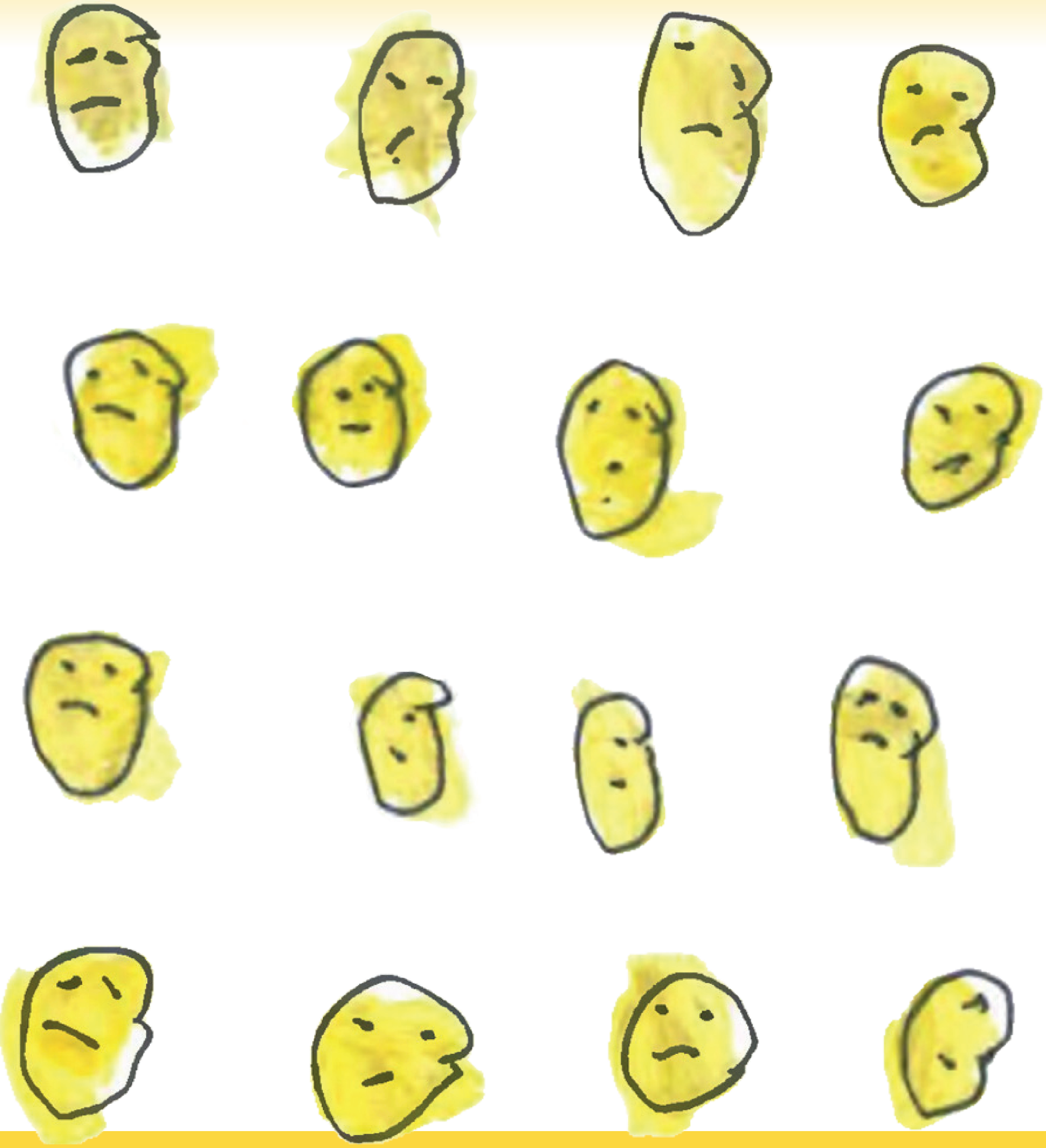
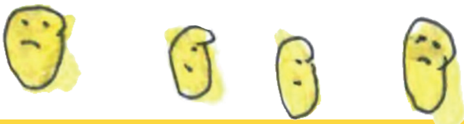
Why ask them? It's about me!

Tailoring self-report instruments to the needs, preferences, and capabilities of people with intellectual disabilities





Roel Kooijmans

Why ask them? It’s about me!

Tailoring self-report instruments to the needs, preferences, and capabilities of people with intellectual disabilities

Roel Kooijmans

De mooie, lieve en krachtige illustraties bij dit proefschrift zijn gemaakt door mensen die bij Uniek in Maasbracht en de Piahoeve in Heel mooie dingen maken als onderdeel van hun dagbesteding. Ze verkopen allerlei leuke zelfgemaakte kunstwerken en bakken heerlijke taart. Ga eens langs voor een bakkie en een praatje!

Het onderzoek in dit proefschrift is uitgevoerd in opdracht van en in samenwerking met Koraal. Dat is een organisatie die ondersteuning, zorg en onderwijs biedt aan mensen van alle leeftijden met ernstige (verstandelijke) beperkingen en complexe gedrags- en/of psychiatrische problemen.

Print&Bind verzorgde het drukwerk.

Why ask them? It's about me!

Tailoring self-report instruments to the needs, preferences, and capabilities of people with intellectual disabilities

Academisch Proefschrift

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College van Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op vrijdag 24 oktober 2025 te 16.00 uur

door Roel John Lily Kooijmans

geboren te Tilburg

Promotiecommissie

Promotores

prof. dr. X.M.H. Moonen	Universiteit van Amsterdam
prof. dr. P.E. Langdon	University of Birmingham

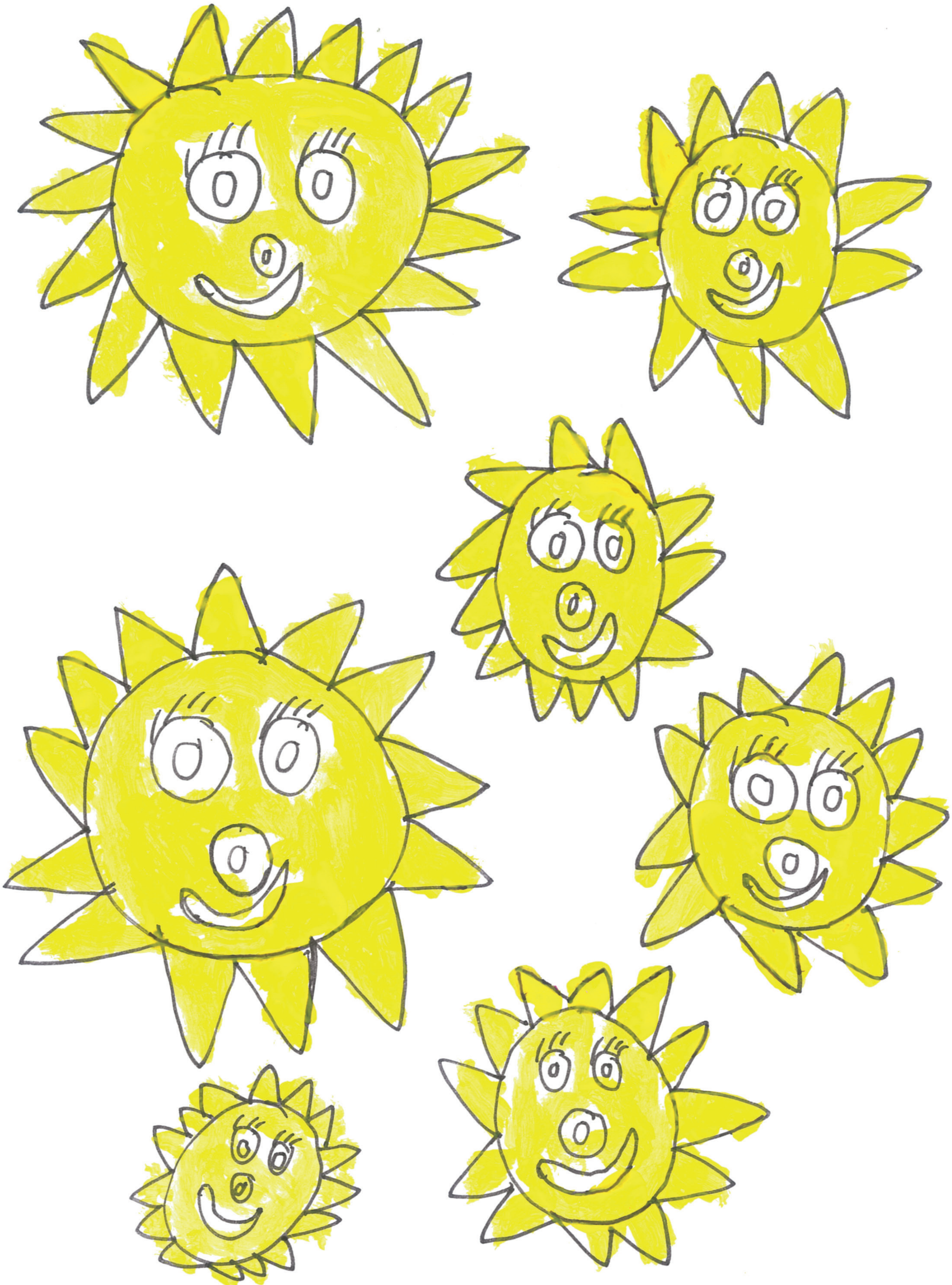
Overige leden

prof. dr. P.S. Sterkenburg	Vrije Universiteit Amsterdam
dr. N. Frielink	Tilburg University
prof. dr. M. van Nieuwenhuijzen	Universiteit van Amsterdam
dr. H.E. Creemers	Universiteit van Amsterdam
prof. dr. C.H.Z. Kuiper	Universiteit van Amsterdam

Faculteit der Maatschappij- en Gedragswetenschappen

Contents

Preface



In my early days as a professional at Koraal, I worked as a care worker in a residential treatment group with six young people. These young people all had mild or moderate intellectual disabilities and additional complex psychiatric problems. There was a lot of aggression between young people and towards us, the carers. Most of my colleagues felt it was important to establish authority and suppress any tendency to oppositional behaviour. Carers put themselves above the youngsters and did not tolerate contradiction. The use of physical restraints and seclusion by carers when youths did not conform to authority were the order of the day.

In the quiet moments, during a walk in the forest, playing a game of table tennis or Mario Kart on the WII, I was increasingly amazed at the two faces these youngsters could show. In one-on-one interactions, these defiant and insolent ‘problem youths’ could suddenly turn into vulnerable children with worries, sorrows and also dreams and wishes. All I had to do was to stand *beside* rather than *above* the young person in that moment. Not talking, just being present, not telling them what they should do, but listening to what they had to say. As it is difficult for many young people with cognitive disabilities to put into words what they are thinking and feeling, this meant it was important to be able to listen carefully to the ‘story behind the story’. And in doing so, take for granted that this story can be brought with little nuance (sometimes with a lot of cursing and shouting…).

Convinced that the voice of ‘our young people’ needed to be heard, I applied for the position of youth council coach at De Hondsberg in 2005. In my almost 10 years with the youth council, my colleague Jeroen (later Marion) and I helped the young people to make clear to management what is important for young people living in a group home. The members of the youth council, who generally showed a lot of aggression and destructive tendencies in their group homes, showed that they were perfectly able to discuss annual plans and budgets with management in an orderly and constructive way. The answers you got from the young people did, however, depend in large part on the way you asked the question. Not all colleagues were able to formulate a question in such a way that an individual with a cognitive impairment understood what was meant. Adapting to the other person’s level of communication turned out to be an art in itself. Over the years, we got ever better at it, and we came up with all kinds of ways to facilitate conversation about complex subjects with young people with cognitive impairments. Sometimes using images, sometimes using videos instead of text and always using language adapted to the recipient.

In that time, we contributed to several studies that were supervised by Professor Xavier Moonen. His passion, like ours, was to look for ways to make the voices of young people with cognitive disabilities heard. We supplied experiences from our youths and our daily practice, Xavier worked from a scientific researcher perspective.

Despite our positive experiences, many colleagues appeared to remain quite sceptical about involving youths with cognitive impairments in participatory and policy-making projects. But in our view, their struggle to have meaningful conversations and obtain valid and reliable results was not to be blamed on the youths' inability to provide meaningful answers. Instead, we felt that if they were getting the wrong answers, they weren't asking the right questions, and maybe even more importantly, they were not asking the questions in the right way.

An example of the consequences of asking questions the wrong way emerged in 2018 when the youth client council at De Hondsberg were presented the results of a recent client satisfaction survey. We asked the council members what they thought of the results. They were surprised at how favourable the results were for the organisation. This seemed to contradict the council members' perception of how children and youth generally expressed their views on the quality of care at the facility. They had expected the results to be much less rosy than those reported by the researchers. The client council suggested that the survey had been biased in a positive direction because carers, who assisted the youths while completing the survey, had influenced the answers of children and youth. They also wondered if all participating youths had truly understood what we had asked of them. We promised the youths that we would look into the literature on surveys and what was known about the topic of bias in survey research with people with cognitive limitations. Our first point of reference was of course Professor Xavier. He pointed out that this topic was largely unexplored territory from a scientific point of view and suggested that I delve into this topic by starting a PhD research project. And so I did.

Roel Kooijmans

Chapter 1

Introduction

'States Parties to the present Convention recognize the equal right of all persons with disabilities to live in the community, with choices equal to others, and shall take effective and appropriate measures to facilitate full enjoyment by persons with disabilities of this right and their full inclusion and participation in the community.'

(Article 19 of the Convention on the Rights of Person with Disabilities, United Nations, 2006)

The right to be heard

This is not the first thesis to open with a reference to the United Nations Convention on the Rights of Persons with Disabilities. The right to self-determination, the right to organise your own life and make your own decisions and the right to accessible information and means of communication are at the heart of this convention. The convention is the starting point for many (research) projects aimed at promoting inclusion and participation for people with disability (Steinert et al., 2016). Despite the convention having been ratified 18 years ago, for many people with intellectual disabilities, it is still not self-evident that their views are taken seriously and that their right to self-determination is respected (McCausland et al., 2022).

People with intellectual disabilities face greater challenges in making their voices heard for various reasons. Many people with intellectual disabilities depend upon carers, social workers or guardians for their daily support, who have been trained to take over tasks and responsibilities, instead of promoting self-determination. This stems from a longstanding belief that people with intellectual disabilities are defined primarily by their limitations rather than their abilities (McCusker et al., 2023; Roy et al., 2021). This focus upon limitations, rather than abilities, has led to a promotion of the view that people with intellectual disabilities are unable make their own choices and organise their lives as they wish.

Aside from problems with 'gatekeepers', where opportunities for community participation for people with intellectual disabilities are restricted and regulated by carers and relatives, communicative barriers also limit the ability of people with disabilities to have their voices heard. Authorities and support workers often struggle to adequately tailor their communication to the cognitive and communicative abilities and limitations of people with intellectual disabilities. This limits their opportunity to communicate what they feel, think, and want (Smith et al., 2020). This is problematic in light of the fact that many people with intellectual disabilities experience lower quality of life (Maguire et al., 2023) and are more vulnerable to developing mental health problems and disorders than members of the general population (Patel et al., 2023). If they are unable or are not given the opportunity to share their perspectives, it is highly unlikely that their needs will be met in a way that is desired and beneficial (Walton et al., 2022). For example, being unable to adequately communicate your needs might mean that you cannot communicate symptoms of physical illness, which go unrecognised, until a disease has progressed beyond the point where treatment is possible, shortening lifespan. Many people with more severe disabilities who are non- or minimally verbal exhibit self-injurious and aggressive behaviour because they have fewer abilities to communicate that they experience pain or anxiety (Oliver, 2022).

Self-report versus proxy-informant reports

Using information provided by proxies

One way to ensure that the needs of people with cognitive and communicative impairments are identified is to ask people who know the person well to be their informant. It is considered a good practiceto triangulate information from different informants when assessing a person's level of functioning or when gathering information about a persons' needs and desires (Havercamp et al., 2022; Shogren et al., 2021). When determining the care and support needs of people with intellectual disabilities, parents, carers, or other relatives can sometimes complement the perspective of the person with intellectual disability very well, and their views can serve as a 'second opinion'. There is debate under what conditions proxy perspectives can be used to *substitute* the first-person perspective of a person with intellectual disability (Webb et al., 2024). This becomes highly relevant when a person with intellectual disability is considered not to be able to act in their own interest.

When would it be a good idea to use proxy reports?

Proxies have proven to be good at estimating more or less objectively determinable indicators. Examples of these include a person's ability to do certain work or the nature of a person's physical limitations (Balboni et al., 2013). In addition, there are circumstances in which people with intellectual disabilities *themselves* are arguably less able to provide robust estimates than proxies. For example, proxies are better than people with intellectual disabilities at estimating how often something occurred over a longer period of time (Santoro et al., 2022), and are better at rating adolescent social skills than adolescents with intellectual disabilities themselves (Fisher et al., 2014).There is some evidence to indicate that people with intellectual disabilities have a tendency to overestimate their own competence (Snell et al., 2009; Golubović, & Škrbić, 2013).

It has been suggested that the ability of people with intellectual disabilities to make accurate assessments of their own functioning decreases as their degree of intellectual disability increases (Webb et al., 2024). Even if self-report measures are adapted to account for the cognitive challenges faced by people with intellectual disabilities, the demands placed upon reasoning and comprehension skills will often exceed the abilities of individuals with more severe forms of intellectual disability (Emerson et al., 2013). In these cases, proxy assessments are generally the only viable option for assessment and act as a 'second-best' substitute for first-person assessments, even for more subjective issues. There is evidence that caregivers provide reliable estimates of quality of life for non-verbal individuals with severe levels of intellectual disability (Balboni et al., 2013).

When should we ask people with intellectual disabilities directly?

Inviting proxies to provide responses about an individual with mild intellectual disabilities, while not asking the person with intellectual disabilities themselves, is obviously problematic. It has negative implications for the validity of the outcomes of diagnostic procedures and needs assessment and it actively prevents the inclusion and participation of people with intellectual disabilities.

Considering clinical practice, excluding people with intellectual disabilities during clinical assessments is troublesome as their own account of their own experience will be excluded. The 'inner world' of the person with intellectual disability, including thoughts, feelings, pain, and beliefs about health and satisfaction, cannot be directly observed by a third party and are coloured by a unique and personal mix of context, previous experiences and personal beliefs (Noonan, 2016). Many people with intellectual disabilities find it difficult to share their inner world with those around them, either because they have difficulty verbalising thoughts and feelings or because they do not trust the people who care for them enough to share their experiences.

Proxy assessments and first-person perspectives of internal states can be quite divergent as a result. In a study by Scott and Havercamp (2018), proxies reported mental health problems in 24% of the people with an intellectual disability they assessed. When people with intellectual disabilities themselves were asked to rate their mental health, almost 68% reported mental health problems. Similarly, the ratings of proxies and people with intellectual disabilities do not converge for a number of other themes, including quality of life (Simões & Santos, 2016), social functioning (Schützwohl et al., 2018) and physical well-being (Wilson et al., 2016). This raises the question whether 'outsiders' are able to adopt the perspective of the person with intellectual disability in assessments. This is relevant for our understanding of mental health and well-being; we risk overlooking an individual's subjective experiences by not asking people with intellectual disabilities about these experiences directly (Maguire et al., 2023; Mileviciute & Hartley, 2015; Patel et al., 2023). This can lead to a mismatch between the care and support needs of the person with intellectual disability and the support that is actually provided (Havercamp et al., 2022; Walton et al., 2022).

Further, prioritizing the viewpoint of individuals with intellectual disabilities is good practice as it promotes participation and inclusion. Not only does directly involving people in assessment provide the purest perspective on many issues, being able to indicate your own needs promotes autonomy and the right to self-determination, all of which in turn promote health and wellbeing (Shogren et al., 2006). Taking an active role in individual assessments gives clients a voice in the decisions that are made at the individual level (McCausland et al., 2022). Another way for people with intellectual disabilities to have their voices heard is by actively participating in research, advocacy, and policy initiatives. In this way, participants can also advocate for the *collective* interests of people with intellectual disabilities and actively influence policy and the political agenda (McDonald et al., 2018; Walton et al., 2022).

Why it is important to hear both sides

It is evident that we should ask people with intellectual disabilities themselves what they feel and think. However, we do not want to give the impression that proxies are unable to make judgements from another person's perspective. For a large number of topics, we know that proxies and people with intellectual disabilities have different views, but it is not clear who is 'right'. For example, there is no objectively determinable gold standard in determining a person's quality of life (Simões & Santos, 2016). And, when assessing skills, the 'verdict' may also depend on the assessor's normative judgment of what behaviour is seen as adequate (or not). Does a self-confident person with an intellectual disability overestimate their own ability (and according to whom?), or does the assessor focus mainly on the person's disability and do they overlook their qualities? Therefore, researchers looking at differences between self-reported and proxy-reported data have concluded that in most cases no preference can be given to either (see, for example, Santoro et al., 2022), and that if possible, both perspectives should be included and compared in assessment and research (e.g. Patel et al., 2023).

However, this is not to say that both perspectives should be given equal weight. Recent research has generally indicated that whenever possible, one should ask people with intellectual disabilities directly about their opinions, wishes, and needs first (e.g. Havercamp et al., 2022; Shogren et al., 2021; Walton et al., 2022).

Self-report measures for people with intellectual disabilities

Why self-reporting can be problematic for people with intellectual disabilities

If we prioritize a first-person perspective in clinical and research practice, self-report measures are an important means for collecting data about internal states, thoughts and opinions. In Tourangeau's (2018) model of survey response, the process of answering self-report questions consists of several steps: comprehension, retrieval, judgment, and estimation and reporting. These cognitive processes are challenging for most people with intellectual disabilities associated with the nature and degree of the disability, including problems with reasoning, verbal expression, reading, abstract thinking, and judgement (Schalock et al., 2010; American Psychiatric Association, 2013). Difficulties with memory or recall can impede the information retrieval process of response formation, lower reading levels may lead to misunderstanding of questions, and an impaired ability to recall detailed or distant events may bias estimation and reporting (Santoro et al., 2022; Shogren et al., 2021).

Improving the cognitive accessibility of self-report measures

Problems with comprehension, retrieval, judgment and estimation can be overcome, at least in part, by designing self-report instruments that are *cognitively accessible* for as many people with intellectual disabilities as possible. According to Kramer and Schwartz (2017), cognitive accessibility is present 'when assessment design anticipates respondent variability in cognitive abilities, and to the greatest extent possible, reduces cognitive demands and/or supports cognitive processes to enable respondents with a range of cognitive abilities to interpret and respond to assessment items as intended.' (p. 1705). There is growing consensus that individuals with mild to moderate intellectual disability can reliably provide valid self-reports for a variety of concepts such as internalizing symptoms and overall well-being if cognitively accessible instruments are used (Havercamp et al., 2022). For example, research has shown that many individuals with intellectual disability can reliably respond to questions using rating scales with visual and textual supports (O'Keeffe et al., 2019). Making measures more cognitively accessible can also remove access barriers and improve motivation to participate in assessment and research (Hanson et al., 2023). Stancliffe et al. (2014) demonstrated much higher response rates (83 % vs 25 %) on a measure of loneliness when using a more cognitively accessible tool with respondents who had mild to moderate intellectual disability

Cognitive accessibility of self-report-measures

This raises the question how cognitive accessibility can be achieved. At the start of this PhD research in 2018, there was little evidence-informed practical guidance on how to attune self-report measures to the needs of people with intellectual disabilities. Most tips and tricks came from 'good practice' and the expertise of clinicians and researchers with a history of working with people with intellectual disabilities. Articles describing the development of 'ID-inclusive' measures rarely justified the reasons or empirical grounds for specific adaptations. The only more or less systematic review that compiled all the existent research literature was a review by Finlay and Lyons (2001) that addressed a wide range of themes associated with the problems of using self-reports for people with intellectual disabilities. An example of a review that addressed specific sub-themes is Hartley and MacLean's (2006) review of the use of Likert-type scales. All other reviews or compendiums were of a more narrative nature (e.g. Bell et al., 2018). In the course of this PhD project, several studies were published that addressed this theme, all of which were of a narrative or scoping in nature rather than a systematic review with a thorough quality assessment of included studies (e.g. Havercamp et al., 2022; Shogren et al., 2021; Witwer et al., 2022).

Bias

The limited cognitive accessibility of self-report instruments does not only impede the opportunities for people with intellectual disabilities to participate and meaningfully self-report, but it is also a threat to the validity and reliability of results. Several factors and mechanisms have been identified that can negatively influence the reliability of answers given by people with intellectual disabilities when completing a self-report instrument, directing results in such a way that they no longer represent the 'true' opinions and feelings of people with intellectual disabilities. The inadvertent distortion of responses is known as *response bias* and occurs in a variety of circumstances and has a variety of causes. For example, bias can occur when people misunderstand the meaning of a question, have trouble remembering what was asked, have misconceptions of what is expected of them, or when sensitive or taboo topics are addressed (Finlay & Lyons, 2002). People with intellectual disabilities have been shown to be particularly susceptible to some forms of response bias. Commonly occurring forms of response bias are acquiescence, recency, and primacy biases and suggestibility.

Acquiescence is the tendency to answer affirmatively to questions regardless of content. It is especially prevalent when individuals do not know the answer to a question (Emerson et al., 2013). Less frequently, *'nay-saying'*, a disposition to deny or answer 'no' to every question, is encountered in interviews with people with intellectual disabilities (Finlay & Lyons, 2002). *Recency bias* is the tendency to select the last option mentioned in multiple-choice questions, irrespective of one's true opinion. Conversely, *primacy* reflects a bias to select the first option. Both occur in the case of working memory overload, if the respondent cannot keep all response options in working memory at once. They may occur more frequently for people with intellectual disabilities as memory deficits are one of the core cognitive impairments associated with intellectual disabilities (American Psychiatric Association, 2013). *Suggestibility* refers to a willingness to change answers following suggestions from another person (Bell et al., 2018). The presence of someone else is also of influence when topics of a sensitive or even taboo nature are addressed. In this case, respondents may be reluctant to admit to socially or culturally unaccepted behaviour. Socially desirable answering is common, especially if the relationship between respondent and the person assisting them is non-symmetrical and there is a power imbalance (Krumpal, 2013).

In sum, response biases can emerge from problems with *understanding* and from complex interpersonal *interactions*.

Assisting people with intellectual disabilities to complete a self-report measure

As can be seen in the nature and origin of response biases, many forms of bias occur in the context of interpersonal dynamic relationships. These are especially relevant as people with intellectual disabilities frequently require assistance, or other adaptations when completing self-administered measures (Finlay & Antaki, 2012). Although this may introduce or enhance bias, helping people with intellectual disabilities to complete a self-report has benefits. Providing assistance can help people with intellectual disabilities to engage in assessment and research. An interviewer assisting with administration can address reading problems, limited writing ability, and clarify questions if needed (Santoro et al., 2022). As some people with intellectual disabilities have attention difficulties, it is often helpful to recap and summarise what the respondent has said to re-engage and focus their attention (Havercamp et al., 2022). Assisting respondents with intellectual disabilities can help overcome problems with cognitive accessibility without modification of an original and unadapted measure, providing that the interviewer or assistant is acutely aware of the ways in which their presence may shape respondent answers and has strategies to minimize their influence.

Several suggestions can be found in the literature on how to avoid this kind of bias as much as possible through adjustments in the process of self-reporting. Recommendations that are mentioned include taking time to get acquainted and an informal conversational style, emphasising that there are no right and wrong answers, adding an ‘I don’t know’ answer category, avoiding questions that make respondent doubt their own answer (‘are you sure…?’) and ensuring that the assistant is a ‘neutral’ person (Havercamp et al., 2022; Jen-Yi et al., 2015; O’Keeffe et al., 2019; Sigstad & Garrels, 2018). The form in which questions are presented can also make a difference; research with neurotypical populations has found that respondents are more likely to report sensitive behaviours on a computerised survey compared to ‘paper-and-pencil’ measures, because they are more confident that results will be processed anonymously (Gnambs & Kaspar, 2014). Researchers have recommended using pre-administration screening of understanding and responsiveness to detect bias and indicate whether someone is able to complete a self-report measure (Emerson et al., 2013; Finlay & Lyons, 2002; Perry & Felce, 2002). O’Keeffe et al. (2019) added that pre-assessment also provides an opportunity to develop rapport with participants and practice response formats. An alternative to pre-assessment screening to detect bias is to integrate bias-detecting items in the questionnaire structure. The results of bias detection screening can be used to establish validity of the results at the individual and population level and to exclude participant data after data completion. Bias is usually detected by using one of three types of bias-detecting items: (a) nonsense questions to which the answer should be negative ("I like eating rotten foods"); (b) using pairs of questions that are opposite in meaning (e.g. 'I am mostly happy', followed by an item that states 'I am mostly sad' and comparing answers); or (c) using equivalent pairs of questions asked in different formats, for example, in a yes-no format or in an either-or format. There is little to no consensus what tools should be used to detect bias and there is debate whether the proposed tools really detect bias or other artifacts associated with the nature of the format (Finlay & Lyons, 2002) .

Aims of the current PhD research project

A recap of the 'why', and what's missing

One of the ways in which people with intellectual disabilities can voice their opinions is through active participation in assessment, research, and advocacy. The tools to elicit their feelings, thoughts and beliefs through self-report require adaptations to accommodate the cognitive and communication challenges associated with the disability. Self-report measures should be made cognitively accessible, so that as many people with intellectual disabilities as possible are able to complete them. This means they must be developed with the needs of people with intellectual disabilities in mind, or adaptations must be made to measures that were not originally developed for use with people with intellectual disabilities. The recommendations for improving cognitive accessibility for people with intellectual disabilities that can be found in the research literature tend to be non-specific (e.g. "use visual supports") and not well-substantiated by empirical evidence. For some topics, the results from different studies are conflicting; for example, on the topic of which response options are most suitable. For example, there is both support for and evidence against the suitability of yes-no answers. Yes-no answers are understood by the largest proportion of persons with ID (Stancliffe et al., 2014), but they may promote acquiescent answering (Heal, 1995). For some recommendations, the empirical evidence is absent; for example, there is no evidence about which type of visual supports promote the comprehension of written language. At the start of this PhD research there was no comprehensive overview of recommended adaptations and the strength of the evidence supporting them had never been assessed. This lack of guidance means that there has been an absence of cognitively accessible and validated instruments about many aspects of (mental) health, quality-of-life and other areas of interest for use by people with intellectual disabilities (Nicolaidis et al., 2020; Vlot-Van Anrooij et al., 2018). If the respondent is not able to complete the measure independently, assessment procedures have to take into account the possibility that interpersonal dynamics between respondent and assistant bias the results and they should detail strategies to minimize this risk. The what and how of these strategies are not well-defined as we have only begun to understand the mechanisms and cognitive processes underlying the several forms of bias arising from respondent-assistant dynamics.

Research questions

At the outset of this PhD project there was insufficient empirical knowledge of what makes a measure cognitively accessible for people with intellectual disabilities, and how to address issues arising from interpersonal dynamics during assessment. The aim of this PhD project was to contribute to a solid evidence base by assessing the state of the field, filling in some of the gaps and applying new and existing knowledge to clinical and research practice. We set out to answer the following questions.

Assessing the evidence base

1. What are empirically validated recommendations to make self-report assessments more accessible for people with mild to moderate intellectual disability?
    1.1. How should self-report measures be constructed to overcome the cognitive challenges associated with self-report for people with intellectual disabilities?
    1.2. What does the available empirical evidence tell us about interpersonal dynamics that introduce response bias and how to address them?
    1.3. What is the quality of the evidence on these topics?
    1.4. What are gaps or inconsistencies in the research literature?

Expanding the evidence base

2. Looking at what we *don't* know: can we fill in the blanks for some of the topics identified under 1.4?

Applying new and existing knowledge

3. Can our knowledge of what makes a measure more cognitively accessible help us to assess the suitability of existing self-report measures for people with intellectual disabilities?
4. When we apply the existing and newly added evidence-informed recommendations to the process of constructing a new measure or adapting an existing self-report measure, does this lead to an instrument
    a. that is demonstratively more cognitively accessible?
    b. that produces more valid and reliable results?

Research *about* people, *with* people (but not all people…)

Why ask them? It's about me! Because participation and inclusion play such a central role in the topic of this PhD research, it was self-evident to ask people with intellectual disabilities to actively participate in our research as much as possible. Especially in the experimental designs, where data collection takes place 'close to the source', we felt it was the people with disabilities themselves who should be given the role as experts. Using participatory action research methods not only empowers individuals with intellectual and developmental disabilities, it will also improve the quality of information that is gathered in research (Havercamp et al., 2022).

Our basic principle was that people should be able to participate according to their own ambitions (what they want), in consideration of their talents and limitations (what they can handle). A consequence of the latter is that we think that it is not possible to include all people with intellectual disabilities as participants or co-researchers. The term 'people with intellectual disabilities' is broad and spans all levels of cognitive functioning, temperament, stamina and curiosity. As described above, the cognitive load of reading, understanding and responding to questionnaires places certain demands on the cognitive abilities of the respondent. We acknowledge, albeit reluctantly, that even with the most accessible tools and inclusive participatory designs, it will never be possible to obtain accurate self-reports from people across the entire spectrum of cognitive functioning (Emerson et al., 2013; Finlay & Lyons, 2001). For this PhD study, we therefore delimit the target group of co-researchers and participants to 'people with mild to moderate intellectual disabilities'. Where possible, we describe to what extent the results may also be generalisable to other levels of cognitive and adaptive functioning.

Thesis outline

Assessing the evidence base

In **Chapter 2**, 'The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review', the first four research questions are addressed. Using a systematic review methodology, we searched the peer-reviewed research literature from the year 1996 onwards for recommendations on how to attune self-report instruments to the needs of people with intellectual disabilities. The quality of the evidence was assessed with the GRADE-CERQual tool. The results are presented in a Summary of Findings table, according to a five-stage model of instrument development. In the Discussion, areas for further exploration are proposed. Examples of concrete topics that needed further exploration were an operationalisation of what constitutes adequately simplified language, what types of visualization should be used to support written text, and ways to detect and prevent bias.

Expanding the evidence base

In **Chapter 3**, ‘Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people’s views’, we looked at how self-reported experiences of youths with intellectual disabilities were influenced by the presence of a person assisting them. We looked at differences in scores between youths who were assisted by a carer with whom they were in a hierarchical relation versus youths who were assisted by a neutral person and youths who completed a survey independently. We followed up on the quantitative findings with focus groups with youths with intellectual disabilities. We discussed the findings and explored possible explanations.

In **Chapter 4**, ‘Does adding pictures to easy-to-read texts benefit comprehension for people with reading difficulties? A meta-analytic review’, we examined the evidence for the use of use visual supports to facilitate comprehension of written texts. We aggregated the results of empirical studies on this topic in a 3-level meta-analysis.

Applying new and existing evidence

In **Chapter 5**, ‘Self-report stress measures to assess stress in adults with mild intellectual disability—a scoping review’, we combined the results from the systematic review in Chapter 2 with expert panel input to assess the suitability of a number of self-report stress measures for people with mild intellectual disability.

In **Chapter 6**, ‘Does adapting a self-report instrument to improve its cognitive accessibility for people with intellectual disabilities result in a better measure? - A cognitive interview study’, we tried to answer the question ‘Does it really matter?’ We applied the recommendations from our earlier studies to improve the cognitive accessibility of an existing and widely-used self-report measure. We examined if this led to a measure that was perceived to be more accessible by people with intellectual disabilities. We did this by interviewing 18 adults with (mild) intellectual disabilities while they completed the measure, using cognitive interviewing methodology. By quantitatively analysing the results and comparing the results from self- and proxy-reports, we explored the differential reliability and validity of the adapted measure.

In **Chapter 7**, ‘Summary and discussion’, the results of the project are summarised and I critically reflect on the question to what extent this project has advanced the field of intellectual disability research and practice.

References

Balboni, G., Coscarelli, A., Giunti, G., & Schalock, R. L. (2013). The assessment of the quality of life of adults with intellectual disability: The use of self-report and report of others assessment strategies. *Research in Developmental Disabilities*, *34*(11), 4248–4254. https://doi.org/10.1016/J.RIDD.2013.09.009

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. In *Psychology, Crime and Law* (Vol. 24, Issue 4, pp. 362–378). Routledge. https://doi.org/10.1080/1068316X.2017.1298761

Emerson, E., Felce, D., & Stancliffe, R. J. (2013). Issues concerning self-report data and population-based data sets involving people with intellectual disabilities. *Intellectual and Developmental Disabilities*, *51*(5), 333–348. https://doi.org/10.1352/1934-9556-51.5.333

Finlay, W. M. L., & Antaki, C. (2012). How staff pursue questions to adults with intellectual disabilities. *Journal of Intellectual Disability Research*, *56*(4), 361–370. https://doi.org/10.1111/j.1365-2788.2011.01478.x

Finlay, W. M. L., & Lyons, E. (2002). Acquiescence in Interviews With People Who Have Mental Retardation. In *American Association on Mental Retardation* (Vol. 40). http://meridian.allenpress.com/idd/article-pdf/40/1/14/2046313/0047-6765

Fisher, M. H., Mello, M. P., & Dykens, E. M. (2014). Who reports it best? A comparison between parent-report, self-report, and the real life social behaviors of adults with Williams syndrome. *Research in Developmental Disabilities*, *35*(12), 3276–3284. https://doi.org/10.1016/j.ridd.2014.08.011

Gnambs, T., & Kaspar, K. (2014). Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behavior Research Methods*, *47*(4), 1237–1259. https://doi.org/10.3758/s13428-014-0533-4

Hanson, S. L., Bruyere, S., Forber-Pratt, A., Reesman, J., & Sung, C. (2023). Guidelines for Assessment and Intervention With Persons With Disabilities: An Executive Summary. *American Psychologist*. https://doi.org/10.1037/amp0001150

Havercamp, S. M., Barnhill, L. J., Bonardi, A., Chapman, R. A., Cobranchi, C., Fletcher, R. J., Rabidoux, P., Seeley, J. R., & Tassé, M. J. (2022). Straight from the horse's mouth: Increasing self-report in mental health assessment in individuals with intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, *35*(2), 471–479. https://doi.org/10.1111/jar.12952

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. https://doi.org/10.1016/j.alter.2015.07.006

Kramer, J. M., & Schwartz, A. (2017). Reducing Barriers to Patient-Reported Outcome Measures for People With Cognitive Impairments. In *Archives of Physical Medicine and Rehabilitation* (Vol. 98, Issue 8, pp. 1705–1715). W.B. Saunders. https://doi.org/10.1016/j.apmr.2017.03.011

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. In *Quality and Quantity* (Vol. 47, Issue 4, pp. 2025–2047). Kluwer Academic Publishers. https://doi.org/10.1007/s11135-011-9640-9

Maguire, S., Davison, J., McLaughlin, M., Simms, V., & Bunting, B. (2023). Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review. In *Journal of Applied Research in Intellectual Disabilities*. John Wiley and Sons Inc. https://doi.org/10.1111/jar.13110

McCausland, D., Murphy, E., McCarron, M., & McCallion, P. (2022). The potential for person-centred planning to support the community participation of adults with an intellectual disability. *Journal of Intellectual Disabilities*, *26*(3), 603–623. https://doi.org/10.1177/17446295211022125

McCusker, P., Gillespie, L., Davidson, G., Vicary, S., & Stone, K. (2023). The United Nations Convention on the Rights of Persons with Disabilities and Social Work: Evidence for Impact? In *International Journal of Environmental Research and Public Health* (Vol. 20, Issue 20). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ijerph20206927

McDonald, K. E., Conroy, N. E., Olick, R. S., Carroll, A., Cuddy, M., Feldman, M. F., Flanigan, D., Fratangelo, P., Gonzalez, L., Kennedy, M., King, K., Mansfield, C., McGowan, D., Romer, R., Turk, M., Velez, S., Walker, P., & Worral, P. (2018). A quantitative study of attitudes toward the research participation of adults with intellectual disability: Do stakeholders agree? *Disability and Health Journal*, *11*(3), 345–350. https://doi.org/10.1016/J.DHJO.2017.12.004

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

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. https://doi.org/10.1089/aut.2019.0074

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

Patel, M., Lee, J. Y., & Scior, K. (2023). Psychometric properties of measures designed to assess common mental health problems and wellbeing in adults with intellectual disabilities: a systematic review. In *Journal of Intellectual Disability Research* (Vol. 67, Issue 5, pp. 397–414). John Wiley and Sons Inc. https://doi.org/10.1111/jir.13018

Perry, J., & Felce, D. (2002). Subjective and Objective Quality of Life Assessment: Responsiveness, Response Bias, and Resident:Proxy Concordance. In *MENTAL RETARDATION* (Vol. 40).

Roy, A., Courtenay, K., Odiyoor, M., Walsh, P., Keane, S., Biswas, A., Marston, G., Thirulokachandran, S., & Munir, K. (2021). Setting priorities for people with intellectual disability/intellectual developmental disorders across the lifespan: a call to action by the World Psychiatric Association. *BJPsych International*, *18*(3), 54–57. https://doi.org/10.1192/bji.2021.6

Santoro, S. L., Donelan, K., & Constantine, M. (2022). Proxy-report in individuals with intellectual disability: A scoping review. In *Journal of Applied Research in Intellectual Disabilities* (Vol. 35, Issue 5, pp. 1088–1108). John Wiley and Sons Inc. https://doi.org/10.1111/jar.13013

Schützwohl, M., Voß, E., Salize, H. J., Stiawa, M., Puschner, B., & Koch, A. (2018). Self- and proxy-rated needs in adults with mild to moderate intellectual disabilities: Perspective matters. *Journal of Applied Research in Intellectual Disabilities*, *31*(2), 285–295. https://doi.org/10.1111/jar.12399

Shogren, K. A., Bonardi, A., Cobranchi, C., Krahn, G., Murray, A., Robinson, A., Havercamp, S. M., Andridge, R., Arnold, L. E., Barnhill, J., Bodle, S., Boerner, E., Bonardi, A., Bourne, M. Lou, Brown, C., Buck, A., Burkett, S., Chapman, R., Cole, C., … Witwer, A. (2021). State of the Field: The Need for Self-Report Measures of Health and Quality of Life for People With Intellectual and Developmental Disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, *18*(4), 286–295. https://doi.org/10.1111/jppi.12386

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. https://doi.org/10.1080/08856257.2017.1413802

Simões, C., & Santos, S. (2016). The quality of life perceptions of people with intellectual disability and their proxies†. *Journal of Intellectual and Developmental Disability*, *41*(4), 311–323. https://doi.org/10.3109/13668250.2016.1197385

Smith, M., Manduchi, B., Burke, É., Carroll, R., McCallion, P., & McCarron, M. (2020). Communication difficulties in adults with Intellectual Disability: Results from a national cross-sectional study. *Research in Developmental Disabilities*, *97*, 103557. https://doi.org/10.1016/J.RIDD.2019.103557

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. https://doi.org/10.1111/jir.12024

Tourangeau, R. (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education*, *26*(2), 169–181. https://doi.org/10.1108/QAE-06-2017-0034

United Nations (2006). *Convention on the Rights of Persons with Disabilities (CRPD)*. United Nations.

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*(1). https://doi.org/10.1186/s12874-018-0539-1

Walton, K., Krahn, G. L., Buck, A., Andridge, R., Lecavalier, L., Hollway, J. A., Davies, D. K., Arnold, L. E., & Havercamp, S. M. (2022). Putting "ME" into measurement: Adapting self-report health measures for use with individuals with intellectual disability. *Research in Developmental Disabilities*, *128*. https://doi.org/10.1016/j.ridd.2022.104298

Webb, K., Cuskelly, M., & Owen, C. (2024). The use of proxy measurement of internal states in people with intellectual disability: A scoping review. *Journal of Applied Research in Intellectual Disabilities*, *37*(4). https://doi.org/10.1111/jar.13241

Wilson, K. S., Wiersma, L. D., & Rubin, D. A. (2016). Quality of life in children with Prader Willi Syndrome: Parent and child reports. *Research in Developmental Disabilities*, *57*, 149–157. https://doi.org/10.1016/j.ridd.2016.06.016

Chapter 2

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

Published as:

Kooijmans, R., Mercera, G., Langdon, P.E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clinical Psychology: Science and Practice, 29*(3), 250-271.

Abstract

People 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.

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.

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.

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 intellectual disability are presented.

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 intellectual disability to voice their opinions, feelings and thoughts. In clinical work and research, assessing the perspectives of people with intellectual disability 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 intellectual disability, 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 intellectual disability 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 intellectual disability 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 intellectual disability is especially poor for personal or sensitive topics (Mileviciute & Hartley, 2015; White-Koning et al., 2005).

Suitability of Self-Report Measures for People with Intellectual Disability

Whereas there is virtually no debate for researchers and clinicians as to whether the opinions, feelings and thoughts of people with intellectual disability 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 intellectual disability will struggle to understand and respond to questions (Nicolaides et al., 2020). In order to address the personal experiences of people with intellectual disability using self-report measures, developers and researchers have turned to one of three options: (a) use measures that can be administered to persons with intellectual disability in an unaltered form, (b) use measures that can be administered to persons with intellectual disability after making adaptations to the original, and (c) use measures specifically designed to be administered to persons with intellectual disability.

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 intellectual disability in a fundamentally different way than for the general population. Several measures have been identified that can be administered to persons with intellectual disability 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 intellectual disability, 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 intellectual disability. 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 Intellectual Disability

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 intellectual disability. 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 intellectual disability 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 intellectual disability. 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 is presented as applicable to ‘persons with intellectual disability’, 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 intellectual disability, 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 from participation (Hartley & MacLean, 2006). This reduces the applicability of many recommendations to people with borderline intellectual functioning or mild intellectual disability - a limited proportion of the total ‘people with intellectual disability’ 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 intellectual disability. 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 intellectual disability 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 intellectual disability 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 intellectual disability 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 intellectual disability. The aim is to provide professionals with evidence-based guidelines for developing, adapting or using measures suitable to use by people with intellectual disability. Since the vast majority of studies on this topic excluded persons with more severe levels of intellectual disability, it is expected that most results pertain to persons with borderline intellectual functioning and mild to moderate intellectual disability. However, those with severe or even profound levels of intellectual disability 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 intellectual disability 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

Eligibility Criteria Studies were eligible for inclusion if they reported on studies including persons with intellectual disability, outcomes included recommendations for the construction or adaptation of measures designed for people with intellectual disability 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 intellectual disability presented by Bell and colleagues (Bell et al., 2018). The model represented in Table 1 resulted from the synthesis of these three models and was used to guide the initial deductive coding process.

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 on page 44.

Table 1

A Priori Coding Framework

Overarching factor	**Subfactors / themes**	**Examples include**
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, lay-out, 'survey flow'.
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 validity and reliability

Results

Study Selection

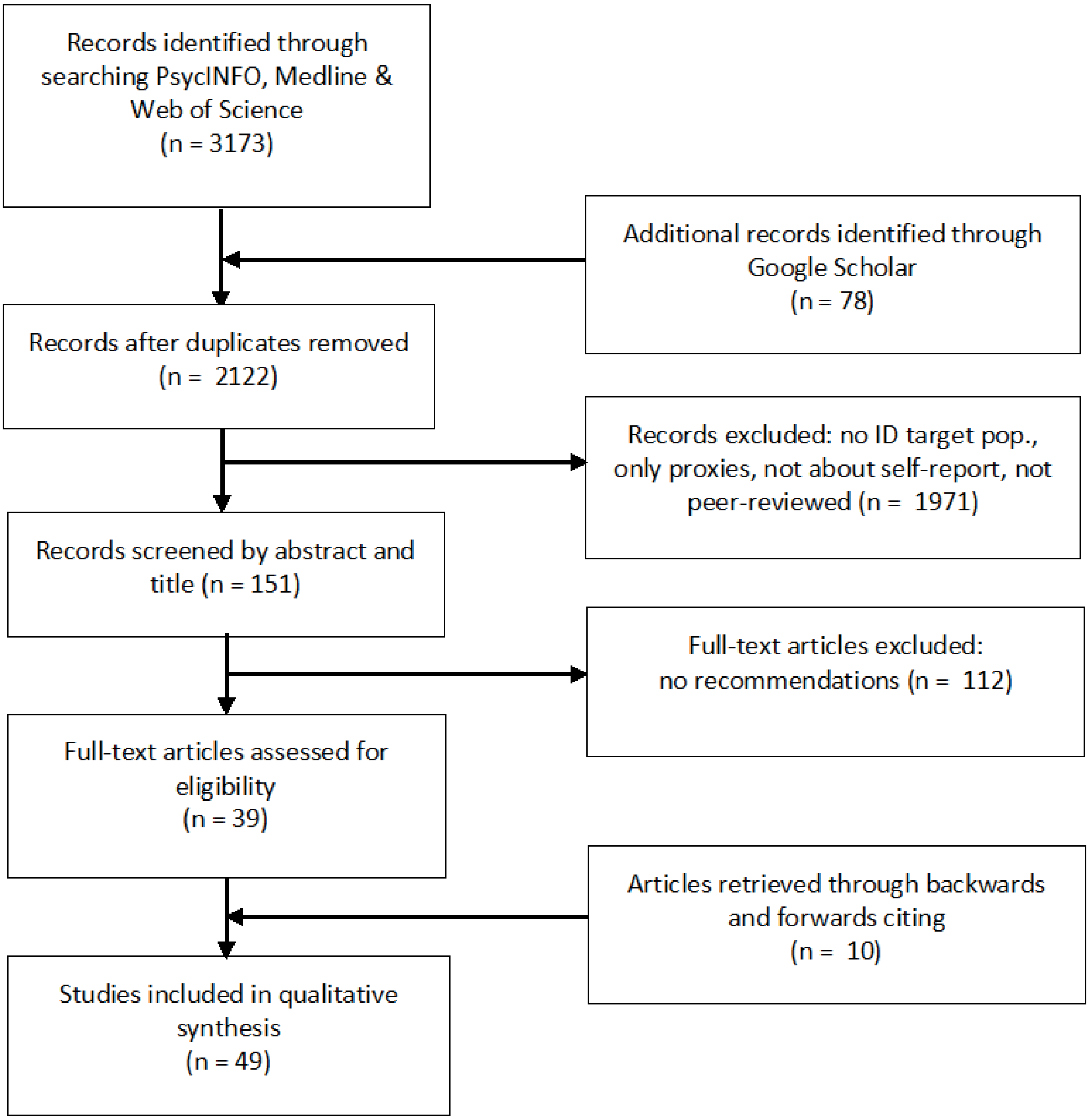
The selected databases were searched in February 2020 and updated in February 2021. A total of 3173 records were found. After deduplication 2122 articles were scanned for eligibility by reading the titles and abstracts. The first 100 records were reviewed by the first two authors (RK and GM) simultaneously. 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 intellectual disability in general, but recommendations were generally made for specific levels of cognitive functioning within the broader 'intellectual disability 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 intellectual disability of participants was not specified or described in very general terms. The resulting summary of recommendations can therefore not be applied for 'persons with intellectual disability' 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 intellectual disability, often because persons with moderate or severe levels of intellectual disability 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'.

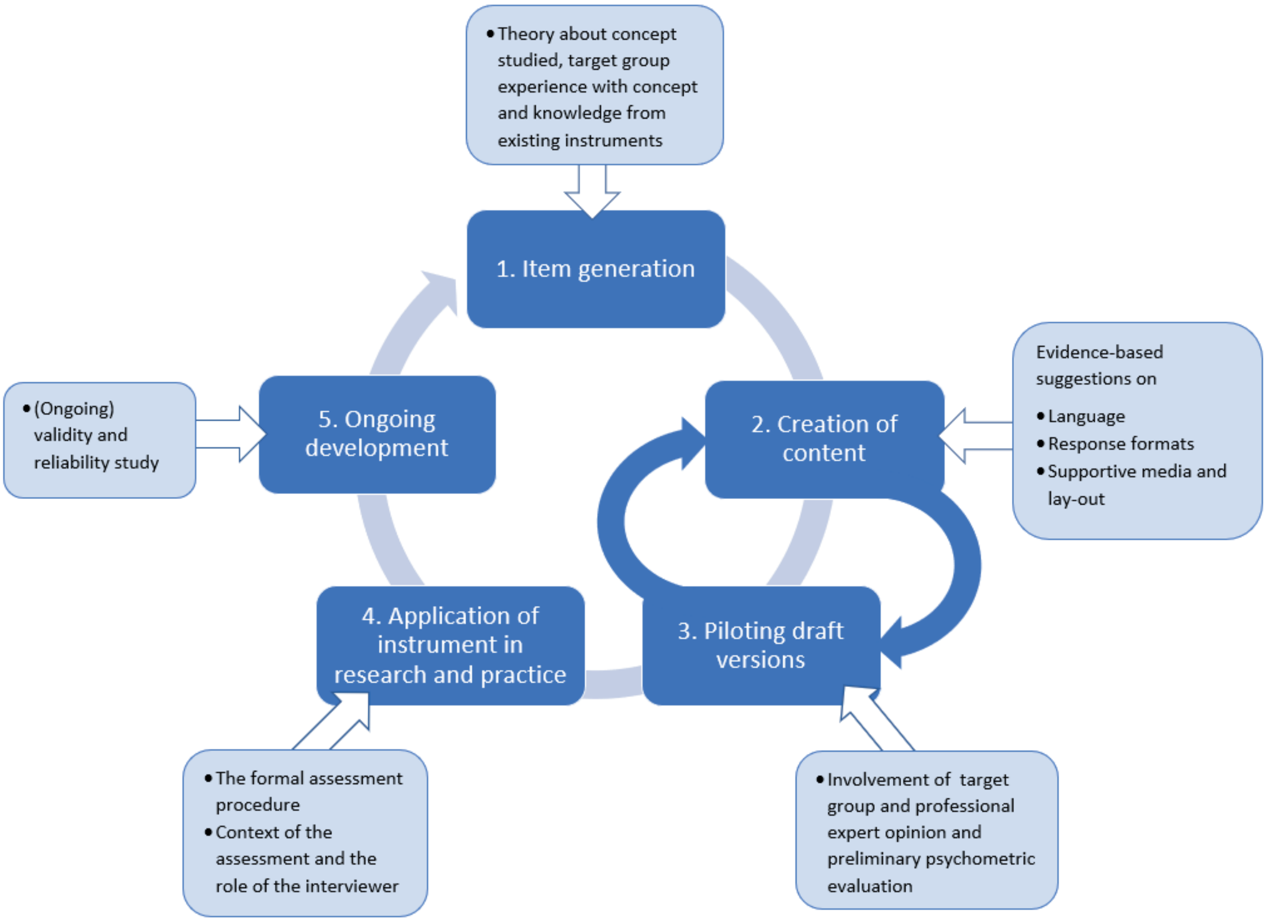
For many studies that used qualitative methods to make inferences, the process of data synthesis was not (clearly) described. The conclusions drawn from the data on some occasions seemed to be highly dependent on the researcher's interpretation, which is of course an artefact of the qualitative nature 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 Measure Development Stage

In the Summary of Findings (Table 2 on page 34), 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 borderline intellectual functioning to moderate intellectual disability. Where relevant, distinctions are made between recommendations for different levels of intellectual disability. If recommendations extend across all levels of functioning, including more severe levels of intellectual disability, 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 intellectual disability 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 people with intellectual disability.

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 intellectual disability 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 intellectual disability 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 intellectual disability participants. However, research trying to replicate both the higher prevalence of acquiescence in persons with intellectual disability 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 intellectual disability and no more than 5 options for persons with borderline intellectual functioning to mild intellectual disability (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 intellectual disability 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 Knegt et al., 2017) and there were some issues with generalising findings from specific subpopulations (children with intellectual disability, persons with Down Syndrome) to the broader intellectual disability 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 intellectual disability 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 intellectual disability can be performed with a high level of standardisation, interviewing persons with varying degrees of intellectual disability 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; Zabalia, 2013).

Table 2

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

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
1. Item generation			
Participation of people with intellectual disability	Use focus groups to gather insights in cognitive processes, personal experiences and idiom used by intellectual disability 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.
2.1 Creation of content: format and language			
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.

Table 2

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

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
1. Item generation			
Participation of people with intellectual disability	Use focus groups to gather insights in cognitive processes, personal experiences and idiom used by intellectual disability 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.
2.1 Creation of content: format and language			
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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	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 intellectual disability population.
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) memory.	Bell 2018, Finlay 2001, Jen-Yi 2015, Scott 2018	High confidence Coherent evidence base form diverse studies.
	Don't ask to retrieve detailed information over longer periods of time.	Vlot van Anroij 2018	Moderate confidence Replication needed for broader intellectual disability population.
	Ask respondents to reflect on what they *currently* feel and think. Stay in the here and now.	Williams 2007	Moderate confidence Replication needed for broader intellectual disability population.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
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.
2.2 Creation of content: Response format			
Dichotomous answer options	Yes/no type questions are understood by the largest proportion of persons with intellectual disability.	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.

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 intellectual disability.	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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
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.
	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 intellectual disability.	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 intellectual disability 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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
2.3 Creation of content: Supportive media and lay-out			
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).
	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 Knegt 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 Knegt 2017	Moderate confidence

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
			Compelling evidence but from a specific and small sample of participants. Need for replication.
Lay-out	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.

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 intellectual disability to report internal states autonomously or with minimal support.	Clark 2017, de Knegt 2017	Moderate confidence Computerised testing offers great advantages, but research on the opportunities and pitfalls for people with intellectual disability is scarce. Furthermore, this recommendation lacks specificity.
3. Piloting draft versions			
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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	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)
4.1 Application in practice: formal assessment procedure			
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 Knegt 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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
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. .
	or		
	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.
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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	Group items on related topics.	Jen-Yi 2015	Low confidence Although common sense, the recommendation reflects the professional opinion of the researcher.
	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	Moderate confidence 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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
4.2 Application in practice: Role of assessor in case of assisted administration or structured interview			
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.
	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, prescripted 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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
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.
	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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	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.
5. Ongoing development			
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.
	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.

Development stage and corresponding issues	Recommendation	Contributing studies	GRADE-CERQual assessment
	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)

Discussion

Research on the ability of persons with intellectual disability 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 intellectual disability 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 sensical 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 intellectual disability 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 intellectual disability 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 ‘intellectual disability 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 intellectual disability by Jahoda et al. (2017), which offers interesting insights into the ability of people with intellectual disability to self-report cognitions and feelings. Or the chapter on interviewing people with intellectual disability by Prosser and Bromley (1998), that offers guidance on conducting interviews with persons with intellectual disability. 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 intellectual disability 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 intellectual disability’, 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 intellectual disability. 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 intellectual disability (Emerson et al., 2013). Acknowledging these limitations leads to the question of how to involve people with more severe levels of intellectual disability.

Third, many recommendations from the current review appear to be based upon common sense and not specific to intellectual disability 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 intellectual disability 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 (Nicolaides 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 intellectual disability, and these results need to be replicated in empirical studies involving participants with intellectual disability. 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 intellectual disability, meaning that most of the recommendations made within the current systematic review are in relation to those with borderline to mild intellectual disability.

Acknowledging the need to make self-report research and practice accessible for people with intellectual disability 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 intellectual disability-inclusive measures, that would at least be suitable for most persons with mild intellectual disability, 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 intellectual disability 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 intellectual disability community and extend to persons with other impairments.

Acknowledgements

The authors thank Janneke Staaks at the University of Amsterdam, for her invaluable guidance with drawing up and executing the search strategy, and Martina de Witte for providing very valuable feedback on earlier versions of the article.

References

Studies marked with an asterisk (*) are included in the systematic review.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.).

*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. https://doi.org/10.1177/104973239900900402

*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. 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. https://doi.org/10.1371/journal.pone.0235390

*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. https://doi.org/10.1080/1068316X.2017.1298761

*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. https://doi.org/10.1080/19315864.2011.593696

*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. https://doi.org/10.1111/j.1468-3148.2007.00404.x

*Bonham, G. S., Basehart, S., Schalock, R. L., Marchand, C. B., Kirchner, N., & Rumenap, J. M. (2004). Consumer-based quality of life assessment: The Maryland ask me! Project. *Mental Retardation, 42*(5), 338–355. https://doi.org/10.1352/0047-765(2004)42,338:CQOLAT.2.0.CO;2

*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. https://doi.org/10.1002/acp.2953

Bredel, U., & Maass, C. (2016). *Leichte sprache. Theoretische grundlagen? Orientierung für die Praxis* [Easy language. Theoretical base? Orientation for practical use]. 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. https://doi.org/10.1016/j.jbusres.2015.10.138

Carroll, C., Booth, A., Leaviss, J., & Rick, J. (2013). "Best fit" framework synthesis: Refining the method. *BMC Medical Research Methodology,13*(1), 37. https://doi.org/10.1186/1471-2288-13-37

Chachamovich, 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. https://doi.org/10.1016/j.jclinepi.2008.03.002

*Clark, L., Pett, M. A., Cardell, E. M., Guo, J. W., & Johnson, E. (2017). Developing a health-related quality-of-life measure for people with intellectual disability. *Intellectual and Developmental Disabilities, 55*(3), 140–153. https://doi.org/10.1352/1934-9556-55.3.140

Cockerill, H. (2002). Supporting communication in the child with a learning disability. *Current Paediatrics, 12*(1), 72–76. https://doi.org/10.1054/cupe.2001.0251

Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual context effects in web surveys. *Public Opinion Quarterly, 71*(4), 623–634. https://doi.org/10.1093/poq/nfm044

*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. https://doi.org/10.1111/j.1468-3148.1997.tb00017.x

Cummins, R. A. (2002). Proxy responding for subjective well-being: A review. *International Review of Research in Mental Retardation, 25*, 183–207. https://doi.org/10.1016/S0074-7750(02)80009-X

*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. https://doi.org/10.3109/13668250.2013.832734

*Cuthill, F. M., Espie, C. A., & Cooper, S. A. (2003). Development and psychometric properties of the Glasgow Depression Scale for people with a Learning Disability. Individual and carer supplement versions. *The British Journal of Psychiatry, 182*(4), 347–353. https://doi.org/10.1192/bjp.182.4.347

*Dagnan, D., & Ruddick, L. (1995). The use of analogue scales and personal questionnaires for interviewing people with learning disabilities. *Clinical Psychology Forum, 79*, 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. https://doi.org/10.1080/1068316X.2017.1284217

*de Knegt, 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. Https://doi.org/10.1016/j.ridd.2013.07.016

*de Knegt, 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. https://doi.org/10.3109/07434618.2015.1100677

*Emerson, E., Felce, D., & Stancliffe, R. J. (2013). Issues concerning selfreport data and population-based data sets involving people with intellectual disabilities. *Intellectual and Developmental Disabilities, 51*(5), 333–348. https://doi.org/10.1352/1934-9556-51.5.333

*Fang, J., Fleck, M. P., Green, A., McVilly, K., Hao, Y., Tan, W., Fu, R., & 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*(6), 537–549. https://doi.org/10.1111/j.1365-2788.2011.01401.x

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

*Finlay, W. M., & Lyons, E. (2002). Acquiescence in interviews with people who have mental retardation. *Mental Retardation, 40*(1), 14–29. https://doi.org/10.1352/0047-6765(2002)040,0014:AIIWPW.2.0.CO;2

Freedman, R. I. (2001). Ethical challenges in the conduct of research involving persons with mental retardation. *Mental Retardation, 39*(2), 130–141. https://doi.org/10.1352/0047-6765(2001)039,0130:ECITCO.2.0.CO;2

Fujiura, G. T. (2012). Self-reported health of people with intellectual disability. *Intellectual and Developmental Disabilities, 50*(4), 352–369. https://doi.org/10.1352/1934-9556-50.4.352

*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. https://doi.org/10.16993/sjdr.615

*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. https://doi.org/10.1023/A:1022282521625

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-IDs). *British Journal of Clinical Psychology, 53*(3), 315–332. https://doi.org/10.1111/bjc.12048

*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*(11), 813–827. https://doi.org/10.1111/j.1365-2788.2006.00844.x

*Heal, L. W., & Sigelman, C. K. (1995). Response biases in interviews of individuals with limited mental ability. *Journal of Intellectual Disability Research, 39*(Pt 4), 331–340. Https://doi.org/10.1111/j.1365-2788.1995.tb00525.x

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Nicolau, B. (2018). *Mixed methods appraisal tool (MMAT)* (Version 2018). 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. Https://doi.org/10.1111/j.1468-3148.2011.00624.x

Huus, K., Granlund, M., Bornman, J., & Lygnegård, F. (2015). Human rights of children with intellectual disabilities: Comparing self-ratings and proxy ratings. *Child: Care, Health and Development, 41*(6), 1010–1017. https://doi.org/10.1111/cch.12244

*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. https://doi.org/10.1007/s12187-015-9307-5

Jahoda, A., Stenfert-Kroese, B., & Pert, C. (2017). *Cognitive behaviour therapy for people with intellectual disabilities*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-47854-2

*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. Https://doi.org/10.1016/j.alter.2015.07.006

*Jobson, L., Stanbury, A., & Langdon, P. E. (2013). The Self- and Other-Deception Questionnaires-Intellectual Disabilities (SDQ-ID and ODQID): Component analysis and reliability. *Research in Developmental Disabilities, 34*(10), 3576–3582. Https://doi.org/10.1016/j.ridd.2013.07.004

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. https://doi.org/10.1007/s10803-016-2852-z

*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. https://doi.org/10.1080/13668250701378538

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. https://doi.org/10.1111/j.1468-3148.1999.tb00088.x

*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. Https://doi.org/10.1016/j.rasd.2018.06.010

Kramer, J. M. (2009). A mixed methods approach to building validity evidence: The child occupational self assessment. Dissertation Abstracts International: 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. https://www.jstor.org/stable/30033265

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

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. https://doi.org/10.1016/j.dr.2010.01.001

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 GRADECERQual to qualitative evidence synthesis findings: Introduction to the series. Implementation Science, 13(2, Suppl. 1), 2. https://doi.org/10.1186/s13012-017-0688-3

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. https://doi.org/10.5042/amhid.2010.0320

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. https://doi.org/10.1111/j.1365-2788.1988.tb01440.x

Lindsay, W. R. (2002). Research and literature on sex offenders with intellectual and developmental disabilities. Journal of Intellectual Disability Research, 46(Suppl. 1), 74–85. https://doi.org/10.1046/j.1365-2788.2002.00006.x

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(1), 55–68. https://doi.org/10.1348/135532505X85882

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

*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. https://doi.org/10.1352/0047-6765(1997)035,0075:AIQIWA.2.0.CO;2

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

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: International Journal of Methodology, 45(4), 801–815. https://doi.org/10.1007/s11135-010-9370-4

Mindham, J., & Espie, C. A. (2003). Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID): Development and psychometric properties of a new measure for use with people with mild intellectual disability. Journal of Intellectual Disability

Research, 47(1), 22–30. https://doi.org/10.1046/j.1365-2788.2003.00457.x

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 GRADECERQual to qualitative evidence synthesis findings-paper 3: How to assess methodological limitations. Implementation Science, 13(1, Suppl. 1), 9. https://doi.org/10.1186/s13012-017-0690-9

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. https://doi.org/10.1016/j.pcd.2012.11.001

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: Lessonslearned and recommendations. Autism in Adulthood: Challenges and Management, 2(1), 61–76. https://doi.org/10.1089/aut.2019.0074

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

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. Systematic Reviews, 5(1), 210. https://doi.org/10.1186/s13643-016-0384-4

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 studiesreview. International Journal of Nursing Studies, 49(1), 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Page, M. J., Moher, D., 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., . . . McKenzie, J.E. (2020). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. MetaArXiv. 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. https://doi.org/10.1002/cpp.412

*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. https://doi.org/10.1111/j.1741-1130.2006.00092.x

*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. https://doi.org/10.1352/0047-6765(2002)040,0445:SAOQOL.2.0.CO;2

*Power, M. J., & Green, A. M. (2010). Development of the WHOQOL disabilities module. Quality of Life Research, 19(4), 571–584.

Prosser, H., & Bromley, J. (2012). Interviewing people with intellectual disabilities. In E. Emerson, C. Hatton, K. Dickson, R. Gone, A. Caine, & J. Bromley (Eds.), Clinical psychology and people with intellectual disabilities (pp. 107–121). John Wiley & Sons. https://doi.org/10.1002/9781118404898.ch6

*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. https://doi.org/10.1007/s10882-008-9120-x

*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. https://doi.org/10.1016/j.ridd.2006.01.001

*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. https://doi.org/10.1007/s10882-005-6617-4

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. https://doi.org/10.1002/(SICI)1099-1298(199608)6:3,207::AID-CASP370.3.0.CO;2-T

*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. https://doi.org/10.1007/s11031-008-9117-x

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

39. https://doi.org/10.1186/s13643-020-01542-z

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. https://doi.org/10.1016/j.evalprogplan.2008.02.001

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. American Association on Intellectual and Developmental Disabilities.

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

*Scott, H. M., & Havercamp, S. M. (2018). Comparison of self- and proxyreport of mental health symptoms in people with intellectual disabilities. Journal of Mental Health Research in Intellectual Disabilities, 11(2), 143–156. https://doi.org/10.1080/19315864.2018.1431746

Sentell, T. L., & Ratcliff-Baird, B. (2003). Literacy and comprehension of Beck Depression Inventory response alternatives. Community Mental Health Journal, 39(4), 323–331. https://doi.org/10.1023/A:1024072108467

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), 53–58.

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

*Stancliffe, R. J., Wiese, M. Y., Read, S., Jeltes, 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. https://doi.org/10.1111/jar.12358

*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. https://doi.org/10.1111/jir.12024

*Stancliffe, R. J., Tichá, 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. https://doi.org/10.1352/1934-9556-53.3.163

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. https://doi.org/10.1111/j.1468-3148.1998.tb00054.x

Taal voor Allemaal. (n.d.). De Taal voor Allemaal methode [The Language for All method]. Retrieved February 5, 2021, from 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. https://doi.org/10.1093/poq/nfh035

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. https://doi.org/10.1093/poq/nfx035

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

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. 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. In N. Beail (Ed.), Psychological therapies and people who have intellectual disabilities (pp. 115–139). British Psychological Society, Division of Clinical Psychology.

*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(1), 80. https://doi.org/10.1186/s12874-018-0539-1

*White-Koning, M., Arnaud, C., Bourdet-Loubère, S., Bazex, H., Colver, A., & Grandjean, H. (2005). Subjective quality of life in children with intellectual impairment—How can it be assessed? Developmental Medicine and Child Neurology, 47(4), 281–285. https://doi.org/10.1017/S0012162205000526

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. https://doi.org/10.1111/j.1365-2788.2011.01440.x

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

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. https://doi.org/10.1002/mrdd.20133

*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. https://doi.org/10.1080/10683160601060739

Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2004). Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. Research on Social Work Practice, 14(3), 191–200. https://doi.org/10.1177/1049731503257882

Zabalia, M. (2013). Beyond misconceptions: Assessing pain in children with mild to moderate intellectual disability. Frontiers in Public Health, 1, 23. https://doi.org/10.3389/fpubh.2013.00023

Chapter 3

Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people's views.

Published as:

Kooijmans, R., Langdon, P. E., & Moonen, X. (2022). Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people's views. *Methods in Psychology*, *7*, 100102.

Abstract

Many children and youth struggle to complete surveys and questionnaires by themselves. They are often assisted when asked to give their opinion. From discussions with youth from the client council at a residential treatment facility for youths in The Netherlands, the notion emerged that interaction factors may impact the results of surveys, especially when sensitive topics are addressed.

Using a mixed methods design, we explored the question if and how survey results are influenced by the presence of an assistant during assessment. 120 children and youth that reside at one of the treatment facilities of Koraal, a Dutch multi-site care facility, completed a survey about the perceived quality of care at the facility. They were randomly assigned to one of three conditions: (a) unassisted, (b) assisted by their care worker, or (c) assisted by a research assistant. The resulting scores in each condition were compared quantitatively. In successive focus groups with children and youth, the results and possible explanations were discussed.

Participants in the Assisted by care worker condition exhibited significantly higher satisfaction scores than participants in each of the other two conditions. Results from the focus groups indicated that complex client-carer interaction dynamics contribute to these differences. Several explanatory mechanisms and implications for practice were suggested by the participants.

These results suggest that bias may be introduced when children and youth need assistance to complete surveys or diagnostic measures. This requires careful consideration on the part of researchers working with these vulnerable participants.

Introduction

Despite worldwide efforts to prevent out-of-home placements, millions of children and youth continue to move to live in residential group care settings because the preconditions for growing up at home cannot be met satisfactorily (van IJzendoorn et al., 2020). Children and youth admitted to group care are at risk of physical and emotional abuse, and might have unmet medical needs (Desmond et al., 2020). Potential adverse consequences include delays in physical growth, cognition, attention, socioemotional development and difficulties with mental health (Van IJzendoorn et al., 2020).

One potential way to mitigate the potential adverse effects of living within residential group care settings is to explicitly consider the views of children and youth about the quality of care in participative research and care improvement programs (Goldman et al., 2020; ten Brummelaar et al., 2018). An example of an instrument that was purposefully designed to this end is the My Opinion (Wissink and Kooijmans, 2020) survey. The My Opinion survey was constructed to accommodate the challenges associated with self-report measures for vulnerable populations, including children, persons with reading problems and persons with intellectual disabilities (ID; Kooijmans et al., 2022). Examples of 'ID-inclusive' features include Easy Read language, 3- to 5-point Likert scale options with supportive visualisations and digital assessment with a read-aloud function for questions.

At Koraal, it is standard practice for staff and children to engage in joint reflection upon the results of the My Opinion survey. Recently, survey results were discussed with the Koraal client council, incorporating youth, and it was noted by council members that reported satisfaction with care was high. This seemed to contradict the council members' perception of how children and youth generally expressed their views on the quality of care at the facility. Consequently they expected the results to be much less rosy than those reported via the survey. The Koraal client council suggested that the survey may have been biased in a positive direction because care workers assisted many children and youth with completing the survey. As a consequence, the council recommended (a) the completion of a brief literature review to explain why biased responding may occur with reference to acquiescence, social desirability and dependency effects, and (b) an exploration of mechanisms that may introduce bias within self-report assessments leading to the generation of hypotheses and an associated methodology which informed the development of the current study.

Factors that may explain overreporting of satisfaction in self-report client surveys

There are numerous potential sources of bias that threaten the reliability and validity of self-report measures (Havercamp et al., 2021). It has been suggested that as much as 40% of the variance in self-reported data can be explained by response biases (McCrae, 2018). These biases may be more marked amongst vulnerable populations, including children, whose cognitive, communicative and social skills have not fully developed (Bell, 2007) and persons with limited cognitive abilities (Nicolaidis et al., 2020). Difficulties with understanding may lead to incorrect or incomplete responses, while the introduction of support from another person when completing self-report measures may introduce certain types of bias, such as socially desirable responding, as a consequence of respondent-assistant dynamics (Finlay and Antaki, 2012; Kramer et al., 2010).

How misunderstanding can inflate satisfaction scores

A variety of different sources of bias may affect responses to self-report measures when used with vulnerable populations. Research on children, people with ID, and low literate persons suggests that sources of bias may include (a) acquiescence which is the tendency to say yes to questions regardless of content, (b) recency bias which is the tendency to select the last option mentioned in multiple-choice questions, irrespective of one's true opinion, (c) nay-saying which means saying no to every question, and (c) suggestibility (Bell et al., 2018) which refers to willingness to change answers following suggestions from another person. These response biases are more prevalent when individuals do not know the answer to the question (Emerson et al., 2013). For children, the variance that can be attributed to acquiescent responding can be twice as large than for adults (Soto et al., 2008).

Returning to consider the My Opinion survey (Wissink and Kooijmans, 2020), acquiescent responding may have occurred as all questions are positively phrased; negatively worded questions or questions using double negatives were not used as they tend to be confusing and lead to more errors (Payne and Jahoda, 2004). Acquiescent responding can occur when positively phrased questions are misunderstood leading to an increase in the frequency of affirmative answers.

Interaction as a source of bias

The My Opinion questionnaire was designed to be accessible for those who have difficulties with reading and understanding information, and while it is recommended that children and youth should be given the opportunity to complete the survey by themselves, in practice this happens rarely. This is because many children and youth may seek help from others or are offered unsolicited help by adults involved in their day-to-day care.

It cannot be assumed that unassisted completion will yield the same results as assisted assessment. Pioneers in research on complex carer-client interactions, such as Antaki and Rapley (Antaki, 2013; Rapley and Antaki, 1996), have demonstrated conclusively that these interactions can heavily influence the outcomes of discussions about support needs. Garton and Copland (2010) showed that any prior relationship between interviewee and interviewer turns objective accounts of the interviewee's reality into an interactional event where meaning is constructed jointly. Interviewers may willingly or unwillingly direct respondents towards certain answers by the way they react to respondents' verbal and nonverbal expressions. For instance, by nodding to favorable answers or frowning upon criticism.

Answers are also shaped by the respondent's expectations about how the assistant will react to their answer. When the relationship is non-symmetrical, general *submissiveness* may contribute to acquiescent responding (Finlay and Lyons, 2002). Submissiveness is the tendency to conform to the opinions of people with authority (Finlay and Lyons, 2002).

Social desirability may occur when questions on sensitive or even taboo subjects are asked (Krumpal, 2013). The respondent may be reluctant to admit to socially or culturally unaccepted behavior in the presence of an authority figure (Bell, 2007; van de Mortel, 2008).

Pleasing occurs when people will answer the question in a certain way because they think it is the 'right' answer the interviewer would like to hear (Rapley and Antaki, 1996). Children may especially report more socially desirable behavior (or less socially undesirable behavior) when they fear that this information is shared with their parents or other adult authorities, a phenomenon that is known as 'the bystander effect' (Havermans et al., 2015). In residential facilities, despite efforts to create more egalitarian relationships between care workers and clients, children and youth are still taught to conform to their care workers' expectations (De Valk et al., 2019). Children and youth who openly criticise care worker competence or other 'sensitive' aspects of quality of care may be seen as noncompliant or oppositional and may face repercussions as a consequence. This may lead to an increased chance of accepting suggestions and advice from others (Douma et al., 2012).

An important construct that may have an impact upon response style is the nature and degree of *dependency* children and youth have upon adults. They depend upon their carers to organize and provide support to fulfil their physical and psychological needs, while they are also the conduit to the "outside world", including family and friends (Moonen, 2019). As a consequence, children and youth in residential care may work to maintain positive relationships with care workers and avoid offering any criticism.

The present study

In the current study we explore response bias that is introduced when children and youth with MBID are assisted completing a satisfaction questionnaire. To investigate this, we used a sequential explanatory mixed-method design (Creswell, 2016).

In the quantitative phase, participants were allocated to one of three groups and invited to complete the My Opinion questionnaire either (a) unassisted, (b) with assistance from a care worker, or (c) with assistance from someone unknown to the participant. Group differences were analysed using inferential statistics. Based on the literature and the input from the client council we hypothesised that the participants in the assisted conditions would return higher satisfaction scores than those in the unassisted condition.

In the qualitative phase, the findings from the quantitative study were followed up within a series of online focus groups with children and youth from the target population. In line with a participatory research approach (Groundwater-Smith et al., 2014), we regarded our participants as coresearchers who were actively involved in the explanatory process. Participants discussed possible explanations for the findings from the quantitative study, to provide first-person views on the topic of response bias.

The aim of the integration of quantitative and qualitative results in an explanatory sequential design was to advance the knowledge on the occurrence and nature of response bias in survey research, leading to recommendations for future practice.

Methods

Mixed-method research design overview

An explanatory sequential design (Creswell, 2016) was used to guide the methodology of this study. In such a design, quantitative data are analysed statistically, but limited inferences are made from the data. The qualitative part of the design is used to explore possible explanations for the quantitative results (Walker and Baxter, 2019). First, we performed quantitative analyses on the results from the survey, comparing the three experimental conditions. We then followed-up the quantitative survey analysis with online focus group interviews because (a) we could not adequately explain the results from the quantitative analyses by referring to the extant literature, as the explanations offered in the literature were diverse and inconclusive; and (b) because we as researchers did not want to make inferences about the subjective justifications of children and youth. Instead, we wanted participants to explain the results to us themselves and in their own words.

For the quantitative study, a randomized experimental design was used. Participants were randomly assigned to one of three conditions, using simple randomisation with an online random number generator (www.random.org). In the first condition, participants completed the My Opinion survey unassisted, in the second condition they were assisted by a care worker, and in the third condition they were assisted by a research assistant. The results were analysed to test if satisfaction scores differed between conditions.

To explore explanations for the results found in the quantitative study, children and youth participated in online focus groups and individual interviews. In the structured focus group interviews, participants reflected on the quantitative study results and explored possible explanations for the results.

The Journal Article Reporting Standards (JARS) for Quantitative (JARS-Quant; APA, n.d.) and Qualitative (JARS-Qual; APA, n.d.) Research were used to guide the reporting of the quantitative and qualitative study components respectively. The Mixed-Method Article Reporting Standards (MMARS; APA, n.d.) were used to report the integration of quantitative and qualitative data.

Quantitative study

Participants

An a priori power analysis using the G-power computer programme (Faul and Erdfelder, 2007) indicated that a total sample of between 66 and 159 respondents would be needed to detect medium to large main effects (η2 = 0.06 to 0.14) with 80% power using an ANOVA with alpha at .05. We set out to include a minimum of 160 participants, but despite efforts to motivate more participants for inclusion and higher attrition rates than expected (see Participant Recruitment and Sampling) we had to settle for a final study sample of 120 participants. This meant that the minimum population sample requirements to detect a large effect were met, but in case the main effect proved to be in the direction of a medium-sized effect, the design was slightly underpowered.

Participants were children and youth aged 11–23 years that temporarily resided in one of three residential treatment facilities of Koraal, an organisation for care and education based in the south of The Netherlands. Most of the children and youth admitted to these facilities have a mild intellectual disability or borderline level of general intellectual functioning (MBID) and were admitted to the facility by formal referral. Of participants for whom a recent Full Scale IQ was available, 50% of participants had a total IQ score below 70, 49% had a total IQ score between 70 and 84 and one participant had an IQ score of 86. The level of adaptive functioning was not formally assessed for any of the included participants.

There were no group differences in terms of demographic characteristics across conditions. For Age, $F(2, 117) = 0.233$, $p = .792$, for IQ, $F(2, 61) = 1.396$, $p = .255$. For Gender, $X^2(2, 120) = 2.129$, $p = .345$. Participant characteristics are displayed in Table 1.

Table 1.

Characteristics of Quantitative Study Participants

	Unassisted	Assisted by		Total
		Care Worker	Research Assistant	
	(N=49)	(N=40)	(N=31)	N=120
Age				
Mean (SD)	16,0 (2,9)	16,2 (2,8)	15,7 (3,2)	16,0 (2,9)
Gender				
Female (%)	20 (41%)	16 (40%)	8 (26%)	44 (36,7%)
Male (%)	29 (59%)	24 (60%)	23 (74%)	76 (63,3%)
IQ[a]				
Mean (SD)	70 (9,5)	67 (7,8)	71 (7,9)	69 (8,5)

[a] Total N for IQ is 64. For 56 participants (47%) their recent total IQ score was unknown or they or their parents did not consent to share recent IQ scores.

Participant Recruitment and Sampling

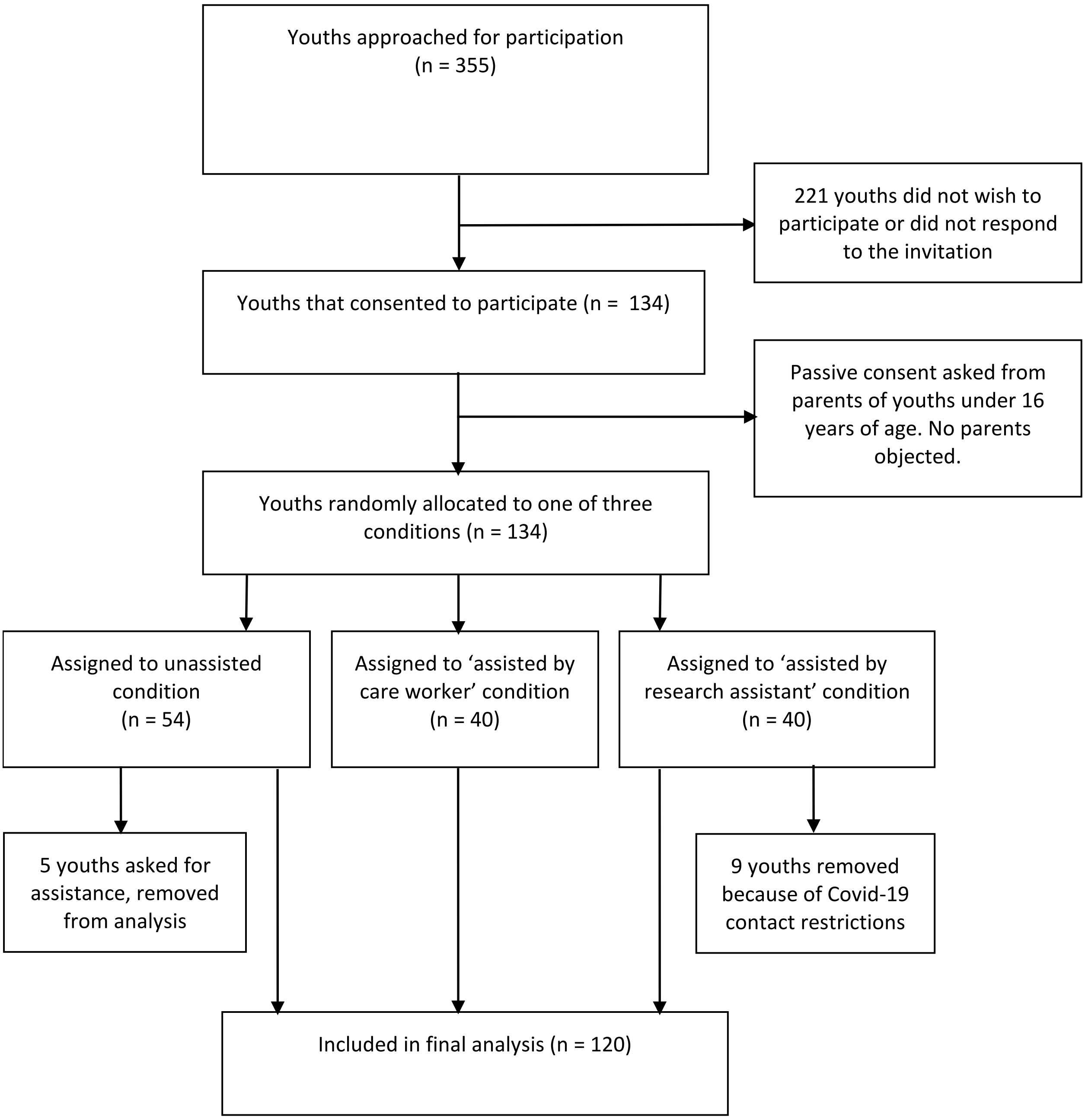
For the purpose of the present study, potential participants received a flyer, information letter, consent form with study details and requirements for participation. Because all participants are part of a vulnerable low-literate target group, information and consent forms were formatted according to Easy-Read guidelines. Of 355 children and youth that were approached, 134 (38%) consented to participate in the research. Parents and guardians of children and youth under 16 years of age were asked to consent to the participation of their child. No parent denied participation for their child. All participants were rewarded with a gift card to the value of €5.

During the actual survey period, due to rising numbers of Covid-19 infections, the enforcement of strict limitations in contact between clients and persons from outside the institution limited the access of research assistants to participants. As a result, nine participants were dropped from the *Assisted by research assistant* condition. They were offered to complete the survey unassisted or assisted by a care worker, and their results were omitted from the analyses. Participant recruitment and sampling flow can be found in figure 1.

The quantitative part of this study was reviewed and approved by the Ethics Review Board from the University of Amsterdam (registration number 2019-CDE-10133).

Figure 1.

Participant Recruitment Flow Diagram



Instruments and procedures

My Opinion client satisfaction survey. The My Opinion questionnaire measures a client's satisfaction with aspects of received care and quality of life on 17 items. Assessment results in scores on the subscales client-carer relation, autonomy, leisure & physical environment and group climate and in a total satisfaction score. The questions are presented in a digital format, one question per page. The language for items and responses are formatted according to Dutch Easy-Read guidelines (Moonen, 2021). Responses are given on a 5-point Likert scale. A read-aloud function and visualised response options assist participants who are less able to read. The My Opinion questionnaire has been validated in samples of children and youth with MBID (De Meyer, Van Dam and Delsing, 2016) and adults with mild to moderate intellectual disabilities (Wissink and Kooijmans, 2020). The internal consistency of the My Opinion questionnaire is satisfactory (Cronbach's α' s range from 0.78 to 0.83 across populations). Adequate convergent validity was observed when comparing My Opinion results to the results of a comparable youth client survey, the so-called C-test (Dutch: C-toets; Franssen and Jurrius, 2005). Cronbach's Alpha for the My Opinion 17-item total score in this study was 0.84.

Assessment procedure and conditions. After randomisation, all participants and their care workers received instruction on how to complete the survey.

In the *Unassisted* condition participants were instructed to choose a quiet room where there was minimal chance of disruption and completed the questionnaire by themselves. Follow-up contact with respondents indicated that all participants in the *Unassisted* condition had indeed managed to complete the survey without help.

In the *Assisted by care worker* condition, care workers were instructed to schedule a moment for the assessment with the participant and complete the survey together at the appointed time. The care worker read aloud every question and the participant was subsequently invited to answer the question. The care worker was instructed to stick to the procedures outlined in the survey manual, which are meant to standardise the assessment as much as possible.

In the *Assisted by research assistant* condition, the same procedure as in the *Assisted by care worker* condition was followed. The research assistant was not familiar to any of the participants beforehand.

Quantitative data analysis

For all analyses, the independent variable was the assessment condition and the dependent variable was the mean client satisfaction score across all items.

Homogeneity of variance was tested using Levene's test. The error variance of the dependent variable was equal across all groups in the model $F(17, 102) = 1.47$, $p = .122$. A Shapiro-Wilk test of normality revealed that scores in the Unassisted and Assisted by researcher conditions were distributed normally, $W(49) = 0.96$, $p = .059$, and $W(31) = 0.98$, $p = .834$, respectively, but the scores in the Assisted by care worker conditions were not, $W(40) = 0.91$, $p = .005$. Because the assumptions about normality of the distribution of scores could not be met, a Kruskal- Wallis test was performed to assess differences in the total satisfaction scores under the three conditions. Dunn's post-hoc tests were performed to further explore any statistical differences in group means. Qualitative study

Participants

Participants for the qualitative study were children and youth that lived in group care homes at Koraal. Seventeen children and youth participated in individual or focus group interviews. Characteristics of the interview participants are comparable to the quantitative study participants in terms of age and general cognitive capacities (see Table 2).

Table 2.

Characteristics of Qualitative Study Participants

	N	Percentage	Mean (min-max)	SD
Age	16[a]		15,8 (11-21)	2,46
Gender	17			
Female	2	12%		
Male	15	88%		
IQ	15[1]		70,5 (61-88)	8,34

[a] For 1 participant, age was not shared. For 2 participants a recent Full Scale IQ was not known or participants/parents did not consent to share IQ scores.

Participant Recruitment and Sampling

Participants were recruited from the three facilities' respective client council members. Participants were asked to participate in a small (maximum of 6 participants) focus group. At their own request or for logistic reasons, some of the children and youth participated in individual interviews. After the first interviews and focus groups, the data were analysed. In a series of iterative steps more interviews and focus groups were conducted until no new information was put forth by the participants. The final sample consisted of 17 participants. Of these, 13 were members of the client council. Four more participants were not council members but requested to participate voluntarily. Interview participants did not receive a monetary remuneration, but were given drinks and treats during the interview as a token of appreciation.

The focus group study was approved by the University of Amsterdam Ethics Review Board (registration number 2019-CDE-11604).

Instruments and procedures

Online focus groups. Due to COVID-19 contact restrictions, opportunities for organizing in-person focus groups were limited. As an alternative to in-person focus groups, we resorted to online meetings using Teams (Microsoft, 2018) as an online platform. In comparison with in-person focus groups, richness of data and themes that emerge in online focus groups are comparable. Researchers comparing both modalities even suggest that participants may be more candid when they discuss sensitive topics online than they would in-person (Woodyatt et al., 2016).

The main goal of the focus group meetings was to let participants reflect on the outcomes of the quantitative study. Before presenting the results, probing questions were asked to engage the participants and help stimulate the process of reflection. They were asked to predict outcomes using the online Mentimeter web polling application (www. mentimeter. com). Upon sharing the actual results, the participants were then asked if the results fitted their preconceived ideas. Mentimeter questions were further used to ask a multi-response question about possible mechanisms underlying the results (e.g. "Why do you think this result is observed? Because explanation a., explanation b., etc."). They were then asked to use their own lived experience of everyday life in the facility to elaborate on their explanations and forward examples to illustrate. Three main questions were the subject of study in the focus groups:

1. Satisfaction scores proved to differ between conditions. In what condition do the survey results reflect the children and youth's 'true' opinions?
2. Why do results differ between different conditions? Does the presence of a person assisting you distort the answers you give? What cognitions, feelings and expectations account for these distortions?
3. Knowing that assistance influences results, and acknowledging that many children and youth require assistance to complete the survey, how should researchers adjust survey assessment procedures to optimally reflect children and youth's true opinions?

These questions were broken up into sub-questions and reworded to take into account the verbal understanding skills of the children and youths under study. Questions were presented on screen and read aloud by the researchers. Each question was addressed by first asking 'How does this work for you?", then "How do you think it might work for others (and why?)" and finally "Knowing this, what should we do with the outcomes?". In case participants had different viewpoints on topics, they were encouraged to challenge each other's opinions through respectful discussion. The researchers served as mediators. When all questions surrounding a topic were answered, the researchers concluded by providing a summary of the participants' responses and asked them to corroborate or correct the proposed conclusions.

The interview procedure was scripted in detail to ensure a high level of standardisation. The scripts for individual interviews and focus groups were identical. The interview structure and online focus group format were pilot-tested with representatives from the client council (who were not part of the eventual study sample). Minor revisions to the script were made to clarify language, visualisations of survey results and the Mentimeter questions.

Three focus groups were held, one at each participating location, with four, six and two participants respectively, and a further five participants were interviewed individually at their own request. Each session lasted between 32 and 53 min and was recorded and transcribed verbatim. All sessions were moderated by the first author (RK) and a research assistant.

Qualitative data analysis

To analyse the transcripts from the focus groups and interviews, thematic analysis was applied following the steps outlined by Braun and Clarke (2012). The first author (RK) and a research assistant familiarised themselves with the data by viewing and reviewing the recorded interviews, comparing notes on the interviews and reading and rereading the transcribed interviews. The data were then restructured in a data charting form, where all relevant citations were arranged under the corresponding interview topic. The answers to the closed Mentimeter questions were seen as sensitising questions and were not formally analysed. Next, all text fragments, arranged per topic, were uploaded in Atlas.ti 8 for initial coding by the first author and a research assistant. A behavioral scientist with extensive clinical and research experience working with adolescents with MBID was added to the coding team. She was not involved in any part of the design or implementation of the survey or focus groups and was therefore seen as an 'impartial outsider', with no preconceived notion of the topic.

For the purpose of additional external validation, a second 'impartial researcher' was asked to review the raw data and the inferences that were made from them. This researcher was a seasoned qualitative research expert in the field of youth care research, working in an unaffiliated university. Her findings were used to corroborate, refute or finetune our own conclusions.

Integration of quantitative and qualitative data

The thematic analysis resulted in a set of tentative explanations for the results and recommendations for future practice, all forwarded by the children and youth themselves. In the Discussion, the results from the quantitative analyses were integrated with the explanations and recommendations offered by the participants. Conclusions were mapped against the interaction factors mentioned in the Introduction, to connect their observations and opinions with the extant literature. Recommendations and implications for practice were formulated based on the integrated results.

Results

Results of the quantitative analyses

Because non-parametric tests were used to test differences between conditions, both median and mean total scores are presented. Median and mean total satisfaction scores per condition can be found in Table 3. There was a significant difference between the three groups on satisfaction scores, *H*(2) = 12.2, *p* = .002, η2 = 0.09. Post-hoc testing revealed that children and youth assigned to the 'assisted by care worker' group exhibited significantly higher satisfaction scores than children and youth in the 'unassisted', *p* = .009, and 'assisted by researcher' groups, *p* = .006, while there was no significant difference between the 'unassisted' and 'assisted by researcher' group, *p* = 1.00.

Table 3.

Median and Mean Total Satisfaction Scores per Condition

	N	Median	Minimum – maximum[a]	Mean (SD)
Unassisted	49	2,53	0.65 – 3.71	2,49 (0,65)
Assisted by care worker	40	3.00	1.18 – 3.59	2,88 (0,52)
Assisted by researcher	31	2.41	1.18 – 3.88	2,45 (0,71)
Total	120	2,71	0.65 – 3.88	2,61 (0,65)

[a]possible minimum total score = 0, maximum total score = 4

Findings from the interviews and focus groups

All participating children and youth expressed that they had enjoyed discussing the findings with the researchers and with each other. Participants shared some differing but also very similar experiences of daily life in a treatment facility and gave very powerful insights in the complex dynamics between those receiving and providing care.

Below, the results from the interviews and focus groups are summarised under each of the three questions that are mentioned in the Methods section under Online Focus Groups. Per question, emergent themes from the thematic analyses are discussed.

Under what conditions do children and youth voice their 'true' opinions?

When discussing this topic, a notion emerged that was replicated throughout all interview topics: many respondents tended to make a distinction between what they themselves would think and do and how their peers might think and act. Specifically, some participants suggested that they would be honest under all conditions and that it wouldn't matter if someone was present. But others might find it difficult to give honest answers in the presence of an assistant. M (20 yrs, female): "I'm always honest, doesn't matter who's there ... I expect others to be honest too, no matter what... but maybe your answers aren't truthful if you feel it's something to be embarrassed of ... ?".

When the results of the quantitative experiment were presented, demonstrating that participants assisted by a care worker showed higher satisfaction scores, most participants said they were not surprised. Next, we presented the participants with two options as to what the 'true' value of the mean population satisfaction scores was: either respondents give unrealistically high scores in the presence of their care worker or they give unrealistically low scores in the other conditions. On the one hand, participants expected respondents to feel free to offer open and honest criticism in the 'unassisted' condition, whilst being hesitant to open up to their care workers. On the other hand, some participants stressed that children and youth may be tempted to exaggerate their criticism of the facility when they are alone or with a stranger, hence deflating 'true' mean satisfaction scores across conditions. J (17 yrs, male): "I for one like to be honest. I don't like lying, I really dislike it... but when they are on their own, I think they might give very low scores on purpose." In this case, the respondents' scores in the 'assisted by care worker' condition would more truly reflect the population satisfaction scores. One respondent stressed that there was really no way of knowing if the scores in the 'unassisted' condition are true scores as there is no one to follow-up on their answers to see if they understood the question and check if this is really how they felt: "They can answer just about anything, just to get it over with".

Why do results differ between conditions?

We further explored underlying mechanisms that could account for the observed difference between scores. Some very plausible, yet not conclusive nor mutually exclusive, explanations emerged that provide insight in how children and youth experience the complexities of carer - client dynamics in residential care. After sensitising the participants with a multiple-choice Mentimeter question, three main explanatory mechanisms emerged from the reflexive discussions on the results: (a) avoiding tension and conflict; (b) empathic reactions; and (c) dependence. A fourth theme pertained to (d) the quality of the relation of children and youth with their care worker as a mediating factor.

Avoiding tension and conflict. Almost all children and youth indicated that providing open and honest critical feedback in the presence of a care worker can be quite daunting. In the perception, and occasionally the direct experience of participants, being critical about the quality of care they receive can easily be mistaken as insolence by care workers.

J (15 yrs, male): “… for instance, if you had an argument with this person before, you may not want to tell the truth. Because you’re scared that the person gets angry with you.” Sometimes expressing an opinion will even be seen as a testimony of a youth’s ignorance - ‘they don’t know what’s good for them’. A (17 yrs, female): “Once, I made a decision that did not agree well with what my care worker thought was smart and then he and everyone around me got mad at me. And then I thought I should have been compliant with what they think is best. So that sort of thing makes me quite insecure about answering these kinds of questions with my care worker, they might get mad again.” Most participants stated that they can imagine that their peers may be apprehensive to provide negative feedback. But when asked if they had experienced at first hand that care workers reacted angrily if they expressed criticism, very rarely could they provide examples. On the contrary, many examples were put forward of care workers encouraging children and youth to be open and honest. A (15 yrs, male): “… generally they will remain calm, you can just give your honest opinion. They will say it’s okay to be honest.“. So this apprehension to provide critical opinions because of a fear of retribution on the part of the care worker seems to be quite generic, while there does not seem to be a direct justification for this fear.

Attunement. Participants made it very clear that completing a questionnaire is more than answering questions in a social vacuum. Children and youth consider the perceived expectations, thoughts and feelings of the person assisting them when thinking about what answer would be the most appropriate. Often, they may choose a response that aligns with what they think is expected of them, compromising between what they themselves feel and what they think is ‘the right answer’ in the eye of the care worker. A (17 yrs, female): “When you’re on your own you just think ‘this is it!’. But when you’re with your care worker you might think ‘Will she be ok with this?’. And then you’re considering what she may think, so it’s much better when you’re on your own.” J (17 yrs, male): “When you are with your care worker, you usually agree with him, you sort of say what the care worker wants you to say.”

Some participants are very considerate about the care worker’s feelings. They may fear that being critical about the quality of care is perceived to be a direct assault on the person providing care, projecting feelings of rejection or disappointment on the care worker. J (13 yrs, male): “I found it quite difficult that the care worker was there with me. I you are with them… well, you don’t want to offend them or anything.” M (20 yrs, female): “When there’s someone next to you, you take their feelings into account. You may worry that your care worker feels bad when you offer criticism.”

This consideration with what someone else thinks and feels does not apply to all situations where an adult is present. Participants expressly attributed this effect on satisfaction scores to the connection they have with their care workers. Attachment bonds with care workers form when children and youth are placed in a group home and this shapes the way they will respond in each other’s presence. In the absence of emotional ties with an impartial researcher, it is easier for children and youth to take a more rational stance. J (13 yrs, male): “I think it’s best to do it with an unfamiliar person, because he or she is sort of independent. So you can be more honest, because your care worker ..., well you want to keep him happy.”

Dependence. Not only do children and youth often experience an emotional bond between youth and care worker (Harder et al., 2013), children and youth in residential care often learn that from a rational point of view, it is in their best interest to collaborate with care workers. Placement in residential care often implies partly handing over your autonomy to care workers. Children and youth may rely on their care workers to organize or facilitate contact with their families, organize daily activities such as school and provide access to therapy. Care workers are often the prime informants for staff on how the youth is doing in terms of disruptive or adaptive behavior and goal attainment. As most children and youth are well aware that presenting a positive image of themselves greatly enhances the chances of terminating their stay in the facility and returning to their families, establishing a positive working relationship with their care workers is essential. Not complaining, keeping it positive, refraining from ‘being a nuisance’ contributes to maintaining this relationship and improves your chances of eventually going home. D (15 yrs, male): “I’m not complaining, what if they think I’m not doing ok in here?“.

The quality of the relation as a mediator

Participants offered many different explanations as to why they tend to be more positive in the presence of a care worker. But the impact of the proposed mechanisms is seen to depend greatly on the quality of the relation between youth and care worker. If a young person has established a secure connection with his or her care worker, if they experience mutual trust, respect and empathy, the explanations mentioned above more or less lose their pertinence. The opposite is actually true for many participants. J (20, male): “I think [I’d want to do the survey] with my care worker, because he kind of knows you, what you are like.” Having a secure bond with your care worker makes it easier to share and be honest. Moreover, the care worker is in the position to actually do something about your grievances, whereas an impartial researcher may not be able to provide a solution directly. M (20 yrs, female): “[... completing a survey with] my care worker, I feel most comfortable. And she arranges things for me.” When completing the survey by yourself, there would rarely be a guarantee that staff will follow-up on your complaints and do something about them. Especially when the results are processed anonymously. However, in the experience of some, opening up to your care worker does not necessarily imply that things will change for the better. J (17 ys, male): “So I talked to her and I said that the workers at my group home should change the way they approach me, come chat with me now and then, ask how I am... But she didn’t share it with the team, so nothing changed ...“.

How to optimise survey assessment procedures?

In a final reflection exercise, the participating children and youth were asked how the results should be translated to practice. How should the next survey be organized, knowing that many children and youth need assistance on the one hand, and seeing that it really does matter who is assisting you on the other? Not surprisingly, most participants asserted that children and youth should be offered a choice as to who assists them. There are many different preferences across children and youth. The defining factor is perceived to be the level of trust they have in their care workers. If the youth is not sure if he can trust his care worker, or is uncertain as to how the care worker will react to criticism, assistance from an impartial assistant can help them to give an honest opinion without feeling pressured into giving socially desirable answers. If there is trust and a general positive bond between youth and care worker, participants think assistance from their care worker is the best option. Mainly because of the opportunity to directly act upon the feedback, transforming complaints into an opportunity to make things better for the youth. Participants stressed once more that if staff invites children and youth to give an opinion, this means that they should be prepared to do something constructive with the results.

What was somewhat surprising to the researchers, was that not many participants advocated that it should be made possible for all children and youth to complete the survey unassisted. When asked why this option was not appealing to them, participants observed that many of their peers were unable to read and struggled to voice their opinions unaided. Although unassisted completion of the survey would minimize the chances that bias distorts the results, it was not deemed to be feasible for a large part of the population of MBID children and youth. This highlights the need to use measures that are adapted to accommodate for literacy problems and cognitive impairments.

Some participants mentioned that they would prefer trained peers to help them. Peers share the same perspective and may therefore better understand what children and youth are going through. They can be trusted, because they are ‘on the same side’.

In regard to the ‘other’ category, one participant expressed that he thought parents could help as well. They are the persons many children and youth trust the most, and who generally have a natural position to advocate the youth’s interests.

Discussion

In survey studies with children and youth, respondents are often in need of assistance to complete the survey. In this study, we set out to explore whether the results of a client satisfaction survey were influenced by the presence of an assistant. Based on the literature and the input from a client council we assumed that the respondents who were helped would return higher satisfaction scores than those who were not. In a subsequent focus group study we aimed to explore possible mechanisms underlying the results.

The quantitative analyses of the survey data showed that those who were assisted by a care worker had the highest satisfaction scores relative to children and youth who completed the survey unassisted or those who were helped by an assistant unacquainted to the participant.

Results from the focus groups and solo interviews revealed that several mechanisms might have contributed to the difference between conditions. Conscious deliberations and subconscious processes contribute to biased answering tendencies, reflecting both *functional* and *empathic* components.

In terms of *functionality*, most considerations for children and youth to choose an answering strategy refer to the possibility that offering open and honest criticism may have negative consequences for their stay at the facility. Some children and youth fear that care workers may get mad when they are told that they are not 'doing it right'. They perceive that it is in their best interest to keep the relationship positive at all times. This seems to reflect elements of a power dynamic; participants expressed that they feel they depend on the care worker to maintain contact with friends and relatives and to organize support to meet current and future support needs. Being positive about the quality of care at the facility in the presence of a care worker, may contribute to presenting a positive image of oneself (Van de Mortel, 2008), which may be beneficial for creating a positive working alliance. A positive working alliance in turn contributes to achieving adolescents' goals (Orsi et al., 2010). Children and youth who experience warm and trusting relationships with their care worker on the other hand, assert that they have nothing to fear and being critical is seen as an opportunity to improve care by both youth and care worker.

Aside from the functional aspect of the youth-carer relationship, some children and youth let empathic deliberations weigh in on their satisfaction ratings. They may report favorable satisfaction ratings because they may feel sorry for the care worker if they are being critical. As far as we know, empathy as a source of response bias has not been noted before in research on response biases (e.g., see Finlay and Lyons, 2002). The qualitative design, in which participants were asked to reflect on their motivations directly, may account for the emergence of this finding. This finding is somewhat surprising given the difficulties with mentalizing abilities – i.e. the ability to recognise and reflect on mental states of others and self, such as feelings and thoughts – that are observed in many children and youths with MBID (Allen et al., 2008). A general submissiveness is also observed, where children and youth may provide the answers they think care workers will want them to give (Borgers et al., 2000).

The extent to which these factors influence scores is subject to two moderating variables: (a) interpersonal variance; and (b) the quality of the relationship. Many participants acknowledged that they adjusted their answers to the person who happened to sit by them. Others maintained that they were not affected by these factors at all, and that they had no problem being open and honest in any situation. Regarding the quality of the relationship, the presence of a care worker at assessment generally inflates satisfaction scores, but predominantly so if children and youth are in a non-trusting relationship with their care worker. In that case, bias factors mentioned in the Introduction manifest themselves most clearly.

Strengths and limitations

Complex client-carer interactions have been observed to influence the outcomes of dialogues between those who receive and provide care (e.g. Finlay and Antaki, 2012). To our knowledge, the current study is the first that attempts to quantify the effect of assistance on survey outcomes in a population of children and youth. A marked strength to this study is the use of a mixed method design where qualitative methods were used to help us understand our results jointly with our participants.

An important limitation of this study is that the absence of a 'gold standard' for satisfaction limits the inferences that can be drawn from the results. Because the scores in the three conditions cannot be compared to an undisputed 'true' satisfaction measurement, there is no conclusive answer to the question if scores in the 'assisted by care worker' condition can be regarded as unrealistically high, or if the scores in the other two conditions are lower than they should be. Objectively, the scores can only be compared relative to one another, without referring to one as 'right' and the other as 'wrong'. Most explanations for the observed differences seemed to indicate that the impact of interpersonal dynamics is greatest in the presence of the care worker. For several reasons, participants expressed that most children and youth would be hesitant to be critical when assisted by their care worker, as opposed to being assisted by an impartial assistant or on their own. The proposed mechanisms suggest that scores in the 'assisted by care worker' condition are subject to inflation, as a result of submissive responding tendencies. But the current qualitative design does not permit this conclusion to be drawn with any undisputed certainty and this requires further exploration.

As always, qualitative analysis calls for careful reflection of the process of making inferences from statements by a limited number of participants to general statements and even theories. We have tried to ensure a certain degree of methodological integrity by letting several unaffiliated researchers participate in the process of analysis and perform checks on the conclusions we drew from the data. This has not led to major adjustments in our conclusions but helped to nuance some findings. It was especially helpful in that we were constantly reminded by the 'outsiders' to stay close to the data and be wary to make inferences from reading between the lines.

Another limitation in regard to the qualitative phase of this study relates to our decision to sample our participants from members of the client board. They may be more invested in the research than others living in the facility, thus raising the question of representativeness. On the other hand, the very reason that they were elected member of the board is that they are deemed (and trained) to represent the opinion of their constituency. When discussing topics, they are generally more prone to take into account different perspectives from their own than other residents of the facility. Regarding the transferability of our results to more general populations, we cannot be sure that the results from our study with MBID children and youth in a residential setting will be applicable to all children and youth that are asked to complete surveys. The cognitive impairments associated with MBID may have magnified the impact of some of the response biases. A greater tendency for acquiescent responding is observed in both cognitively impaired study populations (e.g. Finlay and Lyons, 2002) and children (Havermans et al., 2015) and there may be a confluence in this study. In a direct comparison of children with and intellectual disabilities, Ramirez (2005) did not find evidence for differential acquiescent responding. Further, children and youth with MBID in residential care may be especially sensitive to client-carer dynamics as a result of their dependency of carers to provide for their everyday needs (Harder et al., 2013). Obviously, the client-carer dynamics in residential care facilities do not apply directly to children and youth living with their parents or independently. Nevertheless, dependencies exist for these children and youth too, and the assistance from parents or social workers may unwittingly shape responses to surveys for children and youth outside residential care as well.

Directions for further research

This study offers some interesting yet experiential insights in the way children and youth take various interpersonal considerations into account when completing a survey with assistance. Many possible mechanisms are proposed by the participants, leading to tentative conclusions and recommendations, but these mechanisms were not tested empirically. Specifically, throughout the accounts of the participants, the quality of the client-care worker relation emerged as a possible mediator for the impact of bias. It is presumed by the participants that more valid scores are obtained if the respondent has an open and trusting relation with the care worker.

Similarly, the current design has allowed us to only make statements about relative differences between conditions. Scores in one condition were higher than in the other two. But no inferences could be made about which scores more accurately reflect true satisfaction. In subsequent studies, we intend to use cognitive interviewing techniques (Miller et al., 2014) to gain further insight into what is actually happening when a respondent completes the survey.

In this study, respondents were assigned to a condition randomly, receiving assistance on the basis of their allocated condition, rather than an observed or expressed need for assistance. In practice, many children and youth, especially with MBID, receive help unsolicited, often because they are deemed incapable of completing a survey by themselves. To promote a sense of autonomy, and reduce possible bias from interaction factors, unassisted completion could be beneficial in many instances. In order to allow more children and youth with MBID to complete surveys by themselves, more research on the boundaries of giving valid self-reports is needed. What can they do by themselves, how can inclusive design features support them and where is assistance really needed?

Implications for practice

From the results of this study, it is very likely that it matters who provides assistance. Researchers conducting a survey should not assume that helping children and youth to complete a survey is always beneficial if they want to elicit the true opinion of children and youth. When thinking about conducting a survey, especially children and youth that are placed in residential care, the option of letting respondents themselves choose who can assist them might be considered. However, this study has demonstrated that pairing participants to different modes of assessment based on their preference might lead to different results for different participants (i.e., those not assisted versus those who are assisted). In the design stage of a study, researchers should at least be aware of the possible distorting effect of offering assistance. They should carefully balance out the added value of tailoring assistance to the needs and preferences of participants and the value that is put on standardized research procedures. The first maximizes the validity of answers, especially in self-reported information on sensitive topics, while the second may be essential if reliability requirements call for standardized procedures. Where research designs involve offering respondents assistance, and there is a pre-existing relationship between a pair, then researchers may need to take this into account when designing studies and analyzing findings.

If for logistic or other practical reasons this is not feasible, it would be preferrable to arrange for ‘impartial outsiders’ to assist if needed. If there is no alternative to letting care workers assist participants, they should at least be offered the opportunity to choose the person they trust the most – or distrust the least.

Another suggestion that was offered by participants was to use trained peers as assistants. This might be especially helpful for children and youth who have a history of adverse life experiences involving adults and who have developed a general distrust against adults. Careful consideration is warranted, as peer assistants are usually not in the position to help turn results into actions and may struggle to provide after-care in reaction to strong emotional and behavioral reactions to the assessment, especially when covering sensitive topics.

The most reliable way to minimize interaction factors and ensure that participants will answer truthfully however, is to stimulate that as many children and youth as possible can complete the survey unaided. This calls for an effort to make the survey and the survey procedure as inclusive as possible. Evidence-based guidelines for adapting self-report instruments to persons with intellectual disabilities are provided by Kooijmans et al. (2022). These include the use of Easy Read guidelines, visualization, modified response options and the use of digital media such as text-to-speech and speech-to-text options. These recommendations extend beyond the field of intellectual disability research and are applicable to other vulnerable populations that benefit from inclusive research practices.

Further efforts are needed to connect survey outcomes from unassisted assessments to quality improvement measures that benefit the individual respondent. When children and youth experience that they are asked to give their opinion, but the results are not followed-up by tangible actions, they may be reluctant to contribute the next time we ask them.

The implications of this study may reach well beyond research with children and youths to other vulnerable populations. Generally, our results suggest that the nature of a relationship between a respondent and the person asking questions may introduce bias. While this has implications for research involving children and young people, it also has implications for others who struggle to complete questionnaires by themselves, including those with intellectual disabilities, reading difficulties, non-native English speakers or those with physical disabilities. ‘Dependency’ appears to be a major contributing factor to the occurrence of response bias, and other vulnerable populations may be equally dependent on maintaining positive relations with caregivers or other helpers. Although our study does not permit us to draw conclusions about the applicability of the results to other marginalized populations, there are implications for other studies where a respondent is asked questions by an interviewer.

Acknowledgments

The authors wish to thank all children and care workers at Koraal who chose to participate in this study.

References

Águila-Otero, A., González-García, C., Bravo, A., Lázaro-Visa, S., & del Valle, J. F. (2018). Children and young people with intellectual disability in residential childcare: Prevalence of mental health disorders and therapeutic interventions. *International Journal of Social Welfare, 27*(4), 337-347. https://doi.org/10.1111/ijsw.12351

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

American Psychological Association (n.d.). Journal Reporting Standards - Mixed Methods Research. Retrieved November 11, 2021, from https://apastyle.apa.org/jars/mixed-methods

American Psychological Association (n.d.). Journal Reporting Standards - Qualitative Research. Retrieved November 11, 2021, from https://apastyle.apa.org/jars/qualitative

American Psychological Association (n.d.). Journal Reporting Standards - Quantitative Research. Retrieved November 11, 2021, from https://apastyle.apa.org/jars/quantitative

Antaki, C. (2013). Two conversational practices for encouraging adults with intellectual disabilities to reflect on their activities. *Journal of Intellectual Disability Research, 57*(6), 580–588. https://doi.org/10.1111/j.1365-2788.2012.01572.x

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 and Law, 24*(4), 362–378. https://doi.org/10.1080/1068316X.2017.1298761

Charmaz, K. (2014). *Constructing grounded theory*. London: Sage.

Conlon, C., Timonen, V., Elliott-O'Dare, C., O'Keeffe, S., & Foley, G. (2020). Confused About Theoretical Sampling? Engaging Theoretical Sampling in Diverse Grounded Theory Studies. *Qualitative Health Research, 30*(6), 947–959. https://doi.org/10.1177/1049732319899139

De Meyer, R., Van Dam, C., & Delsing, M. (2016). *Rapportage psychometrisch onderzoek MijnMening* [psychometric evaluation report My Opinion]. Nijmegen: Praktikon.

Desmond, C., Watt, K., Saha, A., Huang, J., & Lu, C. (2020). Prevalence and number of children living in institutional care: global, regional, and country estimates. *The Lancet Child & Adolescent Health*, *4*(5), 370-377. https://doi.org/10.1016/S2352-4642(20)30022-5

De Valk, S., Kuiper, C., van der Helm, G. H. P., Maas, A. J. J. A., & Stams, G. J. J. M. (2019). Repression in Residential Youth Care: A Qualitative Study Examining the Experiences of Adolescents in Open, Secure and Forensic Institutions. *Journal of Adolescent Research, 34*(6), 757–782. https://doi.org/10.1177/0743558417719188

Douma, J., Wit, M., & Moonen, X. (2012). A guideline towards more effective interventions for youth with mild intellectual disabilities. *Journal of Intellectual Disability Research*, *56,* 7-8.

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.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191. https://doi.org/10.3758/BF03193146

Franssen, J., & Jurrius, K. (2005). De C-toets getoetst. Ervaringen, ontwikkelingen en plannen rondom de C-toets [Testing the C-test. Experiences, developments and plans for the C-test]. *Nederlands Tijdschrift voor Jeugdzorg, 9,* 293-296.

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. M. L., & Antaki, C. (2012). How staff pursue questions to adults with intellectual disabilities. *Journal of Intellectual Disability Research, 56*(4), 361–370. https://doi.org/10.1111/j.1365-2788.2011.01478.x

Finlay, W. M., & Lyons, E. (2002). Acquiescence in interviews with people who have mental retardation. *Mental retardation, 40*(1), 14-29. https://doi.org/10.1352/0047-6765(2002)040<0014:AIIWPW>2.0.CO;2

Fox, F. E., Morris, M., & Rumsey, N. (2007). Doing synchronous online focus groups with young people: Methodological reflections. *Qualitative Health Research, 17*(4), 539–547. https://doi.org/10.1177/1049732306298754

Goldman, P. S., Bakermans-Kranenburg, M. J., Bradford, B., Christopoulos, A., Ken, P. L. A., Cuthbert, C., Duchinsky, R., Fox, N. A., Grigoras, S., Gunnar, M. R., Ibrahim, R. W., Johnson, D., Kusumaningrum, S., Agastya, N. L. P. M., Mwangangi, F. M., Nelson, C. A., Ott, E. M., Reijman, S., van IJzendoorn, M. H., ... Sonuga-Barke, E. J. S. (2020). Institutionalisation and deinstitutionalisation of children 2: policy and practice recommendations for global, national, and local actors. *The Lancet Child and Adolescent Health, 4*(8), 606–633. https://doi.org/10.1016/S2352-4642(20)30060-2

Han, J., Torok, M., Gale, N., Wong, Q. J. J., Werner-Seidler, A., Hetrick, S. E., & Christensen, H. (2019). Use of web conferencing technology for conducting online focus groups among young people with lived experience of suicidal thoughts: Mixed methods research. *JMIR Mental Health*, 6(10). https://doi.org/10.2196/14191

Havercamp, S. M., Barnhill, L. J., Bonardi, A., Chapman, R. A., Cobranchi, C., Fletcher, R. J., Rabidoux, P., Seeley, J. R., & Tassé, M. J. (2021). Straight from the horse's mouth: Increasing self-report in mental health assessment in individuals with intellectual disability. *Journal of Applied Research in Intellectual Disabilities,* 1-9. https://doi.org/10.1111/jar.12952

Kooijmans, R., Mercera, G., Langdon, P.E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clinical Psychology: Science and Practice, 29*(3), 250-271. https://doi.org/10.1037/cps0000058

Kramer, J. M., Kielhofner, G., & Smith, E. v. (2010). Validity evidence for the child occupational self assessment. *American Journal of Occupational Therapy, 64*(4), 621–632. https://doi.org/10.5014/ajot.2010.08142

Maïano, C., Coutu, S., Morin, A. J. S., Tracey, D., Lepage, G., & Moullec, G. (2019). Self-concept research with school-aged youth with intellectual disabilities: A systematic review. *Journal of Applied Research in Intellectual Disabilities, 32*(2), 238-255. https://doi.org/10.1111/jar.12543

Matthijssen, M. (2018). Gespreksleider focusgroepen [Moderating focus groups]. Rotterdam: author.

Microsoft (2018). Welcome to Microsoft Teams. Available at: https://docs.microsoft.com/en-us/microsoftteams/teams-overview

Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (Eds.). (2014). *Cognitive interviewing methodology*. John Wiley & Sons.

Moonen, X. (2019). Participatief onderzoek met mensen met verstandelijke beperkingen in de weg staan: de rol van poortwachters en onderzoekers [Participative research with people with intellectual disabilities: the role of gatekeepers and researchers]. *Kwalon, 24*(2), 9-12.

Moonen, X. (2021). Easy language in the Netherlands. In: Camilla Lindholm and Ulla Vanhatalo (eds.) *Handbook of Easy Languages in Europe* (pp. 345-370). Berlin: Frank & Timme.

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. https://doi.org/10.1089/aut.2019.0074

Orsi, M. M., Lafortune, D., & Brochu, S. (2010). Care and control: Working alliance among adolescents in authoritarian settings. *Residential Treatment for Children and Youth, 27*(4), 277–303. https://doi.org/10.1080/0886571X.2010.520637

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 and Psychotherapy, 11*(4), 265–274. https://doi.org/10.1002/cpp.412

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. https://doi.org/10.1007/s10882-005-6617-4

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. https://doi.org/10.1002/(SICI)1099-1298(199608)6:3,207::AID-CASP370.3.0.CO;2-T

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. Journal of Applied Psychology, 84(5), 754. https://doi.org/10.1037/0021-9010.84.5.754

Strijbosch, E. L. L., Wissink, I. B., van der Helm, G. H. P., & Stams, G. J. J. M. (2019). Building a positive group climate together: How monitoring instruments are part of an improvement process in residential care for children. *Children and Youth Services Review, 96*, 266–277. https://doi.org/10.1016/j.childyouth.2018.11.051

Taal voor Allemaal (n.d.). De Taal voor Allemaal methode [The Language for All method]. Retrieved January 10, 2022, from https://www.taalvoorallemaal.com.

Ten Brummelaar, M. D. C., Harder, A. T., Kalverboer, M. E., Post, W. J., & Knorth, E. J. (2018). Participation of youth in decision-making procedures during residential care: A narrative review. *Child and Family Social Work, 23*(1), 33–44). https://doi.org/10.1111/cfs.12381

Van de Mortel, T. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40.

Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Duschinsky, R., Fox, N. A., Goldman, P. S., Gunnar, M. R., Johnson, D. E., Nelson, C. A., Reijman, S., Skinner, G. C. M., Zeanah,

C. H., & Sonuga-Barke, E. J. S. (2020). Institutionalisation and deinstitutionalisation of children 1: a systematic and integrative review of evidence regarding effects on development. *The Lancet Psychiatry, 7*(8), 703–720). https://doi.org/10.1016/S2215-0366(19)30399-2

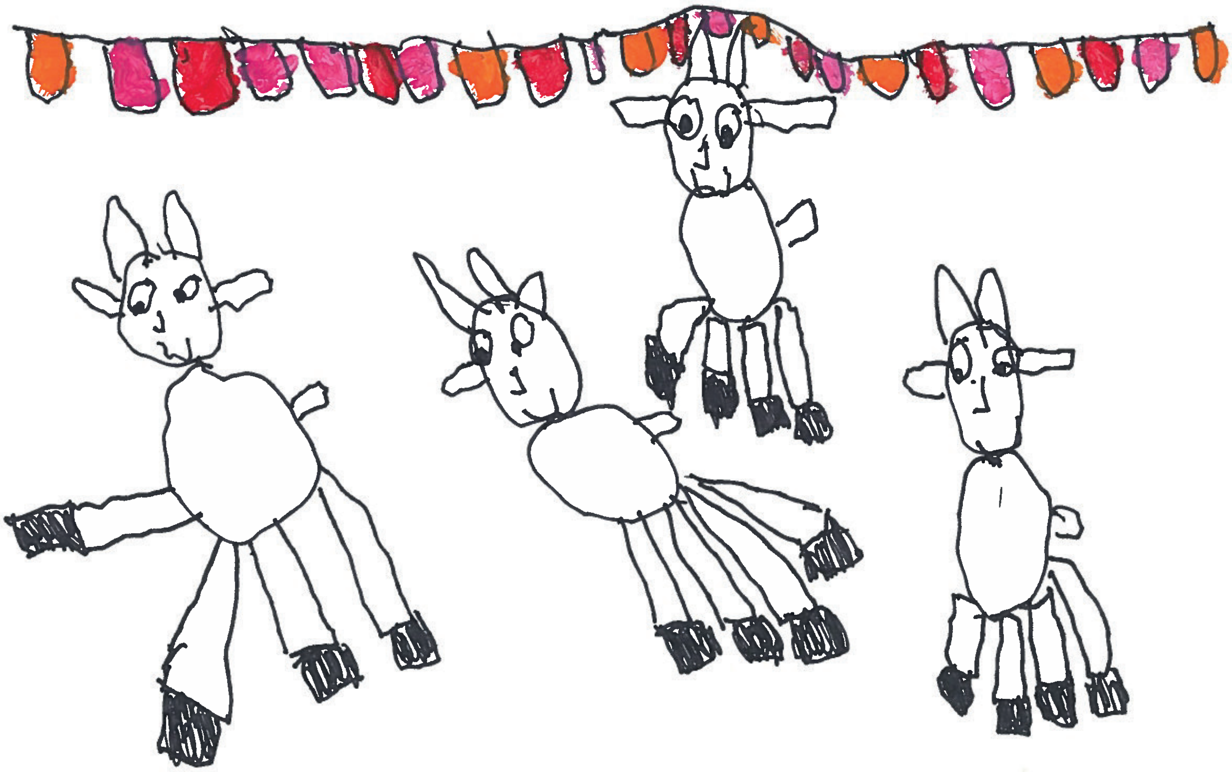
Walker, C., & Baxter, J. (2019). Method Sequence and Dominance in Mixed Methods Research: A Case Study of the Social Acceptance of Wind Energy Literature. *International Journal of Qualitative Methods, 18*. https://doi.org/10.1177/1609406919834379

Wissink, I., & Kooijmans, R. (2020). *Psychometrische analyses cliëntervaringsinstrument Mijn Mening* [psychometric analyses client satisfaction survey instrument My Opinion]. Amsterdam: University of Amsterdam.

Woodyatt, C. R., Finneran, C. A., & Stephenson, R. (2016). In-Person Versus Online Focus Group Discussions: A Comparative Analysis of Data Quality. *Qualitative Health Research, 26*(6), 741–749. https://doi.org/10.1177/1049732316631510

Chapter 4

Does adding pictures to easy-to-read texts benefit comprehension for people with reading difficulties? A meta-analytic review.



Accepted for publication:

Kooijmans, R., Dalemans, R., Van der Spek, J., Langdon, P. E., & Moonen, X. (in press). Does adding pictures to easy-to-read texts benefit comprehension for people with reading difficulties? A meta-analytic review. *Evidence-Based Communication Assessment and Intervention*. https://doi.org/ 10.1080/17489539.2025.2551910

Abstract

Individuals with reading problems may experience communication problems in everyday life. Creating accessible information for people with reading difficulties is imperative to facilitate inclusion and participation in society. Guidelines for creating accessible information generally include the use of visualizations. However, a synthesis of the empirical evidence on the benefits of adding visualization to text on comprehension was lacking.

We aimed to systematically review the research literature on this topic, with the intention to provide a quantitative (meta-)analysis of quantitative experimental results of included studies. Eight studies met our eligibility criteria and 13 effect sizes were extracted and analyzed in a 3-level meta-analysis, following PRISMA guidelines. Quality of included studies was assessed by using the RoB-2 risk of bias assessment with added considerations regarding the quality of the visualizations used in the studies. The methodological quality of most studies was questionable. Further, visualization used in the included studies was diverse, generally of questionable quality and justification for the choice of visualization was unclear.

The findings from the meta-analysis did not support the assumption that adding visualizations to easy-to-read text improves understanding for people with reading difficulties. These findings are surprising given the fact that adding visualizations to texts for persons with reading difficulties is recommended in most guidelines on augmentative communication.

Introduction

Many individuals experience difficulties in day-to-day life because of reading problems as a result of an intellectual disability, aphasia, dyslexia, a visual disability or other difficulties (Conners, 2003; Van Ewijk et al., 2017; Sorber, 2021). Other potential readers struggle to understand and use written language due to poor education or because they have acquired it as a second language (Verhoeven et al., 2019). Reading skills are needed to participate in many everyday activities, such as navigating public transport, communicating through social media, shopping in the supermarket, reading medicine labels, or undertaking tasks within the workplace. Being able to read gives individuals the opportunity to connect to the social world, learn new skills, and increase job opportunities. For individuals with reading difficulties, the ability to connect with the outside world is limited.

The right to have access to clear information is one of the core elements of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD; united Nations, 2006) and is included in legislative documents worldwide. Providing information in a more accessible format minimizes (health) inequalities for people with reading difficulties, and promotes inclusion and self-determination (Chinn & Homeyard, 2017). Creating accessible information for people with reading difficulties has consequently become a priority for governments, researchers, advocacy groups, and health authorities (Yaneva et al., 2015; Scheffers et al., 2021).

Plain Language and Easy-Read Text

To improve the accessibility of written information, several strategies and interventions have been proposed that include reducing grammatical and semantic complexity, use of visualization, and incorporating design features such as accessible typology and spacing (Foundation for People with Learning Disabilities, 2023; Nomura, Nielsen, & Tronbacke, 2010).

In 2023, the International Organization for Standardization (ISO) published the ISO 24495-1:2023 standard to provide a universal standard for creating ‘plain language’ (ISO, 2023). Plain language is defined as ‘communication in which wording, structure, and design are so clear that intended readers can easily find what they need, understand what they find, and use that information’ (International Plain Language Federation, 2023). Plain language is written at CEFR level B1 and is intended to cater for people with intermediate levels of reading fluency and understanding (Council of Europe, 2023).

People with lower levels of literacy require further support to be able to understand written texts. For the purpose of writing texts for persons with (very) low literacy skills, so called ‘Easy-Read’ guidelines were developed. These are not universal and differences exist between countries, cultural contexts and the organizations that produce them. Examples of such guidelines include the British ‘Easy-read guidelines’ (UK Department of Health, 2010), Dutch *Taal voor allemaal* guidelines (Koraal, 2020), Finnish *Selkokieli* guidelines (Tuohimetsä, 2020), and the German *Leichte Sprache* guidelines (Leichte Sprache, 2021). These guidelines usually advocate writing at CEFR levels A2 or even A1. The creation of easy-read texts has increased the availability of accessible information significantly in recent years, especially in the field of health care (Chinn & Homeyard, 2017).

The Use of Visualizations in Relation to Understanding

To guide the augmentation of written communication beyond simplification at the semantical and grammatical level, guidelines for creating plain language and easy-read texts usually include the use of visualizations to support meaning. In the context of creating easy-to-understand texts, visualization refers to adding images that represent information in the text to help people with reading difficulties understand what is written (Inclusion Europe, 2023).

Many different complementary skills are required to understand written text, ranging from technical reading skills such as decoding and breaking words down into syllables and phonemes, to more cognitive processes and connecting what is read to background knowledge. For diverse reasons, any of these abilities involved in written language comprehension may be impaired. These include neurodegenerative diseases such as aphasia, congenital or acquired cognitive impairments or simply because the person has never had the opportunity to learn to read. The processing of visual information takes place in regions of the brain that are connected but separately executed from those processing verbal information (Hibbing & Rankin-Erickson, 2003). People respond to and process visual data better than any other type of data. The human brain processes images 60,000 times faster than text and 90 percent of information transmitted to the brain is visual. Because we are visual by nature, we can use this skill to enhance data processing (Eisenberg, 2018). The use of visualization as a broad recommendation is therefore universally recommended in accessible communication guidelines, regardless of the nature or etiology of the reading difficulty.

Different theories have been developed regarding the use of visualizations in understanding written information. Visualizations have a crucial role in expanding the cognitive system when trying to understand written language (Ware, 2004). The combination of visualizations and texts can reduce the cognitive load to understand the material (Mayer, 2009). This enhances understanding, facilitates remembering, and promotes learning (Hibbing & Rankin-Erickson, 2003; Jee & Li, 2014; Meppelink, 2015) and at the same time helps people with reading difficulties to engage with the text (Doak et al., 1996). Critics of this theory oppose that presenting photographs and text in conjunction can create cognitive overload. This renders the working memory unable to process either modality efficiently (Hurtado et al., 2014). This may be particularly detrimental for people with working memory deficits, such as some people with an intellectual disability or neurodegenerative conditions.

One of the influential rationales behind using visualizations and text together is Paivio’s (1991; 2013) dual-coding theory, which states that mental representations consist of two distinct knowledge systems: (a) nonverbal visual systems which are related to holistic processing of data, and (b) verbal systems which are related to abstract and sequential processing of data. According to Paivio (2013), the non-verbal visual subsystem is able to construct representations of knowledge associated with the verbal subsystem when visualizations are added to written texts. In other words, visualizations combined with verbal information can reinforce verbal comprehension. When reading text and related visualizations are presented at the same time, verbal and nonverbal data are processed simultaneously in different cognitive systems.

Different Types of Visualization to Support Texts

Guidelines for easy writing generally advocate the use of visualization, but very rarely offer guidance beyond the recommendation that “Images should be selected to represent each section of text where possible” (Foundation for Learning Disabilities, 2023). Two notable exceptions are the German *Leichte Sprache* (2021) guidelines, which are accompanied by their own image library consisting of purpose-made line drawings, and the Dutch *Taal voor Allemaal* guidelines (Koraal, 2020) which have a separate section on how to use visualizations and what types of images can be used.

Visualizations are intended to symbolize persons, objects, feelings, activities, and situations. According to DeLoache (1998) iconicity generally facilitates symbol use. Iconicity is defined as the perceptual resemblance between a symbol and its referent, with symbols that are highly iconic (e.g. colour photographs) labeled as transparent, moderately iconic symbols (e.g. black-and white line drawings) labeled as translucent and symbols with little or no resemblance to their referent (e.g. icon, Makaton symbols, written words) as opaque (Fuller 1997; de Rijdt 2013).

All levels of visualizations can be used to support text. Photographs can be helpful to better understand text if they are a familiar and recognizable representation of reality (Oskam & Scheres, 2016). Colored photos with a high contrast that are strongly related to what is to be depicted are highly valued by people with an acquired brain injury (Dalemans et al., 2021).

In some instances, drawings are considered to be clearer visualizations for people with reading difficulties than photographs because drawings carry less possibly confounding information and may cause less sensory overload (Rijdt, 2013). Drawings display the necessary features of the concept or situation and are more symbolic than photographs (Sutherland & Isherwood, 2016). Pictograms are simple drawings that represent only the defining features of the concept or situation (De Rijdt, 2013).

When pictures are used to accompany written information, there are three levels of support intensity (Poncelas, 2007). The most basic is where one picture represents the topic or essence of a whole sentence or even a paragraph. The next level is to use pictures and key symbols to convey all the meaning of key concepts (nouns, verbs, emotions), but not the grammatical items. Finally, the most sophisticated level is where a symbol is used for every word and linguistic element in a sentence. Makaton Symbol and Widgit Rebus Symbol systems are examples of such 'symbol reading' operationalizations.

Current Evidence for the Beneficial Effects of Visualization on Comprehension

In research with people without reading difficulties, adding visualization to text has been shown to facilitate understanding. Especially in the field of health literacy, the beneficial effects of pictorial information are well-documented, for instance to promote medicine intake (Katz et al., 2006).

Although adding visual support to text is generally advocated to improve understanding for readers with lower literacy levels, empirical evidence on the beneficial effects of visualizations is scant. Much of the supporting evidence appears to come from the experience of experts and service-users (Sutherland & Isherwood, 2016). Few studies investigate this topic with experimental designs that include service users directly. Schubbe et al. (2020) assessed the effect of pictorial health information on patients' and consumers' health behaviors by means of a systematic review and meta-analysis. In a subgroup analysis they investigated if the uptake of health-related information improved for people with 'lower health literacy' when written information was supported by pictures. Both understanding and recall were found to significantly improve in this subgroup. The definition and operationalization of 'lower health literacy' was very diverse between studies and the included study samples did not appear to include clinically impaired readers.

In their narrative review on the comprehensibility of easy-read texts for people with intellectual disabilities, Sutherland and Isherwood (2016) included a review of the effect of adding symbols or photographs. Mixed results were found regarding the benefit of adding pictorial representations to text. Photographs appeared to be more effective than line drawings or symbols because photographs require less symbolic processing than symbols or line drawings.

The Present Study

Existing reviews on the effect of visualization on understanding of written texts for people with reading difficulties were of a narrative nature and produced inconclusive results (Sutherland and Isherwood, 2016) or did not have a clear definition of the level of reading impairment for the included study samples (Schubbe et al., 2020). Although adding visualization to texts intended to be read by people with reading difficulties is common practice, the supposed beneficial effect on reading comprehension has not been reviewed empirically. We therefore set out to systematically review the research literature on this topic, with the intention to provide a quantitative (meta)analysis of all quantitative experimental results of included studies.

The main research question for our review was ‘Do individuals with reading difficulties better comprehend plain language or easy-read texts when visualization is used, compared to a text without visualization?’

We followed up on the main question by performing subgroup analyses to explore whether the nature or origin of the reading impairment moderated the effect of the use of visualization and performed moderator analyses with the type of visualization used in the study and the difficulty of the texts used as potential moderators.

Methods

This systematic meta-analytic review has been conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). 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 CRD42021289225).

Eligibility criteria

Studies were included if they (1) examined the effect of adding visualizations to an easy-to-read text (see criteria below), (2) focused on reading comprehension as an outcome measure, (3) included adults (18 years or older) with reading difficulties, and (4) used a quantitative or mixed-methods experimental study design.

The means of presentation of experimental stimuli could be either paper or digital. The type of visualizations used had to be described, in order to determine which type of visualization was potentially effective. There were no restrictions on the type of visualization used.

The readability of the text had to be formally assessed with the use of readability tools like the Flesch-Kincaid Readability tool (Flesch, 2007; Kelly, 2017), the Gunning Fox index (Kelly, 2017), the Fry Readability Graph (Kelly, 2019) or the Scientific Research Associates (SRA) reading tools. These tools are used to calculate the readability of a text by quantifying structural elements of the text such as the average number of words per sentence and the average number of syllables per word. The (US reading proficiency) grade level of the text had to be 6 or lower, as a Flesch-Kincaid Readability grade 6 or lower is considered to be 'easy-to-read' (Flesch, 2007; Kelly, 2017; Witwer, 2022). Only studies were included that used a paper or digital same-level text without visualizations as a control condition.

Regarding the outcome eligibility criteria, the level of reading comprehension had to be assessed quantitatively. Studies that only focused on topics like acceptability of visualizations or emotional responses to visualizations were excluded.

Studies were included if the study sample consisted of adults with reading difficulties, with no restrictions regarding the origin of the reading difficulties. Preferably, reading skills and/or verbal understanding were assessed using validated instruments.

For Study type, only experimental studies, including those with mixed methods, were included.

Search Strategy

A systematic literature search was conducted using PsycINFO (Ovid), Medline (Ovid), ERIC (Ovid), and the Web of Science Core Collection. The final search was completed on November 20, 2023. The search yielded 2,803 results. Search strings were created by combining search terms for (1) easy to read texts, (2) visualizations, (3) study type, and (4) adults. No limit was set on publication year, publication type, or language during this phase. All systematic search activities were guided and co-performed by an information specialist from the University of Amsterdam.

The parameters for the search strategy, including search terms, and strings for all databases are provided in Appendix A.

Study Selection

The protocol for screening and selection of included studies can be accessed at https://www.crd.york.ac.uk/PROSPEROFILES/289225_PROTOCOL_20211103.pdf

Duplicates were removed using Zotero (Ahmed & Dhubaib, 2011) and Rayyan (Ouzzani et al., 2016). Two researchers (RK and JvdS) independently screened the titles and the abstracts of all studies (double-blind in Rayyan) using the protocol for screening and selection. Discrepancies between researchers were discussed until consensus was reached. According to the Landis and Koch (1977) guidelines interrater agreement was observed to be 'moderate', $k = 0.56$, $p < .001$. After discussing discrepancies, the interrater agreement was 100%. After initial screening, the remaining articles were screened full-text by two researchers (RK and JvdS) independently. Interrater agreement for this stage was perfect.

Data Extraction

Relevant data from the included studies were extracted using a data charting form. Author(s) and publication year, participant characteristics (N, type of disability or cause of reading difficulties, reading ability, age), study type, difficulty of the text, type of visualization, dependent outcome variables and main conclusions were charted (see Table 1).

Quality Assessment

Formal risk of bias assessment was performed to assess the quality of the included studies. The included studies were assessed independently by two researchers (RK and JvdS), using the Revised Cochrane Risk of Bias tool for randomized trials (RoB-2) with additional considerations for crossover trials (Higgins et al., 2020). Initial interrater agreement for the risk of bias assessment before discussion was 'fair' (Landis & Koch, 1977), $k = 0.37$, $p = .141$. Because of the less-than-perfect initial agreement, the authors decided to reevaluate the RoB-2 guidelines jointly, using non-included publications as training material. The risk of bias assessment for the included articles was then repeated by both authors independently. This led to near-perfect agreement, $k = 0.81$, $p = .001$. Remaining discrepancies were resolved through discussion.

In the RoB-2 tool, five domains are assessed: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcome, and (5) bias in selection of the reported results. The risk of bias for each domain can be rated as low risk, some concerns, or high risk, based on the aggregate score across several questions per domain. In none of the included studies the study design permitted participants to deviate from the intended intervention. Therefore, the domain 2 criteria were deemed not relevant and all criteria were scored 'not applicable'. An overall Risk of Bias score is determined by combining the domain scores.

A sixth domain was added by the authors to assess the quality of the operationalization of the visualization used in the study. It was reasoned that choosing a certain type of visualization would influence its potential impact. More specifically, well-chosen visual supports that align with best-practices for a certain population have the best chance of producing favorable results. A pilot form was drafted by the authors and reviewed by two independent experts on the use of visualization in communication. For the final version, three questions were formulated to determine if there were concerns regarding the choice of visualization in the study:

1. Did the researchers justify their choice for the means of visualization (based on best practices or prior research)?
2. Was the suitability and appropriateness pilot tested with participants that are representative of the target population?
3. If yes, are the results of the pilot test used to optimize the final intervention?

As with the other RoB-2 domains, the scores across questions were aggregated to obtain an overall rating of no concerns, some concerns, or many concerns.

Data analysis

For each study, reading comprehension scores were extracted for each of the two conditions: simple text with visualizations (intervention condition) and simple text without visualizations (control condition). If a study compared different types of visualization (for example photos and drawings) with the text-only control condition, both effects were analyzed. All analyses were performed using R Statistics (R Core Team, 2022) and SPSS (version 29).

For all studies that compared mean scores between conditions, the standardized mean difference (SMD) effect sizes and the 95% confidence intervals (CIs) were calculated for each reading comprehension comparison. As an alternative to Cohen's d for effect size, Hedge's g was reported, as this is a more accurate effect size for comparisons with smaller numbers of participants (fewer than 20) or when sample sizes are unequal (Ellis, 2010). For the Mansoor (2003) and Rose et al. (2011) studies, which reported proportions as the main outcome, Cohen's h and the 95% confidence interval were calculated. The numerical values of Hedge's g and Cohen's h and their respective confidence intervals have identical interpretive definitions.

A three-level meta-analytic model was applied, with effect sizes nested within studies at the third level to account for dependency between effect sizes from the same study. A mixed-effects model was used to determine possible subgroup and moderating effects. Subgroups were defined based on the cause of reading difficulties (intellectual and developmental disability, aphasia, or English as a second language). Type of visualization was also considered as a moderating variable. Visualizations were categorized based on the level of iconicity (DeLoach et al., 1998). Because most drawing-type visualizations could not clearly be classified as either 'translucent' (medium iconic) or 'opaque' (low iconic), these were collapsed into a 'drawing' category and compared with 'transparent' or highly iconic photographs. Grade level of the text (grade 4 or lower versus grade 5 or higher) was also included as a possible moderator.

The Wibbelink and Assink (2016) guidelines for performing a three-level-meta-analysis were used to calculate the overall effect size, confidence intervals and the prediction interval for the overall effect size and to perform subgroup and moderator analyses.

In case the overall and subgroup effect sizes were non-significant, an equivalence test following the TOST procedure (Lakens, 2017; Schuirman, 1987) was performed using the TOSTER package in R (Lakens et al., 2018). The minimum effect that was deemed worthwhile to consider, referred to as the Smallest Effect Size of Interest (SESOI), was set at 0.50, representing a moderate effect size (Cohen, 1988).

Following the Hunter and Schmidt (2004) guidelines for small-scale meta-analyses, the decision to perform subgroup and moderator analyses was based on the proportion of explained variance at the first level (sampling variance) of the 3-level meta-analysis. They state that it can be meaningful to perform subgroup and moderator analyses if less than 75% of the total variance can be attributed to sampling variance. An estimate of the sampling variance was made by using the formula of Cheung (2014), allowing us to calculate the proportion of explained variance at the first (sampling variance), second (within-study), and third (between-study) level of the meta-analysis. I2 is then determined as the sum of the within- and between-study variance. The forest plot was inspected to look for additional indications of heterogeneity and to identify outlying effect sizes.

Including studies with high risk of bias domains in meta-analyses can lead to invalid conclusions (Marušić et al, 2020). To explore the influence of methodological quality on meta-analysis results, sensitivity analyses were conducted by repeating the primary meta-analysis (a) without the studies that had an overall high risk of bias rating and (b) without the studies for which there were serious concerns regarding the choice of visualization.

Results

Study Selection

In the final search 2,803 studies were found. After the removal of the duplicates, 2,008 studies remained, which were then screened for eligibility. This resulted in 13 studies for full-text assessment. After full-text assessment eight studies remained for inclusion in the meta-analysis. The PRISMA Flowchart in Figure 1 illustrates the search and selection process. The eight studies included in the final selection are indicated with an asterisk in the list of references. The eight included studies and their characteristics are presented in table 1.

Figure 1.

PRISMA 2020 Flow Chart

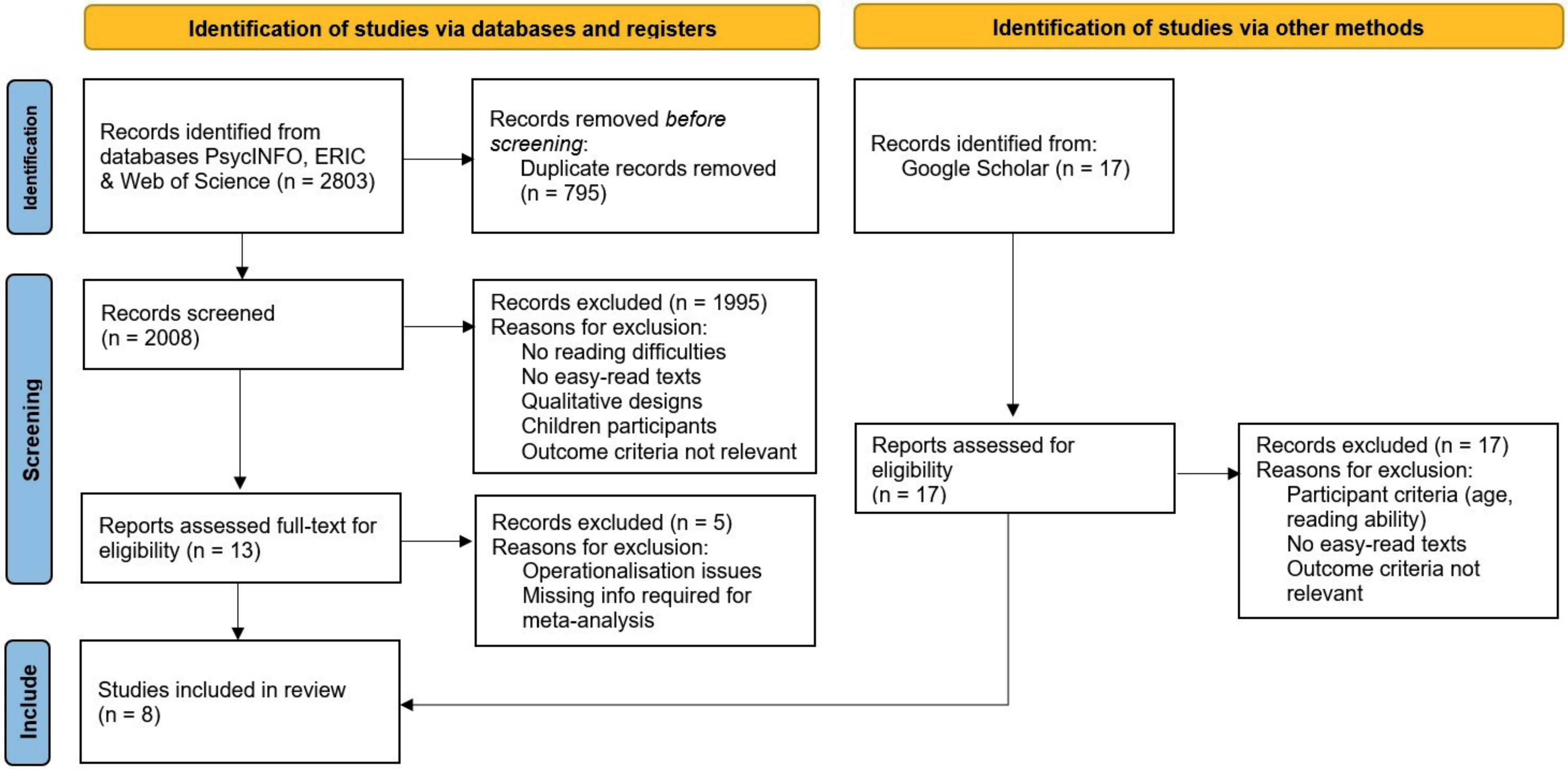


Table 1:

Characteristics of Included Studies and Participants

Study	*Research question(s) relevant to review topic*	*Participants*	*Reading ability*	*Study design*	*Text difficulty*	*Type of visualizations*	*Dependent variable(s)**	*Outcome(s) relevant to review question*
Brennan et al. (2005)	Do people with aphasia comprehend aphasia-friendly paragraphs better than control paragraphs?	Adults between 34 and 70 years old diagnosed with Aphasia (*N*=9)	WAB: 63.1-90 (Aphasia Severity Mild to Moderate)	Cross-over study (within-subjects)	SRA Grade Level 5 and 6	Simple color drawings (Google Images)	**Reading comprehension**	Adding visualization to the paragraphs did not significantly improve comprehension.
Dietz et al. (2009)	Does adding high-context or low-context photographs to easy-read text affect reading comprehension?	Adults with chronic aphasia (*N*=7)	WAB: 31-72; RCBA: 45-84	Cross-over study (within-subjects)	Flesch Kincaid Grade Level 2.2 - 2.5	High-context and low-context photographs	(a) **reading comprehension**, (b) response time, (c) perceptions of image helpfulness.	Both high- and low-context photographs significantly improved participants' comprehension of the paragraphs.
Jones et al. (2007)	Do symbols improve the reading comprehension of	Adults with mild or borderline	WORD: age equivalent <72-114	Cross-over study	Neale Analysis of Reading	Widget Rebus symbols (line drawings)	**Reading comprehension**	Better comprehension scores when symbols were added. Persons

Study	*Research question(s) relevant to review topic*	*Participants*	*Reading ability*	*Study design*	*Text difficulty*	*Type of visualizations*	*Dependent variable(s)**	*Outcome(s) relevant to review question*
	adults with intellectual disabilities?	intellectual disability (*N*=19)	months; TROG: median score 7/20 blocks	(within-subjects)	Ability revised level 1 and level 2 (Grade Level 1 and 2)			with lower reading proficiency benefited more from visualization. Persons who were familiar with the pictograms used benefited more.
Mansoor & Dowse (2003)	Do pictograms improve understanding of easy-text patient information leaflets and medicine labels for low-literate people?	Low-literate patients with English as a second language (*N*=60)	Not formally assessed. Participants 'with low literacy and poor reading skills'	Randomized Control Trial (between-subjects)	Fry's Readability Test Grade 4	Pictograms designed for the study	(a) **Reading comprehension**; (b) Patient acceptability	The pictograms had a positive effect on comprehension of the patient materials.
Poncelas & Murphy (2007).	Do added symbols help people with intellectual disabilities to better understand a political manifesto written in easy-read English?	Adults with intellectual disabilities (*N*=34)	BPVS: around 8 yrs age equivalent; BAS: 59-80	Randomized Control Trial (between-subjects)	Flesch Reading Ease Scores 68-97, Flesch-Kincaid Grade Levels 5-7	Makaton and CHANGE Picture Bank symbols (line drawings)	(a) **Reading Comprehension**; (b) Retention	Participants did not show better understanding of the text when symbols were added.

Study	*Research question(s) relevant to review topic*	*Participants*	*Reading ability*	*Study design*	*Text difficulty*	*Type of visualizations*	*Dependent variable(s)**	*Outcome(s) relevant to review question*
Rose et al. (2011)	Do black and white line drawings or color photographs improve reading comprehension for people with aphasia?	People with aphasia (*N*=22)	WAB: 34-93; RCBA: 58-98	Cross-over study (within-subjects)	Flesch-Kincaid Grade Level 6.26-6.30	Color photographs and colored line drawings (Google Images)	(a) **Reading Comprehension**; (b) Reading times; (c) Participant preference	Adding black and white line drawings or color photographs did not significantly enhance reading comprehension compared to the no visualization condition.
Saletta et al. (2019)	Are there differences in comprehension between passages with colored photographs or black-and-white line drawings compared to text with abstract images?	Young adults (18-25) with intellectual disabilities (*N*=45)	Woodcock-Johnson Passage Comprehension: <40-117	Randomized Control Trial (between-subjects)	Grade 3 materials (“Easy Curriculum-Based Measurements”; University of Oregon, 2018)	Color photographs, black and white line drawings or non-figurative illustrations	**Reading comprehension**	Neither supportive visualisation condition produces better comprehension scores than the control condition

* Outcome variables in **bold** are included in the analyses

Visualization Used in the Studies

There was a wide variety of (largely unsubstantiated) choices regarding the ‘iconicity’ of images, placement of images, and support intensity (how many elements of the text are accompanied by images). A summary of the most important characteristics and examples are provided in Table 2.

Participants

In four out of eight included studies, participants had a confirmed diagnosis of aphasia. Three studies included people with intellectual disabilities. In one study, the sample consisted of non-native English speakers with low literacy. For six studies, the reading ability of participants was formally assessed using a validated tool. The Western Aphasia Battery and Reading Comprehension Battery for Aphasia were used in three of the aphasia studies. In the fourth aphasia study (Wilson & Read, 2016) the reading ability of participants was not explicitly determined. In the studies with people with intellectual disabilities as participants, reading ability was assessed with a variety of instruments; the Reading Comprehension sub-test of the Wechsler Objective Reading Dimensions (WORD) test and the Test for the Reception of Grammar (TROG) test in the Jones et al. (2007) study, the British Picture Vocabulary Scale and British Ability Scales word reading test in the Poncelas & Murphy (2007) study, and the Woodcock-Johnson Passage Comprehension test in the Saletta et al. (2019) study. Outcomes of reading ability assessments are included in Table 1.

Quality Assessment

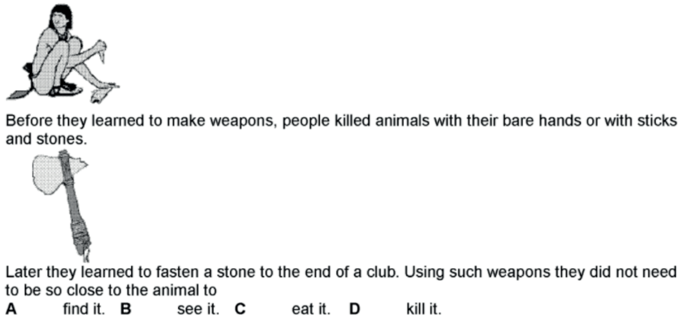
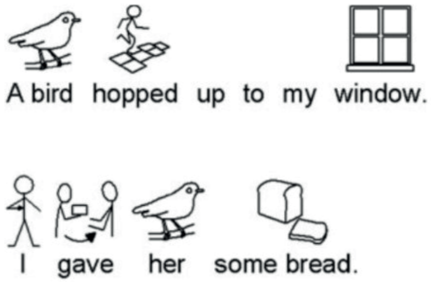
Methodological Quality

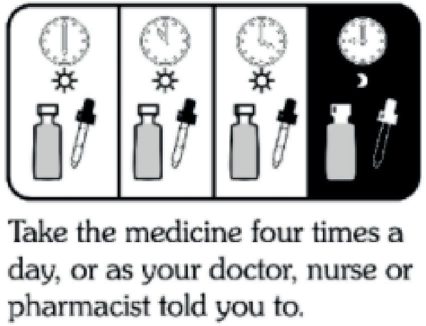
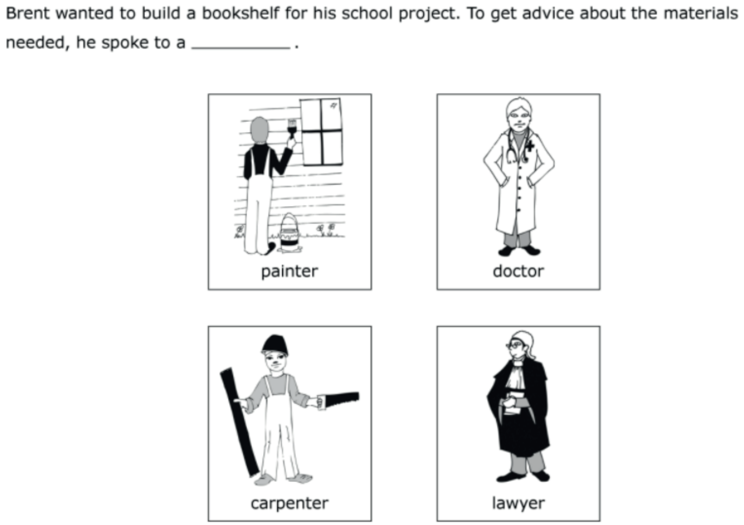
The methodological quality of the seven included studies was assessed with the RoB 2-tool (Higgins, Li, & Sterne, 2020). A summary of the quality assessment for each study with the most significant findings can be found in Appendix B.

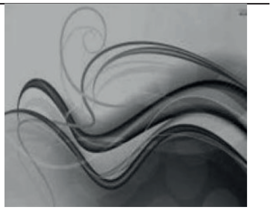
Few problems were encountered for Domain 1 (randomization process) and Domain 2 (deviation from intended interventions). Criteria for Domain 2 were deemed not applicable for any of the included studies, mainly because the ‘interventions’ (reading text with and without visualization) were so non-invasive that no effect of assignment to intervention was to be anticipated. For all studies and participants, all interventions were delivered as intended. Further, in five of seven studies, a cross-over design was used in which participants received both ‘interventions’.

Table 2

Visualization Characteristics per Study

Study	Visualization type	What was visualized?	Placement	Example (if provided in the study publication)
Brennan et al. (2005)	Simple color drawings (Google Images)	Key verbs or nouns or relation between key concepts of a sentence. One per sentence. Multiple-choice answers not visualized.	One picture above each sentence.	 Before they learned to make weapons, people killed animals with their bare hands or with sticks and stones. Later they learned to fasten a stone to the end of a club. Using such weapons they did not need to be so close to the animal to **A** find it. **B** see it. **C** eat it. **D** kill it.
Dietz et al. (2009)	High-context and low-context photographs	Two photographs per paragraph (either low- or high-context, depending on condition). Multiple-choice answers not visualized.	No information	No example provided
Jones et al. (2007)	Widget Rebus symbols (line drawings)	Key verbs or (pro)nouns or relation between key concepts of a sentence. Multiple symbols per sentence. Oral comprehension questions (not visualized).	Each symbol positioned above the word(s) that it accompanies.	 A bird hopped up to my window. I gave her some bread.

Study	Visualization type	What was visualized?	Placement	Example (if provided in the study publication)
Mansoor & Dowse. (2003)	Pictograms designed for the study	One picture represents an action or sequence of actions. Oral comprehension questions (not visualized).	Picture positioned above the corresponding passage.	 Take the medicine four times a day, or as your doctor, nurse or pharmacist told you to.
Poncelas & Murphy (2007)	Makaton and CHANGE Picture Bank symbols (line drawings)	Key verbs or (pro)nouns or relation between key concepts of a sentence. Multiple symbols per sentence. Oral comprehension questions (not visualized).	Each symbol positioned above the word(s) that it accompanies.	
Rose et al. (2011)	Line drawings designed for the study and (stock) color photographs (Google Images)	Target sentences were not illustrated. Multiple-choice response options were illustrated with either line drawings or color photographs (depending on condition).	Picture positioned above each response option.	 Brent wanted to build a bookshelf for his school project. To get advice about the materials needed, he spoke to a ________ .

Study	Visualization type	What was visualized?	Placement	Example (if provided in the study publication)
Saletta et al. (2019)	Photo by research team, line drawings derived form photo, control (Google Images)	Three pictures accompanying each text passage.	Directly above the text passages.	 b/w photo  line drawing  control
Wilson & Read (2016)	Color photographs and colored line drawings (Google Images)	Picture of noun representing a key concept, one per paragraph. Pictures were color photographs or line drawings, depending on condition. Multiple-choice answers not visualized.	No information	No example provided

For Domain 3 (missing outcome data), two studies (Brennan et al., 2005; Rose et al., 2011) reported that more severely impaired participants were removed from the procedure or parts of the analysis because they were deemed unable to provide meaningful answers.

Most concerns were observed with regard to Domains 4 (measurement of the outcome) and 5 (selection of the reported result). For Domain 4, three studies received a ‘high risk’ rating. In the Dietz et al. (2009) study the design was flawed, leading to the introduction of possibly confounding variables influencing the measurement of comprehension. Specifically, participants were required to store the information read in memory before answering the comprehension questions, thereby introducing retention issues as a potential confounder. In the Mansoor and Dowse (2003) study, many comprehension questions referred to information in text that was not supported with visuals in either of the two conditions, so no significant difference can be expected for these questions. Furthermore, there were no clearly described criteria for the comprehension test. In the Rose et al. (2011) study, only the answers to questions about text passages were visualized, not the body of text itself, leading to a question about how participants are supposed to answer correctly if they don’t understand the text. In the Jones et al. (2007) and Poncelas and Murphy (2007) studies, not enough information was given to properly assess the outcome measurement procedure, leading to a ‘some concerns’ rating.

Several, but arguably less severe, concerns were raised for the Domain 5 criteria. Several studies received lower ratings because the analysis plan was not clearly described. For some studies not all measured outcomes seemed to be included in the reported results, possibly pointing to publication bias (underreporting of non-significant results).

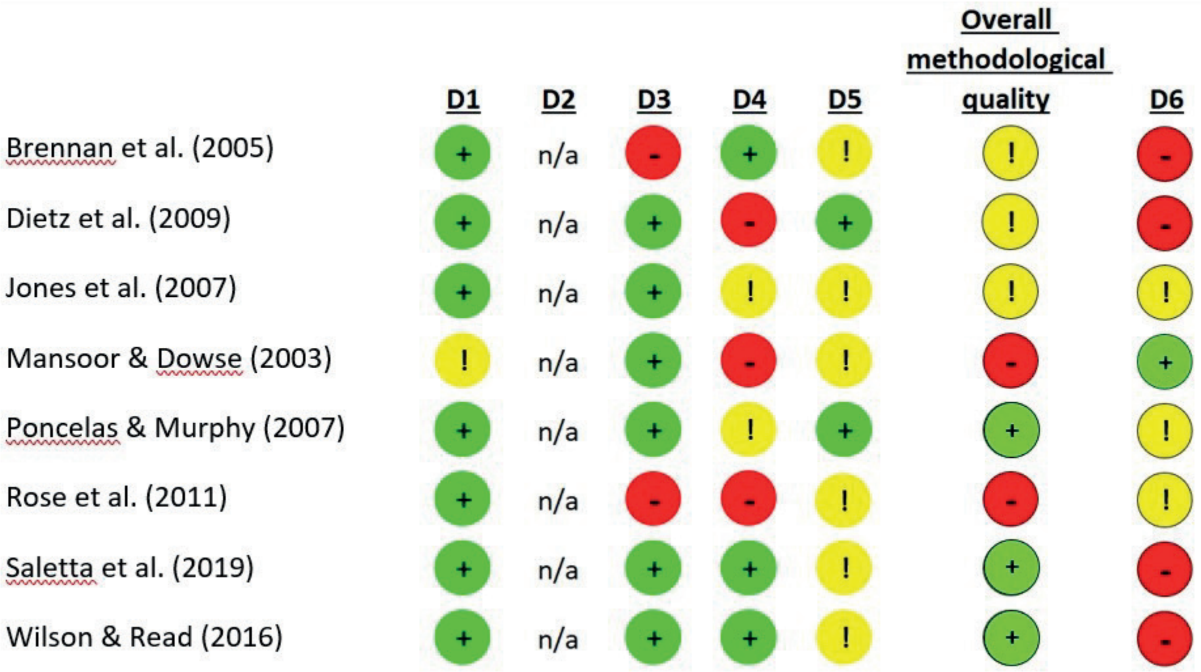
Operationalization of Visualizations in Individual Studies

For only one study (Mansoor and Dowse, 2003) the quality of the visualization used to accompany text was perceived to be of high quality. For other studies, there was no clear rationale given for the choice of visualization. The most frequently encountered problem was that researchers tended to select the pictures they deemed appropriate themselves or with the help from ‘experts’, without consulting members of the target population. Instead of using validated visual support systems some researchers reverted to self-searched google images or clipart pictures (e.g. Wilson & Read, 2016). To what extent the visual supports used were appropriate and clear for the intended use was tested with potential participants only in the Mansoor and Dowse (2003) study. In the Dietz et al. (2009) and Rose et al. (2011) studies, the appropriateness was tested with non-impaired individuals, the results of which clearly do not automatically transfer to people with reading impairments. In some studies, the authors referenced ‘best practice’ standards or existing guidelines to justify their choices but the scientific validation of these standards, and hence their practical merits, remain unclear. In several studies, justification for the means of visualization entailed that previous studies using the same type of visualization were referenced, but most of these studies had shown contrasting results. A breakdown of the sixth domain assessment can be found in Appendix C.

A traffic light plot of the risk-of-bias assessment for all domains, including the added sixth domain is presented visually in figure 2, using the ‘RoBVis’ web application (McGuinness & Higgins, 2020).

Figure 2

Study Quality and Operationalization Quality of Visualization



	D1	D2	D3	D4	D5	Overall methodological quality	D6
Brennan et al. (2005)	+	n/a	-	+	!	!	-
Dietz et al. (2009)	+	n/a	+	-	+	!	-
Jones et al. (2007)	+	n/a	+	!	!	!	!
Mansoor & Dowse (2003)	!	n/a	+	-	!	-	+
Poncelas & Murphy (2007)	+	n/a	+	!	+	+	!
Rose et al. (2011)	+	n/a	-	-	!	-	!
Saletta et al. (2019)	+	n/a	+	+	!	+	-
Wilson & Read (2016)	+	n/a	+	+	!	+	-

Results from Individual Studies and Meta-Analysis

Based on the overall effect size across the eight included studies and 13 corresponding effect sizes, the data did not support a general beneficiary effect of adding visualization to plain language texts on reading comprehension, overall $g = 0.136$; $SE = 0.103$; $t(12) = 1.319$; $p = 0.212$; 95% CI, -0.089, 0.361; 95% Prediction Interval, -0.285, 1.042.

Subgroup Analysis

A breakdown of the total variance into the variance distributed at the three levels of the model revealed that 72% could be attributed to sampling variance, 1% to within-study variance, and 27% to between-study variance. Following guidance by Hunter and Schmidt (2004), this indicates that sufficient within- and between-study variance ($I^2 = 28\%$) remains to warrant subgroup and moderator analyses. Three subgroups were defined based on the cause of reading difficulties. There was no differential effect for the subgroups, *F*(2, 10) = 2.297, *p* = 0.151. The mean effects for each separate subgroup were as follows: Participants with aphasia, *g* = 0.151; *SE* = 0.118; *t*(10) = 1.273; *p* = .232; 95% CI -0.113, 0.415; Participants with ID, *g* = - 0.162; *SE* = 0.184; *t*(10) = -0.877; *p* = .401; 95% CI -0.572, 0.249; Participants with general low literacy, *g* = 0.361; *SE* = 0.160; *t*(10) = 2.256; *p* = .048, 95% CI 0.004, 0.718. There was no significant difference between the aphasia and low literacy subgroup, *t*(10) = -1.058; *p* = 0.315, the ID and low literacy subgroup, *t*(10) = -2.142; *p* = .058, and the aphasia and ID subgroup, *t*(10) = 1.426; *p* = .184, effect sizes. A summary of overall results and subgroup results of individual studies, overall effects and confidence intervals are presented in Figure 3.

Moderator analyses

Neither of the moderator analyses yielded significant results. Type of Visualization used in the study (photo vs drawing), *F*(1, 11) = 0.117, *p* = .739. Grade level of the text used in the study (grade level 1-4 vs grade level 5-6), *F*(1, 11) = 0.114, *p* = .742.

Sensitivity Analyses

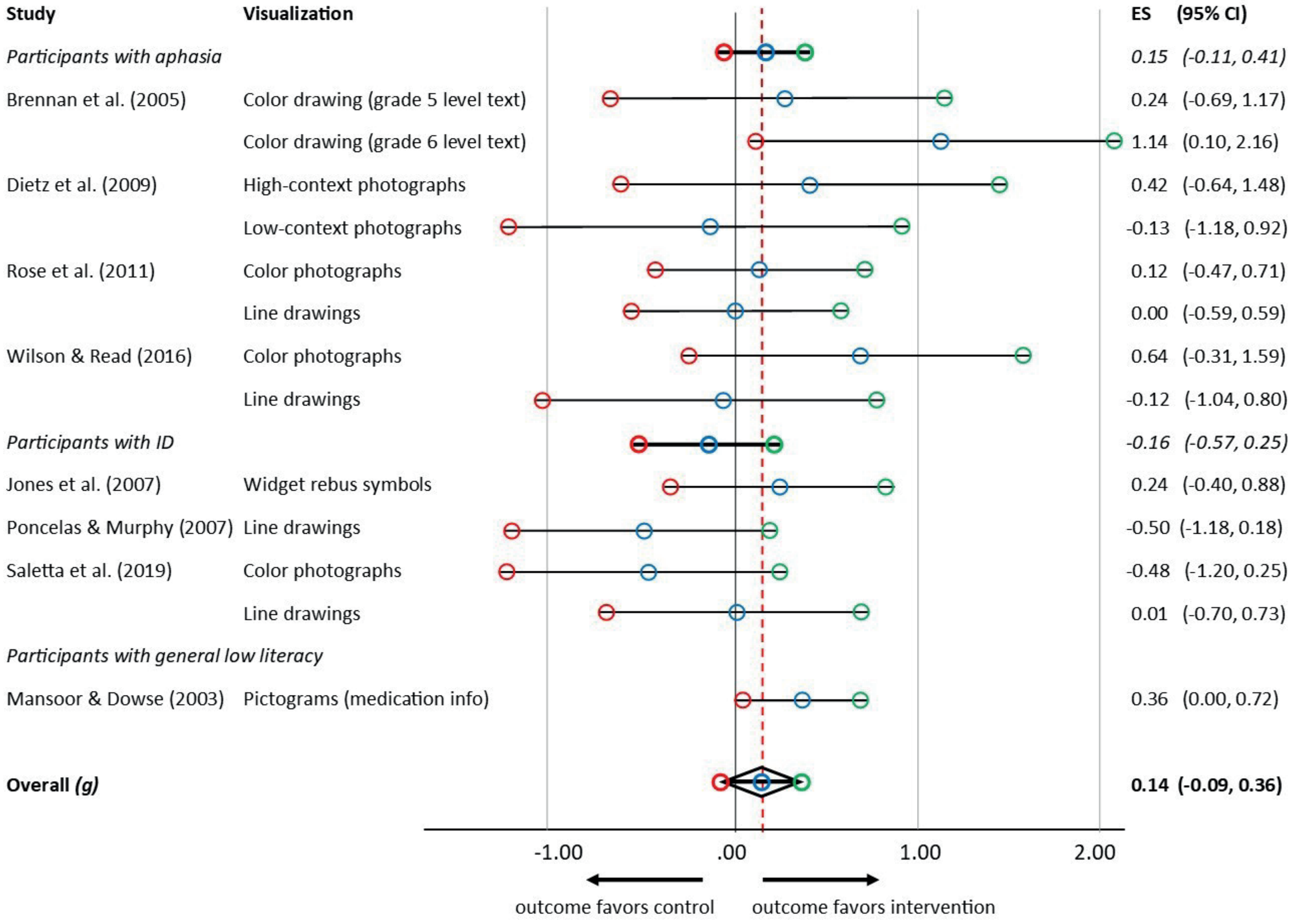
To assess whether the overall effect would prove to be different if only studies of medium-to-high quality studies were included, studies with a high risk of bias (Mansoor & Dowse, 2003; Rose et al., 2011) were excluded from the meta-analysis. The overall effect size was lower when low-quality studies were excluded, *g* = 0.071; *SE* = 0.162; *t*(9) = 0.437; *p* = .673; 95% CI, -0.296, 0.437.

Similarly, excluding the studies for which a poor quality of visualization was observed (Brennan et al, 2005; Dietz et al, 2009; Wilson & Read, 2016) did not improve the significance of the overall effect, *g* = .121; *SE* = .142; *t*(4) = .853; *p* = .442; 95% CI, -0.273, 0.515.

Equivalence

Because non-significance of effect sizes can be attributable to a small number of effect sizes included in small-scale meta-analyses, we performed equivalence tests for each subgroup separately to see if the non-significant effect sizes were practically equivalent to the absence of a true population effect for each subgroup. Equivalence tests were performed for the two subgroups participants with aphasia and participants with intellectual disabilities separately. This was done because they were perceived to be conceptually different, for instance in terms of etiology or cognitive deficits underlying reading problems and contributed the most studies to the overall effect in this meta-analysis.

Figure 3. *Individual and Overall Effect Sizes (ES) and Confidence Intervals (CI)*



Equivalence Test for Participants with Aphasia

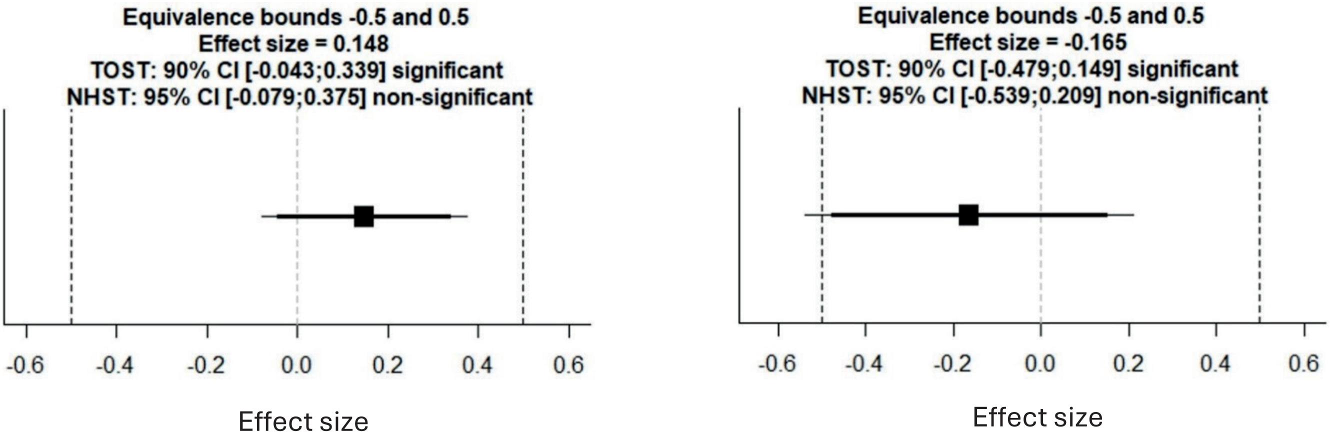
The equivalence test was significant, *z* = -3.034, *p* = .0012, given equivalence bounds of -0.50 and 0.50 and alpha set at .05 (two-tailed). As illustrated in Figure 4, the assumption that the true population effect size approaches zero (equivalence) cannot be rejected, as the confidence interval of the overall effect includes 0. The assumption that a true population effect greater than the SESOI of 0.50 or smaller than -0.50 exists can be rejected, as the confidence interval of the effect does not include values smaller than 0.50 or greater than 0.50.

Equivalence Test for Participants with ID

The equivalence test was significant, *z* = 1.754, *p* = .0397, given equivalence bounds of -0.50 and 0.50 and alpha set at .05 (two-tailed). As illustrated in Figure 4, the assumption that the true population effect size approaches zero (equivalence) cannot be rejected, as the confidence interval of the overall effect includes 0. The assumption that a true population effect greater than the SESOI of 0.50 or smaller than -0.50 exists can be rejected, as the confidence interval of the effect does not include values smaller than 0.50 or greater than 0.50. We can conclude that it is highly unlikely that there is a greater than medium detrimental effect of visualization on comprehension for people with intellectual disabilities.

Figure 4.

Equivalence Tests for Participants with Aphasia and Participants with intellectual disabilities



Participants with aphasia

Participants with intellectual disability

Discussion

In this review we attempted to answer the question if adding visualizations to an easy-to-read text facilitates understanding this text for people who have difficulties reading. We did this by compiling the evidence from empirical studies on this topic in a meta-analysis. The combined effect size of the 13 effect sizes from the eight included studies did not indicate that adding pictures significantly enhanced text comprehension. The non-significance of an effect size can reflect statistical equivalence (i.e. no effect exists) between intervention and control conditions. But alternatively, given the small numbers of participants in the included studies and a limited number of studies that could be included in the meta-analysis, this could also reflect a power problem (Lakens, 2017). In other words, the design of this meta-analysis, in which a small number of studies with a limited number of participants are combined, might have not permitted us to detect small true population effect sizes such as the one found in the current study. Therefore, we performed equivalence tests, to determine if there was a possibility that a true population effect size of greater than 0.50 or smaller than -.50 (a medium effect; Cohen, 1988) might have been missed in this study due to power problems. Looking at the results of the equivalence test, for both participants with aphasia and participants with intellectual disabilities, we cannot reject an effect size of zero. Additionally, it is unlikely that a true population effect size larger than 0.50 or smaller than -0.50 will be detected with the study designs included in this meta-analysis.

Analysis of between-study and within-study variance indicated that there was no moderating effect for any of the subgroups or moderators investigated. There was no differential effect for aphasia versus intellectual disability versus general low literacy as cause of reading difficulties, drawings versus photographs, or for lower grade level versus higher grade level text difficulty. Sensitivity analyses revealed that higher-quality studies did not perform any better than studies of lower methodological quality.

Further Reflections on the Absence of Statistical Evidence for a Positive Effect of Visualization

Using visualization to clarify written content is one of the most widely applied means of augmentative communication. It seems to be an effective strategy on an intuitive level. There are numerous examples of visualizations that genuinely help us better understand. Examples are wayfinding pictograms in airports and visualized instructions for putting together Billy bookcases (Frixione & Lombardi, 2015). Adding pictures to text is a common-sense strategy for enhancing comprehension for low-literate persons and is widely advocated in guidelines for accessible communication. So why does the available empirical evidence not unequivocally support the notion that visualization may be helpful for persons with reading difficulties?

Participant characteristics

An absence of definitive evidence in favor of visualization might be testimony to the 'cognitive overload' theory by Hurtado (2014). Having to process two sources of information (visual and semantical) simultaneously causes the working memory to overflow, hampering the effective processing of information (Poncelas & Jones, 2007). Whereas evidence from studies with typically developed participants seem to indicate better understanding when visualization is provided, the fact that most participants in the studies included in this meta-analysis suffered from aphasia or intellectual disabilities, conditions that are known to negatively affect working memory, lends further explanation for the absence of positive effects.

On a similar note, the 'pictorial competence' of some of the participants may have been compromised. This ability to decipher what abstract conceptual meaning is embedded within a concrete picture is an acquired skill and associated with cognitive abilities (DeLoach et al., 2013). Participants with intellectual disabilities especially may lack the pictorial competence to adequate link a picture's abstract properties to the concept illustrated in the text. As Poncelas and Murphy (2007) demonstrate, promoting pictorial competence by teaching persons with intellectual disabilities the meaning of symbols used in pictogram communication systems greatly enhances the chance that pictures *do* facilitate understanding.

Thirdly, research indicates that visualization is only useful in supporting comprehension if the core meaning of the text itself is understood (Filippatou & Pumfrey, 1998). Pictures that represent concepts that are beyond the reader's ability to understand may actually *interfere* with their comprehension. For persons who have trouble decoding written text and/or struggle to understand written communication on a cognitive and conceptual level, adding pictures may even be detrimental (Dowse et al., 2023). Persons with reading difficulties may use a picture to guess the intended meaning, often incorrectly, and wrongfully think that they understand the message.

Study Quality

Next to statistical and theoretical considerations, a third explanation for the absence of evidence for a positive effect arises from the assessment of the methodological quality of the contributing studies. All included studies suffered from methodological flaws, ranging from relatively benign operationalization imperfections and underpowered designs to serious doubts about the validity of outcome measures and procedures that introduced confounders. Assessing the methodological quality with the established RoB-2 tool (Higgins et al., 2020) revealed that the minority of included studies had a favorable risk-of-bias rating. Distinguishing between moderate-to-high-quality studies and low-quality studies did not show a difference in observed effect size.

Choosing Appropriate Visualizations to Accompany Texts

To maximize the probability that visual supports are effective, we would think that researchers do their utmost to choose the best possible visualizations to support their texts. We would expect their choices to be based on evidence-based practices, incorporating previous research, their clinical expertise, and population preferences. As mentioned earlier in the Results section, for most studies the choice of visualization was questionable for several reasons. Many studies did not use existing best-practice standards, did not report about clarity, clearness, low versus high context, color versus black-white, did not justify the choice of type of visualization (level of iconicity), did not describe the intensity and placement of visualization, and did not check if the intended study population understood the visualizations used. The choices made in the design phase of most of the studies appeared to be based on intuition rather than evidence. In defense of the researchers, no comprehensive guidelines to support the meaning of written text with visuals for less-proficient readers have been developed yet. So many different aspects relating to visualizations can be manipulated that it may well be too complicated to include and evaluate all these different aspects in one study.

Selection Bias

One of the included studies excluded more severely impaired participants beforehand (Dietz, 2009) and some excluded participants who in the course of the study appeared to be less able to read (Brennan et al., 2005; Rose et al., 2011). Earlier and contemporary research on the topic of health literacy showed that the effect of visualization may be greater for more severely impaired readers (Houts, 2006; Dalemans, et al. 2022). It seems somewhat ironic that the people who potentially benefit most from visual supports were not able to participate in some of the studies under investigation.

Text Difficulty and the Use of Readability Formulas

We used the level of text difficulty as one of the exclusion criteria for inclusion in this review as it was expected that our intended study population – i.e. less-proficient readers – would have too much difficulty reading texts beyond a certain level. Different tools to define reading grade levels were used in the different studies (Fry, Flesh Kincaid, Flesh Reading Ease, Neale analysis of Reading Ability revised and SRA grade). There is considerable doubt that these readability formulas can be used interchangeably to reliably assess the difficulty of a given text and there is scepsis about the presumed linear association between grade levels and reading ease (Begeny & Greene, 2013). But as the level of text difficulty was uniformly expressed in terms of grade levels, we had to make do with the operationalization provided by the contributing researchers.

Specifically, texts that exceeded Grade Level 6 were deemed to be too difficult for individuals with reading impairments. This may have introduced a ceiling effect for the impact of visualization on comprehension. It could imply that the text was sufficiently clear and concrete in itself and that illustrations did not cause an added effect above and beyond the beneficiary effect of the textual simplification. Research on health literacy demonstrated that generally, people *do* understand and adhere to medical instructions better if they are visualized, but *only* for difficult texts (Meppelink, 2015). On the other hand, a Grade Level 6 may prove to be too difficult to read and understand by many people with organic or developmental disabilities. As mentioned above, supportive visualization may do more bad than good if a reader does not understand the core message a text intends to convey.

In conclusion, it seems that the less a person is able to understand a body of text (be it because the subject discussed is too complex, the reading level of the text is too difficult or the reading impairment is too high), the more she or he has to rely on visual information for a sense of understanding but the greater the chance that visualization may actually impede comprehension. Especially if not enough attention is paid to ensure the visualization suits the intended reader's communication needs and cognitive level.

If we integrate this conclusion in all the methodological imperfections mentioned above, for most studies, it is questionable if significant and reliable results were to be expected *a priori*.

Recent Developments in Design and Use of Visualization

All studies included in this review are relatively dated, with the majority of studies published before 2010. In the last 10 years, mainly because of increased social media usage, the use of visualization to support meaning has surged (Li & Xie, 2020). These days, most readers will be accustomed to seeing infographics and emoji's alongside texts in journals, advertising, and online content. Not only the exposure to visualization has increased, the way written meaning is visualized has also changed dramatically, with current visuals looking strikingly different from the examples provided in this study. As some recent studies show, new modes of visualizing written or worded content may hold promise in promoting understanding above and beyond 'low-tech' text augmentations such as static visualizations. In a recent study by Kayler and colleagues (2023) for instance, video instructions significantly improved understanding of a medical procedure compared to traditional leaflets with text and pictures.

Limitations

Several limitations in regard to the design and conduct of the meta-analytic review preclude our ability to draw firm conclusions. First, we applied strict inclusion criteria for the studies to be included in the review. This led to a relatively small set of included studies and corresponding effect sizes. This seems problematic in the context of a quantitative synthesis of results, but scholarly discussion on the minimum number of studies to be included in a meta-analysis generally settles on the notion that combining the results of as few as two studies can be meaningful (Valentine et al., 2010). Conversely, expanding our inclusion criteria would have led to a possibly unmanageable heterogeneity in study designs, populations, and outcomes, obstructing our ability to make comparisons between studies.

A second limitation pertains to the quality assessment procedure, the outcomes of which play an important role in the conclusions we permit ourselves to draw (and not draw). While the authors used a well-validated tool from a renowned institute (Cochrane), the assessment procedure leaves some room for interpretation of the individual reviewers. For the assessment of the quality of the visualizations used, no validated tool was readily available, and the criteria were drawn up by the research team members themselves, based on best-practice standards and expert review. The validation of this addendum requires further study. For the assessment procedure we tried to ascertain a certain standard of scientific rigor by using a double-blind assessment procedure and corroborating our findings with the broader review team.

Lastly, we only included ‘comprehension’ as outcome to determine the effect of visualization. The operationalization of text comprehension was diverse and sometimes opaque. Recent advances in language studies suggest that there may be more appropriate and valid ways of assessing comprehension, such as the use of cloze testing (Kleijn et al., 2019). Several other outcome variables are mentioned in the included and comparable studies, that could be indicative of a positive effect. Two outcomes that are used in some studies to complement text comprehension are reading time or text processing speed and appreciation of the visualizations used. Processing speed may be an indication of the effort involved in information processing, with higher processing speeds indicating less demanding and therefore more efficient processing. However, the relation between processing speed and comprehension as an ultimate outcome is not clear. It may well be that it takes more time and effort to simultaneously process text and visual representation, but this effort results in better understanding. Questions relating to whether participants prefer texts to be accompanied by pictures and if they like the pictures used are often asked in the course of studies on visualization. Although preferences or appreciation of participants may not relate directly to comprehension, appealing imagery can be important for engaging people with texts, thereby indirectly influencing the results. Because of these unclear or indirect relations with comprehension, both aspects were not used to determine effect sizes in this review.

Conclusions

This meta-analytic review has not been able to demonstrate that adding visualization to easy-to-read texts leads to better text comprehension for people who have difficulty reading. We offer several explanations why a positive effect was not to be expected *a priori*, considering participants characteristics and methodological issues surrounding the studies included in the meta-analysis.

The findings are in stark contrast with the overwhelmingly widespread practice of accompanying texts with pictures to promote accessibility of written communication. If anything, this study highlights that providing effective visual supports to facilitate understanding may be a potentially effective strategy, but it is not merely a matter of randomly slapping pictures next to a sentence.

In all studies that were included in this review, the aim was to evaluate the quite generic question ‘Does visualization work?’. We suggest that, given the many unanswered questions surrounding the precise working mechanisms of visualization, it is more pertinent to first investigate ‘what might work when for whom’. From our own clinical and research experience, what pictures may actually help can be highly idiosyncratic; what works for one does not always work for all. Factors that may play a role are myriad and include personal preferences, cognitive abilities, familiarity with the visualization system (Dalemans et al., 2021), and the level of ‘pictorial competence’ (DeLoach et al., 2003). There is likely no one-size-fits-all solution; even within intended target groups, variation between individuals may be much greater than variation between target groups. The level of verbal and lexical fluency of persons within the group ‘persons with mild intellectual disability’ varies considerably for instance. If researchers intend to make inferences about what type of visualization works for the majority of persons in an intended population of readers, more attention should be paid to including representative participants in the design of the study. Testing the appropriateness, clarity and acceptability of the visualizations before the execution of the actual study should always be a part of the process.

References

Ahmed, K. M., & Al Dhubaib, B. (2011). Zotero: A bibliographic assistant to researcher. *Journal of Pharmacology and Pharmacotherapeutics*, *2*(4), 303. https://doi.org/10.4103/0976-500x.85940

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, *311*(7003), 485.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. Chichester (UK): Wiley.

* Brennan, A., Worrall, L., & McKenna, K. (2005). The relationship between specific features of aphasia-friendly written material and comprehension of written material for people with aphasia: An exploratory study. *Aphasiology*, *19*(8), 693–711. https://doi.org/10.1080/02687030444000958

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*, 211–229. https://doi.org/10.1037/a0032968

Chinn, D., & Homeyard, C. (2017). Easy read and accessible information for people with intellectual disabilities: Is it worth it? A meta-narrative literature review. *Health Expectations*, *20*(6), 1189–1200. https://doi.org/10.1111/hex.12520

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.

Conners, F. A. (2003). Reading Skills and Cognitive Abilities of Individuals with Mental Retardation. *Language and Communication in Mental Retardation*, 191–229. https://doi.org/10.1016/s0074-7750(03)27006-3

Dalemans, R., Stans, S., von Helden-Lenzen, S., Lemmens, J., Ummels, D., Cornips, S. (2021). *Leidraad Communicatievriendelijk meten.* [Guideline for Communication-friendly Measurement]. Zuyd University of Applied Sciences. Report. https://doi.org/10.25385/zuyd.16782985.v1

Dalemans, R., von Lenzen-Helden, S., & Stans, S. (2022). The futility of asking incomprehensible questions: making questionnaires communication-friendly. In S. Ahrens, R. Schulz, J. Kröger, S. Hernández Garrido, L. Keller, & I. Rink (eds.). *Accessibility – Health Literacy – Health Information*. Frank & Timme.

DeLoache J. S., Pierroutsakos S. L., Uttal D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science, 12*(4), 114–118. https://doi.org/10.1111/1467-8721.01244

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science, 9*(3), 205-210. https://doi.org/10.1111/1467-9280.00039

Department of Health. (2010). *Making written information easier to understand for people* *with learning disabilities*. COI for the Department of Health. Https://webarchive.nationalarchives.gov.uk/20130703133435/http:/odi.dwp.gov.uk/docs/iod/easy-read-guidance.pdf"https://webarchive.nationalarchives.gov.uk/20130703133435/http://odi.dwp.gov.uk/docs/iod/easy-read-guidance.pdf

De Rijdt, C. (2013). *Ondersteunend communiceren: werken met visualisaties* [Supportive communication: working with visualizations]. Maklu.

* Dietz, A., Hux, K., McKelvey, M. L., Beukelman, D. R., & Weissling, K. (2009). Reading comprehension by people with chronic aphasia: A comparison of three levels of visuographic contextual support. *Aphasiology*, *23*(7–8), 1053–1064. https://doi.org/10.1080/02687030802635832

Doak, C. C., Doak, L. G., & Root, J. H. (1996). *Teaching Patients With Low Literacy Skills* (Subsequent ed.). Lippincott Williams & Wilkins.

Dowse, R., Okeyo, S., Sikhondze, S., & Khumalo, N. (2023). Pharmaceutical indication pictograms for low literacy viewers: Health literacy and comprehension. *Health SA Gesondheid, 28*, 2192.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge university press.

Inclusion Europe (2023). *European standards for making information easy to read and understand*. https://www.inclusion-europe.eu/wp-content/uploads/2017/06/EN_Information_for_all.pdf

Flesch, R. (2007). Flesch-Kincaid readability test. *Retrieved October*, *26*(3).

Frixione, M., & Lombardi, A. (2015). Street signs and Ikea instruction sheets: Pragmatics and pictorial communication. *Review of Philosophy and Psychology, 6*, 133-149. https://doi.org/10.1007/s13164-014-0216-1

Foundation for People with Learning Disabilites (2023). *Easy Read*. https://www.learningdisabilities.org.uk/learning-disabilities/a-to-z/e/easy-read

Fuller, D. R. (1997). Initial study into the effects of translucency and complexity on the learning of Blissymbols by children and adults with normal cognitive abilities. *Augmentative and Alternative Communication, 13*, 30–39. https://doi.org/10.1080/07434619712331277818.

Hibbing, A. N., & Rankin-Erickson, J. L. (2003). A picture is worth a thousand words: Using visual images to improve comprehension for middle school struggling readers. *The reading teacher, 56*(8), 758-770.

Higgins, J.P.T., Li, T., & Sterne, J. (2020). Revised Cochrane risk of bias tool for randomized trials (RoB 2) Additional considerations for crossover trials. https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-crossover-trials

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient education and counseling, 61*(2), 173-190. https://

doi.org/10.1016/j.pec.2005.05.004

Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed). Thousand Oaks, CA: Sage.

Hurtado, B., Jones, L., & Burniston, F. (2013). Is Easy Read information really easier to read? *Journal of Intellectual Disability Research*, *58*(9), 822–829. https://doi.org/10.1111/jir.12097

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp International Organisation for Standardisation (2023). ISO/DIS 24495-1(en) Plain language — Part 1: Governing principles and guidelines. https://www.iso.org/obp/ui/#iso:std:iso:24495:-1:dis:ed-1:v1:en

International Plain Language Federation (2023). *Plain language definition*. https://www.iplfederation.org/plain-language/

* Jones, F. W., Long, K., & Finlay, W. M. L. (2007). Symbols can improve the reading comprehension of adults with learning disabilities. *Journal of Intellectual Disability Research*, *51*(7), 545–550. https://doi.org/10.1111/j.1365-2788.2006.00926.x

Kayler, L. K., Keller, M. M., Breckenridge, B., Feeley, T. H., Suboh, J., & Tumiel‐Berhalter, L. (2023). Preliminary feasibility of animated video education designed to empower patients’ referral to kidney transplantation. *Clinical Transplantation, 37*(1), e14838. https://doi.org/10.1111/ctr.14838

Kelly, L. (2017). Readability and the Gunning Fox Index. https://readable.com/blog/the-gunning-fog-index/

Kelly, L. (2017). The Flesch Reading Ease and Flesch-Kincaid Grade Level. https://readable.com/blog/the-flesch-reading-ease-and-flesch-kincaid-grade-level/

Kelly, L. (2019). The Fry Readability Graph. Readable. https://readable.com/blog/fry-readability-graph/

Kleijn, S., Pander Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing, 36*(4), 553-572.

Koraal (2020). Brochure de Regels voor Taal voor Allemaal [Brochure the Rules for Language for All]. Koraal. https://www.koraal.nl/extra-informatie/aanbod/taal-voor-allemaal

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science, 8*(4), 355-362.

Lakens D., Scheel A. M., Isager P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1, 259–269.

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

Leichte Sprache (2021). Was ist Leichte Sprache? Leichte-Sprache [What is Easy Language? Easy Language]. https://leichte-sprache.de

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research, 57*(1), 1-19.

* Mansoor, L. E., & Dowse, R. (2003). Effect of Pictograms on Readability of Patient Information Materials. *Annals of Pharmacotherapy*, *37*(7–8), 1003–1009. https://doi.org/10.1345/aph.1c449

Marušić, M. F., Fidahić, M., Cepeha, C. M., Farcaș, L. G., Tseke, A., & Puljak, L. (2020). Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals. BMC medical research methodology, 20(1), 1-10. https://doi.org/10.1186/s12874-020-00966-4

Meppelink, C. S., Smit, E. G., Buurman, B. M., & Van Weert, J. C. M. (2015). Should We Be Afraid of Simple Messages? The Effects of Text Difficulty and Illustrations in People With Low or High Health Literacy. *Health Communication*, *30*(12), 1181–1189. https://doi.org/10.1080/10410236.2015.1037425

Nomura, M., Nielsen, G. S., Tronbacke, B. I., & International Federation of Library Associations and Institutions. Section of Libraries Serving Disadvantaged Persons. (2010). *Guidelines for Easy-to-read Materials*. IFLA Headquarters.

Oskam, E., & Scheres, W. (2016). *Totale communicatie* [Total communication]. Bohn Stafleu van Loghum.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1). https://doi.org/10.1186/s13643-016-0384-4

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

Paivio, A. (2013). *Imagery and Verbal Processes*. Taylor & Francis.

* Poncelas, A., & Murphy, G. (2007). Accessible Information for People with Intellectual Disabilities: Do Symbols Really Help? *Journal of Applied Research in Intellectual Disabilities*, *20*(5), 466–474. https://doi.org/10.1111/j.1468-3148.2006.00334.x

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R-project. https://www.R-project.org/

*Rose, T. A., Worrall, L. E., Hickson, L. M., & Hoffmann, T. C. (2011). Exploring the use of graphics in written health information for people with aphasia. *Aphasiology*, *25*(12), 1579–1599. https://doi.org/10.1080/02687038.2011.626845

* Saletta, M., Kaldenberg, E., Rivera, K., & Wood, A. (2019). Do Illustrations Promote Reading Comprehension in Adults with Intellectual or Developmental Disabilities?

Education and Training in Autism and Developmental Disabilities, 54(3), 225-236. https://www.jstor.org/stable/26780623

Scheffers, F., Moonen, X., & Van Vugt, E. (2021). Assessing the quality of support and discovering sources of resilience during COVID-19 measures in people with intellectual disabilities by professional carers. *Research in Developmental Disabilities*, *111*, 103889. https://doi.org/10.1016/j.ridd.2021.103889

Schubbe, D., Scalia, P., Yen, R. W., Saunders, C. H., Cohen, S., Elwyn, G., ... & Durand, M. A. (2020). Using pictures to convey health information: A systematic review and meta-analysis of the effects on patient and consumer health behaviors and outcomes. *Patient education and counseling, 103*(10), 1935-1960. https://doi.org/10.1016/j.pec.2020.04.010

Sorber, R. (2021). *Over Toegankelijkheid – Voor wie?* [About Accessibility – for whom?]. Accessibility. https://www.accessibility.nl/over-toegankelijkheid/voor-wie

Stockmeyer, N. O. (2009). Using microsoft word's readability program. *Michigan Bar Journal*, *88*, 46. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1210962

Sutherland, R. J., & Isherwood, T. (2016). The Evidence for Easy-Read for People With Intellectual Disabilities: A Systematic Literature Review. *Journal of Policy and Practice in Intellectual Disabilities*, *13*(4), 297–310. https://doi.org/10.1111/jppi.12201

Tabachnik, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.) Allyn and Bacon.

Tuohimetsä, R. (2020). *Kirjoita selkokieltä!* [Write in plain language] [Presentation slides]. Selkokeskus. https://savonake.fi/wp-content/uploads/2020/05/Kirjoita-selkokieltä-Riikka-Tuohimetsä.pdf

United Nations (2006). *Convention on the Rights of Persons with Disabilities*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html

University of Oregon (2018). *Easy CBM reading*. Retrieved from https://dibels.uoregon.edu/sites/default/files/2021-06/dibels_8_benchmark_3.zip

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247.

Van Ewijk, L., Versteegde, L., Raven-Takken, E., & Hilari, K. (2016). Measuring quality of life in Dutch people with aphasia: development and psychometric evaluation of the SAQOL-39NL. *Aphasiology*, *31*(2), 189–200. https://doi.org/10.1080/02687038.2016.1168919

Verhoeven, L., Voeten, M., & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics*, *50*, 28–36. https://doi.org/10.1016/j.jneuroling.2018.03.002

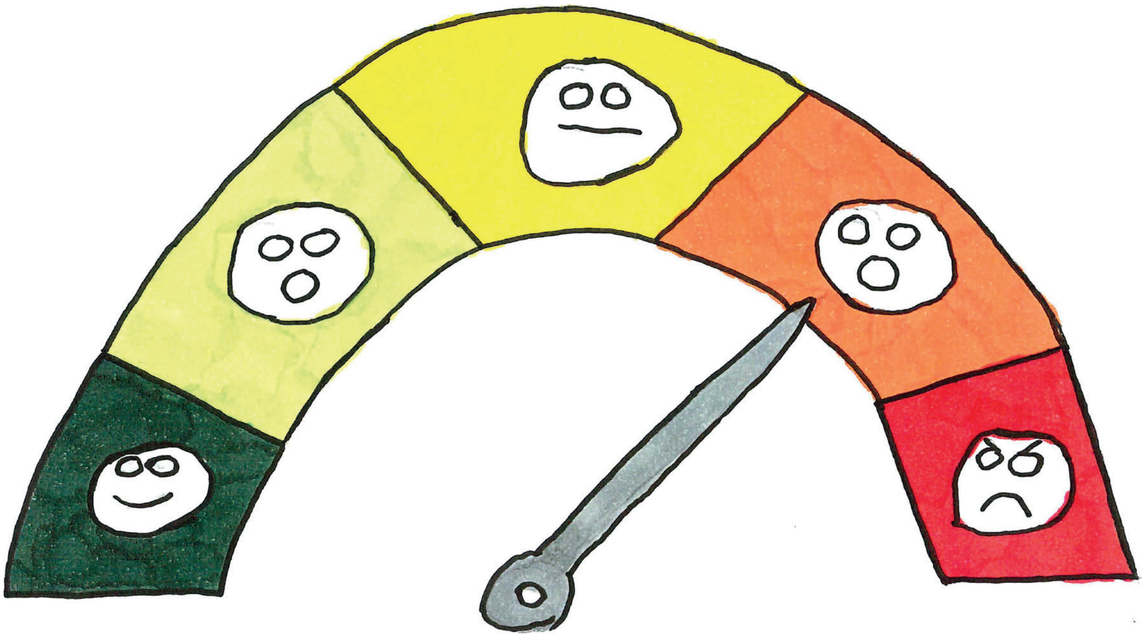
Wibbelink, C., & Assink, M. (2015). *Handleiding voor het uitvoeren van een drie-level meta-analyse in R* [Guide to conducting a three-level meta-analysis in R]. Afdeling Forensische Orthopedagogiek, Universiteit van Amsterdam. https://pure.uva.nl/ws/files/2679897/166572_Handleiding_drie_level_meta_analyse_in_R_versie_3.2_.pdf

*Wilson, L., & Read, J. (2015). Do particular design features assist people with aphasia to comprehend text? An exploratory study. *International Journal of Language & Communication Disorders*, *51*(3), 346–354. https://doi.org/10.1111/1460-6984.12206

Yaneva, V., Temnikova, I., & Mitkov, R. (2015). Accessible Texts for Autism. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*. https://doi.org/10.1145/2700648.2809852

Chapter 5

Self-report stress measures to assess stress in adults with mild intellectual disabilities – A scoping review.



Published as:

de Witte, M., Kooijmans, R., Hermanns, M., Hooren, S. V., Biesmans, K., Hermsen, M., Stams, G.J., & Moonen, X. (2021). Self-report stress measures to assess stress in adults with mild intellectual disabilities—A scoping review. Frontiers in psychology, 12, 742566.

Abstract

Stress has a major negative impact on the development of psychopathology and contributes to adverse physical conditions. Timely recognition and monitoring of stress-related problems are important, especially in populations that are more vulnerable to stress, such as people with mild intellectual disabilities (MID). Recent research on the use of physiological measures to assess stress levels emphasize that, in addition to these measures, self-report instruments are necessary to gain insight into the individual perception and impact of stress on daily life. However, there is no current overview of self-report stress measures that focus on the experience of stress in the present moment or in daily life.

To provide an overview of the existing self-report stress measures for clinicians and researchers, a scoping review was conducted. In addition, to advise clinical professionals on the use of self-report measures of stress for people with MID, the results of an expert consultation were used to refine the preliminary findings.

A systematic scoping literature search resulted in a total of 13 self-reported stress measures that met the final inclusion criteria, of which three were developed specifically for assessing stress in adults with MID (GAS-ID, LI, & SAS-ID). For each included self-report stress measure, the psychometric quality, assessment procedure, & suitability for adults with MID were reported. These were supplemented by the findings from the expert consultation.

Implications for clinical practice on the use of self-report stress measures, particularly for people with MID, are discussed. Recommendations for future research and development are given.

Introduction

Recognizing a person's stress-related problems is increasingly important, as ever more evidence on the adverse effects of stress on health and well-being is accumulated. High stress levels are regarded as an important risk factor for the onset and progression of a wide range of physical and emotional problems, such as cardiovascular diseases, cancer, anxiety disorders, depression, & burnout (American Psychological Association [APA], 2017; 2015; Australian Psychological Society [APS], 2015; Steptoe & Kivimäki, 2012). Nevertheless, the literature reports that it is difficult for many people to both understand the destructive impact of daily life stress experiences (Casey, 2017; de Witte et al., 2020) and to reduce or cope with stress without any professional support (World Health Organization [WHO], 2010). This is especially the case for adults with mild intellectual disabilities (MID), as they experience stress more frequently in daily life than people without intellectual disabilities (Emerson, 2003; Hatton & Emerson, 2004; Schuengel & Janssen, 2006; World Health Organization [WHO], 2010). In addition, people with MID have also been found to have fewer resources to cope with daily life stress experiences (Hartley et al., 2009a; Lunsky & Benson, 2001; Scott & Havercamp, 2014).

Measuring the Concept of Stress

When we use the term "stress" in the present study, we are referring to negative stress experiences, defined by Aldwin (2007) as the quality of an experience produced by a person-environment transaction that, through either overarousal or underarousal, results in psychological or physiological distress (Aldwin, 2007; Riley & Park, 2015). Responses to stress are related to physiological arousal and emotional states, & the underlying systems of both these responses regulate and affect each other in times of stress (de Witte et al., 2020; Linnemann et al., 2017; McEwen & Gianaros, 2010). The physiological response to stress implies the activation of the hypothalamic-pituitary adrenal (HPA) axis and, due to the release of adrenalin and noradrenalin, increased activity of the sympathetic nervous system. This in turn results in increased physiological arousal, such as heart rate (HR), blood pressure, & cardiac output (Bally et al., 2003; Pfaff et al., 2007). Stress-related emotional states can be defined in terms of subjective worry, nervousness, & restlessness (Akin & Iskender, 2011; Cohen et al., 1983; Pittman & Kridli, 2011; Pritchard, 2009), & have many similarities with "state anxiety" as an outcome. Accordingly, many researchers describe state anxiety as an emotional response to an individual's perception of a stressful experience (e.g., Hook et al., 2008; Koelsch, Fuermetz, et al., 2011). In this review, we therefore regard state anxiety as a stress-related outcome. Stress-related outcomes can be measured by means of biomarkers related to physiological arousal (physiological measures) and by assessing people's emotional states related to stress experiences (psychological measures). Empirical studies on stress use either physiological or psychological measurement methods (proxy or self-reports) or a combination of both (Kim et al., 2018) for the measurement of stress-related outcomes.

Although there is a large body of knowledge concerning the immediate effects of stress on physiological arousal, as indicated by several biomarkers like HR, blood pressure, heart rate variability (HRV) and hormone levels (Chandola et al., 2010; Föhr et al., 2017; Kim et al., 2018), increased physiological arousal does not automatically translate to elevated levels of perceived stress. It can also signal, for example, that a person is positively excited or deeply focused (Csikszentmihalyi, 2000; Pfaff et al., 2007; Rheinberg & Engeser, 2018). When examining subjective stress levels, many researchers therefore emphasize the importance of assessing the subjects' perceived emotional state in relation to stress, to help interpret physiological markers of arousal.

Both proxy-reported and self-reported information are used to examine psychological stress-related outcomes, such as people's emotional states (Crawford et al., 2006). Proxy reports refer to information about an individual given by significant others, such as relatives or caretakers. These are often used as an alternative when obtaining self-reported information is not a viable option, for instance when the respondent is not able to communicate verbally (Emerson et al., 2013; Miller & Tucker, 1993; Moore, 1988). Evidence suggests that proxy reports may be less accurate and less sensitive, compared to self-reported information (Scott & Havercamp, 2018; Moss et al., 1996). Self-report measures originally refer to data obtained by questionnaires or interviews in which respondents are asked to report about their personal experiences, values, feelings or thoughts, related to certain contexts and/or circumstances (Chan, 2009). Self-report data are commonly collected on a wide variety of topics in both medical and psychological research on topics such as pain, emotions, & personal preferences (Scott & Havercamp, 2018; Gerald & George, 2010).

Perceived Stress in Adults with Mild Intellectual Disabilities

MID is a neurodevelopmental disability characterized by deficits in intellectual and adaptive functioning skills (American Psychiatric Association [APA], 2013). The term MID generally refers to people with limited intellectual capacities and adaptive skills with IQ scores in the range from 55-70, & may in some definitions include persons with "borderline intelligent functioning" (IQ 70-85; Kaal et al., 2015; Wieland & Zitman, 2016). As in the general population, stress experienced by adults with MID is linked to many negative mental health outcomes (Hartley et al., 2009a; 2009b; Hulbert-Williams & Hastings, 2008; Scott & Havercamp, 2014). Persistent stress in adults with MID can lead to maladaptive coping strategies and detrimental mental and physical health conditions such as depression (Hartley et al., 2009a, 2009b), impaired cognitive functions (Heyman

& Hauser-Cram, 2015), physical health problems (Lunsky, 2008), & substance abuse (Didden et al., 2009).

Adults with MID experience more stress in daily life than people without intellectual disabilities (Bramston & Mioche, 2001; Casey, 2017; De Witte, Spruit, et al., 2020; Emerson, 2003; Hatton & Emerson, 2004; Schuengel & Janssen, 2006; World Health Organization [WHO], 2010). Moreover, they have been shown to have more difficulties coping with their daily stress than adults without intellectual disabilities (APA, 2013), which fits with the theory of Cohen et al. (1983) who defined psychological stress as the extent to which persons perceive that demands exceed their ability to cope. Various explanations have been proposed for this increased risk of stress, including experienced difficulties in social interactions, which appears to be one of the main stressors in their daily lives. This in turn can be explained by their lack of control over minor daily and major life decisions (Dulin et al., 2013; Hartley et al., 2009b; Scott & Havercamp, 2014). Furthermore, adults with MID often seem to lack social support and self-efficacy, important factors for coping with stress (Abbaszadeh & Sardoie, 2016; Everly & Lating, 2019; Seyed et al., 2017).

Psychological Stress Measures for Adults with MID

Because of the superior accuracy and sensitivity of self-reported information over proxy measures, researchers in the field of MID generally prefer self-reporting measures above proxy measures to assess the experience of stress in persons with MID (Lindsay & Skene, 2007; Scott & Havercamp, 2018). Gaining an understanding of someone's personal thoughts, attitudes and feelings can lead to an enriched knowledge base from which opinions can be formed and interventions for stress reduction implemented (O'Keeffe et al., 2019). In this respect, the value of proxy instruments is perceived to be limited, as a proxy can never reliably report on another person's internal mental state (Emerson et al., 2013; Scott & Havercamp, 2018). This is in line with results of studies which show that, when compared to self-reported outcomes, proxies tend to overestimate impairment and underestimate health-related quality of life of people with (M)ID (Andresen et al., 2001; Vlot-van Anrooij et al., 2018).

High quality self-report measures on mental states, including stress, for adults with MID are few and far between (Glenn et al., 2003; Kooijmans et al., 2022; Sams et al., 2006). There are many challenges when collecting self-reported data from people with MID that are associated with the nature of the disability, including problems with reasoning, verbal expression, reading, abstract thinking, & judgment (APA, 2013; Schalock et al., 2010). To accommodate for these challenges, adaptations have to be made to 'standard' instrument language, lay-out, & assessment procedures. Few self-report measures are available that incorporate these adaptations to better suit individuals with intellectual disabilities (Lindsay & Skene, 2007; Scott & Havercamp, 2018). Additionally, Wieland et al. (2012) have identified a number of self-report measurement instruments developed for use in the general population which are suitable for adults with MID.

Purpose of the Present Study

As persistent stress can lead to the development of psychopathology and severe physical conditions, it is becoming increasingly important to recognize stress-related symptoms in populations known to be more vulnerable to stress, like people with MID. It is therefore critical to gain more insights into the way stress can be assessed in this population. Although advances in the use of physiological measures to assess people's stress levels have added substantial value to stress research, it is no substitute for the use of self-report measures, since the individual's perception of stress is directly related to individuals' emotional states. As stated before, physiological and emotional stress are not necessarily directly related (e.g., de Witte, Spruit, et al., 2020; Linnemann et al., 2017; Scott & Havercamp, 2018). In order to provide an overview of the existing self-report stress measures and to provide more information about their suitability for adults with MID, we conducted a scoping review. Moreover, in order to advise clinical professionals on how to correctly use the identified self-report stress measures, expert consultations were held to refine our preliminary findings. Our findings can be applied to research in which stress-related outcomes are measured in both adults with MID as well as those without intellectual disabilities. Results of this scoping review will provide guidance to clinical practitioners to assess perceived stress in adults with MID.

Methods

In order to provide an overview of existing stress self-report measures, we performed a scoping review. A scoping review follows a systematic approach to map evidence or to bundle scientific findings on a topic to identify concepts, theories, sources, & knowledge gaps (Arksey & O'Malley, 2005; Munn et al., 2018; Tricco et al., 2018). Contrary to systematic reviews, scoping reviews can also accommodate grey literature sources, opinions and non-peer-reviewed policy guidelines (Munn et al., 2018). Considering the diverse nature of information sources, risk-of-bias assessment of included sources may often not be appropriate for a scoping review. A scoping view approach matches our research questions, which aim to provide more insights into the different types of self-report measurements and their characteristics, & how they can be used in adults with MID.

For conducting and reporting the review, the authors have followed the guidelines for scoping reviews from the Preferred Reporting Items for Systematic Reviews and Meta-

Analyses, Extension for Scoping Reviews (PRISMA-ScR; Tricco et al., 2018).

Search and Selection Process

Search Terms and Sources

Multiple systematic searches were performed with the help of a university information specialist. Engagement of an information specialist to guide a systematic literature search is associated with significantly higher quality of reported search strategies (Rethlefsen et al., 2015). We conducted a computer-based search of the psychological and medical electronic literature databases, including Medline, Academic Search Complete, CINAHL, Cochrane Library, Web of Science, Wiley Online Library, SpringerLink, PiCarta, Academic Search Premier, ScienceDirect, PsycINFO and Google Scholar. Appropriate key words were identified through exploring the literature on "stress assessment", "stress questionnaires", & "stress measures".

Many previous studies have examined the relationship between state anxiety outcomes and physiological stress-related outcomes (e.g., de Witte, Spruit, et al., 2020; de Witte, Pinho, et al., 2020; Hook et al, 2008; Koelsch, Fuermetz, et al., 2011) and defined state anxiety as a stress-related emotional state (de Witte, Spruit et al., 2020; de Witte, Pinho et al., 2020; Lazarus, 1966; Meijer, 2001; Yang et al., 2011). We have therefore included state anxiety as a stress-related outcome in our current study. In addition, we note that in the literature, the concepts of stress and state anxiety are used interchangeably (Bradt & Dileo, 2014; Lazarus & Folkman, 1984; Ozer et al., 2013; Pittman & Kridli, 2011; Wetsch et al., 2009).

We then combined multiple search terms related to stress or state anxiety with terms referring to psychological testing. Appendix A contains an exemplary search string used for the PsycINFO database. Searches were limited to publication dates from 1980 to April 2020. This time frame is consistent with the consensus within the literature that research concerning psychological measures of stress and / or state anxiety commenced in the 1980s (e.g., Cohen et al., 1983; Spielberger et al., 1983). In addition to the online databases, forward and backward searches were conducted by screening the reference lists of included studies, visiting a university testing library, & consulting research experts for "grey" literature. The initial search resulted in the screening of a total of 3451 studies and an additional 20 measures from forward and backward searches.

Selection of the Self-Report Stress Measures

To identify the self-report stress measures that fit the aims of the present study, we applied several selection criteria in two different selection steps. The first step concerned the screening of the studies found. Titles and abstracts of all the English-language peer-reviewed studies were screened for relevance, which means they had to include the terms “stress” or “state anxiety” related to psychological measures. Psychological measures that did not purely target general stress or state anxiety or stress in daily life were excluded, such as measures specifically assessing work stress, long term stress, parenting stress, or stress within the context of a specific medical diagnosis. At this stage, studies were also included in cases where the abstracts did not explicitly state whether the scale used was specifically a self-report stress measure, or whether the outcome measure concerned stress or state-anxiety in general or in daily life. Studies on self-report stress measures in non-English languages were excluded. This selection step ultimately resulted in 75 self-report measures assessing stress or state anxiety in adults. This reduced the number of studies to 25, which were then full-text screened by at least one author. The table in Appendix B contains the complete overview of the self-report stress measures that resulted from this step one selection.

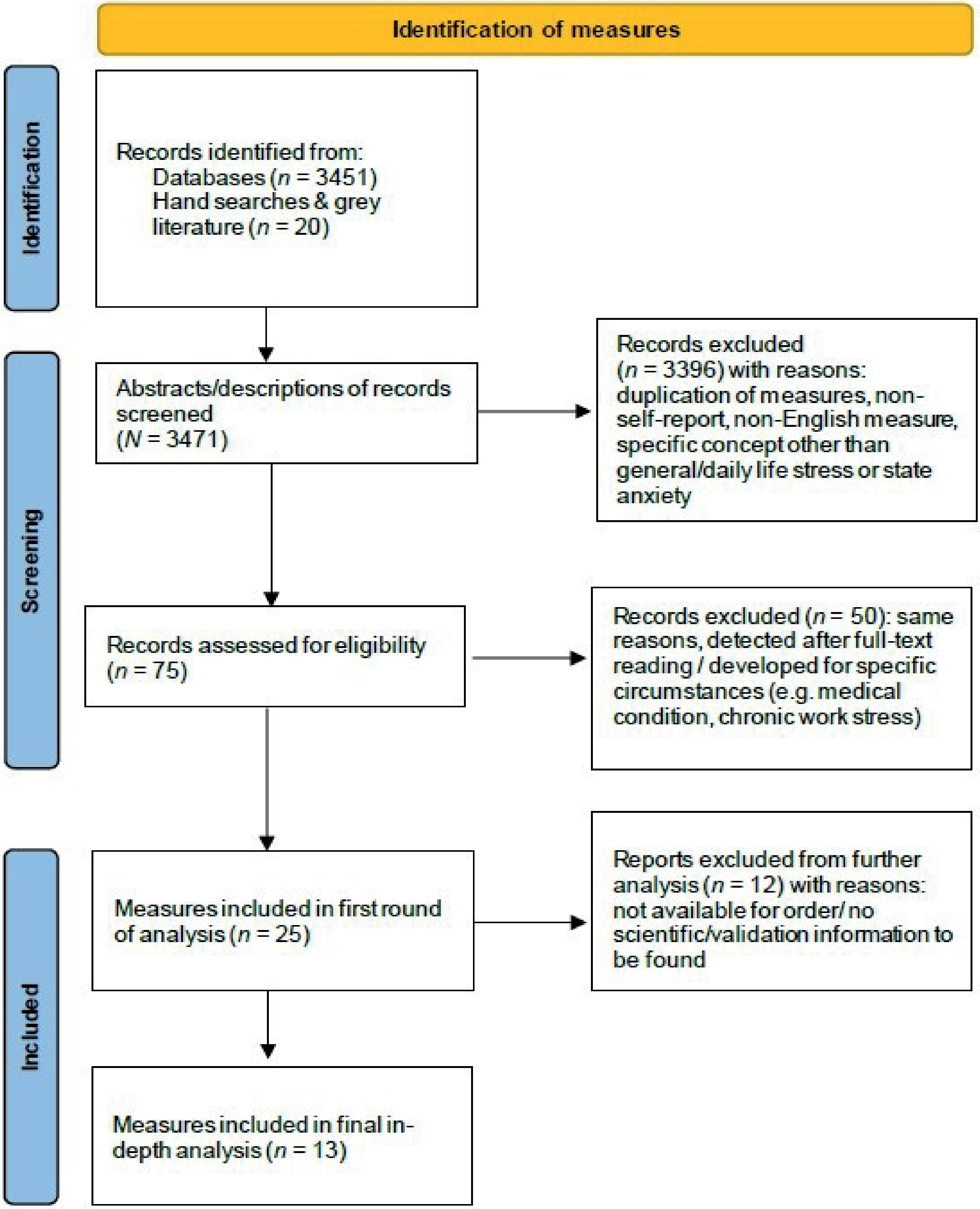
The second selection step concerned the final inclusion of the self-report stress measures. Therefore, we applied the following criteria: instruments had to (1) be available for order in English, (2) have been applied in (clinical) outcome studies published in peer-reviewed scientific journals and, (3) instructions for assessment of the instrument are available. This selection step was performed by the first three authors (MdW, RK and MH) independently. Discrepancies were resolved through discussion. This resulted in consensus on the inclusion of 13 self-report measures for further analysis (see Figure 1: Flow Diagram of the Selection Process).

Evaluation of Included Self-Report Stress Measures

To provide insights into the characteristics and quality of the included self-report stress measures, criteria were formulated to describe their properties. Instrument characteristics relating to the criteria were found in the actual self-report stress measure itself, the user manual, validation studies, & other publications about the self-report measure in peer-reviewed and grey literature. The criteria applied to (1) the psychometric quality of the measure, (2) the assessment procedure of the self-report stress measure and, (3) the suitability for adults with MID. A further definition of the assessment criteria is presented below. Outcomes that relate to each criterion are presented in Table 1 in the Results section for all instruments.

Figure 1.

Flow Diagram of the Selection Process



Psychometric Quality

Reliability and validity are considered the main measurement properties of outcome measures used in clinical practice and research (Frost et al., 2007).

Reliability. A reliable measure is one that measures a construct consistently across time, individuals, & situations. When defining the psychometric quality of measures, three indicators of reliability are generally considered: test-retest reliability (stability over time), internal consistency (coherence of items with the concepts under study), & interrater reliability (equivalence across different researchers or assessors; Salmonds, 2008). Assessing test-retest reliability is typically done by computing Pearson's r. A Pearson's r of .70 or above indicates acceptable alternate-forms reliability (Chiang et al., 2015).

For internal consistency, Cronbach's α is most often reported. An α greater than or equal to .70 is generally considered adequate, & a value of α greater than or equal to .80 is generally considered an indicator of good internal consistency (Allen et al., 2010; Chiang et al., 2015). Interrater reliability concerns the extent to which the different observers are consistent in their judgements. Interrater reliability is often reported as Cronbach's α. For each included self-report measure, we reported the published internal consistency coefficients (Cronbach's α). Manuals were investigated for clear instructions regarding the interpretation of test scores to support objectivity (Moosbrugger & Kelava, 2012).

Validity. The term validity refers to the property of an instrument to measure exactly what it proposes. The main criteria and statistical tests for the assessment of validity are used to determine the content, criterion and construct validity of a measure (Frost et al., 2007). Content validity is evaluated to determine whether the instrument items were generated in accordance with relevant theory. To determine the content validity of the self-report measures, it is important that the self-report stress measure contains a clear description of the measuring construct; all terms related to the target group and outcome measure(s) have to be operationalized. We reported whether the self-report stress measure operationalized the key terms appropriately, such as a description of the characteristics of the type of stress measured, & whether the distinction between stress exposition and stress reaction was described (Chiang et al., 2015; Harkness & Monroe, 2016). Moreover, to provide more insights in the validity of the included self-report measures, we also refer to independent validation research and / or assessments by test commissions. Criterion validity refers to the extent to which the measure agrees with an external standard measure. In the case of stress measurement, the outcomes of psychological self-report measures can be, for instance, compared to physiological measures related to stress responses.

Another relevant form of validity concerns construct validity, which refers to the extent to which scores on a measure correlate with the results of a different test. Concurrent validity is a form of construct validity that determines if the measure correlates highly with an established or widely used test already considered valid (the ‘gold standard’). If there is a high correlation, this gives a good indication that the test measures what is intended. Alternatively, measures that should not be related, should demonstrate low correlations, therefore providing evidence for discriminant validity of the measure.

Assessment Procedure

In addition to its psychometric robustness, the suitability for a stress measure for practical and research purposes can be defined by a number of practical and procedural attributes of the instrument. These include the length of the assessment (determined by the number of the items and procedure), the presentation format (paper/pencil, digital, oral), the role of the assessor (group, guided or individual assessment), & the intended population. These attributes define the context and organizational prerequisites for administration and whether it should be stipulated in the manual.

Suitability for Adults with MID

Review of the Literature. One of the main purposes of this review concerned investigating the suitability of the measure for people with MID. After analyzing each stress self-report measure, we performed a literature search to see if any scientific evidence could be found on the use of the self-report measure in populations that included people with MID. The search was performed in Google Scholar. The following search string was used to guide the search: “learning disabilit*” OR “developmental” OR “mental retard*” OR “intellectual dis* AND [self-report measure]. If a reference was made regarding the suitability of the particular self-report measure in people with intellectual disabilities, learning disabilities, or developmental problems, we reported this.

Expert Consultation. As mentioned, adaptations to standard self-report instruments are generally needed to make them suitable for people with MID. As yet, no comprehensive guidance on how to make these adaptations is available (Kooijmans et al., 2022). To be able to provide more information, we consulted experts in the field of MID research and clinical practice. We used purposive sampling to select internationally renowned researchers in the field of intellectual disability research. The sampling frame was devised from a previously conducted systematic review (Kooijmans et al., 2022). This sample was expanded by probing the authors’ network and asking colleagues in the field of ID research to nominate researchers and clinicians they deemed experts on the topic. We then invited 40 experts from the United States, Europe and Australia to complete an online survey. Of these, 13 experts (33%) from four European countries completed the survey. Participants were academic and clinical staff from the United Kingdom, the Netherlands, Belgium and

Germany with considerable experience in working and conducting research with people with (M)ID. See Table 1 for an overview of the characteristics of the participating experts.

Table 1.

Demographic Characteristics of Experts Consulted

Total N = 13	n (%*)
Country of residence	
The Netherlands	6 (46%)
United Kingdom	5 (38%)
Germany	1 (8%)
Belgium	1 (8%)
Current employment	
Academic setting	9 (69%)
Clinical setting	2 (15%)
Joint academic / clinical	2 (15%)
Years of experience working with people with MID	
1-5	2 (15%)
6-10	4 (31%)
11-20	3 (23%)
20 +	4 (31%)

*percentages not adding up to 100% due to rounding differences

In the survey, the experts were asked to reply to open-ended questions on the subject of how to attune self-report measures to the needs and abilities of people with MID. They were asked to forward suggestions that address the content of self-report stress measures, such as language, response options and supportive media, & procedural issues, such as assessment procedures, questionnaire structure and instructions. Thematic analysis was applied to synthesize the results into general recommendations.

The expert consultation on self-report stress measures was carried out within the context of a larger Delphi study on self-report instruments for persons with (M)ID (Kooijmans et al, unpublished manuscript). As part of the assessment of the suitability of the included self-report stress measures for people with MID, we compared the recommendations from the survey with the published information of the self-report stress measures.

Results

A total of 13 stress-related self-report measures met the final inclusion criteria. Nine of these explicitly focus on stress as an outcome and four on state anxiety as an outcome. Of the included self-report stress measures, the Glasgow Anxiety Scale for Intellectual Disabilities (GAS-ID), the Life Inventory (LI), & the Self-Rating Anxiety Scale for Intellectual Disabilities (SAS-ID) were specifically developed for assessing stress in adults with (mild) intellectual disabilities. First, we share our findings of the analysis of the self-report stress measures included purely from the perspective of the literature. We then discuss the findings of experts consulted, & present the integration of both types of data in Table 3.

Included Self-Report Stress Measures

The characteristics of each individual instrument are described for each of the three criteria: psychometric quality, assessment procedure, & suitability for people with MID based on the consultation of experts and the scientific literature. The findings are summarized in Table 2 and described in more detail below for each instrument (in alphabetical order).

Beck Anxiety Inventory (BAI)

The original publication of the BAI dates back to 1988 (Beck et al.) and it is still widely used today. It measures (state) anxiety symptoms and their level of intensity over the past week. It includes 21 items that target both somatic and more cognitive symptoms of state anxiety, for which respondents rate the intensity on a 4-point rating scale ranging from ‘not all’ to ‘severely’. The total score is rated as minimal, mild, moderate or severe (state) anxiety.

Psychometric Quality. The BAI was found to have high internal consistency (average α coefficients across studies = 0.91; Bardhoshi et al., 2016) and adequate test-retest reliability (test–retest reliability = 0.65; Bhardoshi et al., 2016). It demonstrated both convergent validity with related measures of anxiety (other self-report instruments, diaries, clinical ratings; correlation coefficients ranging from 0.24 to 0.81; Bhardoshi et al., 2016) and moderate discriminant validity with other types of psychopathology (e.g. nonsignificant correlations with a measure of OCD symptomatology; Williams et al., 2013; moderate correlations with the Beck Depression Inventory; average r of 0.59 across studies; Bardhoshi et al., 2016). Both exploratory and confirmatory factor analytic studies generally support a two-factor structure in clinical populations. One factor represents cognitive symptoms of anxiety and the other represents somatic symptoms (Wilson et al. 1999).

Assessment Procedures. The BAI can be self-reported or interviewer-administered. Self-report generally takes a maximum of 10 minutes to complete. It can be administered in paper-and-pencil or interview format, but it is also available online.

Suitability for Adults with MID. The factor structure and other psychometric properties of the BAI were examined in a sample of people with MID (N =108; Mean IQ 67.1; Lindsay & Skene, 2007). To ensure that most people in the sample were able to meaningfully complete the BAI, some adaptations were made. The terminology of some of the items was simplified and the four-point response scale was presented in the form of four bar graph histograms of differing sizes. All questions were read aloud to all respondents by the assessor. On the basis of the analyses in their study, Lindsay and Skene (2007) asserted that people with MID appear to use the BAI reliably and consistently, & that the factors emerging from the sample were similar to those from mainstream populations.

Depression Anxiety Stress Scales (DASS)

The DASS (Lovibond & Lovibond, 1995a) measures three emotional states: depression, anxiety and stress. Three subscale scores for each of the emotional states are obtained that can be compared to norms and clinical cut-offs. For the purpose of this review, the properties of the Stress subscale were considered.

Psychometric Quality. High internal consistency coefficients are reported for each of the subscales of the 42-item and the 21-item versions (Cronbach's α of 0.90 to 0.95 for DASS-Stress; Parkitny & McAuley, 2010). Good evidence has been found for the construct validity through factor analyses (Lovibond & Lovibond, 1995a; Crawford & Henry, 2003) and convergent validity for the anxiety subscales of both the long and short versions of the DASS (correlation between DASS and BAI r = 0.81; Lovibond & Lovibond, 1995a), but the properties of the Stress subscale have been evaluated less extensively. Research in clinical populations has demonstrated responsiveness to treatment effects in, among others, psychiatric patients (Lovibond & Lovibond, 1995a; Ng, 2007) and persons with autistic spectrum disorders (Park et al., 2020).

Assessment Procedures. According to the manual (Lovibond & Lovibond, 1995b), completion takes 10 to 20 minutes for the 42-item version that comprises all three subscales. The shorter 21-item version of the DASS (DASS-21) takes 5 to 10 minutes to complete. A respondent indicates to what extent the statements applied to their lives over the past week on a 4-point scale. The DASS can be administered by paper-and-pencil or computer. The paper-and-pencil questionnaires and scoring forms are available at no cost from the developers' website. No specific training is needed to administer and score the DASS. Numerous officially endorsed translations of the DASS are available in many languages.

Suitability for Adults with MID. No empirical studies involving people with MID were found.

Generally, people with MID were excluded from psychometric studies. The developers state that the DASS should not be presumed valid for some subpopulations, including '[persons with] …low literacy…' (Psychology Foundation of Australia [PFA], 2021). This effectively precludes many people with MID from using the DASS.

Derogatis Stress Profile (DSP)

The DSP is a self-report inventory rooted in interactional stress theory (Derogatis, 1987). Assessment of the DSP results in a detailed profile that identifies stressors on an environmental, personality, & emotional level, in interaction with each other. Cumulative scores provide a quantitative overall summary estimate (global stress score) of the respondent's current stress level.

Psychometric Quality. Strong support for the internal consistency (Cronbach's α > 0.80 for all 'stress domains'), reliability (test-retest coefficients > 0.72 for subscales and total scores) and validity of the DSP (by means of factor analyses) is provided in a small clinical sample and a larger non-clinical sample (Derogatis, 1987). A study on the correlation between several associated stress measures, including physiological correlates, yielded some support for the convergent and construct validity of the DSP (Dobkin et al., 1991).

Assessment Procedures. Respondents are asked to rate 77 statements on a 5-point scale ranging from 'not-at-all true of me' to 'extremely true of me'. According to the information provided on the developer's website, 'the scale takes approximately 12 to 13 minutes to complete under normal conditions, although some individuals may require a few minutes longer.' (Derogatis Testing, 2021).

Suitability for Adults with MID. No empirical studies addressing the suitability of the DSP for people with MID were found. The number of items and the complexity of the measure suggest that assessment may be a challenge for most people with MID (Hartley & McLean, 2006; Bell et al., 2018).

Glasgow Anxiety Scale for People with an Intellectual Disability (GAS-ID)

The GAS-ID (Mindham & Espie, 2003) was specifically developed for people with (M)ID to provide a reliable measure of state anxiety. It targets cognitive and emotional symptoms of state anxiety in the past week, as well as physiological symptoms that are assessed in the here and now.

Psychometric Quality. The GAS-ID showed sufficient methodological quality and excellent reliability (Cronbach's α = 0.96; test-retest r = 0.95) and validity results (ρ correlation coefficient of 0.75 with the BAI; ρ = 0.52 with pulse rate) as reported by the developers themselves (Mindham & Espie, 2003). However, only one external validation study was found (Hermans et al., 2013); the authors concluded that the GAS-ID can be regarded as a reliable self-report measure. High Cronbach's α's (>0.80) and test-retest ICC (0.89) were reported, & the GAS-ID showed satisfactory correlations with related measures (correlation with the HADS-A of r = 0.61).

Assessment Procedures. No manual is available for the GAS-ID. The assessment time is reported to be 5–10 minutes (Mindham & Espie, 2003). The questionnaire is administered as a structured interview. Respondents are asked to rate how often they experienced 27 expressions of fears, worries and physiological symptoms in the past week on a 3-point answer scale (from 'never' to 'always'). Furthermore, respondents are asked whether they experience any physiological symptoms associated with state anxiety in the here and now. Clinical cut-off scores are proposed by Mindham and Espie (2003), but they state that more research is needed.

Suitability for Adults with MID. The GAS-ID is designed specifically to be administered to people with MID. In the process of development, several alternative versions were tested for optimum suitability for people with MID. The resulting measure is perceived by the authors as being suitable for use with those people with MID who demonstrate sufficient ability to communicate verbally in day-to-day interactions (Mindham & Espie, 2003). The GAS-ID is frequently used in research on stress and anxiety with people with MID (e.g. Hartley & MacLean, 2008), is referenced as a preferred diagnostic tool in clinical guidelines for people with MID (e.g. Davis et al., 2008), & is mentioned in several textbooks on diagnostics and treatment of people with (M)ID (e.g. Stavrakaki & Lunsky, 2007).

Index of Clinical Stress (ICS)

The ICS (Abell, 1991) is a self-report questionnaire for individuals older than 12 years. It measures the degree or magnitude of clients' perceptions of personal stress, which is defined by a '... perceived imbalance between the demands of daily living and a person's ability to respond.' The ICS is part of the Walmyr Assessment Scales (WAS), a compendium of more than 25 short-form measurement scales designed for use in assessing the severity or magnitude of a variety of personal and social problems (Walmyr Publishing Company, 2021).

Psychometric Quality. Psychometric evaluation studies were conducted by the developer or researchers affiliated to the WAS (Abell, 1991; Hudson et al., 1995). High Cronbach's α's of .96 (Abell, 1991) and .90 (Hudson et al., 1995) were reported. Evidence for convergent validity was demonstrated by means of significant correlations with associated constructs (mean r =.48) and nonsignificant correlations with discriminant factors (mean r =.08).

Assessment procedures. The respondent is required to respond to the 25 items on the test form by selecting one response from a 7-point scale ranging from ‘none of the time’ to ‘all of the time’. The respondent is expected to fill in the questionnaire unassisted. The WAS manual details no administration times, but is reported to be ‘rapid’. The ICS is available in paper-and-pencil form and can be administered digitally through the publisher’s own digital administration application.

Suitability for Adults with MID. The manual states that those completing the questionnaire must be literate and have no severe cognitive impairment. Readability statistics for the measure are given. The Flesch-Kincaid Grade Level of four suggests that a fourth level reading grade is required to complete the form autonomously. As the majority of people with MID are unable to read beyond grade three level (Conners, 2003), autonomous completion of the ICS may be challenging for many. However, the ICS was developed for individuals from the age of 12 years upwards, meaning that the level of understanding may be appropriate for some people with MID.

Lifestress Inventory (LI). The LI (Fogarty et al., 1997) is a self-report questionnaire designed to measure frequency and impact of stressors in daily life. It was developed specifically for people with MID as an update of the Subjective Stress Scale (SSS) that is no longer available.

Psychometric Quality. In three studies, none of which were conducted by independent authors, the psychometric quality was found to be sufficient (Fogarty et al., 1997; Bramston et al., 1999; Lunsky & Bramston, 2006). For internal consistency, Lunsky & Bramston (2006) found Cronbach’s α to equal .80. In the same study, some evidence was presented for the convergent validity of the LI, by showing significant correlations with related measures (r =.64 to .78). Modest correlations were presented between self-report and informant measures (r =.34 to .70). According to Fogarty et al. (1997), confirmatory factor analysis indicated three underlying factors that impact the experience of stress in daily life. These factors were labeled General Worry, Negative Interpersonal Relations, & Coping.

Assessment procedures. According to the scoring instructions / manual provided by the authors (Bramston & Fogarty, n.d.), the 30 items of the LI are intended to be read aloud. A series of buckets from empty to full can be used as a visual representation for the response options to facilitate understanding. Other possibilities to ensure that an item is understood correctly include repeating or re-wording a question, as well as asking the respondent to elaborate on their answer to make sure they interpreted the question correctly. As an extra response option, “0” indicates that an item/event was not experienced by the respondent; this option helps establish a frequency score. The other response options - from 1 (“no stress”) to 4 (“a great deal of stress”) - indicate the impact of single stressors. Assessment is preferably completed by an experienced psychologist.

Suitability for Adults with MID. The LI has been specifically developed for people with MID and research into validation has been, as quoted above, carried out with people with MID. Notably, the LI was developed by means of focus groups with people with MID and staff members, & was designed to be easily understood and completed by people with MID (Scott & Havercamp, 2018).

Perceived Stress Scale (PSS)

The PSS (Cohen et al., 1983) has become one of the most widely used psychological instruments to measure the degree to which situations in people's lives are appraised as stressful. Cohen et al. (1983) define psychological stress as the extent to which a person perceives that demands exceed his/her ability to cope.

Psychometric Quality. Although scores on the 14-item PSS exhibit good reliability estimates across the literature, four of the items tend to perform poorly when evaluated using exploratory factor analysis (Cohen & Williamson, 1988; Lee, 2012). As a result, the PSS is commonly implemented using the 10-item form. In the review of Lee (2012) on the psychometric qualities of the PSS, it is shown that all included studies (N = 19) reported α coefficients of >.70. The test-retest reliability of the PSS-10 was assessed in four studies, & met the criterion of >.70 in all cases. The PSS correlated significantly and predictably with a range of other measures of stress and pathology (correlations typically in the 0.30 - 0.70 range), such as the Job Responsibilities Scale, HADS and STAI. Additionally, higher PSS scores have been shown to be associated with higher levels of cortisol; a biological indicator of stress (van Eck & Nicolson, 1994).

Assessment Procedures. The PSS is available in a 14 and 10-item form and the average completion time is 5-10 minutes. Items are designed to tap how unpredictable, uncontrollable, & overloaded respondents generally find their lives. The scale also includes a number of direct queries about current levels of experienced stress.

Suitability for Adults with MID. The PSS is designed for use in community samples for those with at least a junior high school education. Although there is no information available on the use of the PSS in adults with MID, some research has been carried out with younger college students with disabilities, such as learning disabilities, ADHD, & autism spectrum disorders (Janusis & Weyandt, 2010). The students with disabilities tended to score higher on the PSS, but the differences did not approach significance. No separate norms for people with disabilities were constructed on the basis of this study.

Perceived Stress Questionnaire (PSQ)

The PSQ (Levenstein et al., 1993) measures the experience or perception of stress by the individual during stressful situations, & is considered valid in the context of a transactional model of stress (Kocalevent et al., 2007). The PSQ was developed for use within the field of clinical psychosomatic research (Levenstein et al., 1993, 2000). There are two forms of the PSQ: the “general“ (the last two years) and the “recent“ (during the last four weeks) form.

Psychometric Quality. The original authors developed the instrument in English and Italian and validated it among 230 subjects (Levenstein et al., 1993). Internal consistency of the original English version (measured by Cronbach’s α) ranges from 0.80 to 0.86 (Kocalevent et al., 2007; Levenstein et al., 1993), & research on test-retest reliability (Pearson correlation coefficients r between 0.80 and 0.86; Levenstein et al., 1993; Sanz-Carillo et al., 2002). The PSQ shows positive associations with compatible self-report measures such as Cohen’s (1983) Perceived Stress Scale (PSS; r = 0.73; Levenstein et al., 1993). Notably, there are some indications that PSQ scores seem to differ between populations of patients and healthy individuals, & that scores seem to be sensitive to change after treatment (Fliege et al., 2005).

Assessment procedures. The PSQ has 30 items. Response options and items of both the PSQ-General (past 1 to 2 years) and the PSQ-recent (past month) are identical. Respondents are asked to estimate how often they deal with stress-related experiences on a 4-point Likert scale. While no extensive manual has been published, free scoring instructions are available to researchers. The administration time is expected to be 5 minutes. Translations along with validation studies are available in Swedish (Rönnlund et al., 2015), Norwegian (Østerås et al., 2018), Spanish, Chinese and German. The instrument is available at no cost under a Creative Commons license.

Suitability for Adults with MID. No information on the suitability for people with MID has been found in previous empirical studies. The PSQ was originally intended for adults, but has also been successfully validated for adolescents aged 15-16 years (Østerås et al., 2018). Mutz and Müller (2016) used the PSQ to assess 14-year-old German upper secondary school pupils, without commenting on the application of the instrument to the target group. The adolescent research projects indicate that research about the usefulness of the instrument for (some) people with MID can be recommended.

Psychological Stress Measure (PSM-9)

The PSM was first published in 1988 (Lemyre & Tessier, 1988) and updated in 2003. The PSM-9 is an abridged nine-item version of the original 49-item assessment of self-reported state stress. Respondents are asked to rate stress symptoms they experienced in the past three to four days on an 8-point Likert scale (from 'not at all' to 'extremely'). The result is a single-factor indicator of perceived stress.

Psychometric Quality. The authors report a wide range of reliability (Cronbach's α's > 0.90; test-retest r's 0.68 - 0.80) and validity coefficients for the 49-item version in a series of publications by the developers of the instrument (Lemyre & Tessier, 1988; Lemyre et al., 1990; Lemyre & Tessier, 2003; Lemyre, Chair, & Lalande-Markon, 2009). The psychometric properties of the short PSM-9 version are reported to be 'the same as the original version' (Lemyre & Tessier, 2003), but only a Cronbach's α of 0.89 is reported for the PSM-9. No external validation studies have been published.

Assessment procedures. The PSM-9 appears to be a short single-page paper-and-pencil questionnaire. No scoring instructions could be retrieved. Only a French version of the manual was published (Lemyre et al., 1990), but it could not be retrieved by the reviewers.

Suitability for Adults with ID. No evidence was found that the PSM-9 would be suitable for people with MID.

Self-Rating Anxiety Scale for Intellectual Disabilities (SAS-ID)

The SAS-ID is an adaptation of the Zung Self-Rating Anxiety Scale for persons with ID by Lindsay and Michie (1988). The SAS is a 20-item self-report assessment instrument for measuring state anxiety. Respondents are asked to indicate to what extent a series of statements apply to themselves within a period of one or two weeks prior to assessment. A total score reflects a general level of state anxiety as experienced by the respondent.

Psychometric Quality. Several researchers have assessed the psychometric quality of the SAS-ID (Lindsay et al., 1994; Masi et al., 2002; Ramirez & Lukenbill, 2008). Psychometric evaluation was conducted by independent researchers and those affiliated to the original developers. Internal consistency coefficients (Cronbach's α) averaged a satisfactory 0.80. Convergent validity was established by finding significant correlations between the SAS-ID and related self-report instruments and diagnostic interviews (correlation coefficients ranging from 0.33 to 0.73).

Assessment procedures. The SAS-ID is a 20-item scale with a yes–no response format. It takes 5-10 minutes to complete. The SAS-ID is presented to respondents orally on an individual basis. Assessors are instructed to rephrase or reword the questions if the respondents appear to lack understanding.

Suitability for Adults with ID. The SAS-ID is an adaptation of the original SAS that is intended for use in the general population. Adaptations are made to ensure that most people with MID are able to meaningfully complete the assessment with assistance.

Adaptations made to the original are the use of a yes-no response format instead of a 4-point Likert-type scale and the rewording of items perceived to be difficult. The SAS-ID has occasionally been used in research involving people with MID (e.g. Carraro & Gobbi, 2012) and is mentioned in textbooks on diagnostics and treatment of persons with (M)ID (e.g. Vargas-Vargas et al., 2019; Hatton & Taylor, 2013).

State-Trait Anxiety Inventory, State Version (STAI-S)

The state version of the STAI (Spielberger, 1983) is one of the most long-standing and commonly used clinical self-rating scales to measure state-anxiety, which is defined as a temporal cross section in a person's emotional stream of life, consisting of subjective feelings of stress, tension, apprehension, nervousness, worry, & activation of the autonomic nervous system (Cattell & Scheier, 1961; Spielberger, 1983). In research, the 20-item STAI subscale is often used to measure state-anxiety before and after an intervention or task. Translated forms of the STAI are now available in more than 60 languages (Spielberger & Reheiser, 2009)

Psychometric Quality. Many psychometric evaluation studies have been published which show that the STAI-S provides excellent psychometric properties: the internal consistency measured using Cronbach's α coefficient ranges from good to excellent (i.e. > 0.70) across several populations (e.g., Creamer et al., 1995; Fonseca Pedrero et al, 2012; Ortuno-Sierra et al., 2016; Spielberger, 1983). Noteworthy, α coefficients are typically higher for the STAI-S when state anxiety is assessed under conditions of psychological stress (Spielberger, 1983; Spielberger & Reheiser, 2009).

Assessment procedures. The STAI-S is a 20-item self-rating inventory which may be given either individually or to groups. The scale is composed of short verbal statements that participants have to rate using a 4-point Likert scale according to the subjective experienced intensity of each described feeling (1 = not at all, 4 = very much so). It is clear that the questionnaire's ease of administration, as well as the simple and straightforward scoring procedure have led many researchers to use this specific instrument (Rossi & Pourtois, 2012).

Suitability for Adults with ID. Although no studies have been published on the applicability of the STAI-S in persons with (M)ID, a STAI child-version (STAI-C) has been developed (Spielberger, 1973), especially constructed for 9 to 12-year old children. The STAI-C manual states that the scale may also be used with older children/adolescents who are below average in ability. In future research, the appropriateness of the STAI-C version for use in people with MID should be investigated.

Stress Arousal Checklist (SACL)

The SACL (Mackay et al., 1978) is a list of mood adjectives intended to measure stress experience as well as arousal. The authors refer back to work by Thayer (1967) and his factor analysis of the Activation-Deactivation Adjective List (AD-ACL). The two-dimensional structure of stress and arousal is explained as follows: “The stress dimension refers to the perceived favorability of the external environment, while arousal refers to ongoing autonomic and somatic activity” (Cox & Mackay, 1985).

Psychometric Quality. In an independent factor analysis, the two-factor structure found by the original authors has been replicated (McCormick et al., 1985). This study also supports the two-dimensional model of stress and arousal operationalized in the SACL. Reliability was found to be relatively high in several studies (> 0.70), especially for the stress scale, while α coefficients showed more variance for the arousal scale (Watts et al., 1983). Evidence for the construct validity of the SACL was found in factor analyses (Fischer & Donatelli, 1987; Fischer et al., 1988; King et al., 1983). However, Hinton et al. (1991) stated that in their view, the stress scale of the SACL does not measure stress as defined by the authors and “is virtually identical to the state version of the STAI”.

Assessment Procedures. There does not seem to be a published manual, but the authors provide scoring instructions and note that “scoring keys are easily made” (Cox & Mackay, 1978, p.284). The 30-item list consists of positive and negative adjectives, for each of which the symbols “++”, “+”, “?” or “-” can be circled by respondents. Responses can be summed up separately for the ‘stress’ and ‘arousal’ subscales (Cox & Mackay, 1978, p.284).

Suitability for Adults with MID. No empirical evidence was found for the suitability of the SACL for people with MID.

Stress Overload Scale (SOS)

The SOS (Amirkhan, 2012) is designed to measure “stress overload”, a state described in stress theories as occurring when demands overwhelm resources. Respondents are asked to answer 30 questions and reflect on the occurrence of stress-related feelings and cognitions in the past week. Total scale scores and scores on two subscales - Personal Vulnerability and Event Load - are calculated. A short 10-item version (the SOS-S) is also available (Amirkhan, 2018).

Psychometric Quality. All psychometric evaluation studies were conducted by the developers (Amirkhan, 2012; Amirkhan et al., 2015; Amirkhan, 2018). They report an excellent internal consistency of the SOS (with Cronbach’s α’s > 0.94 for both subscales and the measure as a whole). Test-retest coefficients averaged 0.75 over a one week period. Convergent validity has been demonstrated in significant correlations with other measures of stress (e.g. correlation coefficient r of 0.45 with the PSS-10) and criterion validity has been shown in the SOS's ability to predict illness following a stressful event. Psychometric properties for the original and short versions are all but identical.

Assessment procedures. Participants are asked to rate feelings and cognitions related to life stress on a 5-point Likert scale (from 'not at all' to 'a lot'). No information on the duration of the assessment of the original or short forms has been published and no manual is available. Scoring instructions are attached to the form.

Suitability for Adults with ID. The development and validation of the SOS made use of community samples. Some attention was paid to make sure that '… Only items that were consistently understood across [a] wide socioeconomic and ethnic spectrum were chosen for the SOS' (Amirkhan, 2012). However, its comprehensibility and general usefulness for people with MID has not yet been demonstrated.

Table 2.

Included Self-Report Stress Measures

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach's α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Beck Anxiety Inventory (BAI), Beck et al. (1988).	State anxiety	Adults	a) yes b) >10 c) average across studies = 0.91	a) 21 items b) 4-point Likert scale c) 10 minutes max.	Yes Adaptation for MID by Lindsey & Skene (2006; 2009)
Depression Anxiety Scales (DASS) (Lovibond & Lovibond, 1995)	Stress + state anxiety	Adults	a) yes b) >20 c) α of 0.84 to 0.92 for DASS-Anxiety, & 0.90 to 0.95 for DASS-Stress	a) 21 short form, 42 regular form b) 4-point Likert scale c) 5-10 minutes short form, 10-20 minutes long form (all 3 subscales)	No Parkitny & McAuley (2010): "certain patient groups (eg ... the developmentally delayed...) may have difficulty understanding the questionnaire items or responding to them in an unbiased manner.

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach's α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Derogatis Stress Profile (DSP) (Derogatis, 1980)	Stress	Adults	a) yes (only 1 study found) b) 2 c) α between 0.83 and 0.88 for different domains	a) 77 items b) 5-point Likert scale + VAS (0-100) for subjective stress experience c) 12-15 minutes	No
Glasgow Anxiety Scale (GAS-ID) (Mindham & Espie, 2003)	State anxiety	"People with an intellectual disability" (age/level not specified)	a) yes (only 1!) b) 2 c) > 0.80	a) 27 b) 3-point Likert scale of frequency c) 5-10 minutes	Yes

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach's α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Index of Clinical Stress (ICS) (Abell, 1991)	Stress	Adults & youths age 12+; Reading level > grade 4	a) no b) 2 (developer + affiliated researchers) c) 0.96 (Abel, 1991)/0.90 (Hudson,1997)	a) 25 b) 7-point Likert scale of frequency c) not specified, but stated as "rapid"'.	Yes Manual: "Persons who are only mildly impaired might be able to complete the WAS scales with considerable accuracy. The major things to watch for are the literacy skills, cognitive development, & ability to integrate affective responses with the item content and meaning of each of the scales." Flesch reading ease: 89 (6th grade level); Gunning Fog Index: 6 (sixth grade level); Flesch-Kincaid Grade Level: 4.

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach’s α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Lifestress Inventory (LI) (Bramston & Fogarty, 1997)	Stress	Age not specified; "suitable for administration to a wide range of people, including the mildly intellectually handicapped”	a) no b) 3 c) 0.80	a) 30 b) 4-point Likert scale + visual aid showing a series of buckets empty through to full can be used to improve understanding of the Likert-type options c) not specified	Yes
Psychological Stress Measure (PSM-9) (Lemyre & Tessier, 1988)	Stress	Adults	a) no b) 4 c) 0.89	a) 9 b) 8-point Likert scale c) not specified	No

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach’s α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Perceived Stress Questionnaire (PSQ) (Levenstein et al., 1993)	Stress	Adults	a) yes b) 6 c) ranging from 0.90 to 0.93	a) 30 regular, 20 short form b) 4-point Likert scale c) 5 minutes	No
Perceived Stress Scale (PSS) (Cohen et al., 1983)	Stress	Adults ("community samples with at least a junior high education"; “accessible to any subpopulation”)	a) yes b) 19 or more c) above 0.70 across studies	a) 10 b) 5-point Likert scale c) 5-10 minutes	No

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach's α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
Stress Arousal Checklist (SACL) (Cox & Mackay, 1978)	Stress	Not specified	a) yes b) 6 or more c) for stress scale ranging from 0.81 to 0.86, lower for arousal scale	a) 30 b) 4-point Likert scale c) not specified	No
Self-Rating Anxiety Scale for adults with Intellectual Disabilities (SAS-ID). (Lindsay & Michie,1988) Adaptation of SAS (Zung, 1971)	State anxiety	People with an intellectual disability (age/level of ID not specified)	a) yes b) 3 or more c) average of 0.80 across studies	a) 20 b) yes-no answer format. c) 5-10 minutes	Yes Adaptations from the original instrument include yes–no response format, rewording of the items, & addition of supplementary items.

Title (author/s, publication date)	**Outcome**	**Target group**	**Psychometric quality:** (a) independent validation research available? (b) number of available validating studies (approximately) (c) internal validity (Cronbach's α)*	**Design**: a) number of items b) response options c) duration of administration	**Published information on applicability with people with MID available?**
SOS Stress Overload Scale (Amirkhan, 2012)	Stress	Adults	a) no b) 3 c) 0.94	a) 30 or 10-item short form b) 5-point Likert scale c) not specified	No
STAI State-Trait Anxiety Inventory (Spielberger, 1981)	State anxiety	Adults (used in ID research)	a) yes b) lots of studies in many different languages c) ranges from good to excellent across several populations	a) 20 items for the State scale (Y1 form) b) 4-point Likert scale . c) approx. 10 minutes for "less educated or emotionally disturbed persons".	Yes: Manual specifies some instructions for the assessor in the case of guided assessment. Norms are based on a sample of 'working persons', which generally will include relatively few persons with MID. The STAI is frequently used in research with persons with MID in unaltered form.

Note. References and additional psychometric results are provided in detailed descriptions of the single measures in the result section.

Results of the Expert Consultation

The experts were asked to reply to open-ended questions on the subject of how to attune self-report measures to the needs and abilities of people with MID. They unequivocally indicated that the factors that improve appropriate use by people with MID in general also apply to the self-reported measurement of stress. Thematic analysis of the answers revealed six general recommendations relevant to the measurement of stress in people with MID.

The first recommendation was to use concrete and easy-to-understand vocabulary, simple grammar, & short sentences. The next was to use relatively short time frames for the retrieval of information. Assessors should not ask to retrieve information over longer periods than one week, as time processing abilities are generally impaired. A third recommendation relates to the use of Likert scales. When designing self-report measures for people with MID, the number of response options in Likert scales should be limited to three for people with moderate ID to MID and five to people with MID to borderline intellectual functioning. Fourth, an 'I don't know' option should be included in both forced-response and open-ended questions to prevent invalid answers from those who do not understand the question. A fifth recommendation was to use visualizations to support the meaning of questions and responses, although how exactly these should be configured was not specified. In regard to the assessment procedures, a sixth recommendation was to use pre-scripted alternative wording if the respondent seems unable to understand the question. Standardization ensures comparability of scores across assessments. The extent to which these factors were reflected in the self-report measures' design and assessment procedures differed across the included instruments. An overview of the suitability of each self-report stress measure for people with MID, according to the experts, is presented in Table 3.

Discussion

The need to measure the degree of stress as accurately as possible in people with MID is reflected in both the literature reviewed and the information of the consulted experts. This can be seen as a response to the fact that people with MID are much more vulnerable to stress (Hatton & Emerson, 2004; Scott & Havercamp, 2014). Persistent stress experiences in people with MID may lead to more impaired information processing (Heyman & Hauser-Cram, 2015) which will adversely affect coping skills. Our study not only provides the first overview and analysis of self-report stress measures, but also provides more insights in how self-report stress measures can be adequately attuned to the needs of people with MID. Of the 13 self-report stress measures found, three measures were specifically designed for use with adults with (M)ID. Five did not appear to be suitable for populations other than the normally gifted population, and the final five could potentially be suitable when applied with consideration.

Table 3.

Factors that Determine the Suitability of Included Self-Report Stress Measures for People with MID According to the Expert Consultation

	use of easy-to-under-stand language	max. 1-week time frame	max. 5 answer options*	"I don't know" answer option	use of visual support	scripted alternative wording
BAI	✓ (adaptation by Lindsay & Skene, 2007)	✓	✓	X	✓ (adaptation by Lindsay & Skene, 2007)	X
DASS	X	✓	✓	X	X	X
DSP	X	✓	✓	X	X	X
GAS-ID	✓	✓	✓	X	X	X
ICS	✓	✓	X	X	X	X
LI	✓	/	✓	X	✓	X
PSM-9	/	✓	X	X	X	X
PSQ	/	X	✓	X	X	X
PSS	/	X	✓	X	X	X
SACL	/	✓	✓	X	X	X
SAS-ID	✓	✓	✓	X	X	X
SOS	/	✓	✓	X	X	X
STAI-S	/	✓	✓	X	X	X

Main Findings

The Lifestress Inventory (LI) was specifically designed for the MID population. Two others, the Glasgow Anxiety Scale for people with Intellectual Disability (GAS-ID) and the Self-rating Anxiety Scale for adults with Intellectual Disabilities (SAS-ID) reported that they were fit for use with people with ID, but the user manuals did not specify the exact intelligence range. As the items concerned mainly refer to insights, feelings and experiences from daily life, participants must be able to grasp these abstract concepts, translate them to their everyday experiences and formulate a meaningful response. This suggests that they are targeted towards adults with MID instead of the total ID population. Generally, these three self-report stress measures have in common that they use items that require a response on simple Likert scales, which could possibly be combined with visual representations of answer alternatives. This is in line with findings reported in previous studies as well as the expert consultations in our study, which show agreement that responses requiring a simple Likert rating scale or only yes/no can lead to appropriate responses from individuals with MID (Hartley & MacLean 2006; Heal & Sigelman 1995; Ramirez 2005). For those individuals in the lower range of MID, pictorial representations of response alternatives could increase the likelihood of gaining appropriate responses (Hartley & MacLean, 2006), which was echoed by the experts consulted.

Our findings also show that some of the other stress self-report measures seem to be more or less suitable for adults with MID. First, some evidence was provided in previous validity studies on populations in which participants with intellectual, learning, or developmental disabilities were also included. This applies to the BAI (see Lindsay & Skene, 2007), the DASS (see PFA, 2021), & the PSS (see Janusis & Weyandt, 2010). Second, other self-report stress measures stated that they could also be used in younger aged populations, which may suggest that, at least in terms of comprehensibility, they may be suitable for people with MID. This applies to the STAI-child version (9-12 years), the ICS (from 12 years), & the PSQ (from 14 years). Hurley (2008) suggests that the use of instruments designed for children may offer a useful basis for adaptation, because the measures use concrete levels of vocabulary and simple sentence structures. This process has also been used by many other researchers (e.g. Guerin et al., 2009; Marshall & Willoughby-Booth, 2007). However, since these stress self-report measures have not been validated specifically for the adult MID population, we recommend thoroughly screening the measurement construct and assessment procedure before using them in clinical practice or in future research (Kooijmans et al., 2022).

The findings from the expert consultations show the importance of adding an extra "I don't know" answer alternative to prevent people with MID who do not understand the question from filling in a random answer (Bell et al., 2018). However, none of the self-report stress measures, even those specifically developed for people with MID, included this option. The Lifestress Inventory (LI) added the answer alternative "actually not experienced", but this refers to the fact that the participant did not experience any stress at all. In addition, none of the self-report measures included "alternative wording" to the questions and/or answer alternatives, as recommended by the expert panel. On the other hand, helping factors such as allowing assessment assistance (SAS-ID) or having someone else read the items (LI and GAS-ID) were not mentioned by any of the experts. Finally, response visualizations seem to be missing from both the GAS-ID and SAS-ID. This is remarkable, as this is considered one of the most important factors with regard to suitability for people with MID, both in the literature and by the experts consulted (e.g. Hartley & MacLean, 2006; Scott & Havercamp, 2018). Of the three self-report stress measures for people with MID, the LI appears to be most consistent with the findings of the experts. However, our findings show that in addition to consulting experts, screening the assessment procedures of existing self-report measures specifically adapted or designed for people with (M)ID is a worthwhile exercise.

The Added Value of Self-Reported Information

Although proxy reports are commonly used in MID, self-report measures prove to be more accurate and more sensitive, even in the MID population (Scott & Havercamp, 2018; Moss et al., 1996). The importance of obtaining self-reported information on subjective stress experiences of people with MID is also reflected in the increased recognition in our society that people with (M)ID are full citizens with the same rights as non-disabled persons, meaning that participation and social inclusion should dominate all organized activities (e.g., Devi, 2014; Giesbers et al., 2019). In other words, including the opinions, feelings, & thoughts of people with MID by using self-report measures, fits the call for knowledge democratization, as citizens increasingly demand their say in policies and research affecting them (Anderson, 2017; Dedding et al., 2020). This is important, because self-determination can be seen as an essential dimension of quality of life and is linked to many positive outcomes for people with (M)ID) (Frielink et al., 2018; Schalock & Verdugo, 2002; Wehmeyer, 2007). Therefore, both the findings of this review and the empirical evidence show that increasing our knowledge of self-report stress measures for people with MID is a highly recommended addition and in line with the contemporary opinion that the voice of people with MID should be included in matters that concern them.

Measuring the Concept of Stress

The way the concept of stress was operationalized by the self-report measures varied according to the theoretical underpinnings and constructs. Different paradigms or stress theories were used, such as the interactional stress model (e.g. the DSP or the PSS), theories on stress as a transitory anxiety state (e.g. the STAI and the BAI), & the tripartite model of anxiety and depression that describes stress as a common symptom for both (e.g. the DASS). Moreover, some of the self-report measures do not seem to have origins in a certain stress theory or model, but were developed empirically, involving expert consensus on the manifestation of stress in clinical practice (e.g. the GAS-ID).

Others are based on the manifestation of stress symptoms described in classification systems of psychiatric disorders (e.g. the SAS-ID). In addition, a distinction can also be seen between self-report stress measures that focus mainly on stress as an experienced psychological and physiological state (e.g. the BAI, the SACL and the STAI) and those that focus on the experience of stress in the context of situations that actually or hypothetically cause stress, such as job related stress or stressful social situations (e.g., the LI, the PSS and the PSQ). To ensure that the concepts being studied are consistent with the design and intended use of the self-report measure, we recommend paying attention to how the concept of stress is theoretically framed when deciding to use a self-report stress measure (Cook et al., 2006).

Implications for Clinical Practice

There is a strong tendency in clinical practice to move away from attributing the symptoms of psychopathology solely to the cognitive deficits of people with MID (known as diagnostic overshadowing; Hagopian & Jennett, 2008; Reiss et al., 1982). Clinicians are becoming increasingly sensitive to the fact that people with MID can also suffer from symptoms of psychopathology. Since the degree of stress is now recognized as a significant factor in the development of severe psychopathology, especially in people with MID, it has become more important to correctly observe/assess stress-related states in clinical practice (Scott & Havercamp, 2014). This review therefore provides a practical basis for determining whether and which self-report stress measures are suitable for people with MID within their own clinical context.

To provide some guidance for clinical practice, we have formulated several recommendations based on our findings. First, with this review, we want to draw attention to the concept of stress and the importance for clinical practice to consider the degree of (daily or present) stress as a crucial factor in the quality of life and course of further psychological treatment in people with MID. In our view, stress assessment should be included as a regular part of the diagnostic phase of clients with MID when consulting clinical practice. Second, as mentioned, we strongly advise clinical practice to always strive to obtain self-reported information in addition to proxy-reports when it comes to medical, psychological, & service decisions involving people with MID. Third, we particularly recommend using the three self-report stress measures specifically designed for adults with (M)ID. These self-report measures are characterized by simple Likert rating scales and/or items requiring yes/no responses. Specifically, the use of simpler wording, fewer response options, & the ability to provide supportive visualization are the main differences with the self-report stress measures developed for the non-ID population. Another significant difference is that self-report measures developed for (M)ID often allow the respondent to be assisted during the assessment (SAS-ID) and that the items can be read aloud by someone else (LI and GAS-ID). Although our assessment of the suitability for MID populations show that, even for MID-specific instruments, there is ample room for improvement, these measures remain a clinician's primary choice.

While there is general consensus that it is necessary to timely assess stress in people with MID, we are also aware that this requires experiential knowledge of clinical professionals working with the MID population. The challenge for clinical practice is to prevent that difficult-to-understand behavior of people with MID too quickly leads to a psychiatric classification, which often has far-reaching consequences (Didden et al., 2016). On the other hand, psychological problems still have to be recognized timely. This requires continuous in-depth behavioral observations and careful consideration by clinical professionals, as people with MID, certainly in combination with additional behavioral/psychological problems, often are unable to clearly request help (Ten Wolde et al., 2006). Decisions made should therefore be adequately aligned with personal and environmental circumstances, as well as with the level of cognitive functioning (Nouwens et al., 2020). Determining and applying suitable self-report measures for clients with MID could contribute to this purpose. Moreover, as indicated earlier, the use of self-report measures is also a way of letting the client's voice speak, & thereby enhances feelings of autonomy, initiative and freedom of choice. In this study, we have attempted to provide a first guide with regard to the use of self-report stress measures.

Limitations of the Present Study

There are some limitations of the present study that should be noted. First, because we strictly followed our inclusion criteria, we may have excluded some self-report measures which could be also suitable for assessing stress in people with MID (see Appendix B). For example, they may have not yet been applied in (clinical) outcome studies published in peer-reviewed scientific journals. Another reason for exclusion was that measures were unavailable in the English language; it is possible that suitable measures exist in other languages.

Second, for the appraisal of the psychometric properties of each measure, we had to rely on the parameters reported by authors in their publications. Nearly all studies report Cronbach's alpha as the main indicator of reliability. Recent advances in psychometric research suggest that this may be a flawed indicator of the internal stability or reliability of a measure. It is stated that other indicators, such as omega, are more robust, & that reliability research should be preceded by Factor Analysis (Crutzen & Peters, 2017).

Third, we would have liked to share more specific information from the expert consultations. However, due to the use of an online survey, there was no opportunity to ask further questions. Therefore, for future research, we recommend adding a more interactive form of data collection when consulting experts on similar questions, such as a multidisciplinary focus group method. Another limitation concerning the expert consultations is that the results reflect the participating experts' professional opinion. Although their clinical and research expertise are highly valued, the experts were not asked to substantiate their statements with references to empirical literature. Therefore, the suggestions by the experts must be valued as tentative and supplementary to the evidence from empirical studies. Finally, the experts' findings were only compared with the published information in the user manuals of the self-report measures, i.e. only with the information already described. An option for follow-up research would be to use a more detailed screening list and to screen the individual instruments with different researchers in the field of MID blinded from each other. This would ensure more accurate statements about the use of existing self-report stress measures in people with MID.

Implications for Future Research

Our study provides an overview of existing self-report stress measures, but can only offer limited guidance on the suitability of the self-report measures for people with MID. Despite many relevant arguments for the use of self-report measures in intellectual disability research, there are few validated self-report measures available, with even fewer for sensitive topics like stressful experiences (Ali et al., 2008; Ruddick & Oliver, 2005). Information on the suitability of a self-report measure for certain subgroups within the general population such as persons with cognitive impairments, limited verbal abilities, or clinical populations, is generally found in the manual or published peer-reviewed validation research. However, in many cases, self-report measures do not have detailed manuals, the manuals are unavailable, or they do not even exist. We therefore strongly advise future researchers to always publish clear user manuals and/or assessment procedures of self-report measures, even if they seem to be simple and easy to use. In addition, for those self-report measures not specifically designed for people with (M)ID, there is no published research on the use in the MID population. Moreover, norm data from validity studies are often based on research that excluded people with MID a priori based on their level of IQ. The relevance/suitability of many of the self-report stress measures found for people with MID therefore still remains unclear. More research is needed on the 'performance' of a measurement instrument in populations including people with MID. Therefore, we recommend that future validation studies of self-report measures always include a subpopulation composed of respondents with MID.

As noted earlier, stress is operationalized by many different theoretical constructs in the self-report stress measures analyzed. This raises the question of whether this could affect the measured results. On the other hand, research also shows that the operationalization of apparently different concepts, such as "stress" and "state-anxiety", essentially measure the same items and therefore can be regarded as the same type of outcome (de Witte, Spruit, et al., 2020; de Witte, Pinho, et al., 2020; Hook et al., 2008). This has led to these concepts being used interchangeably in literature when it comes to outcome studies (Bradt & Dileo, 2014; De Witte, Spruit, et al., 2020; Wetsch et al., 2009). Nevertheless, we think it essential to provide a theoretical framework underpinning the measurement concepts involved. Not only will this offer the necessary background information for future users, like clinicians, but it also increases the content validity of the self-report measure (Higgins & Straub, 2006; Lynn, 1986).

In order to validly and reliably assess stress-related outcomes in people with MID, attempts should be made to make the self-report stress measures more 'MID-inclusive'. However, it is still not entirely clear which specific instrument components or adaptations are required for this purpose. The recent study by Kooijmans et al. (2022) shows that there are still many gaps to fill on this topic. Findings show, for example, that researchers and clinicians assume questions should be read aloud by the assessor in order to assist people with MID. However, there is reason to believe that this may introduce various forms of bias in the results, arising from complex interviewer-interviewee dynamics (Finlay & Antaki, 2012). More research on the impact of assistance on the outcome of self-report measures is needed to decide whether this is an acceptable practice.

Lastly, the literature shows that Likert scales with three to five answer alternatives can be reliably used in research with people with MID (Fang et al., 2011). However, in the field of stress research, more nuanced response formats may be needed to capture the subtle differences in perceived stress over time. Visual Analogue Scales (VAS), for example, may offer an interesting alternative for this and have potential for assessing stress levels in people with MID. The Subjective Units of Distress Scales (SUDS) developed by Wolpe (1969) is an example of this. Notably, Mevissen et al. (2016) show promising results when using the SUDS in the treatment of trauma-related symptoms of people with MID. As many VAS scales differ in form, more research is advised on how to optimally attune these VAS scale formats to the needs of people with MID.

In conclusion, many adults with MID frequently experience stress in daily life and this has a major impact on their wellbeing. This emphasizes the importance of assessing stress levels as part of their support needs assessment. Research suggests that self-report measures are more accurate and sensitive compared to proxy measures. However, this scoping review found that there are few self-report stress measures suitable for this purpose.

This underlines the need for continuing efforts to develop high quality and "MID-sensitive" self-report stress measures.

Acknowledgements

We would like to thank Arjan Doolaar and Thomas Pelgrim from HAN University of Applied Sciences (Nijmegen, Netherlands), for their advice and assistance with the systematic search.

References

Abbaszadeh, M., & Sardoie, G. (2016). Compare Academic Self-Efficacy and Self-Regulation among Students with Learning Disorder and without Learning Disorder in Normal Elementary Schools (Fourth And Fifth Grade) of Kerman. *Biomedical and Pharmacological Journal, 9*(2), 751–759. https://dx.doi.org/10.13005/bpj/999

Abell, J. N. (1991). The index of clinical stress: A brief measure of subjective stress for practice and research. *Social Work Research and Abstracts, 27*(2), 12-15. https://doi.org/10.1093/swra/27.2.12

Akin, A., & Iskender, M. (2011). Internet addiction and depression, anxiety and stress. *International Online Journal of Educational Sciences, 3*, 138-148.

Aldwin, C. M. (2007). *Stress, coping, & development: An integrative perspective* (2nd ed.). Guilford Press.

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

Allen, P., Bennett, K., & King, J. (Ed.) (2010). *PASW statistics by SPSS: A practical guide, version 18*. National Library of Australia.

American Psychological Association. (2017). *Stress in America: Coping with Change*. https://www.apa.org/news/press/releases/stress/2016/coping-with-change.pdf

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.

Amirkhan, J. H. (2012). Stress Overload: A new approach to the assessment of stress. *American Journal of Community Psychology, 49*, 55–71. https://doi.org/10.1007/s10464-011-9438-x

Amirkhan, J. H., Urizar Jr, G. G., & Clark, S. (2015). Criterion validation of a stress measure: The Stress Overload Scale. *Psychological Assessment, 27*(3), 985. https://doi.org/10.1037/pas0000081

Amirkhan, J. H. (2018). A brief stress diagnostic tool: The short Stress Overload Scale. *Assessment, 25*(8), 1001–1013. https://doi.org/10.1177/1073191116673173

Anderson, G.L. (2017). Can Participatory Action Research (PAR) Democratize Research, Knowledge, & Schooling? Experiences from the Global South and North. *International Journal of Qualitative Studies in Education, 30*(5), 427–431. https://doi.org/10.1080/09518398.2017.1303216

Andresen, E.M., Vahle, V.J., & Lollar, D. (2001). Proxy reliability: Health-related quality of life (HRQoL) measures for people with disability. *Quality of Life Research, 10*, 609–619. https://doi.org/10.1023/A:1013187903591

Argus, G. R., Terry, P. C., Bramston, P., & Dinsdale, S. L. (2004). Measurement of mood in adolescents with intellectual disability. *Research in Developmental Disabilities, 25*(6), 493–507. https://doi. org/10.1016/j.ridd.2004.05.001

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology, 8*, 19-32. https://doi. org/10.1080/1364557032000119616

Australian Psychological Society. (2015). *Stress and wellbeing: how Australians are coping with life.* http://www.psychology.org.au/Assets/Files/PW15-SR.pdf

de Ayala, R. J., Vonderharr-Carlson, D. J., & Kim, D. (2005). Assessing the reliability of the Beck anxiety inventory scores. *Educational and Psychological Measurement, 65*(5), 742–756.

Bally, K., Campbell, D., Chesnick, K., & Tranmer, J. E. (2003). Effects of Patient-Controlled Music Therapy During Coronary Angiography on Procedural Pain and Anxiety Distress Syndrome. *Critical Care Nurse, 23*(2), 50–57. https://doi.org/10.4037/ccn2003.23.2.50

Bandelow, B., Reitt, M., Röver, C., Michaelis, S., Görlich, Y., & Wedekind, D. (2015). Efficacy of treatments for anxiety disorders: A meta-analysis. *International Clinical Psychopharmacology, 30*, 183-192. https://doi.org/10.1097/YIC.0000000000000078

Bardhoshi, G., Duncan, K., & Erford, B. T. (2016). Psychometric meta-analysis of the English version of the Beck Anxiety Inventory. *Journal of Counseling and Development, 94*(3), 356-373. https://doi.org/10.1002/jcad.12090

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897.

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 and Law, 24*(4), 362-378. https://doi.org/10.1080/1068316X.2017.1298761

Bradt, J., & Dileo, C. (2014). *Music interventions for mechanically ventilated patients*. The Cochrane Database Of Systematic Reviews, 12 , CD006902. https://doi. org/10.1002/14651858.CD006902.pub3

Bramston, P., & Fogarty, G. (n.d.). *The lifestress inventory.* University of Southern Queensland ePrints. https://core.ac.uk/reader/11047490

Bramston, P., Fogarty, G., & Cummins, R.A. (1999). The nature of stressors reported by people with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 12,* 1–10. https://doi.org/10.1111/j.1468-3148.1999.tb00046.x

Bramston, P., & Mioche, C. (2001). Disability and stress: A study in perspectives. *Journal of Intellectual and Developmental Disabilities, 26*, 233–242. https://doi.org/10.1080/13668250120063403

Carraro, A., & Gobbi, E. (2012). Effects of an exercise programme on anxiety in adults with intellectual disabilities. *Research in developmental disabilities, 33*(4), 1221-1226. https://doi.org/10.1016/j.ridd.2012.02.014

Casey, G. (2017). Stress and disease. *Kai Tiaki: Nursing New Zealand, 23*(6), 20.

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety.* Ronald.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance and R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp.309-336). Routledge.

Chandola, T., Heraclides, A., & Kumari, M. (2010). Psychophysiological biomarkers of workplace stressors. *Neuroscience and Biobehavioral Reviews, 35*, 51–57. https://doi.org/10.1016/j.neubiorev.2009.11.005

Chiang, I. C. A., Jhangiani, R. S., & Price, P. C. (2015). Reliability and validity of measurement. *Research Methods in Psychology*.

Cohen, S., Kamarch, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behaviour, 24*, 385–396. https://doi.org/10.2307/2136404

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th ed.). New York: McGraw-Hill.

Cohen, S., & Williamson, G. M. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan and S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Sage.

Conners, F. A. (2003). Reading skills and cognitive abilities of individuals with mental retardation. *International review of research in mental retardation, 27*(1), 191-229. https://doi.org/10.1016/S0074-7750(03)27006-3

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine, 119*(2), 166-e7. https://doi.org/10.1016/j.amjmed.2005.10.036

Cox, T., & Mackay, C. (1985). The measurement of self-reported stress and arousal. *British Journal of Psychology, 76*(2), 183–186. https://doi.org/10.1111/j.2044-8295.1985.tb01941.x

Crawford, J. R. and Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*(2), 111–131. https://doi.org/10.1348/014466503321903544

Crawford, J. R., Henry, J. D., Ward, A. L., & Blake, J. (2006). The Prospective and Retrospective Memory Questionnaire (PRMQ): Latent structure, normative data and discrepancy analysis for proxy-ratings. *British Journal of Clinical Psychology, 45,* 83-104. https://doi.org/10.1348/014466505X28748

Creamer, M., Foran, J., & Bell, R. (1995). The Beck Anxiety Inventory in a nonclinical sample. *Behaviour Research and Therapy, 33*(4), 477–485. https://doi.org/10.1016/0005-7967(94)00082-u

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.

Davis, E., Saeed, S. A., & Antonacci, D. J. (2008). Anxiety disorders in persons with developmental disabilities: Empirically informed diagnosis and treatment. *Psychiatric Quarterly, 79*(3), 249-263. https://doi.org/10.1007/s11126-008-9081-3

Dedding, C., Goedhart, N. S., Broerse, J. E., & Abma, T. A. (2020). Exploring the boundaries of ‘good’ Participatory Action Research in times of increasing popularity: dealing with constraints in local policy for digital inclusion. *Educational Action Research, 29*(1), 20–36. https://doi.org/10.1080/09650792.2020.1743733

Derogatis, L. R. (1987). The Derogatis stress profile (DSP): Quantification of psychological stress. In G. A. Fava and T. N. Wise (Eds.), *Research paradigms in psychosomatic medicine* (Vol. 17, pp. 30-54). Karger. https://doi.org/10.1159/000414005

Derogatis Testing (2021). *Synopsis of the Derogatis Stress Profile*. https://www.derogatis-tests.com/dsp/

Devi, N. (2014). Supported decision-making and personal autonomy for persons with intellectual disabilities: Article 12 of the UN Convention on the Rights of Persons with Disabilities. *Journal of Law, Medicine, & Ethics, 41*(4), 792–806. https://doi.org/10.1111/jlme.12090

Didden, R., Embregts, P., van der Toorn, M., & Laarhoven, N. (2009). Substance abuse, coping strategies, adaptive skills and behavioral and emotional problems in clients with mild to borderline intellectual disability admitted to a treatment facility: A pilot study. *Research in Developmental Disabilities, 30*(5), 927–932. https://doi.org/10.1016/j.ridd.2009.01.002

Didden, R., Lindsay, W.R., Lang, R., Sigafoos, J., Dab, S., Wierma, J, & Lancioni, G.E. (2016). Aggressive behavior. In N. N. Singh (Ed.), *Clinical handbook of evidence-based practices for individuals with intellectual and developmental disabilities* (pp. 727-750). New York: Springer.

Dobkin, P. L., Pihl, R. O., & Breault, C. (1991). Validation of the Derogatis Stress Profile using laboratory and real world data. *Psychotherapy and psychosomatics, 56*(4), 185-196. https://doi.org/10.1159/000288555

Dulin, P. L., Hanson, B. L., & King, D. K. (2013). Perceived control as a longitudinal moderator of late-life stressors on depressive symptoms. *Aging and Mental Health, 17*, 718-723. https://doi.org/10.1080/13607863.2013.784956

von Elm, E., Schreiber, G., & Haupt, C. (2019). Methodische Anleitung für Scoping Reviews (JBI-Methodologie). *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 143*, 1-7. https://doi.org/10.1016/j.zefq.2019.05.004

Emerson, E. (2003). Mothers of children and adolescents with intellectual disability: social and economic situation, mental health status, & the self-assessed social and psychological impact of the child's difficulties. *Journal of Intellectual Disability Research, 47*(4-5), 385–399. https://doi.org/10.1046/j.1365-2788.2003.00498.x

Emerson, E., Felce, D., & Stancliffe, R. (2013). Issues concerning self-report data and population-based data sets involving people with intellectual disabilities. *Intellectual and Developmental Disabilities, 51*(5), 333–348. https://doi.org/10.1352/1934-9556-51.5.333

Everly, G. S., & Lating, J. M. (2019). Resilience: The final frontier. In G. S. Everly, & J. M. Lating (Eds.). *A clinical guide to the treatment of the human stress response* (pp. 175-187). Springer, New York, NY.

Fang, J., Fleck, M. P., Green, A., McVilly, K., Hao, Y., Tan, W., Fu, R., & 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*(6), 537–549. https://doi.org/10.1111/j.1365-2788.2011.01401.x

Finlay, W. M., & Antaki, C. (2012). How staff pursue questions to adults with intellectual disabilities. *Journal of Intellectual Disability Research, 56*(4), 361-370. https://doi.org/10.1111/j.1365-2788.2011.01478.x

Fischer, D. G. and Donatelli, M. J. (1987). A measure of stress and arousal: Factor structure of the stress adjective checklist. *Educational and Psychological Measurement, 48,* 127–136. https://doi.org/10.1177/0013164487472014

Fischer, D. G., Hansen, R. J., & Zemore, R. W. (1988). Factor Structure of the Stress Adjective Checklist: Replicated. *Educational and Psychological Measurement, 48*(1), 127–136. https://doi.org/10.1177/001316448804800115

Fliege, H., Rose, M., Arck, P., Levenstein, S. and Klapp, B. F. (2001). Validierung des "Perceived Stress Questionnaire" (PSQ) an einer deutschen Stichprobe. *Diagnostica, 47*(3), 142–152.

Fliege, H., Rose, M., Arck, P., Walter, O. B., Kocalevent, R. D., Weber, C., & Klapp, B. F. (2005). The Perceived Stress Questionnaire (PSQ) reconsidered: validation and reference values from different clinical and healthy adult samples. *Psychosomatic Medicine, 67*, 78–88. https://doi.org/10.1097/01.psy.0000151491.80178.78

Fogarty, G. J., Bramston, P., Cummins, R. A. (1997). Validation of the Lifestress Inventory. *Research in Developmental Disabilities, 18*(6), 435–456. https://doi.org/10.1016/S0891-4222(97)00021-8

Föhr, T., Tolvanen, A., Myllymäki, T., Järvelä-Rejionen, E., Peuhkuri, K., Rantala, S., Kolehmainen, M., Korpela, R., Lappalainen, R., Ermes, M., Puttonen, S., Rusko, H., & Kujala, U. M. (2017). Physical activity, heart rate variability–based stress and recovery, & subjective stress during a 9-month study period. *Scandinavian Journal of Medicine and Science in Sports, 27*(6), 612–621. https://doi.org/10.1111/sms.12683

Fonseca-Pedrero, E., Paino, M., Sierra-Baigrie, S., Lemos-Giraldez, S., & Muniz, J. (2012). Psychometric properties of the State-Trait Anxiety Inventory (STAI) in college students. *Behavioral Psychology/Psicologia Conductual, 20*(3), 547–561.

Frielink, N., Schuengel, C., & Embregts, P. J. (2018). Autonomy support in people with mild-to-borderline intellectual disability: Testing the Health Care Climate Questionnaire-Intellectual Disability. *Journal of Applied Research in Intellectual Disabilities, 31*(1), 159-163. https://doi.org/10.1111/jar.12371

Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures?. *Value in Health, 10,* 94-105.

Fujiura, G. T. (2012). Self-reported health of people with intellectual disability. *Intellectual and Developmental Disabilities, 50*(4), 352–369. https://doi.org/10.1352/1934-9556-50.4.352

Gerald, J. H., & George, S. H. (2010). Self-report: Psychology's four-letter word. *American Journal of Psychology, 123*, 181-188. https://doi.org/10.5406/amerjpsyc.123.2.0181

Giesbers, S. A., Hendriks, L., Jahoda, A., Hastings, R. P., & Embregts, P. J. (2019). Living with support: Experiences of people with mild intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 32*(2), 446-456. https://doi.org/10.1111/jar.12542

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

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

Hagopian, L. P., & Jennett, H. K. (2008). Behavioral assessment and treatment of anxiety in individuals with intellectual disabilities and autism. Journal of *Developmental and Physical Disabilities, 20*(5), 467–483. https://doi.org/10.1007/s10882-008-9114-8

Harkness, K.L., & Monroe, S.M. (2016). The assessment and measurement of adult life stress: Basic premises, operational principles, & design requirements. *Journal of Abnormal Psychology, 125*(5), 727-745. https://doi.org/10.1037/abn0000178

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

Hartley, S. L., & MacLean Jr., W. E. (2008). Coping strategies of adults with mild intellectual disability for stressful social interactions. *Journal of Mental Health Research in Intellectual Disabilities, 1*(2), 109-127. https://doi.org/10.1080/19315860801988426

Hartley, S. L., MacLean Jr, W. E., & Abbeduto, L. (2009a). Depression in adults with mild intellectual disability: Role of stress, attributions, & coping. *American Journal on Intellectual and Developmental Disabilities, 114*(3), 147-160. https://doi.org/10.1352/1944-7588-114.3.147

Hartley, S. L., MacLean Jr, W. E., & Abbeduto, L. (2009b). Stressful social interactions experienced by adults with mild intellectual disability. *American Journal on Intellectual and Developmental Disabilities, 114*(2), 71-84. https://doi.org/10.1352/2009.114.71-84

Hastings, R. P., Hatton, C., Taylor, J. L., & Maddison, C. (2004). Life events and psychiatric symptoms in adults with intellectual disabilities. *Journal of Intellectual Disability Research, 48*(1), 42-46.

Hatton, C., & Emerson, E. (2004). The relationship between life events and psychopathology amongst children with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 17*, 109–117. https://doi.org/10.1111/j.1360-2322.2004.00188.x

Hatton, C., & Taylor, J. L. (2013). The assessment of mental health problems in adults with intellectual disabilities. In J. L. Taylor, W. R. Lindsay, R. P. Hastings, & C. Hatton (Eds.), *Psychological therapies for adults with intellectual disabilities* (p. 31–54). Wiley-Blackwell.

Heal, L. W., & Sigelman, C. K. (1995). Response biases in interviews of individuals with limited mental capacity. *Journal of Intellectual Disability Research, 39*, 331–340. https://doi.org/10.1111/j.1365-2788.1995.tb00525.x

Hermans, H., Wieland, J., Jelluma, N., van der Pas, F., & Evenhuis, H. (2013). Reliability and validity of the Dutch version of the Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID). *Journal of Intellectual Disability Research, 57*(8), 728-736. https://doi.org/10.1111/j.1365-2788.2012.01632.x

Heyman, M., & Hauser-Cram, P. (2015). Negative life events predict performance on an executive function task in young adults with developmental disabilities. *Journal of Intellectual Disability Research, 59*(8), 746–754. https://doi.org/10.1111/jir.12181

Higgins, P. A., & Straub, A. J. (2006). Understanding the error of our ways: mapping the concepts of validity and reliability. *Nursing Outlook, 54*(1), 23-29. https://doi.org/10.1016/j.outlook.2004.12.004

Hinton, J. W., Rotheiler, E., & Howard, A. (1991). Confusion between stress and state anxiety in a much used self-report 'stress' inventory. *Personality and Individual Differences, 12*(1), 91–94. https://doi.org/10.1016/0191-8869(91)90136-Y

Hook, L., Songwathana, P., & Petpichetchian, W. (2008). Music therapy with female surgical patients: Effect on anxiety and pain. *Pacific Rim International Journal of Nursing Research, 12*, 259–271.

Howell, M., Bradshaw, J., & Langdon, P. E. (2020). A Systematic Review of Behaviour-Related Outcome Assessments for Children on the Autism Spectrum with Intellectual Disabilities in Education Settings. *Review Journal of Autism and Developmental Disorders, 8*, 67-91.

Hudson, W. W., MacNeil, G., & Dierks, J. (1995). *Six new assessment scales: A partial* *validation*. WALMYR Publishing Co.

Hulbert-Williams, L. and Hastings, R. P. (2008). Life events as a risk factor for psychological problems in individuals with intellectual disabilities: a critical review. *Journal of Intellectual Disability Research, 52*(11), 883–895. https://doi.org/10.1111/j.1365-2788.2008.01110.x

Hurley, A. D. (2008). Depression in adults with intellectual disability: Symptoms and challenging behaviour. *Journal of Intellectual Disability Research, 52*(11), 905–916. https://doi.org/10.1111/j.1365-2788.2008.01113.x

Janusis, G. M., & Weyandt, L. L. (2010). An exploratory study of substance use and misuse among college students with and without ADHD and other disabilities. *Journal of Attention Disorders, 14*(3), 205-215. https://doi.org/10.1177/1087054710367600

Kaal, H. L., Moonen X. M. H., & Nijman H. L. I. (2015). Identifying offenders with an intellectual disability in detention in The Netherlands. *Journal of Intellectual Disabilities and Offending Behaviour,* 6(2), 94–101. https://doi.org/10.1108/JIDOB-04-2015-0008

Kellett, S., Beail, N., Newman, D. W., & Frankish, P. (2003). Utility of the Brief Symptom Inventory in the assessment of psychological distress. *Journal of Applied Research in Intellectual Disabilities, 16*(2), 127–134. https://doi.org/10.1046/j.1468-3148.2003.00152.x

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation, 15,* 235.

King, M. G., Burrows, G. D. and Stanley, G. V. (1983). Measurement of stress and arousal: Validation of the stress/arousal adjective checklist. *British Journal of Psychology, 74,* 473–479. https://doi.org/10.1111/j.2044-8295.1983.tb01880.x

Kocalevent, R. D., Levenstein, S., Fliege, H., Schmid, G., Hinz, A., Brähler, E., & Klapp, B. F. (2007). Contribution to the construct validity of the Perceived Stress Questionnaire from a population-based survey. *Journal of Psychosomatic Research, 63,* 71–81. https://doi.org/10.1016/j.jpsychores.2007.02.010

Koelsch, S., Fuermetz, J., Sack, U., Bauer, K., Hohenadel, M., Wiegel, M., Kaisers, U., & Heinke, W. (2011). Effects of Music Listening on Cortisol Levels and Propofol Consumption during Spinal Anesthesia. *Frontiers in Psychology, 2*, 58. https://doi.org/10.3389/fpsyg.2011.00058

Kooijmans, R., Mercera, G., Langdon, P. E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: A systematic review. *Clinical Psychology: Science and Practice*, *29*(3), 250-271.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. McGraw-Hill.

Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal, & Coping*. Springer.

Lee, E.-H. (2012). Review of the psychometric evidence of the Perceived Stress Scale. *Asian Nursing Research, 6*(4), 121–127. https://doi.org/10.1016/j.anr.2012.08.004

Lemyre, L., Chair, M. R., & Lalande-Markon, M. P. (2009). Psychological stress measure (PSM-9): Integration of an evidence-based approach to assessment, monitoring, & evaluation of stress in physical therapy practice. *Physiotherapy Theory and Practice, 25*(5-6), 453–462. https://doi.org/10.1080/09593980902886321

Lemyre, L., & Tessier, R. (1988). Mesure de stress psychologique (MSP): Se sentir stressé-e. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 20*(3), 302. https://doi.org/10.1037/h0079945

Lemyre, L., & Tessier, R. (2003). Measuring psychological stress. Concept, model, & measurement instrument in primary care research. *Canadian Family Physician, 49,* 1159-1160.

Lemyre, L., Tessier, R., & Fillion, L. (1990). *La Mesure du stress psychologique: Manuel d'utilisation*. Behaviora.

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., &reoli, A., Luzi, C., Arcà, M., Berto, E., Milite, G., & Marcheggiano, A. (2000). Stress and exacerbation in ulcerative colitis: a prospective study of patients enrolled in remission. *American Journal of Gastroenterology, 95*, 1213–1220. https://doi.org/10.1111/j.1572-0241.2000.02012.x

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., & Andreoli, A. (1993). Development of the perceived stress questionnaire: a new tool for psychosomatic research. *Journal of Psychosomatic Research 37*(1), 19–32. https://doi.org/doi:10.1016/0022-3999(93)90120-5

Lindsay, W. R., & Michie, A. M. (1988). Adaptation of the Zung self-rating anxiety scale for people with a mental handicap*. *Journal of Intellectual Disability Research, 32,* 485–490. https://doi.org/10.1111/j.1365-2788.1988.tb01440.x

Lindsay, W. R., Michie, A. M., Baty, F. J., Smith, A. H. W. and Miller, S. (1994). The consistency of reports about feelings and emotions from people with intellectual disability. *Journal of Intellectual Disability Research, 38,* 61–66. https://doi.org/10.1111/j.1365-2788.1994.tb00348.x

Lindsay, W. R., & Skene, D. D. (2007). The Beck Depression Inventory II and the Beck

Anxiety Inventory in people with intellectual disabilities: factor analyses and group data. *Journal of Applied Research in Intellectual Disabilities, 20*, 401-408. https://doi.org/10.1111/j.1468-3148.2007.00380.x

Linnemann, A., Strahler, J., & Nater, U. M. (2017). Assessing the effects of music listening on psychobiological stress in daily life. *Journal of Visualized Experiments, 120*, e54920–e54920. https://doi.org/10.3791/54920

Long, C. G., Krawczyk, K. M., & Kenworthy, N. E. (2013). Assessing the sexual knowledge of women in secure settings: The development of a new screening measure. *British Journal of Learning Disabilities, 41*(1), 51-65. doi:10.1111/j.1468-3156.2011.00722.x

Lovibond, P. F., & Lovibond, S. H. (1995a). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy, 33*(3), 335-343. https://doi.org/10.1016/0005-7967(94)00075-U

Lovibond, S.H., & Lovibond, P.F. (1995b). *Manual for the Depression Anxiety Stress Scales* (2nd. Ed.). Sydney: Psychology Foundation.

Lunsky, Y. (2008). The impact of stress and social support on the mental health of individuals with intellectual disabilities. *Salud Pública de México, 50*(2), 151–153.

Lunsky, Y., & Benson, B. A. (2001). Association between perceived social support and strain, & positive and negative outcome for adults with mild intellectual disability. *Journal of Intellectual Disability Research, 45*(2), 106-114. https://doi.org/10.1046/j.1365-2788.2001.00334.x

Lunsky, Y., & Bramston, P. (2006). A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. *Journal of Intellectual and Developmental Disability, 31*(1), 20–27. https://doi.org/10.1080/13668250500488660

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385.https://doi.org/10.1097/00006199-198611000-00017

Mackay, C., Cox, T., Burrows, G., & Lazzerini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British journal of Social and Clinical Psychology, 17*(3), 283–284. https://doi.org/10.1111/j.2044-8260.1978.tb00280.x

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

Masi, G., Brovedani, P., Mucci, M., & Favilla, L. (2002). Assessment of anxiety and depression in adolescents with mental retardation. *Child Psychiatry and Human Development, 32*(3), 227–237. https://doi.org/10.1023/A:1017908823046

McCance, K. L., Forshee, B., & Shelby, J. (2006). Stress and disease. In K. L. McCance and S. E. Huether (Eds.), *Pathophysiology: The biologic basis for disease in adults and children* (5th ed., pp. 311–332). St. Louis, MO: Elsevier.

McCormick, I. A., Walkey, F. H., & Taylor, A. J. W. (1985). The Stress Arousal Checklist: An Independent Analysis. *Educational and Psychological Measurement, 45*(1), 143–146. https://doi.org/10.1177/0013164485451013

McEwen, B. S., & Gianaros, P. J. (2010). Central role of the brain in stress and adaptation: Links to socioeconomic status, health, & disease. *Annals of The New York Academy of Sciences, 1186*(1), 190–222. https://doi.org/10.1111/j.1749-6632.2009.05331.x

Meijer, J. (2001). Stress in the relation between trait and state anxiety. *Psychological Reports, 88*(3), 947-964. https://doi.org/10.2466/pr0.2001.88.3c.947

Mevissen, L., Didden, R., & de Jongh, A. (2016). Assessment and treatment of PTSD in people with intellectual disabilities. In C. R. Martin, V. R. Preedy and V. B. Patel (Eds.), *Comprehensive Guide to Post-Traumatic Stress Disorder* (pp. 281-299). Springer Reference. https://doi.org/10.1007/978-3-319-08359-9

Miller, L. A. and Tucker, C. (1993). *The Intrahousehold Communications Study: A Typology of Family Cohesion*. American Statistical Association.

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

Moore, J.C. (1988). Self/proxy response status and survey response quality. *Journal of Official Statistics*, *4*(2), 155-72.

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion mit 41 Tabellen*. Springer.

Moss, S., Prosser, H., Ibbotson, B., & Goldberg, D. (1996). Respondent and informant accounts of psychiatric symptoms in a sample of patients with learning disability. *Journal of Intellectual Disability Research, 40*, 457–465. https://doi.org/10.1046/j.1365-2788.1996.792792.x

Munn, Z., Peters, M., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology, 18*, 143. https://doi.org/10.1186/s12874-018-0611-x

Mutz, M., & Müller, J. (2016). Mental health benefits of outdoor adventures: Results from two pilot studies. *Journal of Adolescence, 49*, 105–114. https://doi.org/10.1016/j.adolescence.2016.03.009

Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., & Berk, M. (2007). The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure. *Acta neuropsychiatrica, 19*(5), 304-310. https://doi.org/10.1111/j.1601-5215.2007.00217.x

Nolen-Hoeksema, S. (2011). *Abnormal Psychology* (5th ed.) New York: McGraw-Hill.

Nouwens, P. J. G., Smulders, N. B. M., Embregts, P. J. C. M., & van Nieuwenhuizen, C. (2020). Differentiating care for persons with mild intellectual disability or borderline intellectual functioning: a Delphi study on the opinions of primary and professional caregivers and scientists. *BMC Psychiatry, 20*, Article 57. https://doi.org/10.1186/s12888-020-2437-4

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

Ortuno-Sierra, J., Garcia-Velasco, L., Inchausti, F., Debbane, M., & Fonseca-Pedrero, E. (2016). New approaches on the study of the psychometric properties of the STAI. *Actas espanolas de psiquiatria, 44*(3), 83-92.

Østerås, B., Sigmundsson, H., & Haga, M. (2018). Psychometric Properties of the Perceived Stress Questionnaire (PSQ) in 15–16 Years Old Norwegian Adolescents. *Frontiers in Psychology, 9,* Article 1850. | https://doi.org/10.3389/fpsyg.2018.01850

Ozer, N., Karaman Ozlu, Z., Arslan, S., & Gunes, N. (2013). Effect of music on postoperative pain and physiologic parameters of patients after open heart surgery. *Pain Management Nursing, 14*(1), 20–28. https://doi.org/10.1016/j.pmn.2010.05.002

Park, S. H., Song, Y. J. C., Demetriou, E. A., Pepper, K. L., Thomas, E. E., Hickie, I. B., & Guastella, A. J. (2020). Validation of the 21-item Depression, Anxiety, & Stress Scales (DASS-21) in individuals with autism spectrum disorder. *Psychiatry Research, 291*, Article 113300. https://doi.org/10.1016/j.psychres.2020.113300

Parkitny, L., & McAuley, J. (2010). The depression anxiety stress scale (DASS). *Journal of physiotherapy, 56*(3), 204. https://doi.org/ 10.1016/S1836-9553(10)70030-8

Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare, 13*, 141-146. https://doi.org/10.1097/XEB.0000000000000050

Pfaff, D. W., Martin, E. M., & Ribeiro, A. C. (2007). Relations between mechanisms of CNS arousal and mechanisms of stress. *Stress, 10*, 316-325. https://doi.org/10.1080/10253890701638030

Pittman, S., & Kridli, S. (2011). Music intervention and preoperational anxiety: An integrative review. *International Nursing Review, 58*(2), 157–163. https://doi.org/10.1111/j.1466-7657.2011.00888.x

Pritchard, M. J. (2009). Identifying and assessing anxiety in pre-operative patients. *Nursing Standard, 23*(51), 35–40. https://doi.org/10.7748/ns.23.51.35.s46

Psychology Foundation of Australia (2021). *DASS: Frequently Asked Questions*. http://www2.psy.unsw.edu.au/dass/DASSFAQ.htm

Ramirez, S. Z. (2005). Evaluating acquiescence to yes-no questions in fear assessments of children with and without mental retardation. *Journal of Developmental and Physical Disabilities, 17*, 337–343. https://doi.org/10.1007/s10882-005-6617-4

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*, 573–580. https://doi.org/10.1007/s10882-008-9120-x

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

Rethlefsen, M. L., Farrell, A. M., Trzasko, L. C. O., & Brigham, T. J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology, 68*(6), 617–626. https://doi.org/10.1016/j.jclinepi.2014.11.025

Rheinberg F., & Engeser, S. (2018). Intrinsic Motivation and Flow. In: J. Heckhausen and H. Heckhausen (Eds.), *Motivation and Action* (pp. 579-622). Springer, Cham.

Riley, K. E., & Park, C. L. (2015). How does yoga reduce stress? A systematic review of mechanisms of change and guide to future inquiry. *Health Psychology Review, 9,* 379–396. https://doi.org/10.1080/17437199.2014.981778

Rönnlund, M., Vestergren, P., Stenling, A., Nilsson, L.-G., Bergdahl, M., & Bergdahl, J. (2015). Dimensionality of stress experiences: Factorial structure of the Perceived Stress Questionnaire (PSQ) in a population-based Swedish sample. *Scandinavian*

Journal of Psychology, 56, 592–598. https://doi.org/10.1111/sjop.12214

Rossi, V., & Pourtois, G. (2012). Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: a comparative review. *Anxiety, Stress and Coping, 25*(6), 603–645. https://doi.org/10.1080/10615806.2011.582948

Ruddick, L., & Oliver, C. (2005). The development of a health status measure for self-report by people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 18*(2), 143–150. https://doi.org/10.1111/j.1468-3148.2005.00243.x

Salmond, S. S. (2008). Evaluating the reliability and validity of measurement instruments. *Orthopaedic Nursing, 27*(1), 28–30. https://doi.org/10.1097/01.NOR.0000310608.00743.54

Sams, K., Collins, S., & Reynolds, S. (2006). Cognitive therapy abilities in people with learning disabilities. *Journal of Applied Research in Intellectual Disabilities, 19*, 25–33. https://doi.org/10.1111/j.1468-3148.2006.00303.x

Sanz-Carrillo, C., García-Campayo, J., Rubio, A., Santed, M. A., & Montoro, M. (2002). Validation of the Spanish version of the Perceived Stress Questionnaire. *Journal of Psychosomatic Research, 52*, 167–172. htps://doi.org/10.1016/S0022-3999(01)00275-6

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

Schalock, R. L., Verdugo, M. A., & Braddock, D. L. (2002). *Handbook on quality of life for human service practitioners*. American Association on Mental Retardation.

Schuengel, C., & Janssen, C. G. (2006). People with mental retardation and psychopathology: Stress, affect regulation and attachment: A review. *International Review of Research in Mental Retardation, 32*, 229–260.

Scott, H. M., & Havercamp, S. M. (2014). Mental health for people with intellectual disability: The impact of stress and social support. *American Journal on Intellectual and Developmental Disabilities, 119*(6), 552-564. https://doi.org/10.1352/1944-7558-119.6.552

Scott, H. M., & Havercamp, S. M. (2018). Comparisons of self and proxy report on health-related factors in people with intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 31*(5), 927-936. https://doi.org/10.1111/jar.12452

Seyed, S., Salmani, M., Motahari, Nezhad F., & Noruzi, R. (2017). Self-efficacy, achievement motivation, & academic progress of students with learning disabilities: a comparison with typical students. *Middle East Journal of Rehabilitation and Health Studies, 4*, e44558. https://doi.org/10.5812/mejrh.44558

Spielberger, C. D. (1973). *Preliminary manual for the State-Trait Anxiety Inventory for Children (“How I Feel Questionnaire”)*. Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G.A. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists Press.

Spielberger, C. D., & Reheiser, E. C. (2009). Assessment of Emotions: Anxiety, Anger, Depression, & Curiosity. *Applied Psychology: Health and Well-Being, 1*(3), 271-302. https://doi.org/10.1111/j.1758-0854.2009.01017.x

Stavrakaki, C., & Lunsky, Y. (2007). Depression, anxiety and adjustment disorders in people with intellectual disabilities. In C. Hemmings and N. Bouras (Eds.), *Psychiatric and behavioural disorders in intellectual and developmental disabilities* (pp. 113-130). Cambridge University Press.

Steer, R. A. (2009). Amount of general factor saturation in the Beck anxiety inventory responses of outpatients with anxiety disorders. *Journal of Psychopathology and Behavioral Assessment, 31*(2), 112–118. https://doi.org/10.1007/s10862-008-9098-9

Steptoe, A. and Kivimäki, M. (2012). Stress and cardiovascular disease. *Nature Reviews Cardiology, 9,* 360–370. https://doi.org/10.1038/nrcardio.2012.45

Thayer, E. (1967). Measurement of activation through self-report. *Psychological Reports, 20,* 663–678. https://doi.org/10.2466/pr0.1967.20.2.663

Tricco, A. C., Lillie, E., Zarin, W., O’Brien, K., Colquhoun, H., Kastner, M., Levac, D., Ng, C., Pearson Sharpe, J., Wilson, K., Kenny, M., Warren, R., Wilson, C., Stelfox, H. T., & Straus, S. E. (2016). A scoping review on the conduct and reporting of scoping reviews. *BMC medical research methodology, 16*, Article 15.

Tricco, A. C., Lillie, E., Zarin, W., O’Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Aki, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Lewin, S., ... & Strauss, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of internal medicine, 169*, 467-473. https://doi.org/10.7326/M18-0850

Tricco, A. C., Cardoso, R., Thomas, S. M., Motiwala, S., Sullivan, S., Kealey, M. R., ... and Straus, S. E. (2015). Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. *Implementation Science, 11*(1), 1-20. https://doi.org/10.1186/s13012-016-0370-1

Vargas-Vargas, C., Costa-Vargas, A., & Montalvo-Pérez, D. (2019) Tests for Dual Diagnosis. In J. L. Matson (Ed.), *Handbook of Intellectual Disabilities. Autism and Child Psychopathology Series* (pp. 401-421). Springer International Publishing. https://doi.org/10.1007/978-3-030-20843-1_23

Vlot-van Anrooij, K., Tobi, H., Hilgenkamp, T. I. M., Leusing, 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,* Article 80. https://doi.org/10.1186/s12874-018-0539-1

Watts, C., Cox, T., & Robson, J. (1983). Morningness-eveningness and diurnal variations in self-reported mood. *Journal of Psychology, 113*, 251–256. https://doi.org/10.1080/00223980.1983.9923583

Wehmeyer, M. L. (2007). *Promoting self-determination in students with developmental disabilities*. Guilford Press.

Wetsch, W. A., Pircher, I., & Lederer, W. (2009). Preoperative stress and anxiety in day-care patients and in-patients undergoing fast-track surgery. *British Journal of Anaesthesia, 103*(2), 199–205. https://doi.org/10.1093/bja/aep136

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. doi:10.1111/j.1365-2788.2011.01440.x

Wieland, J., & Zitman, F. G. (2016). Brief Symptom Inventory symptom profiles of outpatients with borderline intellectual functioning and major depressive disorder or posttraumatic stress disorder: Comparison with patients from regular mental health care and patients with Mild Intellectual Disabilities. *Research in developmental disabilities, 51*, 153-159. https://doi.org/10.1016/j.ridd.2016.01.007

Wilson, K. A., de Beurs, E., Palmer, C. A., & Chambless, D. L. (1999). Beck Anxiety Inventory. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 971–992). Lawrence Erlbaum Associates Publishers.

de Witte, M., Pinho, A. D. S., Stams, G. J., Moonen, X., Bos, A. E., & van Hooren, S. (2020). Music therapy for stress reduction: a systematic review and meta-analysis. *Health Psychology Review*, 1-26. https://doi.org/10.1080/17437199.2020.1846580

de Witte, M., Spruit, A., van Hooren, S., Moonen, X., & Stams, G.-J. (2020). Effects of music interventions on stress-related outcomes: a systematic review and two meta-analyses. *Health Psychology Review, 14*(2), 294–324. https://doi.org/10.1080/17437199.2019.1627897

ten Wolde, A. C., Le Grand, B., Slagter, J., & Storms, M. (2006). *Vaardig en Veilig: Behandeling van sterk gedragsgestoorde licht verstandelijk gehandicapte mensen met risicovol gedrag. Kenmerken van de doelgroep, consequenties voor behandeling en de uitwerking hiervan in gespecialiseerde behandelprogramma's*. [Skilful and Safe: Treatment of severely behaviourally disturbed people with mild intellectual disabilites with challenging behaviour. Characteristics of the target group, consequences for treatment and their elaboration in specialised treatment programmes.]. Hoeve Boschoord.

Wolpe, J. (1969). *The Practice of Behavior Therapy*. Pergamon Press.

World Health Organization. (2010). *Global health diplomacy: Negotiating health in the 21st century*. http://www.who.int/dg/speeches/2008/20081021/en/

Yang, Y., Tang, J., Jiang, Y., Liu, X., Sun, Y., Zhu, Y., & Miao, D. (2011). Development of the Acute Stress Response Scale. *Social Behavior and Personality, 39*(5), 713-720. https://doi.org/10.2224/sbp.2011.39.5.713

Chapter 6

Does adapting a self-report instrument to improve its cognitive accessibility for people with intellectual disability result in a better measure? - a cognitive interview study.

Published as:

Kooijmans, R., van Langen, M., Voss, H., Reichrath, E., Maljaars, J., Dalemans, R., Langdon, P.E. & Moonen, X. (2024). Does adapting a self-report instrument to improve its cognitive accessibility for people with intellectual disability result in a better measure?– A cognitive interview study. *Research in Developmental Disabilities, 154*, 104851.

Abstract

We investigated whether improving the cognitive accessibility of a widely used self-report measure leads to better understanding and more accurate answers in a sample of adults with mild intellectual disability and borderline intellectual functioning.

We undertook a series of cognitive interviews before and after adaptation of the instructions and selected items of an existing self-report measure of adaptive functioning. Interview results and participant feedback were supplemented with quantitative comparisons between participant and carer scores.

Adaptation based on participant experiences and preferences combined with evidence-informed guidelines improved understanding and accuracy. Self-report and carer-report scores showed greater convergence after adaptation; this occurred because people with intellectual disabilities appeared to understand the self-report measure more effectively.

The results show that adaptation of the self-report instrument to suit the needs and preferences of people with mild intellectual disability or borderline intellectual functioning leads to a more accessible measure and more reliable and valid results. Results also highlight the importance of complementing proxy reports with a first-person perspective in assessment as clients and informants may differ in their assessment of behavior and skills.

Introduction

The outcome of (psychological) assessments and diagnostic procedures can have broad implications for the person who is assessed. Outcomes can translate into claims about a person's ability for self-determination, can limit or open up career options, and can determine whether someone can have access to support or benefits (Hanson et al., 2023). It is therefore of the utmost importance that the information gathered about a person is a valid and authentic description of their abilities and needs.

First-person views versus proxy ratings

Traditionally, carers and relatives ('proxies') are a primary source of information about a person with intellectual disability (ID) as they are able to provide reliable and valid information provided they know the individual well (Havercamp et al, 2022). Santoro et al. (2022) proposed that proxies can more accurately recall detailed factual information over longer periods of time than people with intellectual disabilities; for example. when recalling detailed medical histories.

However, the accuracy of proxy reports for *internal states*, such as satisfaction or psychological distress, has been questioned (Emerson et al., 2013). Proxies have been shown to underestimate the quality of life of both children and adults with intellectual disability(Schmidt et al., 2010; Vlot-Van Anrooij et al., 2018; Wilson et al., 2016; Zimmermann & Endermann, 2008) and have a limited ability to accurately recognize symptoms of psychological ill-being and distress (Mileviciute & Hartley, 2015; Scott & Havercamp, 2018).

People with intellectual disabilities sometimes have different views about their abilities and support needs than their carers and family member; for example, people with intellectual disability in assisted living conditions tend to have more positive views about their ability to care for themselves (Fisher et al., 2014), and are more confident in their ability to nurture mutually beneficent relationships and perceive fewer risks of exploitation (Schützwohl et al., 2018). This positive first-person outlook about societal functioning does not necessarily reflect an overestimation of abilities (Snell, 2009). It may also represent an underestimation of the perceived functional status of individuals with a disability by carers and relatives (Nota et al., 2007).

Given the less-than-perfect agreement between the views of proxies and the persons they represent, it is important to complement proxy reports with first-person views as much as possible (Havercamp et al., 2022; Shogren et al., 2021; Walton et al., 2022). Moreover, in line with recent societal movements that advocate emancipation and inclusivity, it is important to prioritize the first-person perspectives of people with intellectual disability

(Kramer & Schwartz, 2017).

Challenges in self-reports for people with intellectual disability

Reporting your own internal states and abilities requires introspection and the aggregation of evaluations of personal functioning across many situations (Shogren et al., 2021). The process of answering self-report questions involves a series of steps including comprehension, retrieval, judgment, and estimation and reporting (Tourangeau & Bradford, 2010). These cognitive processes bring challenges for most people with intellectual disability associated with the nature of their disability, including problems with reasoning, verbal expression, reading, abstract thinking, and judgment (Schalock et al., 2010; American Psychiatric Association, 2013). Difficulties with understanding questions and response options within self-report assessments are considered common amongst people with intellectual disability because they have a tendency to give positive answers or acquiesce to positively worded items when they do not fully understand the question (Emerson et al., 2013).

Kramer and Schwartz (2017) proposed that the *cognitive accessibility* of self-report measures must be improved to overcome these difficulties, so that more people with intellectual disability can meaningfully self-report. They state that "cognitive accessibility is present when assessment design anticipates respondent variability in cognitive abilities and, to the greatest extent possible, reduces cognitive demands and/or supports cognitive processes to enable respondents with a range of cognitive abilities to interpret and respond to assessment items as intended" (p. 1705). A review by Kooijmans et al. (2022) lists practice- and evidence-based recommendations for improving cognitive accessibility, including the use of easy language guidelines, the use of Likert scales with a limited number of response options, and using supportive visualizations that are tailored to the needs and preferences of participants. In the process of developing or adapting measures, the importance of involving people with intellectual disability directly in a co-design approach is emphasized (O'Keeffe et al., 2019).

Aim of the current study: putting the recommendations to the test

When we adapt existing measures to the needs of people with ID, we should ideally investigate how the resulting ID-inclusive measure and original instrument compare (Stancliffe et al., 2017). In practice, acceptable reliability statistics are considered a sufficient testament to the adapted measure's adequacy. Occasionally, developers try to estimate if an adapted measure has improved by comparing outcomes with the original; for example, by comparing the number of non-response items, or more indirectly, by using readability formulas such as the Flesch Reading Ease Test (Flesch,

1948) to compare the readability of parallel versions (see Stancliffe et al., 2014). Existing guidance on how to adapt self-report measures for people with intellectual disability is largely based on theoretical considerations, best practices, and (indirect) evidence from psychometric research (Kooijmans et al., 2022; Walton et al., 2022). However, there are no known published studies that have investigated whether attempts to improve the cognitive accessibility of an adapted version of an existing self-report measure leads to improvements in comprehensibility and validity relative to the original version for people with intellectual disability.

In the current study, we applied evidence- and practice-based recommendations for improving the cognitive accessibility of a widely used diagnostic self-report measure to answer the following questions:

1. Do the adaptations result in a measure that is perceived as less difficult and easier to understand for respondents with mild intellectual disability or borderline intellectual functioning?
2. Do the adaptations result in a measure that produces more reliable and valid answers?

To answer our questions, the results from cognitive interviews and quantitative indicators were compared between original and adapted self-report versions as well as proxy informant versions.

Methods

Design

We collected data at two time points; before (Round 1) and after adaptation (Round 2) of an existing self-report measure of adaptive functioning (see *Instruments*). After round 1, an altered scale was created. In Round 2, the adapted self-report measure was completed, and results from the original and adapted measures were compared.

The research plan and statistical analysis plan were reviewed and pre-registered at ClinicalTrials.gov (registration number NCT05857592).

Cognitive interviewing

We used cognitive interviewing techniques to evaluate the cognitive accessibility of a self-report measure (Miller et al., 2014). In a cognitive interview, participants completed a survey while answering questions about the survey questions and response options. Participants can be asked to reflect by using the 'think-aloud' method of questioning or by using 'probing' questions (Beatty & Willis, 2007). In the think-aloud method, participants are literally asked to 'think aloud' as they are processing the question and tell the interviewer how they arrive at a certain answer. The interviewer is expected to intervene as little as possible and let the participant answer in a free-flow format. This technique requires the ability to access and verbalize cognitive processes, which can be challenging for many people with ID. The probing technique involves a more proactive role for the interviewer, where additional, direct clarifying questions about the respondents' answers are asked. In the current study, we used the verbal probing technique to minimize cognitive burden and to evaluate how difficult certain concepts proved to be for the respondents.

There is no consensus on adequate sample sizes for the evaluation of self-report measures, (Beatty & Willis, 2007), but leaders in the field suggest that cognitive interview samples should constitute anywhere between 10 and 50 participants (Miller et al., 2014). For this study, we aimed to recruit 20 participants.

A concept interview guideline was drafted. Questions pertained to the test instructions, item content and formulation, general difficulty of the survey and suggestions how to improve accessibility. Examples of verbal probes were 'Can you tell me how you would clean your bathroom?', and 'Can you tell me what the word 'deadline' means?'. We tested the concept guideline with an expert-by-experience co-researcher to check if the interview procedure was sufficiently clear and not too strenuous for participants. After the review by the expert-by-experience, we made adjustments to the interview protocol and procedures. We reworded interview questions and instructions to improve comprehensibility and shortened the interview to reduce participant strain. No changes were made to the instructions and item wording of the original measure.

Participants

Interview Participants

In line with our study aim, we intended to include 'people with intellectual disabilities', which by definition includes a very broad range of cognitive and adaptive functioning. In the current study, we included people with both mild intellectual disability (Full-Scale IQ 50-69) and those with borderline intellectual functioning (BIF; Full-Scale IQ 70-84) because there is considerable overlap in support needs and personal and environmental characteristics between people with a mild intellectual disability and borderline intellectual functioning (Nouwens et al., 2017; Vervoort et al., 2021). As participants had to take part in a cognitive interview, we did not include participants with moderate to profound intellectual disability (Hartley & MacLean, 2006). Participants were recruited from assisted living facilities for people with cognitive and adaptive functioning impairments in The Netherlands.

A total of 20 people agreed to participate and their characteristics are found in Table 1. Two participants dropped out; one person indicated he found participation too stressful during the Round 1 interview. The other participant dropped out during the Round 2 interview and did not provide a reason. For both participants, their data were retained for analysis with their consent.

Table 1.

Participant Characteristics

	(N=18)
Age (yrs)	
Mean (SD)	39,1 (15,8)
Range (min – max)	20 – 78
Gender	
Female (%) / Male(%)	4 (22%) / 14 (78%)
IQ[a]	
Mean (SD)	64,4 (11,8)
Range (min – max)	50 - 86

[a] Total N for IQ is 15. For 3 participants (%) their recent total IQ score was unknown or they did not consent to share recent IQ scores.

Co-researchers

An expert-by-experience co-researcher with a mild intellectual disability working at the Ben Sajet Center in The Netherlands participated in the development of the cognitive interview protocol. Three experts-by-experience with a mild intellectual disability at Koraal, a Dutch service provider for people with intellectual disability, helped to edit the wording of the original measure to improve readability and advised on adaptations to make the instrument more cognitively accessible.

Carers

Carers of participants were invited to complete the proxy-version of the adaptive functioning measure. They had to be involved in the participant's direct care for at least one year and needed to have a good understanding of the participant's daily functioning.

Instruments and measures

ABAS-3

We chose the Adaptive Behavior Assessment System - Third Edition (ABAS-3; Harrison and Oakland, 2015) as a suitable measure for adaptation. It is a widely used measure in the care for people with intellectual disabilities worldwide and can be used for self-report and proxy-report assessment.

The manual for the Dutch version specifies that a European CEFR reading level B1 (CEFR, 2024) is required to read the questions independently. Questions can be read aloud to people who have difficulties with reading or the measure can be completed as a semi-structured interview.

For this study, we used the Dutch translation (Kreemers et al., 2020) of the adult version of the ABAS-3 which has excellent internal consistency for the self-report, α = .98, and informant, α = .0.99, total aggregate scale. Confirmatory factor analyses indicated excellent model fit for both a 1-factor model (general adaptive composite) and a 3-factor model (three domain scales: conceptual, social, and practical; Kreemers et al., 2024).

Adaptations to the ABAS-3 for the purpose of this study

The original ABAS-3 is comprised of 238 items which was judged as too many to consider within a cognitive interview format with people with intellectual disabilities. Therefore, we used a representative selection of 30 key items, covering all three conceptual domains. The decision to retain or drop items was based on their psychometric qualities and conceptual relevance. A representative selection of 30 items was chosen based on several criteria, a) all 10 skill areas should be included and each skill area should be represented by three items, b) items should have item-rest correlations with other items in the skill area of at least 0.50 and factor loadings of at least 0.60 with the principal factor for the skill area in question, c) the selected items should reflect both easier and more difficult skills as indicated by variation in mean average scores from low to high mean scores in the normative sample, and lastly, d) items should be relevant to the every-day life of people living in assisted living facilities. To stay as close to the original intended purpose of the instrument as possible, the developers of the Dutch version at the KU Leuven were asked to supervise the abbreviation process. The final selection of items included in this study can be found in Appendix A.

For the purpose of this study, we made some modifications to the original ABAS-3 before the round 1 cognitive interviews. The instructions and response scale were left unaltered, but the response option ‘I don’t know’ was added to give respondents the opportunity to flag difficult items. Following suggestions by the expert-by-experience co-researcher we changed this to ‘This question is difficult’, as this would direct the narrative away from ‘I am not able to answer the question’ to a more neutral ‘this question is not right’. Two pretest questions were inserted before the actual ABAS-3 survey questions to assure that the participant had understood the scoring instructions (see Table 3a).

Adapting the measure

After Round 1, we adapted the original version to improve cognitive accessibility. Adaptations were made based on participant feedback from the Round 1 assessment, quantitative results that indicate the level of comprehension at the individual item and overall level, and by applying evidence-based guidelines for improving the cognitive accessibility of self-report measures by Kooijmans et al. (2022) and Dalemans et al. (2021). The ‘Language for all’ guidelines by Moonen et al. (2022) were used to convert the original instructions, questions, and response options to CEFR A2 reading level (Council of Europe, 2001). The step-by-step process of adaptation is detailed in Table 2. The round 1 findings on which adaptations were based are described in detail in the Results section.

The resulting concept measure was reviewed by the experts-by-experience working group, after which final minor revisions were made to the wording. Three different modalities of visualization (pictograms, drawings and photographs) were presented to eight people with mild intellectual disability or borderline intelligent functioning working at a sheltered workshop. All but one expressed a preference for using photographs accompanying the step 1 response options. Adding symbols to represent frequency for the step 2 response options, for example glasses ranging from empty (never) to full (always), were not considered to be helpful.

An example of a question from the adapted version with the new 2-step response scale can be found in Appendix B.

To ensure that the carers received a version of the measure that was equivalent to the self-report version, the proxy-version for round 2 was identical to the self-report version, including the altered response format. The only deviation from the self-report was that the 30 items were formulated in the third person.

Table 2.

Adaptation Process of the ABAS-3

Input	Information used	Adaptations
Round 1 cognitive interviews	What elements and wording were perceived to be difficult by respondents?	Alternative wording and simplified grammatical structure of instructions and questions.
	Respondent suggestions to improve comprehensibility.	Fewer and easier-to-understand words. Shorter instructions.
		Add pictures to answers.
		Lay-out simplifications.
	Researcher observations.	Alternative wording and simplified grammatical structure of instructions and questions.
		2-step response scale.
Round 1 quantitative data	Questions that have a high 'perceived difficulty score'.	Alternative wording and simplified grammatical structure of questions.
	Questions with low comprehension and 'matching' scores.	Alternative wording and simplified grammatical structure of questions.
		2-step response scale.
Evidence-based guidelines for the adaptation of self-report measures	Kooijmans et al. (2022) evidence-based guidelines for adaptation of self-report measures and Dalemans (2021) guidelines for 'communication-friendly measurement'.	Include 'don't know' or 'not sure' option.
		Lay-out suggestions.
		2-step response scale.
		Include practice items.
		Simplify wording according to evidence-based guidelines (see below).
		Add pictures to questions.

Input	Information used	Adaptations
	'Language for all' guidelines (Moonen et al., 2022).	Adaptation of instructions and questions to CEFR A2 level. Lay-out improvements (fewer questions per page, font size, blank lines and headings between paragraphs). Visualization characteristics.
Expert review	Developers of the Dutch translation of the ABAS-3.	Development of the alternative 2-step response scale.
	Experts-by-experience.	Alternative wording and simplified grammatical structure of instructions and questions. Choice of visualization (photos).

Procedure and data collection

Ethical Review and Consent

This study was reviewed and approved by the Ethics Review Board from the University of Amsterdam (reference number FMG-2567). All potential participants were informed about the study set-up in appropriately worded information pamphlets. All participants who agreed to take part in the study provided written informed consent. Participants were compensated for their time to a value corresponding to minimum wage.

Cognitive interviews

The interviews in round 1 were conducted by authors RK and MvL and took place in July and August of 2023. Interview duration was between 42 and 98 minutes.

Interviewers read all instructions, the 30 selected items, and answer options aloud, even if participants indicated they could read for themselves. This was done to minimize the risk that participants would base an answer on misread information.

First, participants were asked to carefully listen to the test instructions and explain to the interviewer in their own words what they should and shouldn't do. They were encouraged to name as many elements as they could recall. The answers were coded to reflect the number of correctly named elements as a proportion of the maximum number of 10 elements. Participants then scored each item on the 0-3 response scale, after which they answered the probing questions. Interviewers could repeat questions and instructions if requested, but did not provide any instructions or clarification beyond the written instructions in the instrument itself. Whenever a participant indicated difficulties with understanding the question, a ‘difficult question’ score of 1 was recorded and the cause of the difficulty was noted. At the conclusion of the interview, participants were asked to rate the overall difficulty of the questions on a scale from 1 (very easy) to 5 (very difficult) and were asked to provide suggestions on how to improve the cognitive accessibility.

The same authors conducted the round 2 interviews, which were planned for a minimum of three months after completion of the Round 1 interviews to reduce the chances of carry-over effects. The procedure was identical to the round 1 interviews. The Round 2 interviews took place in September and October of 2023 and took between 32 and 70 minutes to complete.

Proxy-reported information

Carers completed the original and adapted 30-item versions of the ABAS-3 informant report without assistance.

Data analysis

Analysis of cognitive interviews

Two assessors (Rk and MvL) independently assessed to what extent the answers reflected the level of understanding and the extent to which the answers and examples to the probing questions matched the numerical score. Scores were attributed according to a preconceived scoring guideline, detailing how the answers to each question should be interpreted.

For the level of understanding, answers were scored as indicating that the participants did *not* get the gist of the question (0), had *some* understanding of the key concepts (1) or seemed to have had a *good* understanding of the question (2). If insufficient information was provided, a missing value was recorded. For the match between answer and score, answers were coded as indicating that the participant awarded himself the correct score (0), the participant chose a score that was *lower* than what the probing questions would suggest (-1), or the participant chose a score that was *higher* than what the probing questions would suggest (1). Missing values were recorded if there was insufficient information to make a judgment. Discrepant coding results were discussed and resolved in consensus meetings.

Quantitative analyses

Composite scores and transformations. For the original, unadapted self-report and carer-report versions, a total mean score across all items and mean scores for the three adaptive domains (conceptual, social, practical) were calculated.

Because the original scoring format of the ABAS-3 proved to be confusing (see *Results*) for participants, the response scale was altered structurally, while trying to maintain its conceptual principles. Specifically, the original 1-option response scale was broken down into two steps, reflecting a more intuitively understandable sequence of scoring. To compare original ABAS-3 scores with adapted ABAS-3 scores, a single composite score was calculated from the two-step response process for each item. A transformation matrix was produced in collaboration with the developers of the Dutch version of the ABAS-3 at the KU Leuven to ensure the resulting composite scores would correspond with the scoring system as originally intended by the developers. For this purpose, all combinations of answers for the two steps were plotted against the response option table provided in the instructions of the original measure (Table 3a). The transformation matrix is presented in Table 3b.

Difficulty scores were calculated for each item, based on the number of people who indicated that they found a particular item hard to understand. The total number of items that were marked as ‘difficult’ across all participants was used as a general measure of difficulty. Additionally, a mean overall difficulty score for the general ‘How difficult were the questions?’ question was calculated by averaging Likert scale (1-5) scores across participants.

For both versions of the ABAS-3 self-reports, the proportion of correctly interpreted items and the proportion of matched scores (score matches descriptions) were calculated.

Difficulty and comprehension for the self-report version. Indicators of difficulty and comprehension for the self-report version of the ABAS-3 were compared between the original and adapted versions using paired-sample t-tests for mean scores and chi-square tests for proportions.

Table 3a

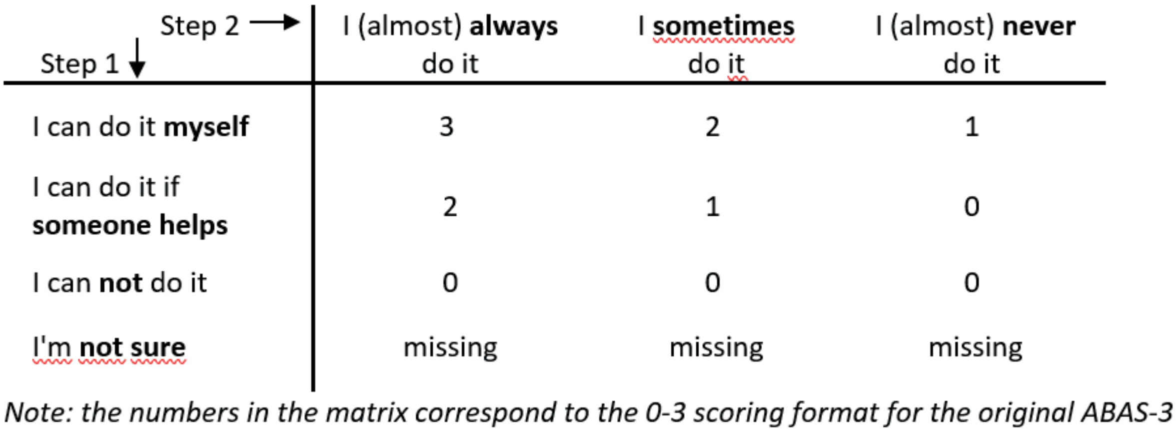
Scoring instructions table from the original ABAS-3 Adult form

Rating	The individual
0 Is not able	• cannot perform the behavior; • has some limitation that prevents performing the behavior; • is too young to have tried the behavior; • does not have the skill to perform the behavior; • has not been taught to perform the behavior; or • has some limitation that prevents performing the behavior.
1 Never (or almost never) when needed	**is able to perform the behavior, but** • never or almost never does it when needed; • never or almost never does it without being reminded; • another person does it for the individual instead of the individual doing it; or • refuses to perform the behavior.
2 Sometimes when needed	**is able to perform the behavior, but** • only does it sometimes when needed; • sometimes does it without help, but sometimes needs help; or • sometimes does it on his or her own, but sometimes needs to be reminded.
3 Always (or almost always) when needed	**is able to perform the behavior, and** • does it most or all of the time without help and without being reminded; or • displayed the behavior at a younger age but has now outgrown it.

Copyright © 2015 by Western Psychological Services, reproduced with permission.

Table 3b

Transformation matrix for transforming scores on the adapted ABAS-3 to corresponding original ABAS-3 scoring



Step 2 → Step 1 ↓	I (almost) **always** do it	I **sometimes** do it	I (almost) **never** do it
I can do it **myself**	3	2	1
I can do it if **someone helps**	2	1	0
I can **not** do it	0	0	0
I'm **not sure**	missing	missing	missing

Note: the numbers in the matrix correspond to the 0-3 scoring format for the original ABAS-3

To provide an objective assessment of reading difficulty, Flesch reading ease scores (Flesch, 1948) and LiNT readability scores (Pander Maat et al., 2023) were calculated for the instructions in the original and adapted versions. Flesch reading ease scores can range from 0 to 100, with higher scores indicating greater reading ease. For the purpose of this study, the Flesch-Douma formula was used. This is an adaptation of the original formula that takes into account semantical and grammatical differences between the English and Dutch language (Douma, 1960). The LiNT readability formula was developed to circumvent conceptual and methodological issues with known readability tools (Begeny & Greene, 2014) and uses insights from contemporary linguistics research. LiNT scores range from 0 (very easy) to 100 (very difficult).

Between-informant comparisons and between-condition interactions. Total, domain, and item scores on the original and adapted versions were compared within dyads (participant – carer) for the original and adapted versions of the ABAS-3. To investigate whether discrepancies between participants and carers differed for the original and adapted versions of the ABAS-3, a two-way mixed-model repeated measures ANOVA was conducted with ‘Informant’ (participant versus carer) as a between-subjects factor and ‘Version’ (original versus adapted) as a within-subjects factor. Between-subjects main effects were tested with post-hoc paired-samples t-tests with Bonferroni corrections to control for family-wise error.

Shapiro-Wilks tests for the normal distribution of residuals of all dependent variables and Levene’s test for equality of variances showed that the assumptions for performing a two-way ANOVA were met.

To assess the level of association between original and adapted ABAS-3 versions, bivariate

Spearman rank correlations correlation was calculated (Table 7). Guidelines by Schober et al. (2018) were used to interpret the strength of the association.

For all comparisons, the corresponding effect sizes were reported; Cohen's d for t-tests, partial η^2 for ANOVAs and Cramér's *V* for Chi square tests.

Results

Round 1 cognitive interviews

Instructions

The Dutch version of the first-page instructions had a Flesch Reading Ease Score of 53, a LiNT readability score of 39, and consisted of 331 words. The instructions proved to be quite difficult to understand for most participants. In the words of one participant: "Too long, too much information, many difficult words. I don't get it.".

When asked to summarize the instructions in their own words, the participants could name only two out of the 10 key elements on average, with a minimum number of 1 element and a maximum of 4 elements correctly named.

Response scale

The response scale proved to be confusing. In the original scale, the respondent is asked to give a 0 to 3 rating for each skill or behavior. The '0' answer ('Not able to do this behavior') is to be chosen if the respondent does not have the ability to show the behavior 'without reminders and without help'. Options 1 to 3 represent a frequency rating. To adequately judge their own ability, a respondent has to combine several different ratings into one. First, they have to ask themselves if they think they are able to do it, then they have to think if they can do it without help and without being reminded, and then how often they would actually do it, but only 'when needed'. The scores and the substantiations respondents gave indicated that they generally collapsed the intricate 4-point scale into a simple frequency rating: 'How often do you do this?'. Many respondents did not factor in the question if they needed assistance to perform the behavior and as a result gave themselves higher scores than they should.

Especially confusing were questions where the respondent has to indicate that they did not engage in an activity and the response scale was reversed. For example, statements starting with "I refrain from…", where respondents are supposed to say they *always* do something to indicate that they *never* actually do it. As in "I *always* refrain from playing computer games because I *never* play them.". These question types received the highest difficulty ratings and were often misinterpreted.

Problematic wording

Most participants handled the passages that were expected to be problematic quite well. They frequently inferred the meaning of the statement from the words they *did* understand and ignored the words or phrases they did *not* understand. Sometimes this caused participants to miss important information, for example when participant NM indicated that he did not know what a 'supervisor' was in the question 'I seek help from a supervisor, as needed, when work-related problems or questions arise', but he nevertheless gave a meaningful answer as it became clear from the probing questions that NM inferred that the item asked if he would 'seek help when there is a problem at work'.

The question that was found to be difficult by the majority of participants was 'I distinguish truthful from exaggerated claims by friends, advertisers, or others', followed by 'I limit time playing computer games or other nonproductive activities.'

Many of the difficulties with ambiguous or difficult wording that are seen regularly in questionnaire design for people with intellectual disability were identified in the ABAS-3, such as problems with double negatives and words that can have more than one meaning. Problems with wording and sentence structure and examples are summarized in Table 4.

Table 4.

Problematic Wording of Original Items

	Item examples	Interpretation problem
Wording		
Figurative vs literal meaning of words	"I make important decisions only after carefully weighing pros and cons, without rushing."	Participant thought the question had something to do with cooking (because of the 'weighing').
	"I write down dates and times for appointments and deadlines."	Participant selected 'Almost never' as a response because she does not write dates down, she *types* them in her agenda.
Infrequent words	Examples of words that were perceived as difficult by some participants: 'adolescents', 'deadline', 'distinguish between', '[to plan something] in logical steps', 'deliberations', 'leisure activities', 'nutrition'.	Participants tended to guess the meaning from context without asking for clarification.

	Item examples	Interpretation problem
Ambiguous meaning and associations	“I commit and cooperate well when I am part of a group or team.”	Some participants associate 'group’ with being part of a group home living arrangement and people who belong to the 'team’, are the people who work at the group home. This is also an example of a 'double-barreled question', asking about more than one concept in a question. One can be productive and not cooperative and vice versa.
	Instruction: “[The ABAS-3] measures important behaviors”	Many residents associated the word 'behaviors’ with 'behavioral problems'. They often have a long history of receiving care on account of 'their behavior’ and therefore think the questions are about how problematic their behavior is.
Sentence structure		
Doubling ('double-barreled questions')	“I show sympathy for others when they are sad or upset.”	'Sad’ and ‘upset’ are not equivalent: “If someone is sad, I will try to comfort him, but if I try to console someone who is upset he may punch me!”
Sentences starting with ‘I refrain...’, ‘I limit...’	“I refrain from saying or doing things that might embarrass or hurt others.”	Participant who says he never does anything to hurt someone puts down a score of ‘Never’ instead of ‘Always’ (“I always refrain myself from...").
	“I limit the time for playing computer games or other nonproductive activities.”	“I Don’t do that sort of stuff, so it's a 'Never’ for me.”

Layout

Some features of the original ABAS-3 caused difficulties. The two-column format of the Instructions page was confusing for some participants, as they did not know where to continue after reaching the bottom of the first column. Secondly, in the Dutch version of the ABAS-3, the response options are displayed at a 90-degree angle, causing many respondents to rotate the questionnaire every time they had to choose an answer. Three participants expressed that they found this to be tiresome and unnecessary and one participant suggested that it should be addressed in the adaptations.

When asked for suggestions on how to improve the layout for the adapted version, four participants suggested that shorter or fewer sentences should be used; “If they are too long, I can’t remember them very well” (participant JE). A suggestion to add pictograms or pictures to the questions or answers for those who have trouble reading was mentioned by four participants.

Round 2 cognitive interviews

Participants generally found the wording of the instructions and questions easier to understand than the original version. Two out of 19 participants indicated they still had trouble understanding the instructions, mainly because of the elaborate scoring instructions. One participant said it was still too much text to remember. Fewer questions were perceived to be difficult. Even though the wording was simplified, some items remained quite challenging: “This is still quite a long sentence don’t you think? Mmm… still a bit hard for me…” (participant JF). Most participants who could remember the first time they completed the original ABAS-3 questions indicated that we did a good job in making the questionnaire easier, “I think this way we can make it possible for everyone to fill in the questions by themselves.” (participant ML). The 2-step response scale seemed to be understood more intuitively, with participants taking less time to select a score and asking fewer clarifying questions.

Although we did not formally assess how strenuous the participants found it to complete both measures, faster completion times and comments by participants illustrated that it required less effort to complete the adapted measure: “This was much more doable than the last time we did this!” (participant JU).

Quantitative analyses

Is the adapted version perceived to be less difficult than the original version?

The written first-page instructions of the adapted-version instructions had a Flesch Reading Ease Score of 86, while the original version had a score of 53. The LiNT readability score was 15 for the adapted version, and 39 for the original version. The scores on both measures confirmed that the adapted instructions had a higher readability rating. The adapted instructions were markedly shorter at 263 words related to the 331 words within the original version.

Questions were almost three times less likely to be marked as ‘difficult’ within the adapted version; 16 questions were marked as difficult in the adapted version versus 44 in the original version. Questions marked as ‘difficult’ most frequently were all part of the ‘Conceptual’ domain. The overall difficulty rating for the adapted version was significantly lower than the difficulty rating of the original version, t (15) = 2.179, p = .046, d = .55 (see Table 5).

Is the adapted version better understood than the original version?

Indicators of difficulty and comprehension for the original and adapted self-report versions are found in Table 5. Participants recalled more elements of the instructions correctly in the adapted version (M = 3.22, SD = 1.80), compared to the original version (M = 1.76, SD = 1.15), t (16) = -4.769, p < .001, d = 1.21.

The comprehensibility of items within the adapted version was significantly greater than items within the original version, X^2 (2, N=432) = 24.26, p < .001, Cramér’s V = 0.17. A significantly higher frequency of scores given by the participant using the adapted version were congruent with their elaboration about the item during the cognitive interview relative to the original version, X^2 (2, N=412) = 321.99, p < .001, Cramér’s V = 0.63.

Table 5.

Overall (Perceived) Difficulty and Comprehension

	Original self-report (N=18)	Adapted self-report (N=17)	
Difficulty			
Flesch reading ease score of the Instructions	53	86	Min 0 - max 100; *higher* scores indicate greater reading ease
LiNT readability score of the Instructions	39	15	Min 0 - max 100; *lower* scores indicate greater reading ease
Number of words in the Instructions	331	263	Word count
Number of questions marked as 'difficult'	44	16	Total number of times a question was marked as difficult; min 0 - max 270
“How difficult did you find the questions?”	2.24 (1.1)	1.57 (0.9)	Average score across respondents; min 0 - max 5; higher scores indicate higher difficulty
Comprehension			
Instruction			
Instruction elements correct	2.0 (1.0)	3.7 (1.4)	Mean (SD) number of elements in the instructions correctly recalled; 0 min - 10 max
Items			
Comprehensibility			Researcher assessment of item comprehension based on his/her elaborations, examples, and other cognitive interview question outcomes
Did not understand	7%	2%	
Some understanding	28%	12%	
Good understanding	65%	86%	
Congruence			Researcher assessment if participant assigned the correct score based on his/her elaborations, examples, and other cognitive interview question outcomes
Self-reported score lower than description	5%	4%	
Self-reported score matches description	71%	88%	
Self-reported score higher than description	24%	8%	

Table 6.

Means, Standard Deviations and Cronbach's Alpha's Original and Adapted Versions of Self-Report and Caregiver-Report

	Original Version (N=18)					Adapted Version (N=16)				
	Self-report	Carer	Difference			Self-report	Carer	Difference		
	Mean (SD)	Mean (SD)	*t*(17)	*P*	Cohen's d	Mean (SD)	Mean (SD)	*t*(15)	*P*	Cohen's d
Total score	2.35 (0.41)	1.70 (0.63)	5.67	< .001*	1.32	2.08 (0.70)	1.84 (0.76)	1.76	.097	0.37
Conceptual Domain	2.21 (0.52)	1.63 (0.71)	4.48	< .001*	1.06	2.09 (0.72)	1.87 (0.66)	1.49	.156	0.37
Social Domain	2.28 (0.49)	1.54 (0.84)	3.59	.002*	0.85	1.98 (0.81)	1.69 (0.92)	1.28	.220	0.31
Practical Domain	2.51 (0.48)	1.82 (0.63)	5.55	< .001*	1.35	2.11 (0.81)	1.86 (0.60)	1.63	.123	0.39
Cronbach's alpha (total scale)	0,68	0,86				0,87	0,85			

** significant with a set at .006 (= .05 / 8) to control for family-wise error*

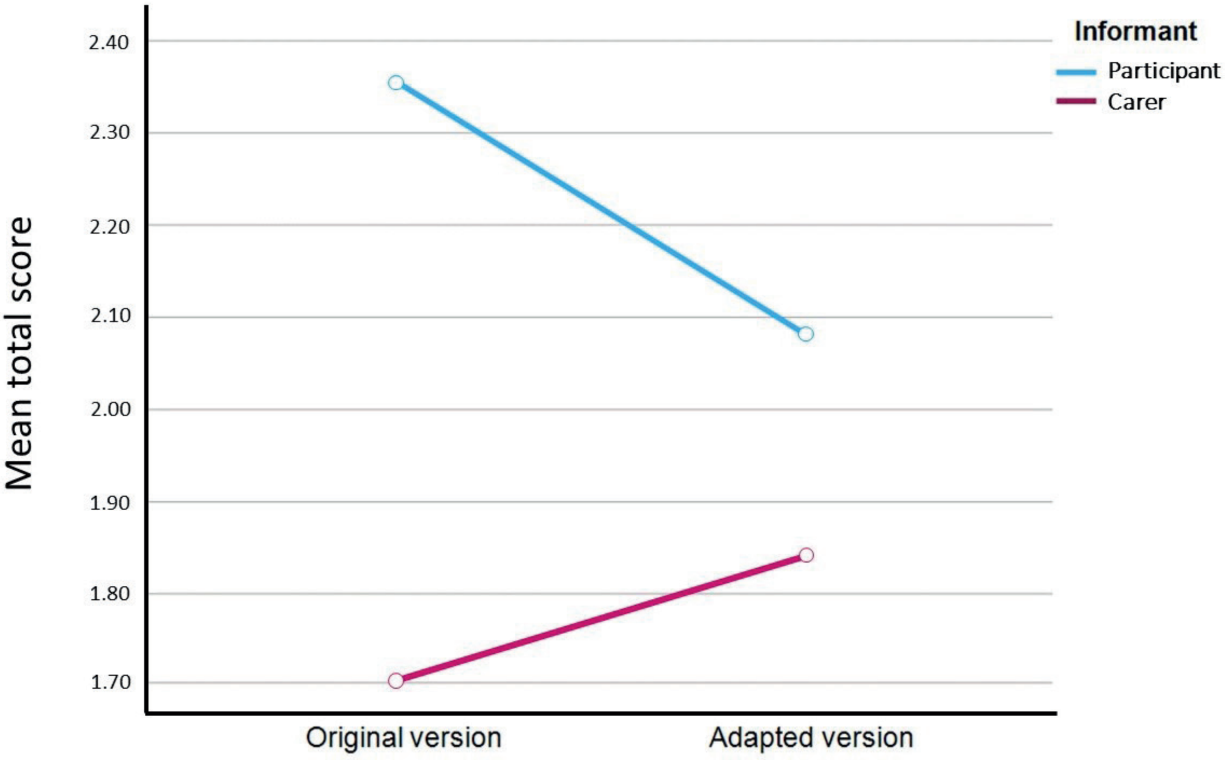
Comparing the original and adapted versions

Means and standard deviations for all four modalities of the questionnaire (self-report and carer scores for the original and adapted version), *t*-values, Bonferroni-corrected significance levels for post-hoc analyses of self-report – carer-report differences, effect sizes (Cohen's d) and internal consistency coefficients can be found in Table 6.

A repeated measures ANOVA was conducted to examine the effect of version (original versus adapted) and type of informant (self-reported versus carer-reported) on the mean total adaptive functioning score for 18 participants. Neither of the within-subject factors had a significant main effect on adaptive functioning scores. The version-by-informant interaction was significant, $F(1) = 5.71$, $p = .023$, $\eta^2 = .15$. Inspection of the profile plot indicated that self-reported and carer-reported adaptive functioning scores converged after the measure was adapted. After adaptation, there was more agreement between participants and carers about the level of adaptive functioning of the participants. This convergence is mainly explained by a significant decrease in self-reported adaptive functioning scores, $t(16) = 2.80$, $p = .006$, d = 0.68. The difference in carer scores between original and adapted versions was not significant.

Figure 1

Profile Plot Interaction Version vs Informant



The internal consistency of the carer versions of both the original and adapted versions was good. For the self-report version the internal consistency was questionable for the original version, but this improved to good after the items were adapted (Cronbach, 1951).

Intercorrelations

Correlations between the adapted and original carer- and self-report versions of the ABAS were moderate, with the exception of correlation between the adapted and original self-report versions, which was strong, ρ (15) = 0.90, $p < .001$.

Table 7.

Correlation matrix: Spearman rank correlations between ABAS-3 scores

Variable	1	2	3	4
1. Mean total ABAS-3 score self-report - original	-			
2. Mean total ABAS-3 score informant report – original	.64**	-		
3. Mean total ABAS-3 score self-report – adapted	.90***	.60*	-	
4. Mean total ABAS-3 score informant report – adapted	.61*	.62**	.61**	-

* $p < .05$ ** $p < .01$ *** $p < .001$

Discussion

In this study, we investigated if improving the cognitive accessibility of a self-report instrument leads to better understanding and more valid answers in a sample of adults with mild intellectual disability and borderline intellectual functioning. A selection of 30 items from the ABAS-3, a widely used instrument to measure adaptive functioning, was used in the study.

Adapting the instrument

The existing measure was adapted, based on participant experiences and preferences combined with evidence-informed guidelines (Dalemans et al., 2021; Kooijmans et al., 2022; Moonen et al., 2022). Although participant suggestions and guidelines overlapped for the most part, the insights offered by participants and experts-by-experiences proved to be crucial in the process of conceptualization and adaptation. Experts-by-experience helped researchers to maintain a balance between study output and participant burden. They also helped refine suggested adaptations by proof-testing concept versions with the researchers and suggest alternative words and formulations. Participant experiences were key in translating abstract guidelines such as ‘simplify wording and grammatical constructions’ to concrete examples. For instance, by pinpointing the exact words and text elements that required reformulation and the grammatical structures that caused the most confusion or ambiguity.

Improved cognitive accessibility results in more accurate self-assessment

Adapting the instrument demonstratively improved the cognitive accessibility of the measure. Participant difficulty ratings decreased, the number of instruction recalled correctly increased, and a higher proportion of items were understood correctly. Improved understanding of the questions and response scale may have led to a more accurate self-report of ability. This result can be explained, at least in part, by evidence that a poor understanding of questions and responses leads to acquiescent responding (Emerson et al., 2013), which can translate to overly positive scores on positively formulated items. Another factor that contributed to a change in scores was likely the use of a clearer 2-step item scoring approach as participant overlooked the fact that lower scores should be awarded if a person needs help to do something using the original response scale (see Appendix B).

Differences and convergence between client and carer scores

When total adaptive scores were compared between the original and adapted versions for both participants and carers, participants rated their own adaptive abilities significantly higher than carers did on the original version of the ABAS-3. After adaptation, participants and carers scores were more similar and this difference was no longer significant. This was due to changes in the responses given by participants rather than changes in the way carers answered items; participants’ scores decreased significantly, whereas carer scores remained more or less constant. Research has found that service providers and family members have a tendency to underestimate the perceived functional status of individuals with a disability (Nota et al., 2007). Other research proposed that differences in perceived abilities between people with intellectual disabilities and proxies were caused by a tendency of people with intellectual disabilities to overestimate their own competence (Snell et al., 2009; Golubović, & Škrbić, 2013). The results from the current study suggest that this may not be caused by poor judgment of the person with intellectual disability’s own ability, but may - at least in part - be caused by a lack of understanding of the questions.

Looking at the rank correlations between versions and informants, a somewhat surprising finding was that the correlation between the original and adapted version was only moderate for carers. This suggests that the relative ordering of assessments has changed between the first and second assessment. Looking into the data in more detail revealed that for most clients the relative order had changed little, but for three clients, the order had changed quite dramatically. Removing these three clients from the analyses increased the rank correlation from 0.62 to 0.83. No plausible explanation why there was such a big difference for these three clients were found. Because the data were processed anonymously, there was no way to ask the carers who submitted the assessment to help explain this finding.

Limitations

For the quantitative analyses, the design was slightly underpowered, increasing the chances of type II errors. Replication of the quantitative part of this study with a larger number of participants is needed to validate the current findings and ensure the robustness of the results.

Another design feature that may have impacted the results from the statistical analyses is the way we operationalized the 2-step response scale for the adapted version. Although breaking down a complex single-step response scale into more manageable elements is suggested to make the response process easier to understand for people with intellectual disability (Ramirez & Lukenbill, 2007), the resulting scale and its transformation of 2-step to 1-step scores were not tested for equivalence; this could be considered in a future study. On the other hand, the main objective was to look at relative informant differences or convergence of scores between informants, and the score pattern does not lead us to believe they are not equivalent.

In regard to the repeated-measures design, learning effects may have contributed to the more favorable difficulty ratings from the participants in the Round 2 assessment and interviews. We purposely planned three months between the two rounds of interviews to minimize the chances of carry-over effects. Still, some participants remembered the preceding interview in detail when they were interviewed the second time, which may have contributed to their perception that the adapted versions was easier to understand, because of a learning effect. In this study, controlling for potential order effects by reversing the order of assessment for half of the population was not possible because the adaptation process was based in large part on the results from the first round cognitive interview. In a subsequent study, the effect of learning and sensitization could be examined and controlled for in the statistical analyses by balancing the order of assessment.

Finally, adjustments were made based on participant suggestions. An example being the mode of supportive visualization used to accompany the response scale. Although this may lead to a measure that suits this particular research sample's *preferences*, it may not necessarily mean that integrating participant preferences improved *understanding*. In this study, the adaptation was a ‘package deal’, incorporating a mix of evidence-informed and participant-informed actions. We cannot therefore make inferences about the differential effect of individual elements.

Conclusions

Adaptation of a self-report measure to promote cognitive accessibility for participants with mild intellectual disability or borderline intellectual functioning improved understanding and decreased perceived difficulty. Improved cognitive accessibility appeared to result in more accurate self-assessment, better agreement between participants and carers and improved internal consistency of the resulting measure. The results of this study cast doubt on the validity of the norms currently used for self-report assessment instruments. These norms are based on scores collected from participants who may have had trouble understanding the questions. This is of particular concern for measures whose outcomes have serious real-life consequences, for example in allocating support resources based on self-reported support needs.

Aside from improved reliability and validity, improved comprehensibility may promote attention to items and reduce fatigue. This allows people with intellectual disabilities to actively contribute to an assessment of their needs, abilities, preferences, and wellbeing. We urge researchers and practitioners working with people with mild intellectual disability and borderline intellectual functioning to make use of evidence-informed guidelines and participant experiences when adapting or constructing measures. The current study shows that the two combined can greatly improve the cognitive accessibility, and hence the reliability and validity of results, of any self-report measure they may use in their practice.

Acknowledgements

The authors wish to thank all the lovely people at Koraal, Ons Tweede Thuis and Cordaan for agreeing to participate in this study. We also wish to acknowledge the invaluable advice offered by our experts-by-experience at the Ben Sajet Center and Koraal.

References

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

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287–311. https://doi.org/10.1093/poq/nfm006

Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials?. *Psychology in the Schools*, *51*(2), 198-215.

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power. *Psicothema*, 35(1), 21-29.

Boeije, H., & Willis, G. (2013). The cognitive interviewing reporting framework (CIRF): Towards the harmonization of cognitive testing reports. *Methodology, 9*(3), 87–95. https://doi.org/10.1027/1614-2241/a000075

Common European Framework of Reference for Languages (2024). *The CEFR Levels*. https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Council of Europe (2001). *Common European Framework of reference for languages: learning, teaching, assessment*. Cambridge University Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Dalemans, R., Stans, S., Von Helden-Lenzen, S., Lemmens, J., Ummels, D., & Cornips, S. (2021). *Guideline for communication-friendly measurement*. Zuyd University of Applied Science. https://www.zuyd.nl/binaries/content/assets/zuyd/onderzoek/factsheets/aenp-guideline-communication-friendly-questionnaires.pdf

De Vries, H., Elliott, M. N., Kanouse, D. E., & Teleki, S. S. (2008). Using pooled kappa to summarize interrater agreement across many items. *Field methods*, *20*(3), 272-282.

Douma, W.H. (1960). De leesbaarheid van landbouwbladen: een onderzoek naar en een toepassing van leesbaarheidsformules [The readability of agricultural magazines: a study into and application of readability formulas]. *Bulletin, 17*.

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 Disability, 51* (5), 333–348.

Fisher, M. H., Mello, M. P., & Dykens, E. M. (2014). Who reports it best? A comparison between parent-report, self-report, and the real life social behaviors of adults with Williams syndrome. *Research in Developmental Disabilities, 35*(12), 3276–3284. https://doi.org/10.1016/j.ridd.2014.08.011

Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*, 221–233.

Golubović, Š,& Škrbić,R. (2013). Agreement in quality of life assessment between adolescents with intellectual disability and their parents. *Research in Developmental Disabilities, 34*, 1863–1869. http.doi.org/10.1016/j.ridd.2013.03.006

Hanson, S. L., Bruyere, S., Forber-Pratt, A., Reesman, J., & Sung, C. (2023). Guidelines for assessment and intervention with persons with disabilities: An executive summary. *American Psychologist*.

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*(11), 813–827. Https://doi.org/10.1111/j.1365-2788.2006.00844.x

Havercamp, S. M., Barnhill, L. J., Bonardi, A., Chapman, R. A., Cobranchi, C., Fletcher, R. J., Rabidoux, P., Seeley, J. R., & Tassé, M. J. (2022). Straight from the horse's mouth: Increasing self-report in mental health assessment in individuals with intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 35*(2), 471–479. https://doi.org/10.1111/jar.12952

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

Kramer, J. M., & Schwartz, A. (2017). Reducing Barriers to Patient-Reported Outcome Measures for People With Cognitive Impairments. *Archives of Physical Medicine and Rehabilitation, 98*(8), 1705–1715. https://doi.org/10.1016/j.apmr.2017.03.011

Kreemers, B., Briers, V., Maljaars, J., Storms, G., Maes, B., & Noens, I. (2024). ABAS-3 for Belgium: standardisation, psychometric properties and guidelines for fair assessment. *International Journal of Developmental Disabilities*, 1-11.

Kreemers, B., Maljaars, J., Briers, V., Maes, B. & Noens, I. (2020). ABAS-3, De langverwachte Nederlandstalige vragenlijst voor adaptief gedrag [ABAS-3, the long-awaited Dutch questionnaire for adaptive behavior]. *Caleidoscoop, 32 (5)*, 45-47.

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

Research, 59(2), 158–169. https://doi.org/10.1111/jir.12075

Miller, K., Willson, S., Chepp, V., & Ryan, J. M. (2014). Analysis. In K. Miller, S. Willson, V. Chepp, & J. Padilla (Eds.), *Cognitive interviewing methodology* (pp. 35–50). John Wiley & Sons.

Moonen, X. (2022). *Handleiding SCAF* [SCAF manual]. Hogrefe Publishing.

Moonen, X., Reichrath, E., Cortenbach, K., Wittelings, T. (2022). *Schrijven in Taal voor allemaal, versie 5* [Writing in Language for all, version 5]. Koraal and Toetsen meten & weten.

Nota, L., Ferrari, L., Soresi, S., & Wehmeyer, M. (2007). Self-determination, social abilities and the quality of life of people with intellectual disability. *Journal of intellectual disability Research*, *51*(11), 850-865.

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

Pander Maat, H., Kleijn, S., & Frissen, S. (2023). LiNT: een leesbaarheidsformule en een leesbaarheidsinstrument [LiNT: a readability formula and a readibility tool]. *Tijdschrift voor Taalbeheersing*, *45*(1), 2-39.

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. https://doi.org/10.1016/j.ridd.2006.01.001

Reichrath, E., & Moonen, X. (2022). Assessing the effects of Language for all. *Nordic Journal of Linguistics*, *45*(2), 232-248.

Santoro, S. L., Donelan, K., & Constantine, M. (2022). Proxy-report in individuals with intellectual disability: A scoping review. *Journal of Applied Research in Intellectual Disabilities, 35*(5), 1088–1108. https://doi.org/10.1111/jar.13013

Schmidt, S., Power, M., Green, A., Lucas-Carrasco, R., Eser, E., Dragomirecka, E., & Fleck, M. (2010). Self and proxy rating of quality of life in adults with intellectual disabilities: Results from the DISQOL study. *Research in Developmental Disabilities, 31*(5), 1015–1026. https://doi.org/10.1016/j.ridd.2010.04.013

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia, 126*(5), 1763-1768.

Schützwohl, M., Voß, E., Salize, H. J., Stiawa, M., Puschner, B., & Koch, A. (2018). Self-and proxy-rated needs in adults with mild to moderate intellectual disabilities: Perspective matters. *Journal of Applied Research in Intellectual Disabilities, 31*(2), 285-295.

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.

Shogren, K. A., Bonardi, A., Cobranchi, C., Krahn, G., Murray, A., Robinson, A., Havercamp, S. M., Andridge, R., Arnold, L. E., Barnhill, J., Bodle, S., Boerner, E., Bonardi, A., Bourne, M. lou, Brown, C., Buck, A., Burkett, S., Chapman, R., Cole, C., ... Witwer, A. (2021). State of the Field: The Need for Self-Report Measures of Health and Quality of Life for People With Intellectual and Developmental Disabilities. *Journal of Policy and Practice in Intellectual Disabilities, 18*(4), 286–295. https://doi.org/10.1111/jppi.12386

Simões, C., & Santos, S. (2016). The quality of life perceptions of people with intellectual disability and their proxies. *Journal of Intellectual & Developmental Disability, 41*(4), 311-323.

Snell, M. E., Luckasson, R., Borthwick-Duffy, W. S., Bradley, V., Buntinx, W. H., Coulter, D. L., ... & Yeager, M. H. (2009). Characteristics and needs of people with intellectual disability who have higher IQs. *Intellectual and Developmental Disabilities, 47*(3), 220-233.

Stancliffe, R. J., Wiese, M. Y., Read, S., Jeltes, 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. https://doi.org/10.1111/jar.12358

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. https://doi.org/10.1111/jir.12024

Tourangeau, R. & Bradburn, N.M. (2010). The psychology of survey response. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of Survey Research* (pp. 315–346). Emerald Publishing.

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*(1). https://doi.org/10.1186/s12874-018-0539-1

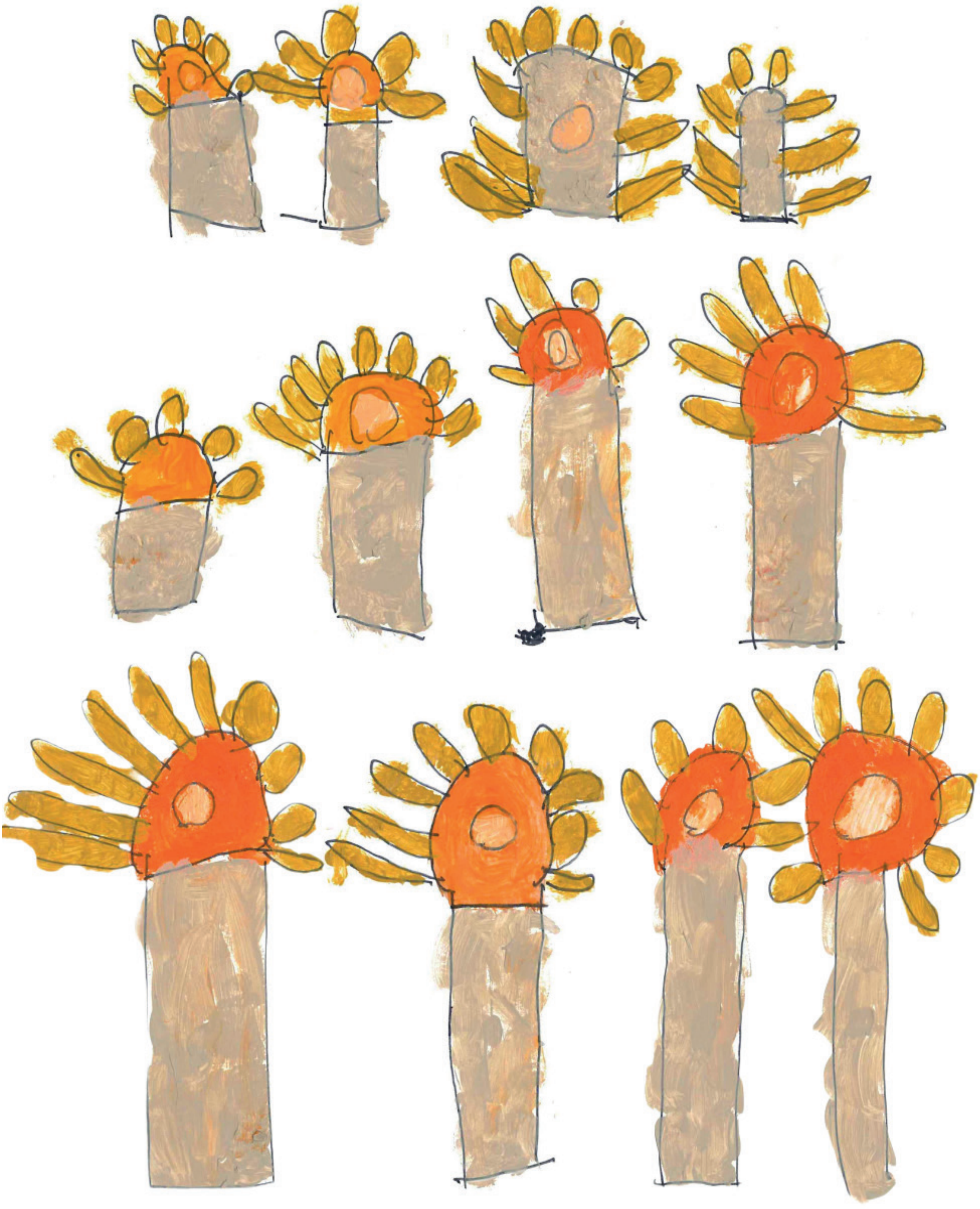
Walton, K., Krahn, G. L., Buck, A., Andridge, R., Lecavalier, L., Hollway, J. A., Davies, D. K., Arnold, L. E., & Havercamp, S. M. (2022). Putting "ME" into measurement: Adapting self-report health measures for use with individuals with intellectual disability. *Research in Developmental Disabilities, 128*. https://doi.org/10.1016/j.ridd.2022.104298

Wilson, K. S., Wiersma, L. D., & Rubin, D. A. (2016). Quality of life in children with Prader Willi Syndrome: Parent and child reports. *Research in Developmental Disabilities, 57,* 149–157. https://doi.org/10.1016/j.ridd.2016.06.016

Zimmermann, F., & Endermann, M. (2008). Self-proxy agreement and correlates of health-related quality of life in young adults with epilepsy and mild intellectual disabilities. *Epilepsy and Behavior, 13*(1), 202–211. https://doi.org/10.1016/j.yebeh.2008.02.005

Chapter 7

Summary and Discussion



The central aim of this thesis was to investigate which factors promote the effective completion of self-report measures by people with intellectual disabilities. Throughout the course of this PhD project, two recurring themes were identified and explored: cognitive accessibility of self-report measures and the interpersonal dynamics of the assessment. In the context of the current work, cognitive accessibility refers to the extent to which assessment design takes into account the cognitive and communication challenges associated with intellectual disability. In addition to issues around assessment design, this project recognizes that interpersonal dynamics affect the outcomes of self-reported research.

This thesis includes work to assess the scientific state of the field in respect to which modifications to self-report measures improve cognitive accessibility and how to take into account or prevent the influence of interpersonal dynamics on results (**Chapter 2**), attempts to expand the evidence base about these two themes (**Chapter 3** and **Chapter 4**) and, finally, applies existing and new knowledge to assess the suitability of self-report measures for people with intellectual disability (**Chapter 5** and **Chapter 6**).

In the following paragraphs, main findings of the previous chapters are summarised, followed by more in-depth considerations and recommendations for clinical practice and future research and development.

Summary of findings

Assessing the evidence base

To date, researchers in most studies have relied on anecdotal evidence, unsubstantiated claims or the findings of a few studies to inform the nature or degree of adaptation of self-report measures for use with people with intellectual disabilities. There have been some previous literature reviews where authors have attempted to summarise findings from different studies to effectively inform how self-report measures can be adapted, but these focused on specific issues or constructs, such as acquiescence (Finlay & Lyons, 2002), or the use of Likert scales (Hartley & MacLean, 2006). Others have conducted narrative reviews that were not systematic (Finlay & Lyons, 2001), including more recent attempts to summarise findings and produce a list of recommendations (Bell et al., 2018).

As a consequence, the first objective of this PhD project was to systematically review the research literature about evidence-based guidance about how to create cognitively accessible self-report measures for people with intellectual disabilities. In **Chapter 2**, 'The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review', a systematic review methodology was used to search the peer-reviewed research literature on this topic from the year 1996 onwards. The methodological quality of included studies was appraised with the Mixed Methods Appraisal Tool (MMAT;

Hong et al., 2018). The quality and strength of the evidence was assessed with the GRADE-CERQual tool (Munthe-Kaas et al., 2018). Most recommendations were based upon only a few studies or clinical and research expertise. There were methodological problems with a substantial proportion of the studies included. The evidence to support most of the resulting recommendations was graded as ‘low’ or ‘moderate’, with very few being given a ‘high’ confidence rating.

The results were presented in a Summary of Findings table, according to a five-stage model of instrument development, from item creation to ongoing development. The more robust recommendations pertained to involving people with intellectual disabilities directly in the creation process, avoiding certain types of words and phrases, and using certain types of answer categories, ‘don’t know’ answer categories, pretests or practice items to establish respondent competence, and certain procedural issues to minimize bias.

It was suggested that the recommendations with a moderate or high confidence rating could at least provide preliminary guidance for developers and researchers. In addition to *assuming* cognitive accessibility based on the application of evidence-based guidelines, it was recommended that researchers must involve a representative sample of intended respondents in the process of creation and evaluation to test and refine the measure.

Finally, several areas for continued research were identified. Examples of concrete topics that needed further exploration were an operationalisation of what constitutes adequately simplified language, what types of visualization should be used to support written text, and ways to detect and prevent bias.

Expanding the evidence base

The results of the review revealed a myriad of topics that required further exploration. However, Two topics were investigated as part of this PhD project. In **Chapter 3** and **Chapter 4**, two of the under-researched topics that emerged from the systematic review were addressed. The topic of ‘bias’ was chosen as starting point for the first study because this was directly related to our own experiences working with young people with intellectual disabilities (see **Preface**).

Assisting children to complete self-report measures introduces bias

In **Chapter 3**, ‘Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people’s views’, we looked into response bias that originates from respondent-interviewer interactions in a residential youth care facility. The research focussed on the assumption that satisfaction ratings were inflated because most youths were assisted by carers when completing the survey. To test this hypothesis, 120 children and youths (aged 11-23 years) with mild intellectual disabilities and borderline intellectual functioning were randomly assigned to one of three conditions: responding (a) unassisted, (b) assisted by their care worker, or (c) assisted by a research assistant. Scores in each condition were compared quantitatively. Youth who were assisted by carers had significantly higher satisfaction scores than youth who were assisted by an impartial researcher and those who completed the survey unassisted.

In successive focus groups with 17 children and youth, the results and possible explanations for the findings were discussed. The focus group participants offered surprising insights in the dynamics between carers and pupils who stay in residential care, that could help explain the observed differences in scores between conditions. The first was *functional dependency*; many youths stated that they depend upon their carers to arrange things for them; for instance, planning a school trip or weekend away from the group home with their parents. Maintaining a positive working relationship with their carer was seen as helpful, and being critical of their carer may damage the relationship. Some youths expected their carer not to invest as much in their wellbeing if they have negative things to say about their carers' functioning. In a similar vein, some youths exhibited a *fear of retribution*. They feared spiteful reactions of a care worker if they were overly critical. They thought they may be viewed as insolent or ungrateful. Some youths showed *submissive tendencies*; they felt they are expected to give what they think might be the 'right' answer. A final mechanism, which had not been described in the literature prior to the current study, was *empathy;* some youths felt sorry for care workers if they were to be criticised or spoken about negatively.

The participating youths stressed that these mechanisms are not universally valid for all youths in residential care. Some are more susceptible to power imbalances or may be less resilient to perceived pressure than others. Some youths stated they had a very trusting relationship with their carer and would have no trouble completing a survey on sensitive topics in their presence, whereas others had no intention to share any information with their care whatsoever. Central to the recommendations was the promotion of self-agency for participants; instead of arranging assistance for each participant, researchers or clinicians should *ask* if they require or wish to be assisted. If there is doubt that a young person can provide responses independently, consider answering the initial questions together to see how they manage. Most well-designed measures will have pre-test practice questions. If the participant needs or wants assistance, ask who they would like to help them. Depending on the circumstances and subject, this may also be a peer or parent. If the respondent has no clear idea of who could help them or if it is suspected that he or she feels obliged to ask their carer, it is best to suggest that an impartial assistant with no prior relation to the respondent could help.

Do pictures improve text comprehension?

The second topic that needed further addressing was decided upon in collaboration with Ruth Dalemans from Zuyd University of Applied Sciences, with whom Koraal collaborates in a lectureship on Accessible Communication. Together with Ruth Dalemans, we identified the lack of specific guidelines about how to use visualisations with easy-to-read text to be a research priority. Although visualisations frequently accompany easy-to-read text to improve comprehension, there is little empirical evidence that they actually improve comprehension.

In **Chapter 4**, 'Does adding pictures to easy-to-read texts benefit comprehension for people with reading difficulties? A meta-analytic review', the research literature on this topic was reviewed. The results of quantitative experimental studies that investigated if easy-to-read texts were easier to understand if the text was accompanied by pictures were aggregated. For this study, not only studies involving people with intellectual disabilities were included. Inclusion criteria also considered studies with people who had reading difficulties for other reasons. Eight studies met eligibility criteria. Four studies included patients with aphasia, three studies included people with intellectual disabilities, and one study addressed participants who were less literate because they learned English as a second language. From the eight studies, 13 effect sizes were extracted and analysed in a 3-level meta-analysis, following PRISMA guidelines. The quality of included studies was assessed by using the RoB-2 risk of bias assessment. A sixth domain was added to the standard RoB-2 domains to rate the quality of the visualizations used within studies. Most studies had poor methodological quality while the visualizations used were diverse, ranging from simple line drawings to colour photographs. The justification for the choice of visualization was generally unclear. The most frequently encountered problem was that researchers tended to select the pictures they deemed appropriate themselves or with the help from 'experts', without consulting members of the different target populations.

The findings from the meta-analysis did not support the assumption that adding visualizations to easy-to-read text improved understanding for people with reading difficulties. The overall effect size was small (g = 0.14) and not significant. Subgroup analyses showed no demonstrable differential effect for different subgroups based on aetiology of the reading difficulty. Subsequent equivalence tests showed that, although it is exceedingly unlikely that any true population effect for any of the subgroups was *positive* above and beyond a 0.50 effect size, it was also unlikely that a true negative population effect greater than 0.50 (a medium effect) exists, based on the studies included in this meta-analysis. Moderator analyses showed that mode of visualization (drawing versus photograph) nor grade level of the text (grade level 1-4 versus grade level 5 or 6) produced contrasting results. Sensitivity analyses showed that removing studies with low overall methodological quality or removing studies with low visualization quality from the analyses did not impact the results in any way. In the Discussion, possible reasons for the absence of an effect were considered including participant characteristics (people with intellectual disabilities may genuinely experience sensory or working memory overload when processing two sources of information simultaneously), study quality (flawed designs lead to invalid results), visualization quality (inappropriate pictures do not help) and selection bias (people who really *need* visualization were generally excluded because they could not read very well).

In the recommendations, two themes were addressed that seemed to recur throughout this PhD project. First, what works for one does not always work for all. Amongst others, personal preferences, cognitive abilities, familiarity with the visualization system (Dalemans et al., 2021) and the level of 'pictorial competence' (DeLoach et al., 2003) of a person determine what type of visualization may or may not work for them. There is likely no one-size-fits-all solution; even within an intended target group, there are large differences in functioning and competence between individuals. Second, the importance of inclusive research practices to address the question of what works are needed. More attention should be paid to including a group of representative participants in the design of the study. Testing the appropriateness, clarity and acceptability of the visualizations before the execution of the actual study should always be a part of the process.

In the light of the limitations above, one of the more unsatisfying conclusions may be that the quality and nature of the included studies is poor and it is therefore currently not possible to draw firm conclusions about whether adding visualizations to easy-to-read text improves comprehension. In the absence of clear and evidence-informed guidance, existing guidelines on accessible communication should be clear about this uncertainty when recommending the inclusion of visualisation alongside easy-to-read text.

Applying new and existing evidence

Although there were still many questions to be answered after our attempts to expand the evidence base, it was agreed upon that there was a large enough body of evidence to put the guidance derived from the earlier studies to the test. First, we investigated if the general knowledge of 'what works' could be used to help determine the suitability of self-report instruments for use in research and practice with people with intellectual disabilities (**Chapter 5**). For the final study of this PhD project, a widely used self-report instrument was adapted to improve its cognitive accessibility for people with intellectual disability. We made adaptations based on the evidence-based guidelines of our review (**Chapter 2**), the findings from **Chapters 3** and **4**, the guideline for communication-friendly measurement by Dalemans et al. (2021) and the 'Language for all' guidelines by Moonen, Reichrath, et al. (2022). The assumption that this would lead to a measure that would enable more people with intellectual disabilities to meaningfully participate in assessment was tested with mixed-methods empirical research **(Chapter 6**)**.**

Using existing guidance to assess a measure's suitability

In **Chapter 5**, 'Self-report stress measures to assess stress in adults with mild intellectual disabilities – a scoping review', a systematic scoping review was conducted to search for self-report measures to assess stress. Each instrument's potential for use in practice and research with people with mild intellectual disabilities was assessed. This assessment was based upon the measure's psychometric properties, practicality of the assessment procedure and suitability for adults with mild intellectual disabilities. To determine a measure's suitability, a two-way strategy was used. First, the literature was searched to see if the instrument had been used with people with intellectual disabilities in research or in clinical practice. If so, the researchers' documented experiences (both subjective and empirical) with using the measure with people with intellectual disability were examined. Second, experts in the field of intellectual disability research were asked what requirements and preconditions these tools should meet to be suitable for people with intellectual disabilities. The questions were based on the topics identified as relevant in our systematic review (Kooijmans et al., 2022).

From the scoping review, 13 self-report measures emerged that tapped into stress-related concepts, three of which were specifically designed for use with adults with intellectual disabilities. Not surprisingly, the three 'intellectual disability-specific' measures emerged as instruments of choice for clinicians and researchers working with people with intellectual disabilities. These measures had adequate psychometric properties and workable assessment procedures, but most of all appeared to have face validity in terms of cognitive accessibility and had been used frequently in published research involving people with intellectual disabilities. In the Discussion, important drawbacks of the recommended measures were mentioned. One was that they lacked sufficiently detailed assessment instructions for assisted assessment. A second limitation was that no norms were available to compare scores with members of the general population. This was a recurring theme in discussions surrounding the question whether it is best to develop 'intellectual disability-specific' measures or measures for the general population that are 'intellectual disability-inclusive' (see *Challenges* below).

This study found that the combined guidance from our review (Kooijmans et al., 2022) and expert opinion can be used to identify self-report measures that are cognitively accessible for people with intellectual disabilities.

Does it matter? Putting the guidance to the test

In Chapter 5, mainly theoretical considerations were used to determine the suitability of the measures under study. It had not been tested empirically if adaptations result in improved cognitive accessibility and lead to better, more reliable and valid results. Nor had it been investigated whether people with intellectual disabilities actually *experience* that adaptations to mainstream measures make it easier or more meaningful for them to participate. These issues were addressed in **Chapter 6**, ‘Does adapting a self-report instrument to improve its cognitive accessibility for people with intellectual disabilities result in a better measure? A cognitive interview study’.

The self-report variant of the Adaptive Behavior Assessment System, third revision (ABAS-3), a widely used questionnaire to measure adaptive functioning, was assessed to determine how accessible it was for people with (mild) intellectual disability. We asked 20 people with mild intellectual disability who were living at one of the three facilities Koraal, Ons Tweede Thuis and Cordaan to complete the selected ABAS-3 questions in the presence of a researcher. Cognitive interview techniques were used to find out which elements of structure and lay-out were the most challenging for respondents. The original version was then adapted to improve the cognitive accessibility of the measure, based upon respondent feedback and recommendations from the earlier studies in this thesis. Changes were made to the instructions, scoring procedure, item language, and lay-out. The language was reformatted to a more reader-friendly version using Taal voor Allemaal (Moonen et al., 2022) guidelines. All adaptations were reviewed and accepted by the developers of the Dutch-language version of the original ABAS-3. Co-researchers at Koraal and the Ben Sajet Centre reviewed all adaptations and made suggestions for improvements.

In the second cognitive interview round, participants completed the adapted version and gave feedback a second time. The adapted version was perceived to be less difficult, and participants were better at remembering the instructions. Comparing the item scores with participant verbal elaborations, there was greater congruence on the adapted version, compared to the original version. This indicated that participants were probably better at accurately self-reporting their abilities when they used a more cognitively accessible instrument. This was reflected in reliability scores, which were higher for the adapted self-report version, compared to the original version.

Care workers completed the selected ABAS-3 questions (original and adapted) about the same person twice. Carer ratings were lower than participant ratings for both versions, but the difference was attenuated for the adapted version. This indicated better agreement between people with mild intellectual disabilities and their carers for the adapted ABAS-3 version.

From the results, it was concluded that adapting self-report measures to the capabilities of people with mild intellectual disability leads to a demonstrably better and more inclusive instrument. From the spontaneous and probed elaborations and examples participants gave about their adaptive behaviour, it appeared that the people with mild intellectual disabilities had more accurate self-reported adaptive skill scores if they understood the questions better. Further work is needed to confirm the concurrent validity of the revised version, by comparing adapted ABAS-3 scores with direct observations of adaptive behaviour, for example. In the recommendations, the importance of involving people with intellectual disabilities when adapting an existing measure to their needs and preferences was stressed once again. Ideally, people with disabilities should be involved even earlier, in the process of construction and standardisation of the original measure. In doing so, no separate 'ID-specific' derivative versions have to be made and people with intellectual disabilities are represented in the norm groups. A second recommendation was to initially invite the person with an intellectual disability to complete the instrument prior to seeking advice from a proxy. Their answers can be compared with proxy-ratings from carers or relatives. Both judgements are perceived truths, and neither is necessarily right. However, when it comes to measuring internal states such as feelings and thoughts, the argument was made that the perception of individual should always take precedence.

General conclusions

Improving the cognitive accessibility of self-report measures for people with mild intellectual disabilities

The results presented in this thesis demonstrate that it is possible to increase the cognitive accessibility of self-report measures by applying knowledge from previous research (chapter 2) and the findings from chapters 3 and 4 of this PhD research project. Improving the cognitive accessibility of self-report measures lowers the barrier for people with mild intellectual disabilities to participate in assessment and research. 'Intellectual disability-inclusive' self-report measurement produces more valid and reliable answers, more agreement between respondents and proxies and lowers cognitive burden.

To ensure that adaptation actually improves cognitive accessibility, it was advised to always involve people with mild intellectual disability when designing or adapting measures. In co-creation, it is possible to check whether questions are understood, answer categories are complicated, and supporting pictures convey the intended meaning.

Interaction factors

From the start of this PhD, it was clear that besides the 'technical aspects' of making adaptations to instruments, the interpersonal dynamics of the assessment procedure should be acknowledged: many people with intellectual disabilities need help to complete self-report measures and guided questioning does not happen in a social vacuum. People who assist other people when completing a self-report measure or when interviewing the other person unwillingly exert a certain influence upon the other; there is always a chance that response bias is inadvertently introduced. An important cause of this type of bias is a power imbalance between respondents and thos who assistant them. People who need care and support because of cognitive and adaptive impairments often depend upon carers or relatives to express themselves and arrange support. This dependence can mean that many people with mild intellectual disability are hesitant to be critical, to give less socially desirable answers, or to be open about sensitive topics in the presence of an assistant when answering questions. By improving cognitive accessibility, more people can be enabled to complete a self-report measure independently so that there is a smaller chance of undue influencing. If a person needs or seeks help when completing the questions, it is desirable to let them choose a person they trust or to help them or arrange for an impartial assistant who has received instruction how to minimize the risk of bias.

Strengths and limitations

The (methodological) strengths and shortcomings of the work in this thesis have been discussed in the individual chapters. Here, several general positive points and some methodological issues and challenges are discussed.

Strengths

Two overall strengths of this PhD research can be identified. First, a comprehensive review of factors that improve the cognitive accessibility of self-report instruments for use with people with mild intellectual disability was completed. Several other authors have provided summaries of best practices and strategies to improve cognitive accessibility using narrative reviews with non-exhaustive literature searches. The work completed as part of the current PhD was a systematic and comprehensive evaluation of the evidence base incorporating validated evaluation tools such as PRISMA methodology for searching and reporting, the MMAT, and RoB tools for methodological quality evaluation, the GRADE-methodology for assessing the strength of evidence, and a 3-way meta-analysis methodology for aggregating results across studies.

A second strength of this PhD is the use of a participative mixed-methods design for the experimental studies (Chapter 3 and 6). Using mixed-methods made it possible to supplement quantitative findings with the direct perspectives and experiences of people with intellectual disabilities.. This has given meaningful insights into the cognitive mechanisms and motivations of participants that help explain the statistical outcomes. It enriched and deepened the conclusions, explanations and recommendations of the findings. Moreover, involving people with intellectual disabilities as co-researchers reinforced our belief that people with intellectual disabilities are capable of working on solutions together with us.

Limitations

There are five overall limitations that should be considered. First, review studies are always subject to risk of publication bias; studies that report significant positive effects are more likely to be published than studies with negative, nonsignificant, or inconclusive findings (Sutton, 2009). Publication bias may have unduly influenced conclusions in several studies included in this thesis (Chapters 2, 4 and 5). Although tools exist to check for the risk of publication bias in quantitative social sciences reviews (Renkewitz & Keiner, 2019), this is not formally assessed in narrative reviews (Chapter 2 and Chapter 5). In the meta-analysis (Chapter 4), a funnel plot was used to visually inspect for the possible occurrence of publication bias. The plot showed a symmetrical distribution of scores across the plot, which indicates that publication bias was probably not threatening the validity of the results. Because of word count limitations, the plot was not included in the published materials but can be supplied upon request.

Second, many of the studies included in the systematic review (Chapter 2) and meta-analysis (Chapter 3) were of poor methodological quality. In the discussion sections of both papers it was mentioned that results should be interpreted with some caution. In particular, it was noted that the absence of significant findings within the meta-analysis may be due to poor study methodology and lack of good-quality visualisations.

Third, the eligibility criteria for the systematic reviews were defined narrowly by limiting ourselves to results of research done with people with intellectual disabilities. Although this is a logical consequence of the demarcation of the scope of this PhD, this may have caused us to overlook interesting recommendations from adjacent domains with different populations, that may be applicable for people with intellectual disabilities. Research with groups with impaired communication (e.g., aphasia) may offer sensible recommendations that benefit work with for people with intellectual disabilities. Moreover, a vast amount of research about the construction and interpretation of surveys for people in the general population is published by marketing and communication scientists. For example, research with the general population has led to the development of a theory on the process of answering self-report questions by Tourangeau and colleagues; this has served as the theoretical underpinning for a number of publications in this thesis.

Fourth, the framework for appraising the suitability of measures for people with intellectual disabilities in Chapter 5 was informed in part by results from a Delphi study. Due to a combination of circumstances, the manuscript for this study was not submitted to a journal, and hence not subjected to peer review. One of the reasons it was decided to not submit was that the results did not add any substantial new insights. However, it must be acknowledged that that the results from the Chapter 5 study were unlikely influenced in any way by this decision.

A final limitation is that there is no objective gold standard for many of the concepts and constructs measured through self-report. Because the self-report outcomes reported in this thesis (e.g. in Chapter 3 and Chapter 6) were not extensively triangulated with observations, scores from other established instruments measuring the same construct, or other means of assessment, it was not possible to report on the concurrent validity of any of the self-report measures. In the absence of a gold standard for most concepts, it remains unclear "who reports it best" (Fisher et al., 2014).

In Chapter 3, it was not possible to objectively determine under which condition the youths' scores were a sincere expression of their satisfaction. This is why the qualitative interpretation of the results in the focus groups was so important to explain the quantitative differences. This approach to establish concurrent validity of the self-reported scores allowed us to make statements about the likelihood of which scores were closest to 'objective reality'. As noted in the discussion of the study reported in Chapter 6, it could only be tentatively assumed that the adapted version provided more valid scores, based on better congruence between participants' scores and self-descriptions of behaviour, and higher reliability indices for the adapted version compared to the original version. But because differences in scores between clients and proxies persisted after adaptation, it was argued that both the perspective of the person with an intellectual disability and their proxies should be considered.

Directions for future research and development

Gaps in the evidence base

At the outset of this PhD project, it was clear that only a modest contribution to the evidence base for this broad area of investigation could be made. There is a myriad of possible adaptations to a wide variety of factors that may improve the cognitive accessibility of self-report measures. The results from the first study in this thesis (Chapter 2) confirmed that there were many factors or areas that were under-researched, and in some instances, the evidence for some adaptations were conflicting or unclear. Examples of under-researched topics were:

- What constitutes helpful visualisation: when do pictures genuinely help and when do they cause confusion or misdirection?
- Is it possible to adapt self-report measures sufficiently for use with people with moderate to severe intellectual disabilities?

Examples of topics for which there are unclear or conflicting results include

- The prevalence of response biases in self-reported information for people with intellectual disabilities. How often are results impacted by response biases such as acquiescence, social desirability or recency/primacy biases? This is disputed among researchers, with some arguing that acquiescent responding is pervasive in samples of people with intellectual disabilities, while others contend that this is an artefactual outcome of experiments conducted in unrealistic lab settings. The most up-to-date review of research on the topic dates from 2002 (Finlay & Lyons) and there has not been an update since. The prevalence and underlying causes of response biases and how they can be addressed remains largely unanswered. The results from the study in Chapter 3 show that people with intellectual disabilities can be very different in their susceptibility for bias and there are many possible underlying cognitive mechanisms that each require a different solution.

- A related topic is the use of bias detecting items and other tools to establish the validly of self-report. There are mixed opinions about which tools are adequate (see Havercamp et al., 2022, for a discussion). A concern is that adding bias-detecting items or tools to self-report measures makes the measure more demanding in terms of cognitive capacity and attention span.
- What response options are appropriate for (most) people with (mild) intellectual disability? There is debate about the most appropriate type of response option which varies according to the topic and respondent characteristics. Different types of response categories have their own strengths and limitations and are associated with different types of bias. Yes/no answers are prone to inducing acquiescent tendencies for instance, while agree/disagree answers have been associated with strong recency effects (Höhne & Krebs, 2018). What options are most appropriate and effective requires further investigation.

Gold standards

Although improving cognitive accessibility may improve participation, some researchers and clinicians remain sceptical about the validity of self-reported information by people with intellectual disabilities. It is clear that people with intellectual disabilities and their proxies often have different perceptions of a construct (Fisher et al., 2014). For behaviours that are directly observable it is possible to develop a gold standard measure. For example, the adaptive behaviour can be observed through targeted tasks: ‘fry an egg for yourself’, ‘invite a group member to play a board game’, ‘figure out how to get from A to B by public transport if you have to be in B at 11.00 o’clock’, etc. The outcomes could then be compared with scores on the self-reported adaptive skills of supervisors and clients. Although this likely provides a more realistic estimate of a person’s true capacities, even behaviour observations are not free from bias and measurement error. The notion of ‘being’ observed’ itself may impact a participant’s behaviour. Many participants under observation will present themselves in an unrealistically positive way, a phenomenon known as the Hawthorne effect (Berkhout et al., 2022).

The promise (and challenges) of adaptive testing

In several of the studies included in this thesis, it was concluded that ‘what works for one, does not necessarily work for all’. People with mild intellectual disability, including those with a similar level of general intellectual functioning, differ greatly in their ability to self-report. While some may experience executive functioning problems (e,g, short-term memory impairments or problems with attention) but are proficient readers, others may have relatively well-functioning memory, but have great trouble understanding written communication. The consequence of this, is that a one-size-fits-all approach is not sustainable. At the same time, we should not settle for self-report measures that suit many but not all people. From the viewpoint of inclusion and participation, the most sensible approach would probably be to develop adaptive testing procedures that cater for a broader range of abilities and preferences. In social sciences research, computer adaptive testing (CAT) is used to ask only items that are relevant to a respondent. Usually, item response theory is used to efficiently determine the items administered (see Cordeiro et al., 2020, for an example). In addition to item selection, CAT algorithms could be used to tailor question and answer characteristics to the (cognitive) level of functioning of the respondent, based on pretest-items. An added advantage to using digital and adaptive measures is that it offers opportunities to add multimedia supports to the measure, such as read-aloud functions or film clips explaining more about the item content. Pilot tests with such platforms have showed that it enabled respondents with limited literacy skills to independently complete surveys and tests (Davies et al., 2017). In a study by Ebenhard and Gebenbeck (2024) students with intellectual disabilities who were assessed with CAT had to complete fewer items, showed reduced bias, and higher accuracy. Although adaptive testing holds promise from the viewpoint of inclusivity, it also poses challenges because many different types of items are needed to suit different levels of functioning. Developing CAT measures is also a resource-consuming endeavour. Further, results from adaptive testing procedures are not readily comparable between respondents if different respondents use different response formats. This also compromises the collection of representative norm data. Generally speaking, many more respondents are needed to ensure standardisation sample representativeness.

From research to practice

Making impact

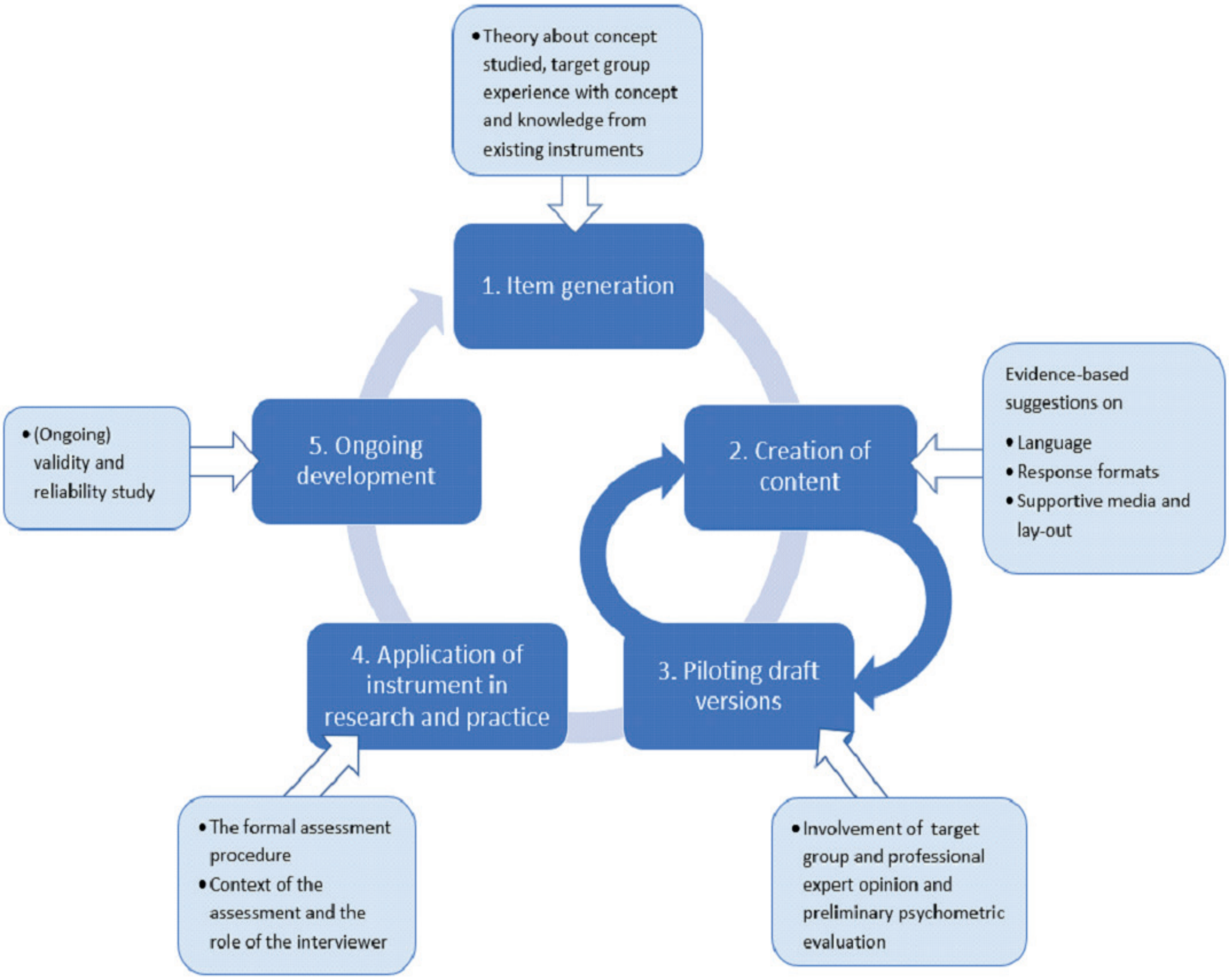
This PhD research offers opportunities for researchers and clinicians to include the views of people with mild intellectual disability in their practice. In the course of this PhD, many of the study findings have already found their way to research and practice. Our results were disseminated with fellow researchers and clinicians through lectures, posters, expert panels and workshops, both in The Netherlands and internationally. The works in this thesis provided substantiation for the recently updated ‘Guidelines for self-report for people with mild intellectual disabilities and borderline intellectual functioning (MBID)‘, that is published by the Dutch Knowledge Centre on MBID (LKC LVB, 2025). As ‘emerging expert’ in communication-friendly measurement, the author of this thesis was asked to provide consultation for many research projects and for the development of several self-reporting tools. A valuable example is the development of the COVID monitor by the GGD (a Dutch public health service organization), where the proceeds from this PhD research were used to ensure that people with mild intellectual disability were included in the national study on the impact of COVID-19 on people’s well-being. For Dutch readers, Zuyd University of Applied sciences developed a guideline for communication-friendly measurement (Dalemans et al., 2021). In 2025, the results from this PhD will be used to provide an update and add scientific substantiation to the recommendations in the guideline.

Practical guidance

Until then the most readily available guidance for researchers and developers comes from the review in Chapter 2. This publication offers practical suggestions for making self-report cognitively accessible, which can be used to adapt or review existing measures and in the process of developing new measures. The Summary of Findings table in Chapter 2 lists a plethora of factors to take into consideration with a rating of the strength of the evidence per recommendation.

The five-stage model for the inclusive design of self-report measures presented in Chapter 2 can be used as a blueprint for the process of developing new instruments.

Stages in 'intellectual disability-inclusive' Self-report Instrument Development



Broadening the scope of cognitive accessibility beyond intellectual disability

Our research assumed that specific adaptations to self-report measures are needed to accommodate the needs and preferences of people with intellectual disabilities. Looking closely at the kinds of adaptations suggested, it can be seen that many of these suggestions might just as well apply to other people with cognitive or communicative challenges. Writing texts in accessible language, for example, is just as beneficial for people with a different first language, children, or people who are less literate because of a lack of education opportunities. It lowers the cognitive load for any person reading a text and may benefit the ease of reading for people without reading impediments as well. Limiting the number of response categories will also help people who experience memory function loss as a result of acquired brain injury or aphasia. It is therefore plausible that, at least to some extent, the recommendations in this thesis are not limited to self-reporting for people with intellectual disabilities. They may well extend to other populations with cognitive impairment or language proficiency problems. That is why in Chapter 4, and in the recommendations in Chapter 6, the choice was made not to limit inclusion to people with intellectual disabilities, but to include ‘people who for whatever reason have difficulty understanding self-report measures’.

Challenges

In this PhD, an inclusive approach to all aspects related to the topic of self-reporting was promoted. Inclusion is a universal right (United Nations, 2006) and the right to express your own opinion is part of this. In the final part of this thesis, two barriers for inclusivity that we encountered during this PhD project are highlighted. One barrier is related to the gatekeeper problem cited in the introduction. The other barrier pertained to the discriminatory view that people with intellectual disabilities should not be afforded the same opportunities as other people due to having an intellectual disability; a view which continues to be very problematic.

The gatekeeper problem

Involving people with mild intellectual disability in needs assessment, policy making, and research is becoming more self-evident (Walton et al., 2022). Nevertheless, barriers for participation are still encountered, both in the literature (McCusker et al., 2023) and in our own experience when trying to involve people with mild intellectual disability directly in our research. Improved cognitive accessibility allows us to ‘technically enable’ people to participate, but it does not mean that people with intellectual disabilities are given the opportunity to participate by gatekeepers. The gatekeeper problem excludes many people with intellectual disabilities from expressing their own opinions independently (McFarland et al., 2024). We experienced several instances where people with mild intellectual disability who in our opinion would be perfectly capable of completing measures themselves were assisted. Although they were invariably helped with the best of intentions, this increased the chance of response bias (as can be seen in Chapter 3). In more worrisome cases, people who were capable of participating through self-report were excluded from participation altogether by gatekeepers. ‘He is not up to that, it’s too difficult and stressful’, ‘it will only give us unreliable information’, ‘he always overestimates himself’ or even ‘he doesn’t know what’s good for him’ is still heard when carers or clinical workers are asked why a client might not participate in our research. In a recent satisfaction survey among Koraal residents with mild or moderate intellectual disability, using a questionnaire specially developed for (and with) the target group, only 1% of clients completed the questionnaire independently; 61% did so with help, while 38% could not give their opinion themselves. For these clients, relatives or supervisors estimated their satisfaction with Koraal’s services.

The unwanted distinction between people with and without intellectual disabilities

This PhD project advocates making adaptations to existing, validated and standardised instruments that are less cognitively accessible for people with intellectual disabilities. By creating ‘intellectual disability-versions’ of self-report measures, a number of issues are raised, both on a conceptual, moral-ethical level and in terms of psychometrics and methodology.

The creation of special intellectual disability-versions of mainstream measures assumes that there are two clearly demarcated types of people: people with and people without intellectual disabilities. This is conceptually untenable; there is no uniformly defined measurable boundary that distinguishes people with intellectual disabilities from neurotypically functioning individuals. Moreover, it is morally questionable to distinguish groups of people solely based upon their cognitive abilities, without valuing their strengths, talents, and personal experiences. So paradoxically, one recommendation from this PhD – make adaptations to mainstream measures – contravenes one of its central tenets – the promotion of inclusivity.

A more practical methodological disadvantage of making separate versions for people with intellectual disabilities is that it becomes challenging to compare the scores of people with intellectual disabilities with general population norms, unless participants across the entire spectrum of cognitive functioning were included in the standardisation samples. An associated risk of making intellectual disability-versions is that measurement invariance of the instrument (Farmer et al., 2024) can be compromised, because alternatively worded versions of the same questionnaire measure may not measure the exact same constructs as the original. The validity and reliability of the adapted version cannot be assumed, even if they were found to be excellent in the original will have to be investigated anew. For example by comparing versions with concurrent validity procedures or examining the construct validity of versions by comparing the outcomes of factor analyses for different versions and groups of participants.

Fortunately, there is an outstanding single solution for both the conceptual and methodological problem. If developers and publishers can be convinced of the importance of including people with intellectual disabilities in the process of development and norm collection, separate versions will no longer be needed. The resulting inclusive measures ensure that as many people with (mild) intellectual disability can participate in assessment and research, while at the same time enabling the comparisons of scores of all respondents with representative norm groups. This would not only greatly advance the participation of a large proportion of the population (those with below-average cognitive functioning), it will also benefit many other people who face challenges completing a self-report measure for other reasons.

Closing remarks

The findings from this PhD demonstrated that it is possible to enable people with intellectual disabilities to meaningfully participate in assessment and research by means of self-report. Not only by making it easier to fill in questionnaires, but also – and maybe most importantly - by recognising that the opinions of people with intellectual disabilities matter and that every person is capable, in their own way, by themselves or with the help of others, of sharing what they think, feel and think. Traditionally, many people with disabilities are used to being cared for, to have challenges taken away, and unfortunately, to experience that their opinions are not valued. For many people with intellectual disabilities, to claim your space and demand to be included will therefore not come naturally. The people who support people with intellectual disabilities can take the first step towards more autonomy and inclusivity by removing the gates, leaving room, having trust in the people's own abilities, giving podium and taking a step back.

There are many soft and quiet voices that deserve to be noticed. I hope the findings of this PhD research will be used to make it easier for people with intellectual disabilities to express themselves and be heard.

References

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. https://doi.org/10.1080/1068316X.2017.1298761

Berkhout, C., Berbra, O., Favre, J., Collins, C., Calafiore, M., Peremans, L., & Van Royen, P. (2022). Defining and evaluating the Hawthorne effect in primary care, a systematic review and meta-analysis. In Frontiers in Medicine (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fmed.2022.1033486

Cordeiro, L., Villagomez, A., Swain, D., Deklotz, S., & Tartaglia, N. (2020). Adaptive skills in FXS: a review of the literature and evaluation of the PEDI-computer adaptive test (PEDI-CAT) to measure adaptive skills. Brain Sciences, 10(6), 351.

Dalemans, R., Stans, S., von Helden-Lenzen, S., Lemmens, J., Ummels, D., Cornips, S. (2021). Leidraad Communicatievriendelijk meten. [Guideline for Communication-friendly Measurement]. Zuyd University of Applied Sciences. Report. https://doi.org/10.25385/zuyd.16782985.v1

DeLoache J. S., Pierroutsakos S. L., Uttal D. H. (2003). The origins of pictorial competence. Current Directions in Psychological Science, 12(4), 114–118. https://doi.org/10.1111/1467-8721.01244 Davies, D. K., Stock, S. E., King, L., Wehmeyer, M. L., & Shogren, K. A. (2017). An accessible testing, learning and assessment system for people with intellectual disability. International Journal of Developmental Disabilities, 63(4), 204–210. https://doi.org/10.1080/20473869.2017.1294313

Ebenbeck, N., & Gebhardt, M. (2024). Differential Performance of Computerized Adaptive Testing in Students With and Without Disabilities – A Simulation Study. Journal of Special Education Technology. https://doi.org/10.1177/01626434241232117

Farmer, C., Kaat, A. J., Edwards, M. C., & Lecavalier, L. (2024). Measurement Invariance in Intellectual and Developmental Disability Research. American Journal on Intellectual and Developmental Disabilities, 129(3), 191–198. https://doi.org/10.1352/1944-7558-129.3.191

Finlay, W. M. L., & Lyons, E. (2002). Acquiescence in Interviews With People Who Have Mental Retardation. In American Association on Mental Retardation (Vol. 40). http://meridian.allenpress.com/idd/article-pdf/40/1/14/2046313/0047-6765

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

Fisher, M. H., Mello, M. P., & Dykens, E. M. (2014). Who reports it best? A comparison between parent-report, self-report, and the real life social behaviors of adults with Williams syndrome. Research in Developmental Disabilities, 35(12), 3276–3284. https://doi.org/10.1016/j.ridd.2014.08.011

Frixione, M., & Lombardi, A. (2015). Street signs and Ikea instruction sheets: Pragmatics and pictorial communication. Review of Philosophy and Psychology, 6, 133-149. https://doi.org/10.1007/s13164-014-0216-1

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

Havercamp, S. M., Barnhill, L. J., Bonardi, A., Chapman, R. A., Cobranchi, C., Fletcher, R. J., Rabidoux, P., Seeley, J. R., & Tassé, M. J. (2022). Straight from the horse's mouth: Increasing self-report in mental health assessment in individuals with intellectual disability. Journal of Applied Research in Intellectual Disabilities, 35(2), 471–479. https://doi.org/10.1111/jar.12952

Höhne, J. K., & Krebs, D. (2018). Scale direction effects in agree/disagree and item-specific questions: a comparison of question formats. International Journal of Social Research Methodology, 21(1), 91–103. https://doi.org/10.1080/13645579.2017.1325566

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Nicolau, B. (2018). Mixed methods appraisal tool (MMAT) (Version 2018). Canadian Intellectual Property Office, Industry Canada.

Kooijmans, R., Mercera, G., Langdon, P. E., & Moonen, X. (2022). The Adaptation of Self-Report Measures to the Needs of People With Intellectual Disabilities: A Systematic Review. Clinical Psychology: Science and Practice, 29(3), 250–271. https://doi.org/10.1037/cps0000058

Landelijk Kenniscentrum LVB (2025). Handreiking zelfrapportage LVB [Guidelines self-report MBID]. LKC LVB. https://www.kenniscentrumlvb.nl/product/handreiking-zelfrapportage-lvb/

McCusker, P., Gillespie, L., Davidson, G., Vicary, S., & Stone, K. (2023). The United Nations Convention on the Rights of Persons with Disabilities and Social Work: Evidence for Impact? In International Journal of Environmental Research and Public Health (Vol. 20, Issue 20). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ijerph20206927

Moonen, X., Reichrath, E., Cortenbach, K., Wittelings, T. (2022). Schrijven in Taal voor allemaal, versie 5 [Writing in Language for all, version 5]. Koraal and Toetsen, meten & weten.

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 GRADECERQual to qualitative evidence synthesis findings-paper 3: How to assess methodological limitations. Implementation Science, 13(1, Suppl.1), 9. https://doi.org/10.1186/s13012-017-0690-9

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research. Zeitschrift für Psychologie, 227(4), 261–279. https://doi.org/10.1027/2151-2604/a000386

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. Hedges, & J. Valentine (Eds.), The handbook of research synthesis and meta-analysis (pp. 435–452). Russell Sage Foundation.

United Nations (2006). Convention on the Rights of Person with Disabilities. Retrieved from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html (accessed on 10 September 2024).

Walton, K., Krahn, G. L., Buck, A., Andridge, R., Lecavalier, L., Hollway, J. A., Davies, D. K., Arnold, L. E., & Havercamp, S. M. (2022). Putting “ME” into measurement: Adapting self-report health measures for use with individuals with intellectual disability. Research in Developmental Disabilities, 128. https://doi.org/10.1016/j.ridd.2022.104298

Chapter 8

Appendices

Appendices Chapter 2

- Appendix A. Characteristics of Included Studies
- Appendix B. GRADE-CERQual Evidence Profile

Appendices Chapter 4

- Search Strategy november 2023
- Risk of Bias Assessment
- Quality assessment domain 6

Appendices Chapter 5

- Exemplary search string for PsycInfo
- Overview of Stress Self-Report Measures Selected in Step 1

Appendices Chapter 6

- ABAS-3 items included in original version
- Example of a question from the adapted ABAS-3

Chapter 2. Appendix A

Characteristics of Included Studies

First author and year of publication	Summary	Target population level of ID (and *n* 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	

Chapter 2. Appendix A

Characteristics of Included Studies

First author and year of publication	Summary	Target population level of ID (and *n* 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	

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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.
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)	JBI Checklist 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)	JBI Checklist for Text and Opinion + MMAT - Qualitative	10/11	No clear description of study population

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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	JBI Checklist 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	JBI Checklist 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.
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	

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
		were persons with ID)			
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.

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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	
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.

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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	
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	

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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.

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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).
Ramirez 2007	Described the development of the Fear Survey for Adults with Mental Retardation.	Qualitative study: Level of ID not specified (focus group, n = 17). Quantitative study: Mild to moderate ID (n=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 (n=11 for focus groups, n=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	

First author and year of publication	Summary	Target population level of ID (and *n* if applicable)	Quality Appraisal Tool	Quality Appraisal score	Quality Appraisal issues
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, n=10) and proxies (n=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 (n=40)	MMAT - Qualitative	4/5	Sample characteristics not clear. Very specific sample population (Special Olympics participants), results may not generalise to broader ID population.

Chapter 2. Appendix B

Evidence profile: recommendations for the construction or adaptation of self-report instruments for people with intellectual disabilities.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
1. Item generation							
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	No or very minor concerns	No or very minor concerns	No or very minor concerns Mix of qualitative and quantitative studies, both primary and secondary research.	Minor concerns Mostly relevant for persons with MID or Borderline Intellectual functioning and adequate verbal abilities.	High confidence Mostly based on practical experience and expert opinion.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Assess the cognitive and language skills of participants involved in the construction process to determine suitability for target populations.	Emerson 2013	No or very minor concerns	No or very minor concerns	Moderate concerns Only one study. Finding based on clinical expertise.	No or very minor concerns	Moderate confidence Based on expert opinion.
2.1 Creation of content: format and language							
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	No or very minor concerns	No or very minor concerns	Moderate concerns Results from a broad range of methodologie s with high combined N’s, but mostly based on	Minor concerns Level of ID not specified for many included studies.	Moderate confidence Mostly based on clinical expertise. What constitutes simple vocabulary is not specified.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
					clinical expertise.		
	Use literal meaning of words, don't use metaphors and proverbs.	Ikeda 2014	Moderate concerns Small scale focus groups, methodology not well described.	Moderate concerns Finding is not clearly traceable to qualitative data origin.	Serious concerns Thin data from only one study.	Moderate concerns Not clear if proxies' opinions contributed to the finding. Studied in children with MID.	Low confidence Finding seems common sense, but this specific study (Ikeda, 2014) 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	Minor concerns Concerns pertain to qualitative study and some unsubstantiated claims in secondary research.	No or very minor concerns	Moderate concerns Mostly secondary research.	No or minor concerns	Moderate confidence. Finding aligns with expert opinions, but not substantiated empirically.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Avoid adding 'no' and 'not' to positive phrasings. Use negative form of words.	Finlay 2001	No or minor concerns	No or very minor concerns	Minor concerns Only secondary research.	No or very minor concerns	High confidence Likely reduces complexity.
Sentence structure	Keep the sentence structure as clear and simple as possible.	Bell 2018, Finlay 2001, Sigstad 2018, White Koning 2015	No or very minor concerns	No or very minor concerns	Moderate concerns Results mostly based on clinical expertise and secondary research.	Minor concerns Level of ID not specified for many included studies and results from children generalised to broader ID population.	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	No or very minor concerns	No or very minor concerns Findings from review and empirical research align.	Minor concerns Low N in quantitative study.	Moderate concerns Stringent exclusion criteria may introduce selection bias.	Moderate confidence

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Use short sentences to minimise the strain on working and short-term memory.	Bell 2018, Gjertsen 2018	Minor concerns Qualitative synthesis not defined in Gjertsen (2018) study.	No or minor concerns	Minor concerns Thin data in qualitative study.	Minor concerns Finding from qualitative study seems to reflect researcher's professional opinion.	Moderate confidence 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	No or very minor concerns	No or very minor concerns	Moderate concerns Only secondary research.	Minor concerns. Level of ID not specified.	Moderate confidence Linguistics studies show that this likely reduces complexity.
	Use active phraseology (as opposed to passive phrasings) as they are structurally less complex.	Finlay 2001, White Koning 2005	No or very minor concerns	No or very minor concerns	Moderate concerns Only secondary research.	Minor concerns Results from one of two included studies generalised from children with ID to broader ID population.	Moderate confidence More empirical research needed.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Avoid double negatives as this can be confusing, especially in combination with positive and negative response options.	Finlay 2001	No or very minor concerns	No or very minor concerns	Moderate concerns Based on just one secondary study.	Minor concerns. Level of ID not specified.	Moderate confidence More empirical research needed.
	Use established guidelines, such as Easy-Read standards, to reword items.	Kent 2018	No or very minor concerns	No or very minor concerns	Moderate concerns Based on only one small-scale study.	No or very minor concerns	Moderate confidence Replication needed.
	Use established evaluative statistics to check for readability (e.g. Flesch Reading Ease, Gunning-Fog Index).	Keeling 2017	No or very minor concerns	No or very minor concerns	Moderate concerns Based on only one small-scale study. Thin qualitative data	Moderate concerns Study population mainly persons with Borderline Intellectual Functioning.	Moderate confidence Replication needed for broader ID population.
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) memory.	Bell 2018, Finlay 2001, Jen-Yi 2015, Scott 2018	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	High confidence

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Don't ask to retrieve detailed information over longer periods of time.	Vlot van Anroij 2018	No or very minor concerns	No or very minor concerns	No or very minor concerns	Moderate concerns Results from specific subpopulation generalised to broader ID population.	Moderate confidence Replication needed for broader ID population.
	Ask respondents to reflect on what they *currently* feel or think. Stay in the here and now.	Williams 2007	No or very minor concerns	No or very minor concerns	Moderate concerns Not clear how researcher arrives from data to conclusion.	Moderate concerns Study population mainly persons with Borderline Intellectual Functioning.	Moderate confidence Replication needed for broader ID population.
Content: other/miscellaneous	Refrain from leading or presumptive questions, e.g. "You do like the food at your home, don't you?".	Bowles 2014	Minor concerns Small sample size, power issues.	No or very minor concerns	Minor concerns Small sample size	Minor concerns Results from MID population generalised to broader ID population.	Moderate confidence Based on only one small-sample study.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Consider rephrasing questions to accommodate for specific cultural issues.	Jen-Yi 2015	Serious concerns No clear description of methodology, sample, and synthesis strategy	No or very minor concerns	Serious concerns Thin data and finding seems to reflect researcher's opinion.	Serious concerns Finding does not seem to originate in participants' statements.	Low confidence Although common sense, the recommendation solely reflects professional opinion of researcher.
2.2 Creation of content: Response format							
Dichotomous answer options	Yes/no type questions are understood by the largest proportion of persons with ID.	Ikeda 2014, Ramirez 2005, Stancliffe 2015	Minor concerns Lower quality qualitative study.	No or very minor concerns	Minor concerns Very large sample sizes in quantitative studies Thin data in qualitative study.	No or very minor concerns	High confidence Decisive results from quantitative studies.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	intellectual functioning to mild ID.						
	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	No or very minor concerns	No or very minor concerns	No or very minor concerns Data from both secondary and primary research.	No or very minor concerns	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	No or very minor concerns	Minor concerns Only one study.	No or very minor concerns	Minor concerns Level of ID not specified.	Moderate confidence Sensible suggestion, but limited empirical evidence base. Based on research expertise from secondary sources.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Be aware of acquiescent responding on yes/no response formats. Consider using an either/or response format instead of yes/no.	Heal 1995	Moderate concerns References in review possibly outdated.	Minor concerns Some contradictory findings	No or very minor concerns	Moderate concerns Results from 25+ years ago may not transfer to current time frame.	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 concerns References to earlier research possibly outdated.	Minor concerns Some contradictory findings	No or very minor concerns	Moderate concerns Findings probably only valid for persons with better verbal abilities.	Moderate confidence Plausible for persons with adequate verbal abilities.
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	Hartley 2006, Bell 2018, Cummins 1997, Fang 2011, Power 2010, Dagnan 1995	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	High confidence Compelling evidence form a broad variety of research designs. Adequate differentiation of findings between subpopulations.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Check validity of answers on multiple choice questions by asking for examples or further (scripted) probing questions.	Finlay 2001, Hartley 2006	No or very minor concerns	No or very minor concerns	No or very minor concerns	Moderate concerns Level of ID not specified. Findings probably only valid for persons with better verbal abilities.	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 demands.	Finlay 2001	No or very minor concerns	Moderate concerns Findings lack specificity to be implemented in practice.	Moderate concerns Based on just one secondary study.	Minor concerns Level of ID not specified.	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 concerns Qualitative research methodology not clear.	No or very minor concerns	Minor concerns Limited contribution to finding by	Minor concerns Findings probably only valid for persons with	Moderate confidence Plausible for persons with adequate verbal abilities.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
					qualitative study.	better verbal abilities.	
	Do not use open-ended questions when the participant or target group in general does not possess adequate productive verbal ability.	Boland 2008	No or very minor concerns	No or very minor concerns	Minor concerns Only one study, but finding has high plausibility.	No or very minor concerns	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 concerns Selection bias.	No or very minor concerns	Moderate concerns Small sample size.	Moderate concerns Generalisability problems as a result of selection bias.	Moderate confidence Use of VAS holds potential, but more research on reliability and validity is needed. Findings

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
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	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	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	Moderate concerns References in review possibly outdated.	Serious concerns Conflicts with other findings that state the importance of uniformity and predictability for participants.	Moderate concerns Based on just one secondary study.	No or very minor concerns	Very low confidence Conflicting results.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
2.3 Creation of content: Supportive media and lay-out							
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 concerns Unclear sample characteristics in studies Flawed methodology in qualitative study.	Moderate concerns Very diverse operationalisation of what constitutes 'supportive visualisation'.	No or very minor concerns Limited support from qualitative research components.	Minor concerns Some studies lack clear descriptions of the population under study.	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	No or very minor concerns	No or very minor concerns	Minor concerns Based on just one secondary study.	No or very minor concerns Level of ID not specified, but findings obviously applicable across all types of populations.	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	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	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 Knegt 2017, Marshall 2007, O'Keeffe 2019	No or very minor concerns	Moderate concerns Uniformity in results but modes of visualisation very diverse across studies.	Minor concerns More quantitative research needed.	Minor concerns No clear definitions of what works for whom.	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 Knegt 2017	Minor concerns Small sample size.	No or very minor concerns	Minor concerns Small sample size but compelling results.	Minor concerns Not clear if results from people with Down Syndrome can be generalised to the broader ID population.	Moderate confidence Compelling evidence but from a specific and small sample of participants. Need for replication.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
Lay-out	Use a clear and attractive lay-out to capture the respondent's attention and reduce clutter and confusion.	White Koning 2005, Bell 2018	No or very minor concerns	No or very minor concerns	Moderate concerns Limited empirical data, observation from research in practice.	No or very minor concerns	Moderate confidence Sensible suggestion but lacks specificity.
	Display one question per page.	Ikeda 2014	Moderate concerns Small scale focus groups, methodology not well described.	Moderate concerns Finding seems to originate in statement from a single participant.	Serious concerns Thin data from only one study.	Moderate concerns Not clear if proxies' opinions contributed to the finding. Studied in children with MID.	Low confidence Recommendation seems self-evident, but no empirical evidence.
	Limit paragraph length for explanatory texts and instructions.	White Koning 2005	Moderate concerns Some statements not substantiated with literature references.	No or very minor concerns	Serious concerns No references specified for this recommendation.	Moderate concerns Results from research on children with ID.	Low confidence Recommendation seems plausible, but origin of the recommendation is not clear, and it lacks specificity.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Use flash cards when a choice from more than 2 response options is asked.	Boland 2009	No or very minor concerns	Moderate concerns Finding based on experience of researchers in practice, no empirical data.	No or very minor concerns	No or very minor concerns	Moderate confidence Based on observation of the researchers in practice.
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 Knegt 2017	Minor concerns Some issues relating to sample size and characteristics in individual studies.	No or very minor concerns	No or very minor concerns	No or very minor concerns	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.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
3. Piloting draft versions							
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	No or minor concerns Some methodological flaws in qualitative design components.	No or very minor concerns	No or very minor concerns	No or very minor concerns	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	No or very minor concerns	No or very minor concerns	Moderate concerns Based on just one secondary study and origin of recommendation not clear.	No or very minor concerns	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	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	High confidence Finding based on large-scale population survey data and sensitive statistical analyses.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Use triangulation procedures with different informants (e.g., with relatives or professionals).	Jen-Yi 2015	Serious concerns No clear description of methodology, sample, and synthesis strategy.	No or very minor concerns	Serious concerns Thin data and finding seems to reflect researcher opinion.	Serious concerns Finding does not seem to originate in participant statements.	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)

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
4.1 Application in practice: formal assessment procedure							
Use of pre-tests practice formats	Use a pre-test to check the participant cognitive and verbal ability and if they understand the response formats and visualisations used.	Bell 2018, Cummins 2002, Cuskelly 2013, de Knegt 2013, Emerson 2013, Finlay 2001, Finlay 2002, Jen-Yi 2015, Townsend-White 2012, White-Koning 2005	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	High confidence Well-established good practice in research and development, but little specific guidance on how to test comprehension.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Use a pre-test as an opportunity for participants to practice with the response formats	Hartley 2006, Jen-Yi 2015, Reid 2009	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	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	Minor concerns Possibly outdated literature for substantiation.	No or very minor concerns	No or very minor concerns	No or very minor concerns	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	Minor concerns Some methodological issues (no clear description method, small sample size).	No or very minor concerns	No or very minor concerns	Minor concerns Level if ID not structurally described.	Moderate confidence Overall, the recommendations are sufficiently substantiated by the contributing studies, but specification is needed to determine how this should be done. .

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	or						
	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	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	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 concerns Some methodological issues in some studies.	No or very minor concerns	No or very minor concerns	Minor concern Level of ID not specified.	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	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	Moderate confidence Not specific enough to guide developers.
Other structural and procedural issues	Make the questionnaire as short as possible to prevent potential problems with attention span or fatigue.	Bell 2018	No or very minor concerns	Moderate concern	No or very minor concerns	Minor concern Level of ID not specified.	Moderate confidence Seems self-explanatory, not very

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
				Some incongruence between sources included in this review.			specific. More empirical research may be needed to determine limits and optional length.
	Group items on related topics.	Jen-Yi 2015	Serious concerns No clear description of methodology, sample, and synthesis strategy.	No or very minor concerns	Serious concerns Thin data and finding seems to reflect researcher's opinion.	Serious concerns Finding does not seem to originate in participant statements.	Low confidence Although common sense, the recommendation reflects the professional opinion of the researcher.
	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	No or very minor concerns	Moderate concern Some incongruence with the literature/sources.	No or very minor concerns	Minor concerns Level of ID not specified.	Moderate confidence More empirical research needed.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Allow for someone to assist the participant and clarify questions if the participant so wishes.	Gjertsen 2019	Moderate concerns No clear description of qualitative methodology.	Moderate concerns Seems to contradict other findings in this review.	Serious concerns Recommendations based on clinical experience from a single researcher.	Minor concerns Level of ID not specified.	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.
4.2 Application in practice: Role of assessor in case of assisted administration or structured interview							
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.	Jen-Yi 2015, Perkins 2007	No or very minor concerns	No or very minor concerns	No or very minor concerns	Minor concerns Level ID not specified and target population	Moderate confidence Recommendation reflects general good research and clinical practice.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	- become familiar with the topic and procedures.					broader than ID.	
	Interviewers should receive strict instructions follow the format to promote standardisation.	Antaki 1999	No or very minor concerns	No or very minor concerns	No or very minor concerns	Minor concerns Target population MID to moderate ID.	High confidence Self-evident to promote standardisation and reduce ambiguity.
Pre-assessment introduction	Take the time to develop rapport with participants.	O'Keeffe 2019	No or very minor concerns	No or very minor concerns	Moderate concerns Recommendation based on experience during implementation.	No or very minor concerns	Moderate confidence Mainly based on experience from research in practice.
	Use a clear, pre-scripted introduction of the assessment and clear instructions to explain the procedure to the participant.	White Koning 2005	No or very minor concerns	No or very minor concerns	Moderate concerns Only secondary research.	Minor concerns Some results of studies with adults applied to recommendations for children.	Moderate confidence Reflects good research practice to minimise ambiguity,

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Stress that information will or will not be shared with carers, and/or state the conditions under which confidentiality may be breached.	Finlay 2001	No or very minor concerns	No or very minor concerns	No or very minor concerns	Minor concerns Level of ID not specified.	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 concerns Limited description of method	No or very minor concerns	Minor concerns Small sample size	Moderate concerns Level of ID not specified	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 concerns Limited information on methodology in qualitative studies.	No or very minor concerns	No or very minor concerns	Moderate concerns No clear description of population or specific ID sub-group sample.	Moderate confidence Limited empirical evidence, but self-evident.
	Routinely and repeatedly offer the opportunity to withdraw when addressing sensitive or taboo topics. Especially when discomfort is noticed.	Stancliffe 2017	Moderate concerns Small sample size for quantitative study.	No or very minor concerns	Moderate concerns Small sample size for quantitative study.	No or very minor concerns	Moderate confidence Reflects respectful clinical and research practice.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
	Be aware of social desirability, acquiescence or other answering patterns when administering questionnaires.	Jobson 2013, Williams 2007	No or very minor concerns	Minor concerns Two studies included, but with a different definition of bias.	No or very minor concerns	Minor concerns Studied in MID and Borderline Intelligence Functioning sample.	Moderate confidence More in-depth research needed on impact of different sources of bias on results.
	Do not repeat questions on which a subject has already answered satisfactorily. For the participant this may imply the answer was ‘wrong’.	Cummins 1997	No or very minor concerns	No or very minor concerns	No or very minor concerns	Moderate concerns References possibly outdated.	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	No or very minor concerns	No or very minor concerns	No or very minor concerns	Minor concerns Level of ID not specified.	High confidence Reflects good clinical and research practice.
	Repeat, paraphrase, and summarise responses.	Sigstad 2018	No or very minor concerns	No or very minor concerns	Moderate concerns Only secondary research.	Minor concerns Level of ID not specified.	Moderate confidence Empirical research needed.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
Role or status of the interviewer	The interviewer should be a 'neutral' person, especially with sensitive questions.	Finlay 2001, Williams 2007	No or very minor concerns	No or very minor concerns Findings from review and empirical research align.	No or very minor concerns	Minor concerns Level of ID not specified.	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	No or very minor concerns	No or very minor concerns	Minor concerns Based on practice experience from one (large scale) study.	No or very minor concerns	Moderate confidence Promising topic for further investigation.

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
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	No or very minor concerns	No or very minor concerns	Moderate concerns Only secondary research.	Minor concerns Some results of studies with adults applied to recommendations for children	Moderate confidence Well-established from research and clinical practice.
5. Ongoing development							
Psychometric evaluation	Examine the reliability, validity, and factor structure of the adapted or newly developed instrument.	Finlay 2001, Blasingame 2011	No or very minor concerns	No or minor concerns Findings from review and empirical research align.	No or very minor concerns	Minor concerns Level of ID not specified.	High confidence Findings are in line with good research practice.
	Responsiveness of items can be used as a measure of comprehensibility of the instrument.	Stancliffe 2015	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns	High confidence Finding based on large-scale population survey

Instrument development stage and corresponding issues	Recommendation	Contributing studies	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual assessment
							data and sensitive statistical analyses.
	Use triangulation procedures with different informants (e.g., with relatives or professionals).	Jen-Yi 2015	Serious concerns No clear description of methodology, sample, and synthesis strategy	No or very minor concerns	Serious concerns Thin data, and finding seems to reflect researcher opinion	Serious concerns Finding does not seem to originate in participant statements	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)

Chapter 4. Appendix A

Search strategies

The searches below yielded 2.803 results on 20 November 2023; PsycInfo (708 references), Medline (676 references), ERIC (211 references) and Web of Science Core Collection (1.208 references)

PsycInfo (Ovid, APA PsycInfo, 1806 to November Week 2 2023)

1. (accessible communicat* OR communicat* friendly OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ4 (text OR texts OR textual OR leaflet*)) OR linguistic simplificat* OR easy to read OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ1 (written OR language OR information OR read*)) OR apprehensible information OR apprehensible text* OR decipherable information OR decipherable text* OR patient information leaflet* OR text comprehension OR text understanding OR understanding text* OR low literacy OR low health literacy OR understanding of instructions OR understanding instructions).ti,ab,id. OR (leichte sprache OR einfache sprache).af.

2. pictorial stimuli/ OR photographs/ OR digital images/ OR animation/ OR (illustration* OR picto* OR pictur* OR photo OR photos OR drawing* OR visual* OR symbol* OR image OR images OR animation* OR gif OR infograph* OR cartoon* OR diagram* OR graphic*).ti,ab,id.

3. (followup study OR longitudinal study OR field study OR interview OR qualitative study OR quantitative study).md. OR (random* OR RCT* OR field exp* OR quasiexp* OR quasi-exp* OR control group* OR control condit* OR control design* OR posttest OR post test OR pretest OR pre-test OR cohen*s d OR qualitat* OR baseline OR follow-up OR followup OR longitud* OR participant* OR questionnair* OR interview*).ti,ab,id.

4. ((infancy 2 23 mo OR preschool age 2 5 yrs OR school age 6 12 yrs OR adolescence 13 17 yrs) NOT (young adulthood 18 29 yrs OR adulthood 18 yrs older)).ag.

5. 1 AND 2 AND 3

6. 5 NOT 4

Key: / = subject heading, ti = title, ab = abstract, id = key concepts (other keywords added by PsycINFO indexers to supplement the subject headings), af = all fields, ag = age group, md = methodology, ADJn = word distance of maximum n words

MEDLINE (Ovid MEDLINE ALL, including epub ahead of print, in-process & other non-indexed citations and Ovid MEDLINE Daily, 1946 to November 17, 2023)

1. (accessible communicat* OR communicat* friendly OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ4 (text OR texts OR textual OR leaflet*)) OR linguistic simplificat* OR easy to read OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ1 (written OR language OR information OR read*)) OR apprehensible information OR apprehensible text* OR decipherable information OR decipherable text* OR patient information leaflet* OR text comprehension OR text understanding OR low literacy OR low health literacy OR understanding of instructions OR understanding instructions OR leichte sprache OR einfache sprache).ti,ab,kf.

2. (illustration* OR picto* OR pictur* OR photo OR photos OR drawing* OR visual* OR symbol* OR image OR images OR animation* OR gif OR infograph* OR cartoon* OR diagram* OR graphic*).ti,ab,kf.

3. randomized controlled trial/ OR longitudinal studies/ OR follow-up studies/ OR qualitative research/ OR (random* OR RCT* OR field exp* OR quasiexp* OR quasi-exp* OR control group* OR control condit* OR control design* OR posttest OR post test OR pretest OR pre-test OR cohen*s d OR qualitat* OR baseline OR follow-up OR followup OR longitud* OR participant* OR questionnair* OR interview*).ti,ab,kf.

4. (infant, newborn/ OR infant/ OR child, preschool/ OR child/ OR adolescent/) NOT (young adult/ OR adult/)

5. 1 AND 2 AND 3

6. 5 NOT 4

Key: / = medical subject heading (MeSH), ti = title, ab = abstract, kf = author supplied keywords, ADJn = word distance of maximum n words

ERIC (Ovid, 1965 to November 2023)

1. (accessible communicat* OR communicat* friendly OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ4 (text OR texts OR textual OR leaflet*)) OR linguistic simplificat* OR easy to read OR ((easy OR comprehensib* OR simpl* OR friendly OR accessib* OR understandable OR suitable) ADJ1 (written OR language OR information OR read*)) OR apprehensible information OR apprehensible text* OR decipherable information OR decipherable text* OR patient information leaflet* OR text comprehension OR text understanding OR understanding text* OR low literacy OR low health literacy OR understanding of instructions OR understanding instructions OR leichte sprache OR einfache sprach).ti,ab,id.

2. pictorial stimuli/ OR visual stimuli/ OR animation/ OR cartoons/ OR illustrations/ OR graphs/ OR picture books/ OR (illustration* OR picto* OR pictur* OR photo OR photos OR drawing* OR visual* OR symbol* OR image OR images OR animation* OR gif OR infograph* OR cartoon* OR diagram* OR graphic*).ti,ab,id.

3. followup studies/ OR longitudinal studies/ OR field studies/ OR interviews/ OR semi structured interviews/ OR structured interviews/ OR qualitative research/ OR randomized controlled studies/ OR quasiexperimental design/ OR questionnaires/ OR (random* OR RCT* OR field exp* OR quasiexp* OR quasi-exp* OR control group* OR control condit* OR control design* OR posttest OR post test OR pretest OR pre-test OR cohen*s d OR qualitat* OR baseline OR follow-up OR followup OR longitud* OR participant* OR questionnair* OR interview*).ti,ab,id.

4. (infants/ OR toddlers/ OR preschool children/ OR young children/ OR children/ OR preadolescents/ OR youth/ OR adolescents/ OR early adolescents/) NOT (late adolescents/ OR young adults/ OR adults/ OR adult students/)

5. 1 AND 2 AND 3

6. 5 NOT 4

Key: ti = title, ab = abstract, id = key concepts (other keywords added by ERIC indexers to supplement the subject headings), ADJn = word distance of maximum n words

Web of Science Core Collection ((Web of Science Core Collection Editions: Science Citation Index Expanded (SCI-EXPANDED), 1975 - present, Social Sciences Citation Index (SSCI), 1975 - present, Arts & Humanities Citation Index (A&HCI), 1975 - present, Emerging Sources Citation Index (ESCI), 2005 15 - present))

1. TS=(“accessible communicat*” OR “communicat* friendly” OR ((“easy” OR “comprehensib*” OR “simpl*” OR “friendly” OR “accessib*” OR “understandable” OR “suitable”) NEAR/3 (“text” OR “texts” OR “textual” OR “leaflet*”)) OR “linguistic simplificat*” OR “easy to read” OR ((“easy” OR “comprehensib*” OR “simpl*” OR “friendly” OR “accessib*” OR “understandable” OR “suitable”) NEAR/0 (“written” OR “language” OR “information” OR “read*”)) OR “apprehensible information” OR “apprehensible text*” OR “decipherable information” OR “decipherable text*” OR “patient information leaflet*” OR “text comprehension” OR “text understanding” OR “understanding text*” OR “low literacy” OR “low health literacy” OR “understanding of instructions” OR “understanding instructions” OR “leichte sprache” OR “einfache sprach”)

2. TS=(“illustration*” OR “picto*” OR “pictur*” OR “photo” OR “photos” OR “drawing*” OR “visual*” OR “symbol*” OR “image” OR “images” OR “animation*” OR “gif” OR “infograph*” OR “cartoon*” OR “diagram*” OR “graphic*”)

3. TS=(“random*” OR “RCT*” OR “field exp*” OR “quasiexp*” OR “quasi-exp*” OR “control group*” OR “control condit*” OR “control design*” OR “posttest” OR “post test” OR “pretest” OR “pre-test” OR “cohen*s d” OR “qualitat*” OR “baseline” OR “follow-up” OR “followup” OR “longitud*” OR “participant*” OR “questionnair*” OR “interview*”)

4. #1 AND #2 AND #3

Key: TS = topic, which includes title, abstract, author keywords and Web of Science Keywords Plus, NEAR/n = word distance of maximum n words

Extra information #4 Adult filter

Studies with an research population age group from 0-18 years old were excluded, *unless* the research population age group was also 18-85 years old. The age filter was only applied for databases that had appropriate age field metadata (.ag. field in PsycINFO, subject headings for ERIC and Medline (aka Mesh)) and did not rely on information from title or abstract.

Chapter 4. Appendix B

Risk-of-Bias Assessment of the 7 Included Studies Using the RoB-2 Tool (Higgins, Li, & Sterne, 2020)

Study	*D1*	*D2*	*D3*	*D4*	*D5*	*Overall*	*Notes*
Brennan et al. (2005)	Low	n/a	High	Low	Some	Some concerns	D3: participants who failed the Grade 5 text-only task did not complete the Grade 6 text-only task. But they did not fail the Grade 5 text-with-picture-task, so they DID complete this task for Grade 6. As a result, the group who completed the Grade 6 text-with-picture task included fewer proficient readers than the group who completed the Grade 6 text-only task. D5: very limited information on the analysis plan.
Dietz et al. (2009)	Low	n/a	Low	High	Low	Some concerns	D4: After reading, text and photos were removed and participants had to answer comprehension questions. This does not only measure comprehension, but also if the information read is retained in memory.
Jones et al. (2007)	Low	n/a	Low	Some	Some	Some concerns	D4: Not enough information on how the outcome was measured, apparently no validated method. D5: In the results, familiarity with the means of visualization (Widget Rebus symbols) is added as an independent variable, but it is not mentioned before and is not operationalized in the Methods section.

Mansoor & Dowse (2003)	Some	n/a	Low	High	Some	High risk	D1: Group differences are reported in table, but there is no formal test of equivalence between groups. D4: The comprehension test (primary outcome variable) was not validated and there were no clearly described criteria. Many questions about information that was not supported with visuals in either of the two conditions, so no significant difference can be expected for these questions. D5: Reading time is mentioned as an outcome measure, but not used for analysis nor reported in the results. There is only statistical output for separate questions, not for an aggregate measure of comprehension of text vs text + visuals.
Poncelas & Murphy (2007)	Low	n/a	Low	Some	Low	Low Risk	D4: Comprehension as the primary outcome measure is determined somewhat subjectively by the assessor. Interrater agreement between assessors was formally evaluated for 20% of participants and found to be high.
Rose et al. (2011)	Low	n/a	High	High	Some	High risk	D3: Three participants with more severe impairments were removed from the analyses. Not clear on the basis of what criteria. Although there may be doubts about validity when including them, these are also the participants who potentially benefit most from visualization. D4: Only the answers to questions about the paragraphs were visualized, not the body of text itself. This leads to the question how participants are supposed to answer correctly if they don’t understand the text? D5: Non-significant results were not reported, implying a potential publication bias.

							The main outcomes (means and SD for number of questions answered correctly per paragraph) are not reported in a table. In the Discussion it is mentioned that the number of correct answers is higher in the Photographs condition, but the magnitude of difference can not be found in the article.
Saletta et al. (2019)	Low	n/a	Low	Low	Some	Low	D1: no information on the randomization procedure. D5: no information on the statistical analyses plan.
Wilson & Read (2016)	Low	n/a	Low	Low	Some	Low Risk	D5: Many comparisons and analyses on a relatively small sample (*n* = 9). Design appears underpowered. High variance for main outcomes (mean number of paragraphs correct).

D1: randomization process; D2: deviations from the intended interventions (not applicable); D3: missing outcome data; D4: measurement of the outcome; D5: selection of the reported result

Domain 6. Quality of visualisation

Study ID	**Did the researchers justify their choice for the means of visualization?** **Based on what is viewed as best practice in the target population and/or known preferences of the target population** **Based on prior research (with appropriate references)**	**Was the suitability and appropriateness pilot tested with participants that are representative of the target population before execution of the study?**	**If yes, are the results of the pilot test used to optimize the final intervention?**	**Assessor's Judgement**	**General notes**
Brennan et al. (2005)	clinical judgment and best practice with reference			High	Appropriateness not tested, use of Clipart pictures based on 'clinical judgment' of what pictures were suitable, no validated visualisation method
Dietz et al. (2009)	Based on prior research with non-impaired target group, but photo's selected from researchers' personal collection?	Tested with non-impaired 'judges'		High	No way to ascertain if the choice of visualization was appropriate for the target population
Jones et al. (2007)	Widget Rebus symbols were used, but not justified, no info on suitability or effectivity for target population	Speech therapists selected the images		Some concerns	Appropriateness not establised

Mansoor & Dowse (2003)	Based on results of prior pilot study (Dowse & Ehlers, 2000).	Pilot study with 20 participants		Low	Well-designed and executed
Poncelas & Murphy (2007)	Based on what is generally used in daycentres in UK (Makaton symbols). All relevant concept in manifesto represented by a symbol.	Not mentioned in the tekst, probably not. Speech therapists consulted.		Some concerns	Method of visualisation not pilot tested
Rose et al. (2011)	Based on earlier research with people with Aphasia (with contrasting results)	Pilot tested with non-impaired persons to provide normative data on the clarity of the pictorial representations	Pictures were found to be suitable	Some concerns	Text itself was not visualised, only the answers.
Saletta et al. (2019)	Photo's made by research team. No reference to existing literature or best practices.	No clients were involved		High	No way to ascertain if the choice of visualization was appropriate for the target population
Wilson & Read (2016)	images selected in order to extend existing research into the effect of supporting images (Dietz et al. 2009, Brennan et al. 2005). --> Brennen linedrawings: no effect, Dietz photos: some effect	Two independent judges deemed the images 'high quality'.		High	Use of a single image representing a key noun to support the meaning of a whole paragraph. Choice of visualisations on dubious grounds and not tested (authors acknowledge this in Limitations).

Chapter 5. Appendix A

Exemplary search string for PsycInfo

(TI (((psychology* N1 (test OR tests)) OR measur* OR scor*) N3 (stress OR “state anxiety”)) OR AB (((psychology*N1 test*) OR measurement) N3 (stress OR “state anxiety”))) NOT post-traumatic.

Chapter 5. Appendix B

Overview of Stress Self-Report Measures Selected in Step 1

Self-Report Measure	**Author(s) and publication date**	**Outcome**	**Target group**	**Exclusion reason**
Beck Anxiety Inventory (BAI) *	Beck et al. (1988)	State anxiety	Adults and people with ID	
Cognitive Anxiety Scale-State	Weinstein & Smith (1987)	Stress	No information available	No published studies or psychometric data available
Computerized Stress Inventory	Press & Osterkamp (1984)	Stress (stress levels and life satisfaction)	Adults	No published studies or psychometric data available
Depression Anxiety Stress Scales (DASS) *	Lovibond & Lovibond (1995).	Stress / state anxiety	Adults	
Derogatis Stress Profile* (DSP)	Derogatis (1980)	Stress	Adults	
Glasgow Anxiety Scale* (GAS-ID)	Mindham & Espie (2003)	State anxiety	“People with an intellectual disability” (age not specified)	
Hamilton Anxiety Rating Scale (HAM-A)	Hamilton (1959)	State anxiety	Adults with diagnosis of anxiety disorder	Developed for adults with a specific diagnosis
Index of Clinical Stress (ICS) *	Abell (1991)	Stress	Adults & youths age 12+	

Self-Report Measure	Author(s) and publication date	Outcome	Target group	Exclusion reason
Lifestress Inventory (LI) *	Bramston & Fogarty (1997)	Stress	Suitable for a wide range of people, including MID	
Pikunas Adult Stress Inventory (PASI)	Pikunas (1984)	Stress	Adults & youths age 16+	No published studies or psychometric data available
Psychological Stress Measure (PSM-9) *	Tessier et al. (1990)	Stress	Adults	
Perceived Stress Questionnaire (PSQ) *	Levenstein et al. (1993)	Stress	Adult	
Perceived Stress Reactivity Scale (PSRS)	Schlotz et al. (2011)	Stress	Adults	Anticipatory & specific situations
Perceived Stress Scale (PSS) *	Cohen et al. (1983)	Stress	Adults	
Recovery-Stress-Questionnaires (RESTQ)	Kallus & Kellmann (2016)	Stress	Adults, separate version for adolescents and children	Specific situations: stress during recreational activities
Stress Arousal Checklist (SACL) *	Mackay et al. (1978)	Stress and arousal	Adults	
Self-Rating Anxiety Scale for adults with Intellectual Disabilities (SAS-ID) *	Zung (1971) ID version: Lindsay & Michie (1988)	State anxiety	Adults with ID	

Self-Report Measure	Author(s) and publication date	Outcome	Target group	Exclusion reason
Stress Evaluation Inventory (SEI)	Kulhavy & Dee-Burnett (1984).	Stress (career, family, and personal-social)	Adults	No published studies or psychometric data available
Stress Overload Scale (SOS) *	Amirkhan (2012)	Stress	Adults	
Stress Response Inventory (SRI)	Koh et al. (2001)	Stress	Adults	No information on English version available
Stress Symptom Checklist (SSCL) *	Schlebusch (2004)	Stress levels	Adults	
Subjective Stress Scale (SSS)	Bramston & Bostock (1994)	Stress	Adults with ID	Updated version of SSS is included (LI)
State version of the State-Trait Anxiety Inventory (STAI-S) *	Spielberger (1981)	State anxiety	Adults	
Stress Analysis System (SAS)	Nelson et al. (1983)	Stress	No information found	No published studies or psychometric data available
Subjective Units of Distress Scale (SUDS)	Wolpe (1969)	Stress	Adults, adolescents and children, including persons with ID	Not a standardized measure, but a flexible clinical and research outcome tool

* included in final analysis

Chapter 6. Appendix A

ABAS-3 items included in the original abbreviated self-report version

Communication

1. Tells parents, friends or others about my favorite activities.
2. Starts conversations on topics of interest to others.
3. Distinguishes truthful from exaggerated claims, from friends, advertising or others.

Community Use

4. I make appointments by phone or internet.
5. Before buying an item in a store, gives careful thought to the need for it and its cost.
6. Walks or rides bike alone to locations within a 1-mile or 5-block radius of home or work.

Functional Academics

7. Records dates and times for appointments and deadlines.
8. Writes and sends letters, personal notes, or emails.
9. Checks the accuracy of charges before paying a bill.

Home Living

10. Cleans his or her room or living quarters regularly.
11. Cooks simple foods on a stove (for example, eggs or canned soup).
12. Folds clean clothes.

Health and Safety

13. Uses tools and equipment safely.
14. Plans meals in order to get necessary nutrition
15. Cares for own minor injuries (for example, paper cuts, knee scrapes, nosebleeds).

Leisure

16. Plans ahead for fun activities on free days or afternoons.
17. Participates in an organized program for a sport or hobby (for example, practices basketball or takes a music class).
18. Invites others to join him or her in playing games and other fun activities.

Self-Care

19. Selects appropriate clothes for different occasions (for example, casual activities or formal events).

20. Eats a variety of foods instead of preferring only one or two.

Self-Direction

21. Controls feelings when not getting his or her own way.

22. Plans home projects in logical steps

23. Makes important decisions only after careful consideration, without rushing

24. Limits time playing computer games or other nonproductive activities

Social

25. Avoids friends and social settings that may be harmful or dangerous.

26. Refrains from saying or doing things that might embarrass or hurt others.

27. Shows sympathy for others when they are sad or upset.

Work

28. Checks own work to determine if improvements are needed.

29. Seeks help from supervisor, as needed, when work-related problems or questions arise.

30. Is productive and cooperative as part of groups or teams.

Note: the original ABAS-3 uses a third-person perspective for all items in the 'Adult form', regardless if the informant is a client or a proxy informant. In the Dutch version that was used in this study, first-person language is used for items in the self-report versions.

Example of a question from the adapted ABAS-3

* formulation of original item: 'I make important decisions only after careful consideration and without rushing'

Summary

Introduction

People with intellectual disabilities face greater challenges in making their voices heard for various reasons. They frequently rely on caregivers or relatives for daily support, who tend to take over tasks and responsibilities rather than promoting self-determination. In addition to limited opportunities for participation—caused by caregivers and family members restricting and regulating access to society—communication barriers also limit many people with intellectual disabilities from making their voices heard.

Traditionally, a focus on limitations rather than capabilities reinforced the belief that people with intellectual disabilities were incapable of making their own choices and living their lives as they wished. However, new scientific insights and advocacy efforts have increasingly recognised that people with intellectual disabilities are fully capable of representing their own interests. One essential condition for encouraging participation and inclusion is ensuring that communication tools are tailored to their abilities.

One way to capture the opinions, concerns, and wishes of people with intellectual disabilities is through self-reporting. Self-report instruments, such as questionnaires and structured interviews, are used in diagnostics, satisfaction surveys, needs assessments, and scientific research. However, self-report tools developed for the general population often fail to consider the cognitive and communicative challenges that characterise many individuals with intellectual disabilities. Consequently, practitioners and researchers often question the validity and reliability of self-reported information from this group.

This PhD research explored how people with intellectual disabilities can meaningfully participate in diagnostics and research through self-reporting. Two recurring themes were identified and studied: 1. the cognitive accessibility of self-report instruments, and 2. the interpersonal dynamics during their administration. In this study, cognitive accessibility refers to the extent to which the instrument's design accounts for the cognitive and communicative challenges associated with intellectual disabilities. Additionally, this project examined how interpersonal dynamics may influence self-report outcomes and explored strategies to minimise this risk.

The studies in this thesis

This thesis consists of three parts. The first part reviews existing knowledge. The second part adds scientific insights through empirical research on less-studied topics. The third part applies existing and new knowledge to assess whether this leads to better outcomes.

Assessing the evidence base

The first part (Chapter 2) systematically reviewed existing research on adaptations needed to make self-report instruments and administration procedures suitable for individuals with intellectual disabilities. The findings were presented in a table of recommendations, detailing the quality of supporting evidence. The discussion identified areas requiring further research, such as effective visual aids (pictures, photos, symbols) to clarify written language and methods to identify and mitigate response bias.

Expanding the evidence base

Chapter 3 examined how the self-reported experiences of young people with mild intellectual disabilities were influenced by the presence of a helper. It compared the scores of youth assisted by a dependent caregiver versus those assisted by a neutral individual or completing the questionnaire independently. Focus groups were conducted to explore these findings. Results showed that participants gave more positive ratings of group living environments when assisted by a familiar caregiver. The youth feared damaging their relationship with caregivers by being critical, leading to less honest responses. The key recommendation was to ensure independent support for sensitive topics in self-report studies.

Chapter 4 investigated whether adding pictures to simplified text improved comprehension for individuals with reading difficulties. A meta-analysis aggregated results from studies involving people with intellectual disabilities, aphasia, or limited language proficiency due to being non-native speakers. No evidence was found to suggest that pictures improved comprehension for any group. However, the low quality of included studies prevented firm conclusions.

Applying existing and new knowledge

In Chapter 5, the suitability of several self-report instruments for measuring stress in people with mild intellectual disabilities was evaluated. A scoping review identified potential tools, which were then assessed for psychometric quality, prior use with the target group, and alignment with findings from Chapter 2. The three most suitable instruments were specifically developed for this population.

Chapter 6 tested whether adapting an existing self-report tool improved comprehension and reliability. Recommendations from earlier chapters were applied to the ABAS-3, a commonly used tool. Cognitive interviews with 18 adults with mild intellectual disabilities assessed the changes. Participants found the adapted tool easier to understand and complete. Quantitative analysis indicated improved reliability, closer alignment between participants' responses and their self-descriptions, and better agreement between participants and proxies.

General conclusion of the studies in this thesis

Improving cognitive accessibility

This thesis demonstrates that improving the cognitive accessibility of self-report tools can reduce barriers for people with mild intellectual disabilities to participate in assessments and research. Such adaptations result in more valid and reliable responses, greater agreement between self- and proxy-reports, and reduced cognitive load.

Involving individuals with intellectual disabilities in designing or adapting instruments is crucial to ensure that questions are understandable, response options are appropriate, and visual aids convey intended meanings.

Interaction factors

Many people with intellectual disabilities need help completing self-report instruments. Caregivers may inadvertently influence responses, creating response bias. This bias often arises from power imbalances, as individuals dependent on caregivers may avoid critical responses for fear of repercussions. The risk of bias can be reduced by enabling independent completion of self-reports, facilitated by improved cognitive accessibility. When help is needed, providing trained, independent support minimises the risk of introducing bias.

Future research and development

The subject of this thesis, ‘self-reporting for people with intellectual disabilities,’ covers a wide spectrum of topics. From all the potential areas of research, only a limited selection could be made for further exploration. As a result, many topics remain underexplored. Examples of areas that have not yet been sufficiently investigated include: which types of visualisations are genuinely helpful and for whom, how people with more severe cognitive impairments can be enabled to express their opinions, how frequently different forms of bias occur and to what extent they distort scores (and how to address this), and which response categories are most suitable under specific circumstances.

The Discussion (Chapter 7) highlights two additional opportunities for future development. The first involves combining (behavioural) observations with self-reported scores to gain a better understanding of the ecological validity of self-reporting and to interpret differences in scores between respondents and proxies (‘who reports it best?’). The second development is the rise of adaptive testing procedures. Through the use of ‘computer adaptive testing’ (CAT), a broader range of respondents’ abilities and preferences can be accommodated. CAT algorithms can automatically tailor the characteristics of questions and answers to the respondent’s (cognitive) level of functioning and language proficiency. This addresses the issue that traditional ‘one-size-fits-all’ instruments, even if they are adapted for people with intellectual disabilities, may suit a large part of the population but are not suitable for everyone. An additional benefit of using digital and adaptive testing procedures is the potential to incorporate multimedia support, such as read-aloud functions or videos that provide further explanations about the content of the item.

Adaptations may benefit a broader audience

When we closely examine the adjustments proposed to support people with intellectual disabilities, it becomes evident that many of these recommendations could equally apply to other individuals with cognitive or communicative challenges. For example, writing texts in accessible language is just as beneficial for people with a different first language, children, or those who struggle with reading due to a lack of educational opportunities. It reduces the cognitive load for anyone reading a text and can also enhance reading ease for people without reading difficulties. Limiting the number of response categories will similarly assist individuals with memory impairments caused by acquired brain injury or aphasia.

It is therefore plausible that, at least to some extent, the recommendations in this thesis are not limited to self-reporting for people with intellectual disabilities. They are equally relevant for other groups with cognitive impairments or language proficiency challenges.

Challenges

The studies in this thesis primarily examined the technical and procedural conditions required for people with intellectual disabilities to express their opinions through self-reporting. In the Discussion (Chapter 7), two contextual factors are identified that may hinder the participation and inclusion of individuals with intellectual disabilities, even when self-report instruments are optimally tailored to their needs.

The first obstacle is the so-called ‘gatekeeper problem.’ This issue arises when those responsible for supporting individuals with disabilities decide on their behalf what they can and cannot do independently. During the research conducted for this thesis, there were instances where individuals with disabilities were unnecessarily and unsolicitedly assisted, increasing the likelihood of response bias. In some cases, potential participants were excluded from participation in research by gatekeepers, who judged that participation would be too challenging, yield insufficiently reliable information, or cause excessive stress. In many of these cases, the possibility of participation and the potential burden were not discussed with the individual in question.

Another factor at odds with the principle of inclusivity is that adapting existing questionnaires often results in the creation of special ‘intellectual disability versions’ of self-report instruments. This assumes that there are two clearly demarcated types of people: people with and people without intellectual disabilities. This is conceptually untenable; there is no uniformly defined measurable boundary that distinguishes people with intellectual disabilities from neurotypically functioning individuals. Moreover, it is morally questionable to distinguish groups of people solely based upon their cognitive abilities, without valuing their strengths, talents, and personal experiences. So paradoxically, one recommendation from this PhD – make adaptations to mainstream measures – contravenes one of its central tenets – the promotion of inclusivity.

A practical drawback of separate versions for people with and without intellectual disabilities is that derivative versions must be revalidated for reliability and validity, and new normative data must be collected.

Fortunately, there is a straightforward solution to both the conceptual and methodological issues. If developers and publishers can be convinced of the importance of involving people with intellectual disabilities in the development and norm collection processes, separate versions will no longer be necessary. The resulting inclusive self-report instruments would allow as many people with (mild) intellectual disabilities as possible to participate in assessments and research while enabling scores to be compared with representative normative groups. This would not only enhance the participation of individuals with intellectual disabilities but also benefit those who face other challenges in using self-report instruments.

Conclusion

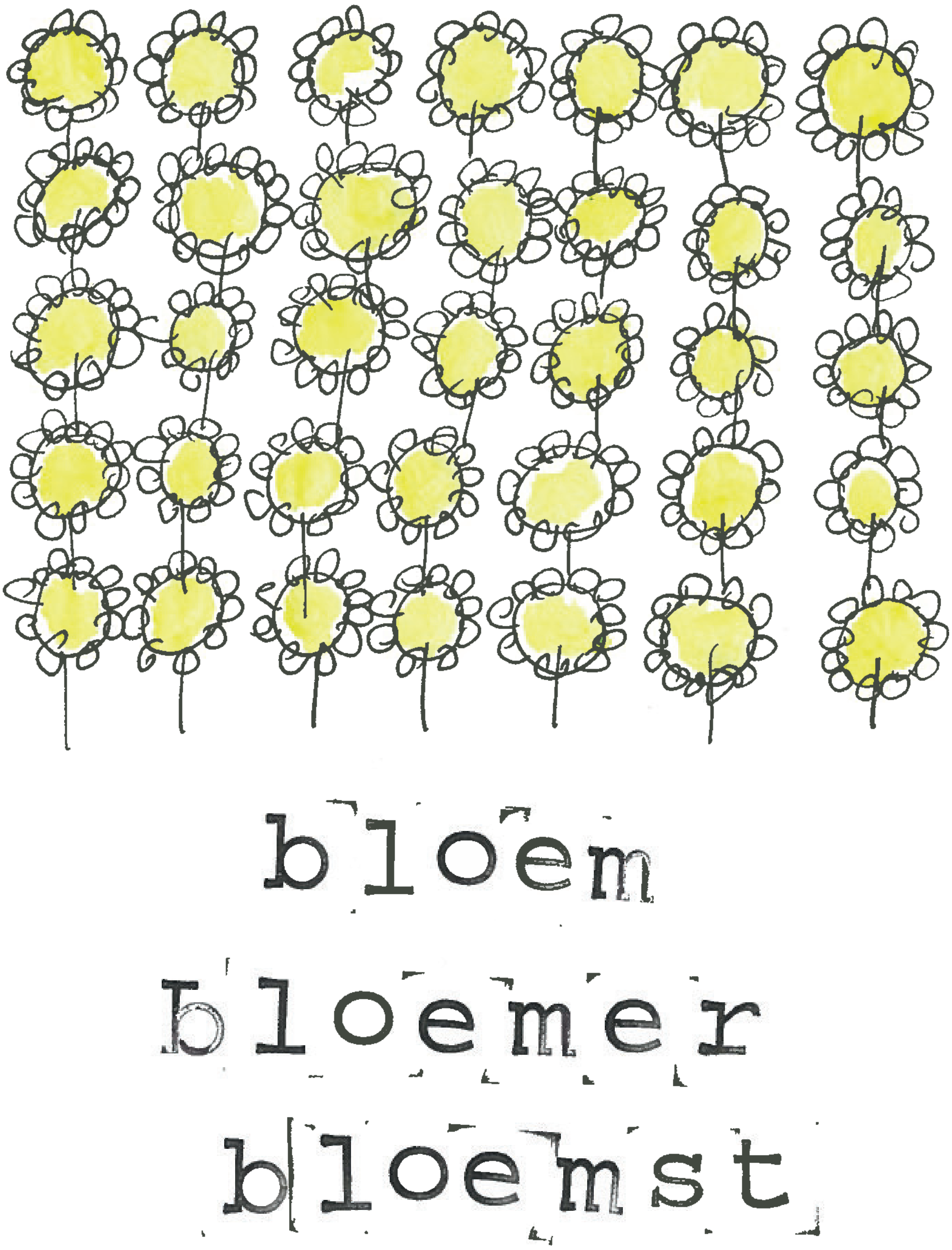
This thesis demonstrates that people with intellectual disabilities are highly capable of participating in assessments and research through self-reporting. This is achieved not only by making questionnaires easier to complete but also—perhaps most importantly—by recognising that the opinions of people with intellectual disabilities matter and that every individual, in their own way, independently or with support, is able to share what they think, feel, and believe.

Many individuals with disabilities are accustomed to being cared for and having challenges removed from their path. As a result, claiming your space and demanding to be included will not come naturally for many people with intellectual disabilities. Caregivers, relatives, and other support figures can take the first step toward greater autonomy and inclusivity by removing barriers, allowing space, trusting in their abilities, providing a platform, and stepping back when necessary.

There are many soft and quiet voices that deserve to be noticed. I hope the findings of this PhD research will be used to make it easier for people with intellectual disabilities to express themselves and be heard.

Nederlandse samenvatting

Deze samenvatting is vooral geschikt voor mensen die al wat van het onderwerp weten. Bijvoorbeeld mensen die onderzoek doen of die zorg verlenen aan mensen met een verstandelijke beperking. Er is ook een samenvatting in makkelijker leesbaar Nederlands. Die staat in het volgende hoofdstuk.



Inleiding

Mensen met een verstandelijke beperking hebben om verschillende redenen meer moeite om hun stem te laten horen. Vaak zijn ze voor hun dagelijkse ondersteuning afhankelijk van begeleiders of verwanten, die gewend zijn om taken en verantwoordelijkheden over te nemen, in plaats van zelfbeschikking te bevorderen. Naast beperkte mogelijkheden voor participatie doordat verzorgers en familieleden de toegang tot het maatschappelijke leven beperken en reguleren, beperken communicatieve barrières het vermogen van veel mensen met een verstandelijke beperking om hun stem te laten horen.

Door een eenzijdige focus op beperkingen in plaats van capaciteiten bestond lang de opvatting dat mensen met een verstandelijke beperking niet in staat zijn om hun eigen keuzes te maken en hun leven in te richten zoals zij dat willen. Door onder andere nieuwe wetenschappelijke inzichten en de inzet van belangenverenigingen is de afgelopen jaren in toenemende mate erkend dat mensen met een verstandelijke beperking prima in staat zijn om hun eigen belangen te behartigen. Een van de voorwaarden voor het stimuleren van participatie en inclusie van mensen met een verstandelijke beperking is dat communicatiemiddelen afgestemd zijn op hun (on)mogelijkheden.

Een middel om de meningen, zorgen en wensen van mensen met een verstandelijke beperking kenbaar te maken is door middel van zelfrapportage. Zelfrapportage-instrumenten, in de vorm van vragenlijsten en gestructureerde interviews, worden onder andere gebruikt in diagnostiek, tevredenheidsonderzoek, vaststellen van ondersteuningsbehoeften en bij deelname aan wetenschappelijk onderzoek. Zelfrapportage-instrumenten die zijn ontwikkeld voor gebruik in de algemene populatie houden in het algemeen weinig rekening met de cognitieve en communicatieve beperkingen die het functioneren van de meeste mensen met een verstandelijke beperking kenmerken. Dit maakt dat onder behandelaars en onderzoekers twijfel bestaat over de validiteit en betrouwbaarheid van zelfgerapporteerde informatie door mensen met een verstandelijke beperking.

In dit promotie-onderzoek werd onderzocht hoe mensen met een verstandelijke beperking in staat kunnen worden gesteld om op een betekenisvolle manier te participeren in diagnostiek en onderzoek door middel van zelfrapportage. In dit proefschrift werden twee terugkerende thema's geïdentificeerd en onderzocht: 1. de cognitieve toegankelijkheid van zelfrapportage-instrumenten en 2. de interpersoonlijke dynamiek van de afname. Cognitieve toegankelijkheid verwijst in dit onderzoek naar de mate waarin het ontwerp van het instrument rekening houdt met de cognitieve en communicatieve uitdagingen die gepaard gaan met de verstandelijke beperking. Daarnaast wordt in dit project onderzocht op welke manier interpersoonlijke dynamiek de uitkomsten van zelfrapportage kan beïnvloeden en wordt gezocht naar manieren om het risico op beïnvloeding te minimaliseren.

De onderzoeken in dit proefschrift

Dit proefschrift is opgebouwd uit drie delen. In het eerste deel werd onderzocht wat eerder al was onderzocht en wat we al wisten. In het tweede deel werd door middel van empirisch onderzoek wetenschappelijk kennis toegevoegd over onderwerpen waarover nog weinig of niets bekend was. In het derde deel werd de bestaande en nieuwe kennis toegepast en werd onderzocht of dat leidde tot betere uitkomsten.

Wat al bekend was

In het eerste deel (Hoofdstuk 2) werd met een systematische review onderzocht wat al uit wetenschappelijk onderzoek bekend was over aanpassingen die nodig zijn om zelfrapportage-instrumenten en afnameprocedures geschikt te maken voor gebruik door mensen met een verstandelijke beperking. De resultaten werden gepresenteerd in een tabel met aanbevelingen, waarbij voor iedere aanbeveling werd aangegeven wat de kwaliteit van het onderzoek was dat eraan ten grondslag lag. In de Discussie werd geïnventariseerd op welke thema’s verder onderzoek nodig was. Voorbeelden van concrete onderwerpen die verder onderzocht moesten worden waren onder andere welke soorten visualisatie (plaatjes, foto’s, picto’s) gebruikt kunnen worden om de betekenis van geschreven taal te verduidelijken en manieren om beïnvloeding (‘bias’) van scores te herkennen, duiden en te voorkomen.

Ontwikkelen van nieuwe kennis

In het onderzoek dat in Hoofdstuk 3 wordt beschreven werd onderzocht hoe zelfgerapporteerde ervaringen van jongeren met een licht verstandelijke beperking werden beïnvloed door de aanwezigheid van een persoon die hen hielp. Er werd gekeken naar verschillen in scores tussen jongeren die werden geholpen door een begeleider met wie ze een afhankelijkheidsrelatie hadden versus jongeren die werden geholpen door een neutraal persoon en jongeren die zelfstandig een vragenlijst invulden. De kwantitatieve bevindingen werden opgevolgd met focusgroepen met jongeren. De bevindingen werden besproken met de jongeren en mogelijke verklaringen werden onderzocht. De resultaten lieten zien dat jongeren positievere beoordelingen van het leefklimaat op een groep gaven in aanwezigheid van een bekende begeleider. Uit de focusgroepen bleek dat jongeren bang zijn dat de relatie met de begeleider wordt geschaad door kritisch te zijn en dat deze hen daardoor minder goed helpt of boos op hen wordt. De belangrijkste aanbeveling was om bij zelfrapportage-onderzoek bij gevoelige onderwerpen te zorgen voor onafhankelijke ondersteuning als een jongere hulp heeft bij het invullen.

In hoofdstuk 4 werd onderzocht of het toevoegen van plaatjes aan eenvoudig geschreven teksten ertoe leidt dat mensen die niet goed kunnen lezen de tekst beter begrijpen. Dit werd gedaan door de resultaten van verschillende onderzoeken naar dit onderwerp te aggregeren in een meta-analyse. Zowel onderzoeken met mensen met een verstandelijke beperking, mensen met afasie en mensen die de taal minder goed machtig waren doordat het niet hun eerste taal was werden geïncludeerd. Voor geen van de groepen werd bewijs gevonden dat plaatjes deze mensen helpt om (eenvoudig) geschreven taal te begrijpen. Maar omdat de kwaliteit van de geïncludeerde onderzoeken in het algemeen mager bleek, was het niet mogelijk om harde conclusies te trekken.

Toepassen van bestaande en nieuwe kennis

In hoofdstuk 5 werd de geschiktheid van een aantal zelfrapportage-instrumenten om stress te meten bij mensen met een licht verstandelijke beperking beoordeeld Door middel van een ‘scoping literature review’ werden potentieel relevante instrumenten geïdentificeerd. De gevonden instrumenten werden beoordeeld op psychometrische kwaliteit, de ervaringen uit eerder onderzoek bij de doelgroep en door de kenmerken van het instrument te vergelijken met de resultaten van de systematische review uit hoofdstuk 2 gecombineerd met input van een panel van experts. Drie instrumenten werden beoordeeld als meest geschikt. Dit waren – niet geheel toevallig - de drie instrumenten die specifiek voor gebruik met de doelgroep mensen met een verstandelijke beperking waren ontwikkeld.

In hoofdstuk 6 werd onderzocht of het aanpassen van een bestaand zelfrapportage-instrument daadwerkelijk leidt tot een beter te begrijpen instrument en meer betrouwbare en valide antwoorden. De aanbevelingen uit de eerdere studies in dit proefschrift werden toegepast om de cognitieve toegankelijkheid van een bestaand en veelgebruikt zelfrapportage-instrument (de ABAS-3) te verbeteren. Onderzocht werd of dit leidde tot een instrument dat als toegankelijker werd ervaren door mensen met verstandelijke beperkingen. 18 volwassenen met een (licht) verstandelijke beperking werden geïnterviewd terwijl ze de vragen invulden, met behulp van cognitieve interviewmethodologie. Door de resultaten kwantitatief te analyseren en de resultaten van zelf- en proxy-rapportages te vergelijken, werd onderzochten of er verschillen in betrouwbaarheid en validiteit bestonden tussen het originele en aangepaste instrument. Participanten vonden het aangepaste instrument makkelijker te begrijpen en in te vullen. Er waren aanwijzingen dat het aangepaste instrument beter presteerde dan het originele instrument bij mensen met een verstandelijke beperking. Zo bleek de betrouwbaarheid van het aangepaste instrument beter, kwamen de gedragsbeschrijvingen van participanten beter overeen met de door henzelf toegekende scores en was er meer overeenkomst tussen de beoordeling van proxy’s en die van participanten zelf.

Algemene conclusie van de onderzoeken in dit proefschrift

Aanpassingen om de cognitieve toegankelijkheid te verbeteren

De resultaten in dit proefschrift laten zien dat het mogelijk is om de cognitieve toegankelijkheid van zelfrapportage-instrumenten te vergroten door het toepassen van kennis uit eerder onderzoek (hoofdstuk 2) en de bevindingen uit de onderzoeken in hoofdstuk 3 en 4. Het verbeteren van de cognitieve toegankelijkheid van zelfrapportage-instrumenten verlaagt de drempel voor mensen met lichte verstandelijke beperkingen om deel te nemen aan assessment en onderzoek. Zelfrapportage-instrumenten die aangepast zijn aan het lees- en begripsniveau van mensen met een verstandelijke beperking levert meer valide en betrouwbare antwoorden op, meer overeenstemming tussen respondenten en proxy's en verlaagt de cognitieve belasting.

Om ervoor te zorgen dat aanpassingen de cognitieve toegankelijkheid daadwerkelijk verbeteren, moeten mensen met een licht verstandelijke beperking altijd betrokken worden bij de constructie of aanpassing van instrumenten. In co-creatie kan gecontroleerd worden of vragen begrepen worden, antwoordcategorieën passend zijn en ondersteunende afbeeldingen de bedoelde betekenis overbrengen.

Scan de QR code voor een poster met tips om vragen makkelijker te maken:

Interactiefactoren

Veel mensen met een verstandelijke beperking hebben hulp nodig bij het invullen van zelfrapportage-instrumenten. Begeleiders oefenen bij het invullen van een zelfrapportagemeting of bij het interviewen van de ander ongewild een zekere invloed uit op de ander, waardoor vertekening van de antwoorden kan ontstaan (in het Engels heet dit 'response bias'). Een belangrijke oorzaak van dit type bias is een machtsongelijkheid tussen respondenten en degenen die hen helpen. Mensen die zorg en ondersteuning nodig hebben vanwege cognitieve en aanpassingsstoornissen zijn vaak afhankelijk van verzorgers of familieleden om zich uit te drukken en ondersteuning te regelen. Deze afhankelijkheid zorgt ervoor dat veel mensen met een licht verstandelijke beperking minder kritisch durven te zijn, dat zij eerder sociaal wenselijke antwoorden geven en terughoudender zijn om gevoelige onderwerpen te bespreken in de aanwezigheid van een begeleider. De kans dat dit soort factoren invloed heeft op zelfgerapporteerde scores kan verkleind worden door mensen zelfstandig een zelfrapportage-instrument in te laten vullen. Het verbeteren van de cognitieve toegankelijkheid kan eraan bijdragen dat zoveel mogelijk mensen met een verstandelijke beperking dit zelfstandig of met minimale hulp kunnen. Als een persoon hulp nodig heeft of zoekt bij het invullen van de vragen, is het wenselijk om hem of haar een persoon te laten kiezen die hij of zij vertrouwt of die hem of haar helpt, of om te voorzien in onafhankelijke ondersteuning die getraind is om ongewenste beïnvloeding van scores te minimaliseren.

Scan de QR code voor tips hoe je ervoor kunt zorgen dat mensen die hulp nodig hebben zo min mogelijk beïnvloed worden bij het invullen van vragenlijsten:

Toekomstig onderzoek en ontwikkelingen

Het onderwerp van dit proefschrift, ‘zelfrapportage voor mensen met een verstandelijke beperking’, bestrijkt een breed spectrum aan thema’s. Uit alle mogelijk te onderzoeken onderwerpen kon in dit promotie-onderzoek een beperkte keuze worden gemaakt om nader uit te werken. Hierdoor blijft een groot aantal onderwerpen nog onderbelicht. Voorbeelden van onderwerpen die nog niet voldoende zijn onderzocht: welke visualisaties daadwerkelijk helpen voor wie, hoe mensen met ernstigere cognitievere beperkingen in staat kunnen worden gesteld om hun mening te geven, hoe vaak verschillende vormen van bias daadwerkelijk voorkomen en in welke mate zij scores vertekenen (en wat daaraan te doen) en welke responscategorieën het meest geschikt zijn onder welke omstandigheden.

In de Discussie (Hoofdstuk 7) worden nog twee kansen voor toekomstige ontwikkeling genoemd. De eerste is het combineren van (gedrags)observaties met zelfgerapporteerde scores om meer zicht te krijgen op de ecologische validiteit van zelfrapportage en om verschillen in scores tussen respondenten en proxy’s beter te kunnen duiden (‘who reports it best?'). Een tweede ontwikkeling is de opkomst van adaptieve testprocedures. Door de inzet van 'computer adaptive testing' (CAT) kan rekening gehouden worden met een breder scala aan vaardigheden en voorkeuren van respondenten. Met CAT-algoritmen kunnen vraag- en antwoordkenmerken geautomatiseerd worden gestemd op het (cognitieve) niveau van functioneren en de taalbeheersing van de respondent. Dit lost het probleem op van de 'one-size-fits-all' instrumenten die, zelfs als ze zijn aangepast voor mensen met een verstandelijke beperking, passend zijn voor een groot deel van de populatie, maar nog lang niet voor iedereen passend zijn. Een bijkomend voordeel van het gebruik van digitale en adaptieve testprocedures is dat het mogelijkheden biedt om gebruik te maken van multimediale ondersteuning, zoals voorleesfuncties of filmpjes die meer uitleg geven over de inhoud van het item.

Aanpassingen voor een bredere doelgroep

Als we goed kijken naar de aanpassingen die worden voorgesteld om mensen met een verstandelijke beperking te ondersteunen, wordt duidelijk dat dat veel van de aanbevelingen net zo goed van toepassing kunnen zijn op andere mensen met cognitieve of communicatieve uitdagingen. Teksten schrijven in toegankelijke taal is bijvoorbeeld net zo nuttig voor mensen met een andere eerste taal, kinderen, of mensen die minder goed kunnen lezen door een gebrek aan opleidingsmogelijkheden. Het verlaagt de cognitieve belasting voor iedereen die een tekst leest en kan ook het leesgemak voor mensen zonder leesproblemen ten goede komen. Het beperken van het aantal antwoordcategorieën zal ook mensen helpen die geheugenbeperkingen hebben als gevolg van verworven hersenletsel of afasie. Het is daarom aannemelijk dat, tenminste tot op zekere hoogte, de aanbevelingen in dit proefschrift niet beperkt zijn tot zelfrapportage voor mensen met verstandelijke beperkingen. Ze gelden net zo goed voor andere doelgroepen met cognitieve beperkingen of taalvaardigheidsproblemen.

Uitdagingen

In de onderzoeken in dit proefschrift werd met name onderzocht welke technische en procedurele randvoorwaarden nodig zijn zodat mensen met een verstandelijke beperking kun mening kunnen geven door middel van zelfrapportage. In de Discussie (hoofdstuk 7) worden twee contextuele factoren benoemd die de participatie en inclusie van mensen met een verstandelijke beperking kunnen belemmeren. Zelfs als zelfrapportage-instrumenten optimaal zijn aangepast aan de behoeften van mensen met een verstandelijke beperking.

Het eerste obstakel is het zogenaamde 'poortwachter probleem'. Dit probleem treedt op als mensen die verantwoordelijk zijn voor de ondersteuning van personen met een beperking voor de ander bepalen wat zij wel en niet zelfstandig kunnen. Tijdens de uitvoering van het onderzoek in dit proefschrift werd een aantal keer ervaren dat mensen met een beperking ongevraagd en onnodig geholpen werden, waardoor de kans op beïnvloeding toenam. Daarnaast werden potentiële participanten in sommige gevallen door poortwachters uitgesloten van deelname aan onderzoek omdat zij dat niet aan zouden kunnen, dit onvoldoende betrouwbare informatie op zou leveren of het de participant teveel stress zou geven. In veel van deze gevallen was de vraag om deelname en de eventuele belasting voor de participant niet met hem of haar besproken.

Een ander gegeven dat op gespannen voet staat met de inclusiviteitsgedachte is dat er door het aanpassen van bestaande vragenlijsten speciale ‘verstandelijke beperking versies’ van zelfrapportage-instrumenten ontstaan. Dit veronderstelt dat er twee duidelijk afgebakende soorten mensen zijn: mensen met en mensen zonder een verstandelijke beperking. Dit is conceptueel onhoudbaar; er is geen uniform gedefinieerde meetbare grens die mensen met een verstandelijke beperking onderscheidt van neurotypisch functionerende mensen. Bovendien is het moreel discutabel om groepen mensen uitsluitend op basis van hun cognitieve capaciteiten te onderscheiden, zonder hun sterke kanten, talenten en persoonlijke ervaringen te waarderen. Dus paradoxaal genoeg is één aanbeveling van deze PhD - aanpassingen maken aan reguliere maatregelen - in tegenspraak met één van haar centrale grondbeginselen - het bevorderen van inclusiviteit. Een praktisch nadeel van separate versies voor mensen met en zonder verstandelijke beperking is dat afgeleide versies opnieuw onderzocht moeten worden op validiteit en betrouwbaarheid en dat nieuwe normgegevens moeten worden verzameld.

Gelukkig is er een eenvoudige oplossing voor zowel het conceptuele als het methodologische probleem. Als ontwikkelaars en uitgevers overtuigd kunnen worden van het belang om mensen met een verstandelijke beperking te betrekken in het proces van ontwikkeling en normverzameling, dan zijn aparte versies niet langer nodig. De resulterende *inclusieve* zelfrapportage-instrumenten zorgen ervoor dat zoveel mogelijk mensen met een (licht) verstandelijke beperking kunnen deelnemen aan assessment en onderzoek, terwijl scores vergeleken kunnen worden met representatieve normgroepen. Dit bevordert daarnaast niet alleen de deelname van mensen met een verstandelijke beperking, het komt ook ten goede aan de participatie van mensen die om andere redenen moeite hebben met het gebruiken van zelfrapportage-instrumenten.

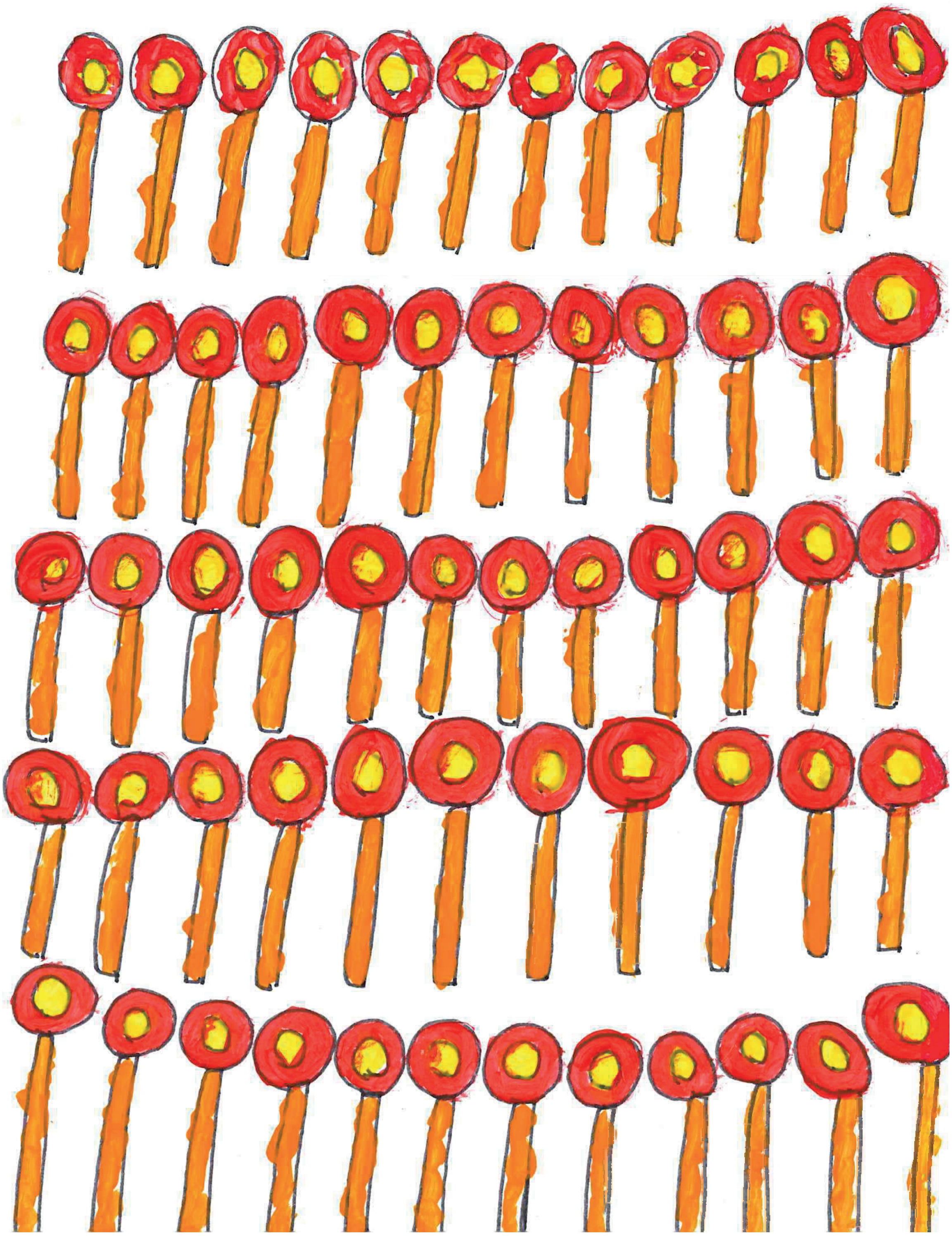
Tot slot

Dit proefschrift laat zien dat mensen met een verstandelijke beperking zeer goed in staat zijn om deel te nemen aan assessments en onderzoek door middel van zelfrapportage. Niet alleen door het invullen van vragenlijsten makkelijker te maken, maar ook - en dat is misschien wel het belangrijkste - door te erkennen dat de mening van mensen met een verstandelijke beperking ertoe doet en dat ieder mens in staat is om op zijn eigen manier, zelf of met hulp van anderen, te delen wat hij denkt, voelt en vindt. Veel mensen met een beperking zijn gewend om verzorgd te worden en dat uitdagingen voor hen weggenomen worden. Voor veel mensen met een verstandelijke beperking is het daarom niet vanzelfsprekend om ruimte in te nemen en te eisen om betrokken te worden. Begeleiders, verwanten en andere steunfiguren van mensen met een verstandelijke beperking kunnen de eerste stap zetten naar meer autonomie en inclusiviteit door hekken weg te halen, ruimte te laten, vertrouwen te hebben in de eigen mogelijkheden, een podium te geven en een stapje terug te doen.

Er zijn veel zachte en stille stemmen die het verdienen om gehoord te worden. Ik hoop dat de opbrengsten van dit promotie-onderzoek gebruikt worden om het voor mensen met een verstandelijke beperking makkelijker te maken om zichzelf te laten horen.

Samenvatting in makkelijker Nederlands

Deze samenvatting is vooral geschikt voor mensen die nog niet zoveel van het onderwerp weten. En ook voor mensen die het fijn vinden om makkelijker leesbare teksten te lezen. Wil je meer details lezen over het onderzoek? Bijvoorbeeld hoe we het uitgevoerd hebben en wat er precies uitkomt? Lees dan de samenvatting met meer moeilijke woorden in het hoofdstuk hiervoor.



Waar gaat het proefschrift over?

Ik heb onderzocht hoe we ervoor kunnen zorgen dat zoveel mogelijk mensen met een lichte of matige verstandelijke beperking zelf vragenlijsten kunnen invullen. Of in interviews kunnen vertellen wat ze belangrijk vinden. Vaak zijn de vragen die we stellen zo ingewikkeld dat veel mensen met een verstandelijke beperking geen antwoord kunnen geven. Dat moet veel makkelijker! Hoe moeten de vragen er dan uitzien? Wat is begrijpelijke taal? Soms hebben mensen hulp nodig om het te vertellen. Hoe kunnen begeleiders dan het beste helpen?

Waarom vind ik het nodig dat dit onderzocht wordt? Omdat het belangrijk is dat mensen met een verstandelijke beperking *zelf* mogen vertellen wat ze vinden en voelen. In plaats van dat groepsleiding of een ouder of broer het vertelt. Mensen met een verstandelijke beperking hebben het *recht* om dat zelf te vertellen. Dat staat in het 'Verdrag inzake de rechten van personen met een handicap' van de Verenigde Naties.

Mensen met een verstandelijke beperking kunnen het meestal ook *beter* zelf vertellen dan anderen. Daar is al veel onderzoek naar gedaan. Vooral als het gaat om gevoelens en gedachten. Niemand anders kan in jouw hoofd kijken.

Maar dan moet je het dus op een goede manier vragen. Op een manier die zoveel mogelijk mensen begrijpen. Ook mensen met een verstandelijke beperking. Daarom heet dit proefschrift "Why ask them? It's about me!", oftewel "Waarom vraag je het aan hen? Het gaat over mij!".

De onderzoeken in dit proefschrift

Hoofdstuk 2: wat hebben andere onderzoekers al ontdekt?

In dit hoofdstuk heb ik opgeschreven wat andere onderzoekers al hebben uitgevonden om vragenlijsten begrijpelijk te maken voor mensen met een VB. Dat heb ik allemaal bij elkaar gezet als een hele lange lijst met tips. Die kunnen mensen gebruiken die vragenlijsten maken of aanpassen voor mensen met een verstandelijke beperking. Voorbeelden van deze tips:

- gebruik Taal voor Allemaal om de vragen makkelijker te maken.
- als je mensen laat kiezen tussen antwoorden: maak de keuzes niet te moeilijk en geef niet teveel keuzes. 3 is meestal genoeg.
- gebruik duidelijke plaatjes om de tekst duidelijker te maken. En test van tevoren of de plaatjes duidelijk genoeg zijn.
- laat mensen met een verstandelijke beperking meedenken bij het maken van vragenlijsten. Dan kun je checken of de vragen en plaatjes duidelijk genoeg zijn.

Er zijn ook onderwerpen waar niet genoeg onderzoek naar gedaan is. Bijvoorbeeld wat voor soort plaatjes helpen om tekst begrijpelijker te maken. En wat er gebeurt als iemand je helpt bij het invullen. Durf je dan nog wel eerlijk te zeggen wat je vindt? Deze vragen hebben we onderzocht in hoofdstuk 3 en 4.

Hoofdstuk 3: hoe kun je mensen het beste helpen als ze het niet helemaal zelf kunnen?

Als mensen hulp nodig hebben bij het beantwoorden van vragen, dan helpen begeleiders of familie soms. Maar is dat wel een goed idee?

We lieten jongeren een vragenlijst invullen over hoe fijn ze het vonden op de groep. Sommige jongeren deden dat met hun begeleider, anderen deden het helemaal zelf zonder hulp of met hulp van iemand die ze niet kenden. We onderzochten of dat iets uitmaakte voor de uitkomst van de vragenlijst. Wat bleek? Jongeren die het samen met hun begeleider deden, gaven veel hogere cijfers dan jongeren die het alleen of met een onbekende invulden. Jongeren vertelden dat ze het moeilijk vonden om eerlijk en kritisch te zijn als hun begeleider erbij zat.

Samen met jongeren keken we naar oplossingen. De belangrijkste oplossing: de vragen zo makkelijk maken dat zoveel mogelijk mensen het helemaal zelf kunnen. En als ze hulp nodig hebben, dan iemand laten helpen die ze niet goed kennen. Of die ze zelf uit konden kiezen omdat ze die persoon vertrouwden. Dan is de kans het grootst dat je een eerlijk antwoord geeft.

Hoofdstuk 4: plaatjes om een tekst duidelijk te maken, helpt dat?

Vaak geven mensen de tip om plaatjes bij een tekst te zetten. Dan kunnen mensen die niet zo goed kunnen lezen de tekst beter snappen. Maar is dat wel zo? Snappen mensen beter wat er staat als je een plaatje bij een tekst zet? We maakten een samenvatting van alle onderzoeken die dit eerder al hadden uitgezocht. De uitkomst: het helpt niet echt. Als de taal al makkelijk gemaakt is, dan maakt een plaatje niet zoveel uit. Mensen begrijpen het dan niet beter. Soms zijn plaatjes zelfs verwarrend. Bijvoorbeeld als er op het plaatje iets anders staat dan wat er in de tekst staat.

We vonden de uitkomsten van het onderzoek best verrassend. Want we denken dat plaatjes soms best kunnen helpen. Waarom komt dat dan niet uit het onderzoek? De meeste onderzoeken die dit hadden onderzocht waren niet zo goed uitgevoerd. De plaatjes waren vaak slecht bedacht. En ze werden voor het onderzoek niet getest samen met mensen met een verstandelijke beperking. Dan weet je eigenlijk niet of het niet helpt, óf dat het komt doordat de onderzoekers het niet goed hebben gedaan. Er moet dus beter onderzoek komen.

Een tip voor mensen die plaatjes willen gebruiken bij het maken van teksten: test van tevoren even of de plaatjes wel goed passen bij de tekst. Doe dat samen met mensen die de tekst later gaan lezen.

Hoofdstuk 5: vragenlijsten om stress te meten

Veel mensen met een verstandelijke beperking hebben vaak last van stress. Er zijn vragenlijsten om te meten hoeveel stress je hebt. Zodat je hulp kunt krijgen als je teveel stress hebt. We hebben alle stress vragenlijsten verzameld die we op internet konden vinden. We hebben gekeken welke vragenlijsten zelf ingevuld kunnen worden door mensen met een verstandelijke beperking. We vonden drie vragenlijsten die speciaal waren gemaakt voor mensen met een verstandelijke beperking. Die waren het meest geschikt. We hebben advies gegeven welke vragenlijsten onderzoekers en psychologen die met gestresste mensen werken het beste kunnen gebruiken.

Hoofdstuk 6: vragenlijsten makkelijker maken: wat levert dat op?

Je kunt dus vragenlijsten makkelijker maken, zodat zoveel mogelijk mensen met een verstandelijke beperking ze zelf kunnen invullen. In de hoofdstukken hiervoor verzamelden we allerlei tips om dat te doen.

In hoofdstuk 6 onderzochten we of die tips nou echt goed werken. Eerst onderzochten we of de ABAS-3 vragenlijst geschikt is voor mensen met een verstandelijke beperking. De ABAS-3 is een hele lange, ingewikkelde vragenlijst die vraagt wat je in het leven allemaal zelf kan.

We vulden een aantal vragen in samen met bewoners van St Anna. We vroegen aan de deelnemers of ze de vragen snapten. De meeste mensen snapten veel vragen niet goed. We vroegen ook aan de begeleiders of ze dezelfde vragen in wilden vullen over de persoon met een verstandelijke beperking. Er zat best veel verschil in de uitkomsten. De begeleiders dachten dat bewoners minder zelf konden dan wat de bewoners zelf dachten.

Toen vroegen we aan de deelnemers wat we konden doen om de vragen makkelijker te maken. We gebruikten hun tips om een makkelijkere vragenlijst te maken over hetzelfde onderwerp. Ook gebruikten we de tips uit hoofdstuk 2. Toen vulden dezelfde bewoners van St Anna hem nog een keer in. Veel meer mensen snapten de vragen nu wel! En de antwoorden van de bewoners lagen dichterbij de antwoorden van de begeleiders. Ze waren het dus meer met elkaar eens.

Als je de vragen makkelijker maakt helpt het dus echt! Meer mensen kunnen de vragen zelf invullen en je krijgt betere antwoorden.

Hoofdstuk 7: conclusie

In het laatste hoofdstuk staat een samenvatting van alle hoofdstukken daarvoor. Hieronder staan belangrijke uitkomsten.

Makkelijker maken van vragenlijsten

De meeste mensen met een verstandelijke beperking kunnen veel zelf. Veel mensen met een lichte of matige verstandelijke beperking kunnen vragenlijsten zelf of met hulp invullen. Ook als het over moeilijke onderwerpen gaat. Als je de vragen maar op de goede manier stelt. Het helpt dus als je de vragen makkelijker maakt. Daarvoor staan er in dit proefschrift veel tips.

Scan deze QR code voor een poster met tips om vragen makkelijker te maken.

Mensen die hulp nodig hebben om de vragen in te vullen

Sommige mensen met een verstandelijke beperking hebben hulp nodig om vragen in te vullen. Bijvoorbeeld omdat ze niet goed kunnen lezen. Of omdat ze extra uitleg nodig hebben hoe ze een antwoord kunnen kiezen. Als je mensen helpt om vragen in te vullen, dan kan het gebeuren dat ze geen eerlijk antwoord durven te geven. Dat gebeurt vooral als de persoon die geholpen wordt afhankelijk is van de persoon die hem of haar helpt.

Scan deze QR code voor tips hoe je ervoor kunt zorgen dat mensen eerlijke antwoorden geven als ze hulp nodig hebben.

Mensen zelf laten proberen

Het gebeurt nog heel vaak dat mensen met een beperking de vragen wel zelf in *kunnen* vullen, maar dat begeleiders denken van niet. En dan vragen ze het niet eens. De begeleiders vullen dan de vragen in voor iemand anders. Of ze vragen het aan familie. We hebben gezien dat dit vaak geen goede oplossing is. Dat moet veranderen!

Tip: vraag eerst aan mensen met een verstandelijke beperking of ze zelf of met hulp de vragen in willen vullen. Laat mensen die de vragenlijst zelf of met hulp in willen vullen het proberen. Als je samen een oefenvraag invult of samen de eerste twee vragen van een vragenlijst invult, dan kun je goed inschatten of iemand het zelf of met hulp kan. Of dat het helemaal niet lukt. Dat moet je dan wel eerlijk bespreken. En samen kijken wie de vragen dan het beste in kan vullen over de persoon.

Tips voor mensen die vragenlijsten maken

Als mensen vragenlijsten maken, moeten ze er rekening mee houden dat mensen met een verstandelijke beperking de vragen *ook* snappen. Het is beter om een vragenlijst voor iedereen te maken. Dat noemen we een inclusieve vragenlijst. Dan kun je de antwoorden van iedereen met elkaar vergelijken. Dus niet één vragenlijst voor mensen *zonder* een verstandelijke beperking, en één vragenlijst voor mensen *met* een verstandelijke beperking.

Mensen met een verstandelijke beperking kunnen heel goed helpen bij het maken van vragenlijsten. Samen kun je dan bedenken hoe zoveel mogelijk mensen met een verstandelijke beperking de vragen in kunnen vullen. Het is nog slimmer om allerlei verschillende mensen te laten helpen. Bijvoorbeeld mensen die een andere moedertaal hebben dan Nederlands. Of mensen die door een niet-aangeboren hersenletsel meer moeite hebben met het begrijpen van taal. Dan kun je samen testen of *iedereen* de vragenlijst begrijpt.

Co-authorship and publications

Chapter 2 was based on a co-authored article, published as: Kooijmans, R., Mercera, G., Langdon, P.E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clinical Psychology: Science and Practice, 29*(3), 250-271.

Authors Roel Kooijmans, Peter Langdon and Xavier Moonen were involved in the conceptualisation of the study's aims, methodology and protocols. Roel Kooijmans executed the search with an information specialist. Roel Kooijmans and Gabriëlle Mercera jointly performed all screening, selection and quality appraisal activities. Data extraction and processing were carried out by Roel Kooijmans with assistance from Gabriëlle Mercera. Roel Kooijmans wrote the manuscript. Drafts were reviewed and edited by all authors. All activities were supervised by Xavier Moonen and Peter Langdon.

Chapter 3 was based on a co-authored article, published as: Kooijmans, R., Langdon, P. E., & Moonen, X. (2022). Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people's views. *Methods in Psychology*, *7*, 100102.

All authors contributed to the study conceptualization and design. Data collection and analysis were performed by Roel Kooijmans. The manuscript was written by Roel Kooijmans and all authors commented on previous versions of the manuscript. All activities were supervised by Xavier Moonen and Peter Langdon.

Chapter 4 was based on a co-authored article, submitted for publication as: Kooijmans, R., Dalemans, R., Van der Spek, J., Langdon, P. E., & Moonen, X. (submitted). Does adding pictures to easy-to-read texts benefit comprehension for people with reading difficulties? A meta-analytic review.

All authors contributed to the study conceptualization and design. The first draft of the manuscript was written by Roel Kooijmans and Judith van der Spek. Roel Kooijmans and Ruth Dalemans wrote final versions of the Introduction. Roel Kooijmans and Janneke Staaks wrote the Methods section. Statistical analyses were conducted by Roel Kooijmans, Judith van der Spek and Peter Langdon. All authors commented on all versions of the manuscript. All activities were supervised by Xavier Moonen and Peter Langdon.

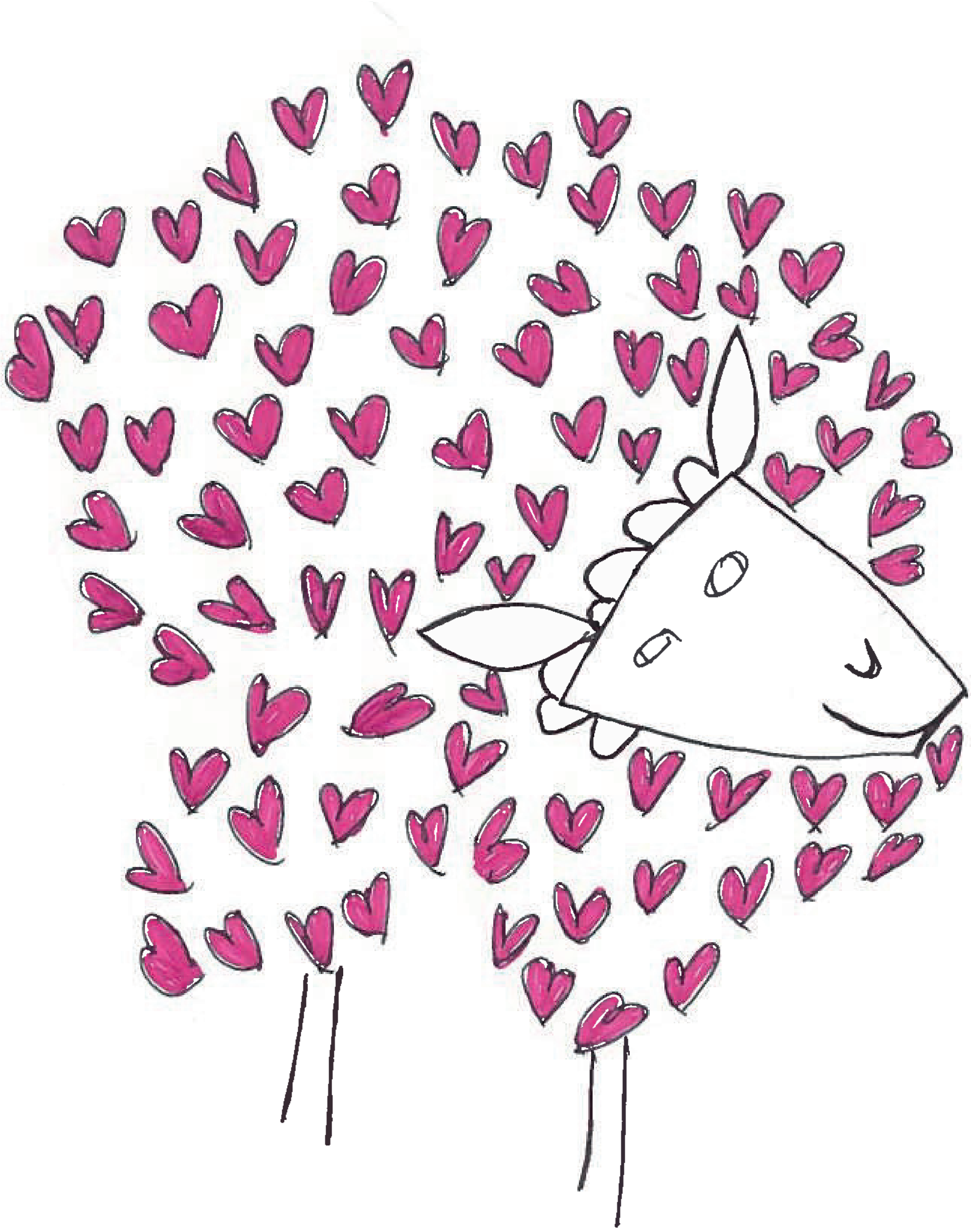
Chapter 5 was based on a co-authored article, published as: de Witte, M., Kooijmans, R., Hermanns, M., Hooren, S. V., Biesmans, K., Hermsen, M., Stams, G.J., & Moonen, X. (2021). Self-report stress measures to assess stress in adults with mild intellectual disabilities—A scoping review. Frontiers in psychology, 12, 742566.

Authors Martina de Witte, Roel Kooijmans and Maria Hermanns were involved in the conceptualisation of the study's aims, methodology and protocols. Maria Hermanns and Martina de Witte executed the search with an information specialist. Data extraction and processing were carried out by Maria Hermanns, Roel Kooijmans and Martina de Witte. Maria Hermanns wrote a first draft of the article. Martina de Witte and Roel Kooijmans shared co-authorship for subsequent versions of the manuscript. All authors commented on final versions of the manuscript. All activities were supervised by Xavier Moonen and Gert-Jan Stams.

Chapter 6 was based on a co-authored article, published as: Kooijmans, R., van Langen, M., Voss, H., Reichrath, E., Maljaars, J., Dalemans, R., Langdon, P.E. & Moonen, X. (2024). Does adapting a self-report instrument to improve its cognitive accessibility for people with intellectual disability result in a better measure?– A cognitive interview study. *Research in Developmental Disabilities, 154*, 104851.

Roel Kooijmans, Myrte van Langen and Hille Voss conceptualised the study. Ruth Dalemans, Enid Reichrath and Jarymke Maljaars advised on parts of the methodology and conceptualisation. Data collection was carried out by Roel Kooijmans with help from Hille Voss and Myrte van Langen. Roel Kooijmans and Myrte van Langen analysed qualitative data. Roel Kooijmans was responsible for all quantitative statistical analyses. Roel Kooijmans and Myrte van Langen wrote first drafts of the manuscript. All authors commented on final versions of the manuscript. All activities were supervised by Xavier Moonen and Peter Langdon.

Dankwoord



Je hoort vaak van mensen die promotie-onderzoek doen of gepromoveerd zijn dat het een *struggle* is, zo'n PhD. Vechten tegen deadlines, weekenden doorhalen, verantwoording afleggen aan subsidieverstrekkers en publicaties die keer op keer afgewezen worden. Partner bijna weggelopen, kinderen verwaarloosd, vervreemd van je hond. Wist je dat 47% van alle promovendi een vergroot risico heeft op mentale klachten en een derde ernstige symptomen van burn-out heeft[1]? Gelukkig herken ik me hier totaal niet in! Integendeel, ik vond het een prachtige reis en vind het jammer dat hij na 7 jaar bijna ten einde is. Waarom ik het zo anders heb ervaren dan minder gelukkige promovendi? De mensen in dit dankwoord zijn een belangrijke bijdrage aan het antwoord.

In de eerste plaats dank aan alle **kinderen**, **jongeren** en **bewoners** van Koraal. Zij zijn voor mij al 25 jaar de reden om iedere dag met plezier en inspiratie aan de slag te gaan. Zoals je in het voorwoord kon lezen waren de ervaringen van de jongeren van de cliëntenraad van Koraal de directe aanleiding om met dit promotie-onderzoek te beginnen. Ik geniet nog steeds elke dag van de pure en directe interacties met de jeugdigen. Vol humor en scherpe observaties zetten ze mij vaak op het verkeerde been als ik weer eens dacht dat ik wel wist hoe het zat.

Als psycholoog werken met jeugdigen en volwassenen met een verstandelijke beperking is één, met en over hen onderzoek doen is iets heel anders! Dankjewel **Xavier** om mij te laten zien hoe mooi het is om onderzoek met en voor jongeren te doen en hoe krachtig het is om hen zelf, op hun eigen manier, te laten vertellen wat telt. Jij hebt mij doordrongen van de waarde van kwalitatief en participatief onderzoek - mét een stevige wetenschappelijke onderlegger!

Ik vind het een voorrecht dat je acht jaar geleden bedacht dat ik het wel in me had om een promotie-onderzoek te starten en dat je mij hierbij wilde begeleiden. Bij de keuze van het onderwerp mocht ik heel dicht bij mijn eigen motivatie en ideeën blijven. Ook in de begeleiding in de loop van mijn PhD heb ik altijd heel veel vrijheid ervaren om de koers (mee) te bepalen. Eerder coachend en begeleidend dan sturend en dwingend en dat voelde als een grote blijk van vertrouwen.

Wat er ook aan heeft bijgedragen dat ik niet bij 'die 47%' ben gaan horen, is dat ik me nooit opgejaagd heb gevoeld. Ik heb nooit te horen gekregen dat ik maar eens op moest schieten, dat het onderhand tijd was voor de volgende publicatie en dat het na zoveel jaar wel welletjes was. Ik heb bij de allereerste afspraak benoemd dat ik niet van plan was om weekenden en avonden met mijn vrouw en kinderen op te offeren voor mijn onderzoek en dat heb jij altijd gerespecteerd.

---

[1] https://hetpnn.nl/actueel/bijna-helft-promovendi-heeft-vergroot-risico-op-mentale-klachten/

Peter, you completed my PhD supervision team and, together with Xavier, formed the ideal duo to guide me. Although you were physically at a distance—and I often left you in the dark for months about what I was actually doing—you were always incredibly quick to respond whenever I asked for your input or feedback. I don't know how you always managed to reply so promptly; I've genuinely wondered at times whether you ever sleep! As a true English (well, Canadian) gentleman you always started and ended your reply with encouraging words and kind compliments, which I have much appreciated.

Your knowledge of methodology and statistics has been of immense value to me. You never let me get away with vague reasoning or general conclusions. Everything had to be thoroughly substantiated and statistically sound. This made each of our publications several notches better in terms of quality. It spared us many difficult reviewer questions and ensured that most of the articles sailed smoothly through the review process.

What impressed me most was your respectful attitude towards the people who participated in our research. You made me deeply aware of the fact that the way we write about people with intellectual disabilities reflects how we treat one another as human beings. Thank you for patiently correcting me, time and again.

Hoe tof was het **Jessica** dat wij in 2018 samen in het PhD schuitje mochten stappen! Beiden waren we een totale leek op wetenschappelijk terrein en samen konden we op ontdekking. Zo fijn om samen onwetend te zijn tijdens onze eerste tripjes naar Leuven en ons eerste echte congres in Glasgow. Niet alleen als collega, maar ook als reis- en stapmaatje heb ik genoten van jouw gezelschap en bij alle leuke dingen die we samen naast het werk hebben gedaan (en nog doen!). Op naar jouw laatste PhD loodjes!

Het illustere O & I duo Roel en Jessica werd al snel een trio toen jij **Gabriëlle** met ons mee mocht komen pionieren bij De Hondsberg. Met z'n drieën hebben we de onderzoekstak van Koraal Jeugd opgezet en stonden we aan de wieg van Koraal Kennis (nu S&KO). We gingen samen op avontuur naar Glasgow, Berlijn en Zagreb. We vertellen samen onze praatjes op allerlei congressen in en buiten Nederland. Ook waren we elkaars co-auteur in sleutelpublicaties van ons beider PhD. Naast een slimme collega ben je vooral een heel fijn maatje. We vinden elkaar in onze waardering voor muziek, lekkere biertjes, slechte cappuccino, pubquizzen en nog veel meer. Ik kan me dan ook geen betere paranimf voorstellen!

Ons onderzoeksclubje werd de afgelopen jaren alsmaar groter. Er valt ook nog zoveel te onderzoeken in de jeugdhulp! Daarom werd ons trio al snel verdubbeld door de komst van (kleine) **Loïs**, **Rianne** en als kers op de taart (grote) **Lois**! We werken keihard om de kwaliteit van de jeugdhulp bij Koraal zichtbaar te maken en te vergroten. En omdat werk ook maar werk is hebben we ook nog de Eetclub samen, hoe leuk!

Over leuke mensen gesproken: wat een mooie club enthousiaste, creatieve en gezellige collega's heb ik bij **S&KO**! Echt een broedplaats van innovatieve ideeën en projecten die de zorg aan de jeugdigen en bewoners van Koraal iedere dag beter maken. Jullie dragen samen met 'ons onderzoekers' het belang van onderzoek doen uit en maken door de fijne samenwerking ons werk zo veel makkelijker en leuker. Speciale dank aan **Jeroen Hoenderkamp**, de 'founding father' van S&KO, **Ilona** en **Edith** die als 'bazen' zoveel van ons werk mogelijk maken en altijd vierkant achter ons staan, **Lara** als lieve moeder van S&KO en creatieve duizendpoot, **Erica** en **Arno** als onderzoeksbloedbroeders, **Enid, Karin** en **Tessie** die bewaken dat ik niet te veel moeilijke woorden gebruik en alle andere lieve leuke collega's om hard mee te werken en hard mee te lachen tijdens teamuitjes. Jij ook bedankt, **Ingrid** die dan wel geen S&KO is, maar wel de verbindende factor tussen strategie, onderzoek en primair proces en zoveel heeft gedaan om TIC tussen de oren en in de harten van de mensen van Koraal te krijgen.

Ik prijs mezelf altijd gelukkig dat ik mezelf 'scientist-practitioner' mag noemen. Waar ik bij S&KO vooral mijn denkende *hoofd* aan mag zetten, werk ik als psycholoog met heel mijn *hart* met de kinderen van **De Hondsberg**. Samen met mijn überslimme en megabetrokken superspecialistische collega's 'van de eerste' en natuurlijk ook 'van de tweede'. Teveel mensen om hier op te noemen, maar allemaal toppers!

Een van de tofste dingen van PhD onderzoek is dat je ook buiten de grenzen van je eigen organisatie mag samenwerken. Wat een bevlogen mensen kom je zo tegen! En soms leiden die ontmoetingen tot mooie projecten en publicaties, waaronder alle publicaties in dit proefschrift. Dank aan alle slimme mensen die hieraan hebben bijgedragen. Dankjewel **Enid Reichrath**, **Jarymke Maljaars, Janneke Staaks, Myrte van Langen** en **Hille Voss** voor jullie denkkracht en scherpe pen**.** Dank **Ruth Dalemans** met wie ik niet alleen mocht schrijven maar ook presenteren en die ik een allround inspirerende onderzoeker vind. Een bijzondere samenwerking had ik met **Martina de Witte**, met wie ik het hoofdstuk over stress schreef. Wat een brok energie en daadkracht! Dankzij haar vloog die publicatie in no time uit onze pen. Wat ontzettend verdrietig dat Martina er niet meer is.

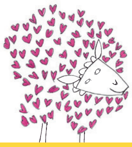
En toen alles was onderzocht en opgeschreven lag daar mijn proefschrift. Ik denk dat de leden van de promotiecommissie kunnen beamen dat compact schrijven niet tot mijn kernkwaliteiten behoort. Dank jullie wel **Paula Sterkenburg, Chris Kuiper, Hanneke Creemers, Noud Frielink** en **Maroesjka van Nieuwenhuijzen** dat jullie hebben toegezegd om het hele boekwerk door te akkeren en van kritisch commentaar te voorzien! Ik zie uit naar de verdediging.

Dit boekje was alleen maar een hoop saaie tekst geweest zonder **Kathleen** en de lieve mensen van **Uniek St. Anna** en de **Piahoeve**. Zij maakten de mooie tekeningen die bij ieder hoofdstuk staan.

Dankjewel **pap** en **mam** voor een warm nest en een opvoeding waarin een onderzoekende houding belangrijk was. Nieuwsgierig zijn werd gewaardeerd en gestimuleerd. Uitblinken in je studie werd niet geëist maar toch wel een beetje verwacht, want zonde om niet alles eruit te halen wat erin zit toch? Ik vond het heel fijn dat jullie vanaf het begin van mijn PhD altijd geïnteresseerd waren en zelfs mijn publicaties lazen, wat niet mee zal zijn gevallen als relatieve buitenstaander…

Arjan, dankjewel dat je samen met G mijn paranimf wilde zijn, al had je geen idee wat dat was. Je bent mijn kleine broertje, maar vooral mijn grote maat. Ik ben heel trots op jouw creatieve talenten en hoe je het doet als vader voor je meiden.

En dan als laatste **Noortje**, **Ciske, Polly** en **Pippa**, de belangrijkste mensen in mijn leven**.** Ik heb altijd geprobeerd om ervoor te zorgen dat jullie vooral geen *last* zouden hebben van mijn ambities. Onderzoekers die zich in de weekenden op zolder opsluiten om hun deadlines te halen: ik snap er niks van. Hoezo zou je achter de laptop gaan zitten als je tijd door kunt brengen met de liefste, knapste en grappigste mensen? Wij zijn het leukste gezin ter wereld heb ik mensen horen zeggen en wie ben ik om dat te ontkennen? Iedere dag laden jullie mijn batterij weer op zodat ik de energie heb om de dingen te kunnen doen die ik doe, thuis en op het werk. Dikke kus!



Koraal
wonen
leren
werken