moderate confidence  Mainly based on clinical and research expertise. Plausible for persons with adequate verbal abilities.
	Reduce the number of response options in orally presented questionnaires to reduce working memory strain.	Finlay 2001	Low confidence.  Needs specification based on empirical research.
Open-ended questions	Open-ended questions can be used when asking for opinions and retrieval of information from memory.	Stenfert-Kroese 1998, Jen-Yi 2015	Moderate confidence  Plausible for persons with adequate verbal abilities.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	Do not use open-ended questions when the participant or target group in general does not possess adequate productive verbal ability.	Boland 2008	High confidence Recommendation seems self-evident.
Visual analogue scales (VAS)	VAS scales can produce reliable and meaningful results in persons with borderline intellectual functioning up to moderate ID.	Dagnan 1995	Moderate confidence Use of VAS holds potential, but more research on reliability and validity is needed. Findings may be valid only for higher functioning adults with ID as a result of selection bias.
Include "don't know" option	Include a response option of 'I don't know' when using forced-choice scales or interviews to minimise acquiescent or random responding.	Bell 2018, Ramirez 2008, Finlay 2001, Finlay 2002	High confidence Well-established from research and clinical practice and backed-up by empirical evidence.
	Use different response formats throughout the measure to check for inconsistencies or bias in answers.	Heal 1995	Very low confidence Conflicting results.
<b>2.3 Creation of content: Supportive media and layout</b>			
Visualisation of content	Use visual representations (of choice objects, relevant people and places, and emotional states) to support the meaning of questions.	Stenfert-Kroese 1998, Finlay 2001, Ikeda 2014, O'keeffe 2019, Reid 2009,	Moderate confidence Recommendation seems plausible but lacks specificity (what exactly works for whom?) and empirical validation. More empirical research needed.
	Only use visualisation that has been proven to facilitate understanding (e.g. in a pilot test with members of the target population).	Finlay 2001	High confidence Limited empirical evidence but in line with good research practice and recommendations elsewhere in the current review (i.e. Stage 3. recommendations).

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Lay-out	Use pictorial representations of response alternatives, such as boxed histogram pictures as a representation of magnitude or frequency and smileys as a representation of (dis)contentment.	Bell 2018, Hartley 2006, Heal 1995, de Knecht 2017, Marshall 2007, O'Keeffe 2019	Moderate confidence  Widely recommended, but types of supportive visualisations are very diverse. Not clear what works best for whom. More empirical research needed.
	Facial representations can be used to depict pain and emotions.	de Knecht 2017	Moderate confidence  Compelling evidence but from a specific and small sample of participants. Need for replication.
	Use a clear and attractive lay-out to capture the respondent's attention and reduce clutter and confusion.	White Koning 2005, Bell 2018	Moderate confidence  Sensible suggestion but lacks specificity.
	Display one question per page.	Ikeda 2014	Low confidence  Recommendation seems self-evident, but no empirical evidence.
	Limit paragraph length for explanatory texts and instructions.	White Koning 2005	Low confidence  Recommendation seems plausible, but origin of the recommendation is not clear, and it lacks specificity.
	Use flash cards when a choice from more than 2 response options is asked.	Boland 2009	Moderate confidence  Observed by the researchers in practice to be helpful for respondents.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Medium of presentation	Consider using computer-administered assessment of the instrument. The multimedia options can assist people with ID to report internal states autonomously or with minimal support.	Clark 2017, de Knecht 2017	Moderate confidence  Computerised testing offers great advantages, but research on the opportunities and pitfalls for people with ID is scarce. Furthermore, this recommendation lacks specificity.
<b>3. Piloting draft versions</b>			
Piloting and participation	Review the wording of items and the intended meaning of supportive media with members of the target population. Develop alternatives together if needed.	Finlay 2001, Gjertsen 2019, Jen-Yi 2015, O'Keeffe 2019, Vlot van Anroij 2018	High confidence  Well-established as good practice in (participatory) research.
	Assess the cognitive and language skills of participants involved in the construction and validation process. to determine the (boundaries of) applicability of the measure.	Emerson 2013	Moderate confidence  Limited evidence but seems to reflect sound research and development practice.
Statistical evaluation	Responsiveness of items can be used as a measure of comprehensibility of the instrument.	Stancliffe 2015	High confidence  Finding based on large-scale population survey data and sensitive statistical analyses.
	Use triangulation procedures with different informants (e.g. with relatives or professionals).	Jen-Yi 2015	Low confidence  Self-report and proxy data cannot be assumed to be interchangeable. Other scholars insist that self-report and proxy data should not be used for triangulation as they may measure entirely different concepts (e.g. Emerson, 2013)
<b>4.1 Application in practice: formal assessment procedure</b>			

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Use of pre-tests practice formats	Use a pre-test to check the participant's cognitive and verbal ability and if he/she understands the response formats and visualisations used.	Bell 2018, Cummins 2002, Cuskelly 2013, de Knecht 2013, Emerson 2013, Finlay 2001, Finlay 2002, Jen-Yi 2015, Townsend-White 2012, White-Koning 2005	High confidence Well-established good practice in research and development, but little specific guidance on how to test comprehension.
	Use a pre-test as an opportunity for participants to practice with the response formats	Hartley 2006, Jen-Yi 2015, Reid 2009	High confidence Familiarity with assessment procedures promotes reliability and validity of answers.
	Do not exclude participants a priori on the basis of client characteristics such as IQ.	Ramirez 2005	Moderate confidence Replication needed to draw firm conclusions.
Bias detection	Add a bias detection measure to the administration process (preferably before administration) to exclude participants from the results or to use as a caution when interpreting results.	Cuskelly 2013, Perry 2002, Hartley 2006, Emerson 2013, Keeling 2017	Moderate confidence Overall, the recommendations are sufficiently substantiated by the contributing studies, but specification is needed to determine how this should be done. .
	Integrate bias-detecting items in the questionnaire structure to establish validity of the results at the individual and population level and to exclude participants after data completion.	Perry 2002, Williams 2007, Townsend-White 2012, Matikka 1997	High confidence Empirical evidence from a range of different studies.
Dealing with difficult items	Allow interviewers to paraphrase and/or expand upon question items or response alternatives according to scripted guidelines for paraphrasing.	Antaki 1996, Hartley 2006, Jen-Yi 2015, Bell 2018, Finlay 2001, O'Keeffe 2019, Sigstad 2018	Moderate confidence Need for flexible testing instructions to include as many persons as possible is at odds with need for standardisation.
	Use pre-questions and examples to introduce difficult concepts and sensitive topics.	Antaki 1996	Moderate confidence Not specific enough to guide developers.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
Other structural and procedural issues	Make the questionnaire as short as possible to prevent potential problems with attention span or fatigue.	Bell 2018	Moderate confidence  Seems self-explanatory, not very specific. More empirical research may be needed to determine limits and optional length. Low confidence
	Group items on related topics.	Jen-Yi 2015	Although common sense, the recommendation reflects the professional opinion of the researcher. Moderate confidence
	If the participant can read, present questions and answers in writing instead of orally as this puts less strain on the memory of the participant. This may also increase the feeling of collaboration and competence on the part of the participant.	Bell 2018	More empirical research needed.
	Allow for someone to assist the participant and clarify questions if the participant so wishes.	Gjertsen 2019	Low confidence  Enhances chance of participation but may introduce various forms of bias arising from interviewer-interviewee dynamics. Impact of support on outcomes not well understood.
<b>4.2 Application in practice: Role of assessor in case of assisted administration or structured interview</b>			
Formal instruction of interviewer	Interviewers should be trained to - practice interviewer skills. - detect bias and other distortions as a result of the interviewer-interviewee relation. - become familiar with the subject and procedures.	Jen-Yi 2015, Perkins 2007	Moderate confidence  Recommendation reflects general good research and clinical practice.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	Interviewers should receive strict instructions to stick to the format to promote standardisation.	Antaki 1999	High confidence Self-evident to promote standardisation and reduce ambiguity.
Pre-assessment introduction	Take the time to develop rapport with participants.	O'Keeffe 2019	Moderate confidence Mainly based on experience from research in practice.
	Use a clear, prescribed introduction of the assessment and clear instructions to explain the procedure to the participant.	White Koning 2005	Moderate confidence Reflects good research practice to minimise ambiguity.
	Stress that information will not be shared with carers, and/or state the conditions under which confidentiality may be breached.	Finlay 2001	High confidence Reflects a suggestion from a moral-ethical point of view.
Interviewer skills and behaviour	Adopt a relaxed, conversational style for interviews. Establish a friendly atmosphere, that promotes the spontaneous sharing of information.	Sigstad 2018, Gjertsen 2019,	Moderate confidence Suggestions based on clinical and research experience.
	Allow sufficient time for responses to allow for slower processing time.	O'Keeffe 2019, Jen-Yi 2015, White Koning 2005, Sigstad 2018	Moderate confidence Limited empirical evidence, but fairly self-evident.
	Routinely and repeatedly offer the opportunity to withdraw when addressing sensitive or taboo topics. Especially when discomfort is noticed.	Stancliffe 2017	Moderate confidence Reflects respectful clinical and research practice.
	Be aware of social desirability, acquiescence or other answering patterns when administering questionnaires.	Jobson 2013, Williams 2007	Moderate confidence More in depth research needed on impact of different sources of bias on results.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	Do not repeat questions on which a subject has already answered satisfactorily. For the participant this may imply the answer was 'wrong'.	Cummins 1997	Moderate confidence Originates in clinical and research practice, no empirical validation.
	Regularly check if the respondent understood the question by asking to clarify or elaborate.	Perry 2002	High confidence Reflects good clinical and research practice.
	Repeat, paraphrase and summarise responses.	Sigstad 2018	Moderate confidence Empirical research needed.
Role or status of the interviewer	The interviewer should be a 'neutral' person, especially with sensitive questions.	Finlay 2001, Williams 2007	High confidence More research is needed on the impact of interviewer-interviewee dynamics on results.
	Consider the use of professionally trained peers to conduct the interviews to minimise hierarchy and social desirability effects.	Bonham 2004	Moderate confidence Promising topic for further investigation.
Role of assessor: Other/miscellaneous	The assessment should take place in a familiar environment where the participant feels at ease and there is minimal chance of distraction or pressure from others.	White Koning 2005	Moderate confidence Well-established from research and clinical practice.
<b>5. Ongoing development</b>			
Psychometric evaluation	Examine the reliability, validity and factor structure of the adapted or newly developed instrument.	Finlay 2001, Blasingame 2011	High confidence Findings are in line with good research practice.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	Responsiveness of items can be used as a measure of comprehensibility of the instrument.	Stancliffe 2015	<p>High confidence</p> <p>Finding based on large-scale population survey data and sensitive statistical analyses.</p>
	Use triangulation procedures with different informants (e.g. with relatives or professionals).	Jen-Yi 2015	<p>Low confidence</p> <p>Self-report and proxy data cannot be assumed to be interchangeable. Other scholars insist that self-report and proxy data should not be used for triangulation as they may measure entirely different concepts (e.g., Emerson, 2013)</p>

## Discussion

Research on the ability of persons with ID to provide reliable and valid accounts of their experiences, feelings and thoughts through self-reported disclosure has yielded many practical suggestions for researchers and clinicians in the past 25 years. This research covered the whole range of topics implicated in self-report instrument design and development, from the generation of relevant items to the implementation of measures in clinical and research practice.

### Quality of the evidence

Generally, the evidence base for the suggestions is not very robust. This is reflected in the very few recommendations for which the level of confidence is rated as 'high'. Some of the observed methodological and validity issues for the studies in this review include unclear sampling procedures and data synthesis strategies, small sample sizes in quantitative experimental studies and very small sample sizes in qualitative studies, contradictory findings, possibly outdated findings and references for some topics, claims that do not always seem to be substantiated by empirical evidence and generalisation of findings from specific subpopulations to the broader ID population. Furthermore, the majority of recommendations are based on clinical experience and subjective interpretations of the researchers. There is little empirical evidence for most of the recommendations, with the possible exception for some of the research on response categories.

Furthermore, some suggestions appear to be sensible at first, but upon closer inspection they are too generic to be put to practice when developing self-report measures. Examples are recommendations to 'use simple language', 'use supportive visualisation' and to 'pay attention to interviewer-interviewee dynamics'. In regard to the latter a remarkable finding is that in daily practice clinicians and researchers almost without exception take the liberty of reading the questions from self-report questionnaires aloud, even if this approach is not formalised in the manual for assessment (Lindsay et al., 2007; Stancliffe et al., 2017). Providing assistance when completing a self-report measure

may introduce various forms of bias, as a result of socially desirable or acquiescent responding, latent tendencies to please the interviewer, and other complex effects of the interaction between interviewer and interviewee, especially when dealing with sensitive topics (Kramer, 2009). The effects of this interaction on the results are not well-researched (Jen-Yi et al., 2015), barring the works of a few pioneer researchers like Antaki (1999) and Rapley and Antaki (1996).

### **Considerations for using the results to guide instrument development**

Should concerns about the ‘scientific robustness’ of the recommendations detain developers from using the results from this review? The suggestions offered in Table 2 are quite conservative in nature and generally reflect practices that are respectful towards the intended persons under study. At least for most recommendations with a high or moderate confidence level rating, the suggestions from the current review can provide *preliminary* guidance. In the absence of definitive guidelines, an obvious solution would be to directly assess the intended respondents’ understanding of the questions that are presented to them. There is support for the positive effect of this so-called teach-back method on comprehension for persons with limited reading abilities, for example of informed consent procedures (Kripalani et al., 2008) and health information (Negarandeh et al., 2013). Alternatively, cognitive interviewing techniques can be used to clarify the thought processes and struggles people with ID face when completing a self-report questionnaire, leading to improvements in the resulting measure (Miller et al., 2011). For all topics regarding content creation, the participation of persons with ID representative of the intended population should be valued as an integral part of instrument construction, as they have a unique position to reflect on the comprehensibility and acceptability of the wording, layout and visual supports of items from an ‘ID person’s viewpoint’.

### **Limitations of the research**

Several factors that limit the general application of the results to the daily practice of researchers and clinicians have been identified. First, we address two potential shortcomings of our

review methodology. Only peer-reviewed articles were included, and while this provides a degree of scientific rigour, some interesting sources of information may have been overlooked. For example, the much-referenced book on cognitive behaviour therapy for people with ID by Jahoda et al. (2017), which offers interesting insights into the ability of people with ID to self-report cognitions and feelings. Or the chapter on interviewing people with ID by Prosser and Bromley (1998), that offers guidance on conducting interviews with persons with ID. Another 'grey' source of information comes from that may have been generated by advocacy groups about accessible communication. Although relevant and interesting, guidelines offered are often not substantiated by underlying scientific research into their effectiveness or impact. Another potential threat to the validity of our results lies in the article selection procedure. For the sake of efficiency we resorted to single-author screening after the initial double-screening of a sample of 100 publications. Although we reached high IRR scores for the sample, and the selection criteria in the screening and selection protocol were clear and unambiguous, using single-author screening always increases the risk of missing key sources. On the other hand, by using backward and forward citation strategies, we feel that any relevant publications that might have been overlooked initially, would have been (and indeed have been) picked up after the screening process.

Second, the lack of information about the level of functioning of participants in some studies makes it difficult to assess the applicability of recommendations across different disability levels. Furthermore, many studies that report on the development of ID specific instruments use 'limited verbal ability' as an exclusion criterion for participants. Consequently, most resulting recommendations are supported only when applied for people with relatively better verbal ability (Hartley & MacLean, 2006; Stancliffe et al., 2014). So even though recommendations are often posited as beneficiary for 'persons with ID', upon closer inspection the recommendations seem to apply mostly to the BIF/MID population, and can be applied with much less certainty to moderate and more severe levels of ID. By routinely excluding persons with lower level of verbal abilities or cognitive functioning from this type of research,

it remains impossible to determine exactly where the boundaries of functioning for providing adequate self-reported information lie. It is clear however, that even after applying all possible adaptations to the measure, the demands placed on reasoning and comprehension skills will exceed the capabilities of persons with the severest levels of ID (Emerson et al., 2013). Acknowledging these limitations leads to the question of how to involve people with more severe levels of ID.

Third, many recommendations from the current review appear to be based upon common sense and not specific to ID participants (e.g. use clear language, involve the target population in the process, thoroughly evaluate psychometric properties, etc), and the reverse could also be true: results from different subpopulations or the general population may be in part, or even largely applicable to the ID population. Examples are research on supportive communication (Cockerill, 2002; Wilkinson & Hennig, 2007), the effect of computers and tablets in survey research (Tourangeau et al., 2017), the use of visual design in consumer research (Couper et al., 2007; Tourangeau et al., 2004), research involving people with low literacy (Chacharnovich et al., 2009; Sentell & Ratcliff-Baird, 2003), research involving children (Keefer et al., 2017; Woolley et al., 2004), and research involving persons with autism (Nicolaidis et al., 2020). Especially interesting is the field of consumer evaluation research in the general population, which has yielded plenty of insights in the use of scales and the lay-out of surveys (Cabooter et al., 2016; Velez & Ashworth, 2007). Of course, results from other study populations should never be assumed to be equally valid for persons with ID, and these results need to be replicated in empirical studies involving participants with ID. Integrating these findings from neighbouring topics appears to be a herculean endeavour, but it has the potential to greatly advance the field of self-report instrument development.

## Conclusions

It is now well established that many people with intellectual disabilities are able to provide reliable, unbiased, and valid information, through the use of self-report measures in research and clinical practice (Emerson et al., 2013). The findings from the current systematic review led to a series of recommendations about self-report instrument construction and adaptation. However, quite a few recommendations are based upon only a few studies or studies where there were methodological problems, and continued research is required. For example, what constitutes adequately simplified language, supportive visualisation or helpful interviewer support needs to be addressed. There is also a marked lack of research involving the use of self-report measures in people with moderate to more severe ID, meaning that most of the recommendations made within the current systematic review are in relation to those with borderline to mild ID.

Acknowledging the need to make self-report research and practice accessible for people with ID may lead to a Solomon's judgement for developers: should they attune the measure to the needs of the intended target population, without being able to compare results with those found in research in a broader population or based on validated norms, or should they stick to using original instruments, potentially excluding a large proportion of intended participants? The solution to this might be to aim for *ID-inclusive* measures, that would at least be suitable for most persons with mild ID, instead of *ID-specific* versions of instruments. All of the proposed adaptations to measures included in this review can easily be applied without losing the measure's suitability for the use in the general population. This would not only benefit the interests of persons with ID but would also mean the measure could be more suitable for persons with other characteristics that may impede their ability to complete self-report measures. For instance, persons with low literacy levels, dyslexia or acquired brain impairments. Adopting an ID-inclusive approach for the construction process could prove to be advantageous to both developers and the ID community and extend to persons with other impairments.

## Acknowledgements

The authors wish to thank Ms. Janneke Staaks, information specialist at the University of Amsterdam, for her invaluable guidance with drawing up and executing the search strategy. They further wish to thank Ms. Martina de Witte PhD, for providing very valuable feedback on earlier versions of the manuscript.

## References<sup>1</sup>

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington VA: Author.

\*Antaki, C. (1999). Interviewing persons with a learning disability: How setting lower standards may inflate well-being scores. *Qualitative Health Research, 9*(4), 437-454.

\*Antaki, C., & Rapley, M. (1996). Questions and answers to psychological assessment schedules: Hidden troubles in 'quality of life' interviews. *Journal of Intellectual Disability Research, 40*(5), 421-437.

Aromatis, E., & Munn, Z. (2017). Joanna briggs institute reviewer's manual. Retrieved from <https://reviewersmanual.joannabriggs.org/>

Barker, M.S., Bidstrup, E.M., Robinson, G.A., & Nelson, N.L. (2020). "Grumpy" or "furious"? Arousal of emotion labels influences judgments of facial expressions. *PLoS ONE 15*(7): e0235390.

\*Bell, N., Tonkin, M., Chester, V., & Craig, L. (2018). Adapting measures of social climate for use with individuals with intellectual developmental disability in forensic settings. *Psychology Crime & Law, 24*(4), 362-378.

---

<sup>1</sup> Studies marked with an asterisk (\*) are included in the systematic review.

- \*Blasingame, G. D., Abel, G. G., Jordan, A., & Wiegel, M. (2011). The utility and psychometric properties of the abel-blasingame assessment system for individuals with intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities, 4*(2), 107-132.
- \*Boland, M., Daly, L., & Staines, A. (2008). Methodological issues in inclusive intellectual disability research: A health promotion needs assessment of people attending irish disability services. *Journal of Applied Research in Intellectual Disabilities, 21*(3), 199-209.
- \*Bonham, G., Basehart, S., Schalock, R., Marchand, C., Kirchner, N., & Rumenap, J. (2004). Consumer-based quality of life assessment: The maryland ask me! project. *Mental Retardation, 42*(5), 338-355.
- \*Bowles, P. V., & Sharman, S. J. (2014). The effect of different types of leading questions on adult eyewitnesses with mild intellectual disabilities. *Applied Cognitive Psychology, 28*(1), 129-134.
- Bredel, U., & Maass, C. (2016). *Leichte Sprache. Theoretische Grundlagen*. Berlin: Duden.
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research, 69*(7), 2574-2584.
- Carroll, C., Booth, A., Leaviss, J., & Rick, J. (2013). "Best fit" framework synthesis: Refining the method. *BMC Medical Research Methodology, 13*(1), 37.
- Chacharnovich, E., Fleck, M. P., & Power, M. (2009). Literacy affected ability to adequately discriminate among categories in multipoint Likert Scales. *Journal of Clinical Epidemiology, 62*(1), 37-46.

\*Clark, L., Pett, M. A., Cardell, E. M., Guo, J., & Johnson, E. (2017). Developing a health-related quality-of-life measure for people with intellectual disability. *Intellectual and Developmental Disabilities, 55*(3), 140-153.

Cockerill, H. (2002). Supporting communication in the child with a learning disability. *Current Paediatrics, 12*(1), 72-76.

Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual context effects in web surveys. *Public Opinion Quarterly, 71*(4), 623-634.

\*Cummins, R. (1997). Self-rated quality of life scales for people with an intellectual disability: A review. *Journal of Applied Research in Intellectual Disabilities, 10*(3), 199-216.

Cummins, R. A. (2002). Proxy responding for subjective well-being: A review. *International review of research in mental retardation, 25*, 183-207.

\*Cuskelly, M., Moni, K., Lloyd, J., & Jobling, A. (2013). Reliability of a method for establishing the capacity of individuals with an intellectual disability to respond to likert scales. *Journal of Intellectual & Developmental Disability, 38*(4), 318-324.

\*Cuthill, F. M., Espie, C. A., & Cooper, S. (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.

\*Dagnan, D., & Ruddick, L. (1995). The use of analogue scales and personal questionnaires for interviewing people with learning disabilities. *Clinical Psychology Forum, 21-24*.

- Daniel, M. R., Sadek, S. A., & Langdon, P. E. (2018). The reliability and validity of a revised version of the How I Think Questionnaire for men who have intellectual disabilities. *Psychology, Crime & Law, 24*(4), 379-390.
- \*de Knecht, N. C., Evenhuis, H. M., Lobbezoo, F., Schuengel, C., & Scherder, E. J. A. (2013). Does format matter for comprehension of a facial affective scale and a numeric scale for pain by adults with down syndrome? *Research in Developmental Disabilities, 34*(10), 3442-3448.
- \*de Knecht, N. C., Lobbezoo, F., Schuengel, C., Evenhuis, H. M., & Scherder, E. J. A. (2016). Self-reporting tool on pain in people with intellectual disabilities (STOP-ID!): A usability study. *Augmentative and Alternative Communication, 32*(1), 1-11.
- \*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.
- \*Fang, J., Fleck, M. P., Green, A., McVilly, K., Hao, Y., Tan, W., . . . Power, M. (2011). The response scale for the intellectual disability module of the WHOQOL: 5-point or 3-point? *Journal of Intellectual Disability Research, 55*, 537-549.
- \*Finlay, W., & Lyons, E. (2001). Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. *Psychological Assessment, 13*(3), 319-335.
- \*Finlay, W., & Lyons, E. (2002). Acquiescence in interviews with people who have mental retardation. *Mental Retardation, 40*(1), 14-29.

Freedman, R. I. (2001). Ethical challenges in the conduct of research involving persons with mental retardation. *Mental Retardation*, 39(2), 130-141.

Fujiura, G. T., Andresen, E., Cardinal, B. J., Drum, C. E., Fujiura, G. T., Hall, T., . . . RRTC Expert Panel Hlth Measurement. (2012). Self-reported health of people with intellectual disability. *Intellectual and Developmental Disabilities*, 50(4), 352-369.

\*Gjertsen, H.G. (2019). People with intellectual disabilities can speak for themselves! a methodological discussion of using people with mild and moderate intellectual disabilities as participants in living conditions studies. *Scandinavian Journal of Disability Research*, 21(1), 141-149.

\*Glenn, E., Bihm, E. M., & Lammers, W. J. (2003). Depression, anxiety, and relevant cognitions in persons with mental retardation. *Journal of Autism and Developmental Disorders*, 33(1), 69-76.

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-ID s). *British Journal of Clinical Psychology*, 53(3), 315-332.

\*Hartley, S. L., & MacLean, W. E., Jr. (2006). A review of the reliability and validity of likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research*, 50, 813-827.

\*Heal, L. W., & Sigelman, C. K. (1995). Response biases in interviews of individuals with limited mental-ability. *Journal of Intellectual Disability Research*, 39, 331-340.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., . . . Nicolau, B. (2018). Mixed methods appraisal tool (MMAT), version 2018. *Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.*

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.

Huus, K., Granlund, M., Bornman, J., & Lyngnegard, F. (2015). Human rights of children with intellectual disabilities: Comparing self-ratings and proxy ratings. *Child Care Health and Development, 41*(6), 1010-1017.

\*Ikeda, E., Kraegeloh, C., Water, T., & Hinckson, E. A. (2016). An exploratory study of self-reported quality of life in children with autism spectrum disorder and intellectual disability. *Child Indicators Research, 9*(1), 133-153.

Jahoda, A., Stenfert-Kroese, B., & Pert, C. (2017). *Cognitive Behaviour Therapy for People with Intellectual Disabilities*. London, UK: Palgrave Macmillan.

\*Jen-Yi, L., Krishnasamy, M., & Der-Thanq, C. (2015). Research with persons with intellectual disabilities: An inclusive adaptation of Tourangeau's model. *Alter, 9*(4), 304-316.

\*Jobson, L., Stanbury, A., & Langdon, P. E. (2013). The self- and other-deception questionnaires- intellectual disabilities (SDQ-ID and ODQ-ID): Component analysis and reliability. *Research in Developmental Disabilities, 34*(10), 3576-3582.

Keefer, A., Kreiser, N. L., Singh, V., Blakeley-Smith, A., Duncan, A., Johnson, C., Klinger, L., Meyer, A., Reaven, J., & Vasa, R. A. (2017). Intolerance of uncertainty predicts anxiety outcomes following CBT in youth with ASD. *Journal of Autism and Developmental Disorders, 47*(12), 3949-3958.

- \*Keeling, J. A., Rose, J. L., & Beech, A. R. (2007). A preliminary evaluation of the adaptation of four assessments for offenders with special needs. *Journal of Intellectual & Developmental Disability, 32*(2), 62-73.
- Kellett, S. C., Beail, N., Newman, D. W., & Mosley, E. (1999). Indexing psychological distress in people with an intellectual disability: Use of the symptom checklist-90-R. *Journal of Applied Research in Intellectual Disabilities, 12*(4), 323-334.
- \*Kent, E. C., Burgess, G. H., & Kilbey, E. (2018). Using the AQ-10 with adults who have a borderline or mild intellectual disability: Pilot analysis of an adapted AQ-10 (AQ-10-intellectual disability). *Research in Autism Spectrum Disorders, 54*, 65-75.
- Kramer, J. M. (2009). A mixed methods approach to building validity evidence: The child occupational self assessment. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69*(12-B), 7458.
- Kripalani, S., Bengtzen, R., Henderson, L. E., & Jacobson, T. A. (2008). Clinical research in low-literacy populations: using teach-back to assess comprehension of informed consent and privacy information. *IRB: Ethics & Human Research, 30*(2), 13-19.
- Langdon, P. E., Clare, I. C., & Murphy, G. H. (2010). Developing an understanding of the literature relating to the moral development of people with intellectual disabilities. *Developmental Review, 30*(3), 273-293.
- Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M.A., Tunçalp, Ö., Colvin, C.J., Garside, R., Carlsen, B., Langlois, E.V., & Noyes, J. (2018). Applying GRADE-CERQual

to qualitative evidence synthesis findings: introduction to the series. *Implementation Science*, 13 (2), 1-10.

Lewis, C., & Morrissey, C. (2010). The association between self-report and informant reports of emotional problems in a high secure intellectual disability sample. *Advances in Mental Health and Intellectual Disabilities*, 4(2), 44-49.

Lindsay, W. R., & Michie, A. M. (1988). Adaptation of the zung self-rating anxiety scale for people with a mental handicap. *Journal of Mental Deficiency Research*, 32(6), 485-490.

Lindsay, W. (2002). Research and literature on sex offenders with intellectual and developmental disabilities. *Journal of Intellectual Disability Research*, 46, 74-85.

Lindsay, W. R., Whitefield, E., & Carson, D. (2007). An assessment for attitudes consistent with sexual offending for use with offenders with intellectual disabilities. *Legal and Criminological Psychology*, 12, 55-68.

\*Marshall, K., & WilloughbyBooth, 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.

\*Matikka, L. M., & Vesala, H. T. (1997). Acquiescence in quality-of-life interviews with adults who have mental retardation. *Mental Retardation*, 35(2), 75-82.

**Mileviciute, I., & Hartley, S. (2015). Self-reported versus informant-reported depressive symptoms in adults with mild intellectual disability. *Journal of Intellectual Disability Research*, 59(2), 158-169.**

- Miller, K., Mont, D., Maitland, A., Altman, B., & Madans, J. (2011). Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality & quantity, 45*(4), 801-815.
- Mindham, J., & Espie, C. (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.
- Munthe-Kaas, H., Bohren, M. A., Glenton, C., Lewin, S., Noyes, J., Tunçalp, Ö., Booth, A., Garside, R., Colvin, C.J., Wainwright, M., Rashidian, A., Flottorp, S., & Carlsen, B. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 3: how to assess methodological limitations. *Implementation Science, 13*(1), 25-32.
- Negarandeh, R., Mahmoodi, H., Noktehdan, H., Heshmat, R., & Shakibazadeh, E. (2013). Teach back and pictorial image educational strategies on knowledge about diabetes and medication/dietary adherence among low health literate patients with type 2 diabetes. *Primary care diabetes, 7*(2), 111-118.
- Nicolaidis, C., Raymaker, D. M., McDonald, K. E., Lund, E. M., Leotti, S., Kapp, S. K., Katz, M., Beers, L.M., Kripke, C., Maslak, J., Hunter, M., & Zhen, K. Y. (2020). Creating accessible survey instruments for use with autistic adults and people with intellectual disability: Lessons learned and recommendations. *Autism in Adulthood, 2*(1), 61-76.
- \*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.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan: a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210.

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International journal of nursing studies*, 49(1), 47-53.

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., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.

<https://doi.org/10.31222/osf.io/gwdhk>

\*Payne, R., & Jahoda, A. (2004). The glasgow social self-efficacy scale--A new scale for measuring social self-efficacy in people with intellectual disability. *Clinical Psychology & Psychotherapy*, 11(4), 265-274.

\*Perkins, E. A. (2007). Self-and proxy reports across three populations: Older adults, persons with alzheimer's disease, and persons with intellectual disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, 4(1), 1-10.

\*Perry, J., & Felce, D. (2002). Subjective and objective quality of life assessment: Responsiveness, response bias, and resident: Proxy concordance. *Mental Retardation*, 40(6), 445-456.

Prosser, H., & Bromley, J. (2012). Interviewing people with intellectual disabilities. In Emerson, E., Hatton, C., Dickson, K., Gone, R., Caine, A., & Bromley, J. (Eds.), *Clinical psychology and people with intellectual disabilities (pp 107-121)*. Hoboken, New Jersey: John Wiley & Sons.

\*Ramirez, S. Z., & Lukenbill, J. (2008). Psychometric properties of the zung self-rating anxiety scale for adults with intellectual disabilities (SAS-ID). *Journal of Developmental and Physical Disabilities, 20(6)*, 573-580.

\*Ramirez, S. Z., & Lukenbill, J. F. (2007). Development of the fear survey for adults with mental retardation. *Research in Developmental Disabilities, 28(3)*, 225-237.

\*Ramirez, S. (2005). Evaluating acquiescence to yes-no questions in fear assessment of children with and without mental retardation. *Journal of Developmental and Physical Disabilities, 17(4)*, 337-343.

Rapley, M., & Antaki, C. (1996). A conversation analysis of the acquiescence of people with learning disabilities. *Journal of Community & Applied Social Psychology, 6(3)*, 207-227.

\*Reid, G., Vallerand, R. J., Poulin, C., & Crocker, P. (2009). The development and validation of the pictorial motivation scale in physical activity. *Motivation and Emotion, 33(2)*, 161-172.

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic reviews, 10(1)*, 1-19.

Rose, J., Willner, P., Shead, J., Jahoda, A., Gillespie, D., Townson, J., Lammie, C., Woodgate, C., Stenfert Kroese, B., Felce, D., MacMahon, P., Rose, N., Stimpson, A., Nuttall, J., & Hood, K. (2013). Different

factors influence self-reports and third-party reports of anger by adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 26(5), 410-419.

Schalock, R. L., Bonham, G. S., & Verdugo, M. A. (2008). The conceptualization and measurement of quality of life: Implications for program planning and evaluation in the field of intellectual disabilities. *Evaluation and Program Planning*, 31(2), 181-190.

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

\*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.

\*Scott, H. M., & Havercamp, S. M. (2018). Comparison of self- and proxy report of mental health symptoms in people with intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities*, 11(2), 143-156.

Sentell, T. L., & Ratcliff-Baird, B. (2003). Literacy and comprehension of beck depression inventory response alternatives. *Community Mental Health Journal*, 39(4), 323-331.

Sigelman, C. K., Budd, E. C., Spanhel, C. L., & Schoenrock, C. J. (1981). When in doubt, say yes: Acquiescence in interviews with mentally retarded persons. *Mental Retardation*, 19(2), 53.

\*Sigstad, H. M. H., & Garrels, V. (2018). Facilitating qualitative research interviews for respondents with intellectual disability. *European Journal of Special Needs Education*, 33(5), 692-706.

- \*Stancliffe, R. J., Wiese, M. Y., Read, S., Jeltos, G., & Clayton, J. M. (2017). Assessing knowledge and attitudes about end of life: Evaluation of three instruments designed for adults with intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 30*(6), 1076-1088.
- \*Stancliffe, R. J., Wilson, N. J., Bigby, C., Balandin, S., & Craig, D. (2014). Responsiveness to self-report questions about loneliness: A comparison of mainstream and intellectual disability-specific instruments. *Journal of Intellectual Disability Research, 58*(5), 399-405.
- \*Stancliffe, R. J., Ticha, R., Larson, S. A., Hewitt, A. S., & Nord, D. (2015). Responsiveness to self-report interview questions by adults with intellectual and developmental disability. *Intellectual and Developmental Disabilities, 53*(3), 163-181.
- Stenfert-Kroese, B. S., Dagnan, D., & Loumidis, K. (1997). *Cognitive-behaviour therapy for people with learning disabilities* (pp. 13-27). Oxford: Routledge.
- Stenfert-Kroese, B., Gillott, A., & Atkinson, V. (1998). Consumers with intellectual disabilities as service evaluators. *Journal of Applied Research in Intellectual Disabilities, 11*(2), 116-128.
- Taal voor Allemaal (2021). <https://www.taalvoorallemaal.com>.
- Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68*(3), 368-393.
- Tourangeau, R., Maitland, A., Rivero, G., Sun, H., Williams, D., & Yan, T. (2017). Web surveys by smartphone and tablets: Effects on survey responses. *Public Opinion Quarterly, 81*(4), 896-929.

\*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.

UK Department Of Health. (2010). Making written information easier to understand for people with learning disabilities guidance for people who commission or produce easy read information – revised edition 2010. Retrieved from <https://webarchive.nationalarchives.gov.uk/20130703133435/http://odi.dwp.gov.uk/docs/iod/easy-read-guidance.pdf>

Velez, P., & Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods, 1*(2) 69-74.

Vlissides, N., Golding, L., & Beail, N. (2016). A systematic review of the outcome measures used in psychological therapies with adults with ID. *Psychological Therapies and People Who have Intellectual Disabilities, 115-139*.

\*Vlot-van Anrooij, K., Tobi, H., Hilgenkamp, T. I. M., Leusink, G. L., & Naaldenberg, J. (2018). Self-reported measures in health research for people with intellectual disabilities: An inclusive pilot study on suitability and reliability. *Bmc Medical Research Methodology, 18*, 80.

\*White-Koning, M., Arnaud, C., Bourdet-Loubere, S., Bazex, H., Colver, A., & Grandjean, H. (2005). Subjective quality of life in children with intellectual impairment - how can it be assessed? *Developmental Medicine and Child Neurology, 47*(4), 281-285.

Wieland, J., Wardenaar, K. J., Fontein, E., & Zitman, F. G. (2012). Utility of the brief symptom inventory (BSI) in psychiatric outpatients with intellectual disabilities. *Journal of Intellectual Disability Research, 56*(9), 843-853.

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.

Wilkinson, K. M., & Hennig, S. (2007). The state of research and practice in augmentative and alternative communication for children with developmental/intellectual disabilities. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 58-69.

\*Williams, F., Wakeling, H., & Webster, S. (2007). A psychometric study of six self-report measures for use with sexual offenders with cognitive and social functioning deficits. *Psychology Crime & Law, 13*(5), 505-522.

Woolley, M., Bowen, G., & Bowen, N. (2004). Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. *Research on Social Work Practice, 14*(3), 191-200.

Zabalia, M. (2013). Beyond misconceptions: Assessing pain in children with mild to moderate intellectual disability. *Frontiers in Public Health, 1*, 23.

## Appendix A

### *Characteristics of Included Studies*

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
Literature reviews					
Bell 2018	Reported challenges in adapting self-report measures for people with ID as part of constructing a living climate questionnaire for people with ID.	Level of ID not specified	JBI Checklist for Text and Opinion	5/6	Some incongruence with the literature/sources Some recommendations from advocacy groups, not substantiated by empirical research
Cummins 1997	Review on QOL instruments for people with ID.	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	Many references possibly outdated.
Emerson 2013	Examined two methodological issues regarding ways of obtaining and analysing outcome data for people with ID: (a) self-report and proxy-report data and (b) analysis of population-based data sets.	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	
Finlay 2001	Review on methodological issues when using self-reports for people with ID	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	
Finlay 2002	Addressed the question of acquiescence in people with ID.	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	
Heal 1995	Reviewed several experiments to assess acquiescence in people with ID.	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	Many references possibly outdated.

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
Kroese 1998	Review of the possibilities for people with ID to actively contribute to consumer evaluations of services.	Not specified, but includes people with very limited verbal abilities	JBI Checklist for Text and Opinion	5/6	Some contradictory findings, many references possibly outdated.
Perkins 2007	Described the differences between self-reports and proxy-reports for different subpopulations.	Persons with Alzheimer's, ID and 'older adults'	JBI Checklist for Text and Opinion	5/6	References to possibly outdated literature whilst overlooking more then-recent studies.
Schalock 2002	Recommendations about the conceptualization, measurement and application of Quality of Life for persons with ID by an international panel of experts (IAADD).	Level of ID not specified	JBI Checklist for Text and Opinion	5/6	Several claims not substantiated with then-recent literature references.
Sigstad 2018	Explored what researchers can do to facilitate communication between interviewers and respondents.	Level of ID not specified	JBI Checklist for Text and Opinion	6/6	
White Koning 2005	Review on conceptual and methodological issues in relation to self-report assessment in children with cerebral palsy and cognitive impairments.	Children with cerebral palsy and ID (level not specified)	JBI Checklist for Text and Opinion	5/6	Some results of studies with adults applied to recommendations for children

**Literature reviews combined with qualitative study**

Gjertsen 2019	Discussion of the methodological challenges of carrying out living conditions studies involving persons with ID, on the basis of literature review and researcher experiences from the survey.	Level of ID not specified (n=93)	JBI Checklist for Text and Opinion + MMAT - Qualitative	6/11	Some recommendations based on clinical experience from a single researcher. Some recommendations from review not substantiated by reference to literature. No clear description of qualitative method.
---------------	--	----------------------------------	---	------	--

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
Jen-Yi 2015	Expanded on Tourangeau's model of survey responses to facilitate use of self-reports for people with ID.	Level of ID not specified for review, Persons in qualitative study: MID (n=106)	JBIChecklist for Text and Opinion + MMAT - Qualitative	8/11	Method description for qualitative research offers limited information on study sample. No clear qualitative synthesis strategy. Study findings from young persons with MID are generalised to the broader ID population.
O'Keeffe 2019	Described the process of developing a grief scale for people with ID.	Level of ID not specified (n=16)	JBIChecklist for Text and Opinion + MMAT - Qualitative	10/11	No clear description of study population
Systematic reviews					
Hartley 2006	Review of the reliability and validity of Likert-type scales for people with ID.	Adolescents (>11 yrs) and adults with borderline intelligent functioning to profound ID	JBIChecklist for Systematic Reviews	9/11	No clear description of search strategy. No quality appraisal of included studies.
Townsend-White 2012	Systematic review of QOL measures for people with ID.	MID to Moderate ID	JBIChecklist for Systematic Reviews	10/11	No quality appraisal of included studies.
Quantitative experimental studies					
Bowles 2014	Examined the effect of misleading information in interview questions on accuracy of responses.	Mild ID (n=41)	MMAT – Quantitative Non randomized Studies	4/5	Persons with MID are assumed to represent the ID population as a whole. Small sample size.

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
Jobson 2013	Investigated the psychometric properties of the SDQ-ID and ODQ-ID, and several aspects of social desirability.	Mild to borderline ID (n=100)	MMAT – Quantitative Non-randomized Studies	5/5	
Keeling 2007	Described the adaptation of 4 measures for use with sexual offenders, many of whom have intellectual disabilities.	Borderline intellectual functioning to mild ID (n=69)	MMAT – Quantitative Non-randomized Studies	5/5	
Knegt 2013	Compared different modalities for the assessment of pain in people with Down Syndrome.	Persons with Down Syndrome, mostly moderate ID (n = 106)	MMAT – Quantitative Non-randomized Studies	5/5	
Knegt 2017	Usability study on a self-reporting tool for pain (STOP-ID) in a sample of persons with Down syndrome.	Mild to Severe ID (n=40)	MMAT – Quantitative Non-randomized Studies	4/5	Not clear if results in study with persons with Down syndrome are representative of the ID population as a whole. Small sample size.
Marshal 2007	Described the modification of a self-report routine outcome measuring instrument for people with ID.	Mild to moderate ID (n=22)	MMAT – Quantitative Non-randomized Studies	3/5	Equivalence of participants between groups was not established. Small sample size. Statistical inferences made on basis of visual analyses (power too low for inferential testing).
Payne 2004	Described the development of the GSSES self-efficacy scale.	Level of ID not specified, no autism, verbally able (n=20 for focus groups, n=77 for main study of whom 38 were persons with ID)	MMAT - Mixed Methods	3/5	Stringent exclusion criteria may introduce nonresponse bias. Small sample size for the quantitative study.

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal	Quality Appraisal score	Quality Appraisal issues
Power 2010	Described the 'disabilities version' of the WHO QOL survey; the WHOQOL-Dis.	Level of ID not specified (n=491)	MMAT – Quantitative Non-randomized Studies	5/5	
Ramirez 2005	Examined the supposed acquiescence response set present in survey data for people with ID on yes/no questions.	Children with ID (borderline intellectual functioning to moderate ID, n=75) and without ID (n=240) were compared	MMAT – Quantitative Non-randomized Studies	5/5	
Reid 2009	Described the development of a pictorial motivation scale for adolescents and adults with ID.	Young persons with ID who have difficulties reading (IQ not disclosed to authors, estimated mild to moderate ID; n=240)	MMAT – Quantitative Non-randomized Studies	3/5	Participants were 'deemed to have an intellectual disability' because of their placement in a special education setting, but this was not formally assessed. Marked differences in subgroup characteristics between EFA and CFA analyses that are not accounted for.
Quantitative observational studies					
Blasingame 2011	Described the construction and validation of the ABID, an instrument for evaluating sexual behaviour problems among individuals with intellectual disabilities.	Mild to moderate ID (n=495)	MMAT - Quantitative descriptive	5/5	
Bonham 2004	Reported on a project in which people with developmental disabilities were trained to survey other consumers' perceived quality of life.	Borderline intelligent functioning to profound ID (n=923)	MMAT - Quantitative descriptive	5/5	

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal	Quality Appraisal score	Quality Appraisal issues
Cuskelly 2013	Examined the reliability of a method for establishing the capacity of individuals with an intellectual disability to respond to Likert scales.	ID, not specified (n=33)	MMAT - Quantitative descriptive	4/5	A convenience sample was used without comparing sample characteristics to the ID population as a whole. Small sample size.
Cuthill 2003	Development of a scale for depressive symptoms for people with ID.	Mild to moderate ID (n=65)	MMAT - Quantitative descriptive	5/5	
Dagnan 1995	Investigated the reliability of analogue scales and personal questionnaires.	Mild to moderate ID (n=29)	MMAT - Quantitative descriptive	3/5	Respondents deemed eligible for inclusion by carers were selected. Small sample size.
Fang 2011	Examined the psychometric properties of 3- and 5-point scales on the WHOQOL.	Mild to moderate ID (n=329)	MMAT - Quantitative descriptive	5/5	
Glenn 2003	Assessment of depression, anxiety, and relevant cognitions in persons with ID by administering modified versions of several self-report measures.	Borderline to moderate mental retardation (n=46)	MMAT - Quantitative descriptive	3/5	Respondents forwarded by director of facility on basis of 'willingness to cooperate'. Sample strategy not in line with research question.
Matikka 1997	Study into the prevalence of acquiescent responding in QOL research.	Borderline intellectual functioning to profound ID (n=662)	MMAT - Quantitative descriptive	4/5	Questions about parts of the operationalisation.
Perry 2002	Detailed analyses of a QOL survey, that focussed on response bias and responsiveness for varying degrees of disability.	Persons with lower adaptive functioning (ABS) (n=154)	MMAT - Quantitative descriptive	5/5	
Ramirez 2008	Described the psychometric properties of the Zung Self-Rating Anxiety Scale for adults with intellectual disabilities (SAS-ID).	Mild to moderate ID (n=137)	MMAT - Quantitative descriptive	5/5	

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if Tool applicable)	Quality Appraisal	Quality Appraisal score	Quality Appraisal issues
Scott 2018	Compared self-report and proxy measures of mental health problems in people with ID.	'Broad range of ID functioning' (n=90)	MMAT - Quantitative descriptive	5 / 5	
Stancliffe 2014	Compared mainstream instrument to measure loneliness with ID specific instrument.	Mild to moderate ID (n=56)	MMAT - Quantitative descriptive	5 / 5	
Stancliffe 2015	Examined the responsiveness to self-report interview questions by adults with ID	Mild to moderate ID (n=11.391)	MMAT - Quantitative descriptive	5 / 5	
Williams 2007	A psychometric study of six adapted self-report measures for use with sexual offenders with cognitive and social functioning deficits	Borderline intellectual functioning to moderate ID (n=211)	MMAT - Quantitative descriptive	5 / 5	
Mixed-method studies					
Boland 2008	Described methodological issues when conducting inclusive research with ID Clients in health care.	Mild to moderate ID (n = 12 for focus groups, n=247 for quantitative study)	MMAT - Mixed Methods	15/15	Detailed description of methodology used
Clark 2017	Described the development of a new measure of health-related quality of life (HRQOL) for individuals with ID.	Level of ID not specified (n=26 for qualitative study, n=103 for quantitative study)	MMAT - Mixed Methods	13/15	No description of sample characteristics in qualitative study. No data synthesis strategy described for qualitative study.
Kent 2018	Compared psychometric properties of original ASS self-report screener with Easy-Read version .	Mild to severe ID (n=6 for focus group, n = 52 for quantitative study)	MMAT – Mixed-methods	14/15	Small sample size for quantitative study (although deemed appropriate by authors on the basis of power analysis).

First author and year of publication	Summary	Target population level of ID (and <i>n</i> if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
Ramirez 2007	Described the development of the Fear Survey for Adults with Mental Retardation.	Qualitative study: Level of ID not specified (focus group, <i>n</i> = 17). Quantitative study: Mild to moderate ID ( <i>n</i> =138)	MMAT - Mixed Methods	12/15	No clear description of focus group method and outcomes.
Stancliffe 2017	Described the development of End of Life Questionnaires for people with ID.	Mild to moderate ID ( <i>n</i> =11 for focus groups, <i>n</i> =78 for main study of whom 38 persons with ID)	MMAT - Mixed Methods	14/15	Small sample size for quantitative study.

#### Qualitative studies

Antaki 1996	Conversational analysis of QOL interviews with people with ID.	Mild to moderate ID	MMAT - Qualitative	5/5	
Antaki 1999	Conversational analysis of QOL interviews with people with ID.	Mild to moderate ID	MMAT - Qualitative	5/5	
Ikeda 2014	Described key themes for the development of QOL/HRQOL self-reports in children with ASD and ID based on interviews and focus groups.	Children with ASD and ID (IQ>50, <i>n</i> =10) and proxies ( <i>n</i> =17)	MMAT - Qualitative	3/5	Small scale focus groups. Recommendations predominantly from statements by proxy participants, limited contribution by participants with ID.
Vlot van Anroij et al., 2018	An inclusive pilot study on suitability and reliability of several self-reported measures in health research.	MID ( <i>n</i> =40)	MMAT - Qualitative	4/5	Sample characteristics not clear. Very specific sample population (Special Olympics participants), results may not generalise to broader ID population.