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

Roel Kooijmans [a,b,*,1], Myrte van Langen [c], Hille Voss [c], Enid Reichrath [b], Jarymke Maljaars [d], Ruth Dalemans [e], Peter E. Langdon [f], Xavier Moonen [a,b,c]

[a] *University of Amsterdam, Research Institute of Child Development and Education, P.O. Box 19268, Nieuwe Achtergracht 127, Amsterdam 1018 WS, the Netherlands*
[b] *Koraal Center of Expertise, P.O. Box 10102, Oisterwijk 5060 GA, the Netherlands*
[c] *Ben Sajet Center, Zwanenburgwal 206, Amsterdam 1011 JH, the Netherlands*
[d] *KU Leuven, Parenting and Special Education Research Unit, Leuven B-3000, Belgium*
[e] *Zuyd University of Applied Sciences, Faculty of Healthcare, PO Box 550, Heerlen 6400 AN, the Netherlands*
[f] *Centre for Research in Intellectual and Developmental Disabilities (CIDD), University of Warwick, Coventry CV4 8UW, United Kingdom*

ARTICLE INFO

ABSTRACT

Background and aims: 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.
Methods and procedures: 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.
Outcomes and results: 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.
Conclusions and implications: 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.

1. Introduction

The outcome of (psychological) assessments and diagnostic procedures can have broad implications for the person who is assessed.

---

* Corresponding author at: University of Amsterdam, Research Institute of Child Development and Education, P.O. Box 19268, Nieuwe Achtergracht 127, Amsterdam 1018 WS, the Netherlands.
E-mail address: rkooijmans@koraal.nl (R. Kooijmans).
[1] ORCID ID: https://orcid.org/0000-0003-1172-0305

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.

1.1. 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).

1.2. 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 & Bradburn, 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 c*ognitive 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 respondents. 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).

1.3. 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.

2. Methods

2.1. 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).

2.2. 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.

2.3. Participants

2.3.1. 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 2

2.3.2. 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.

2.3.3. 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.

2.4. Instruments and measures

2.4.1. 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 (Common European Framework of Reference for Languages, 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, $\alpha = .98$, and informant, $\alpha = .0.99$, 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).

2.4.2. 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.

2.4.3. 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, Reichrath, 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.

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.

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.
	'Language for all' guidelines (Moonen, Reichrath, 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).

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.

2.5. Procedure and data collection

2.5.1. 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.

2.5.2. 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.

2.5.3. Proxy-reported information

Carers completed the original and adapted 30-item versions of the ABAS-3 informant report without assistance.

2.6. Data analysis

2.6.1. 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.

2.6.2. Quantitative analyses

2.6.2.1. 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.

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.

2.6.2.2. 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.

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).

2.6.2.3. 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.

3. Results

3.1. Round 1 cognitive interviews

3.1.1. 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.

3.1.2. Response scale

The response scale proved to be confusing. In the original scale, the respondent is asked to give a 0–3 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 1–3 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 a*lways* refrain from playing computer games because I *never* play them.". These question types received the highest difficulty ratings and were often misinterpreted.

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

3.1.3. 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.

3.1.4. 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.

3.2. 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

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.
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."

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).

3.3. Quantitative analyses

3.3.1. 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).

3.3.2. 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.

3.3.3. 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. Fig. 1

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

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
Item 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 Mean (SD)	Carer Mean (SD)	Difference *t*(17)	*P*	Cohen's d	Self-report Mean (SD)	Carer Mean (SD)	Difference *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 α set at.006 (=.05 / 8) to control for family-wise error

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. Table 7

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).

3.3.4. 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, $p <.001$.

4. 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.

4.1. 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, Reichrath, 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
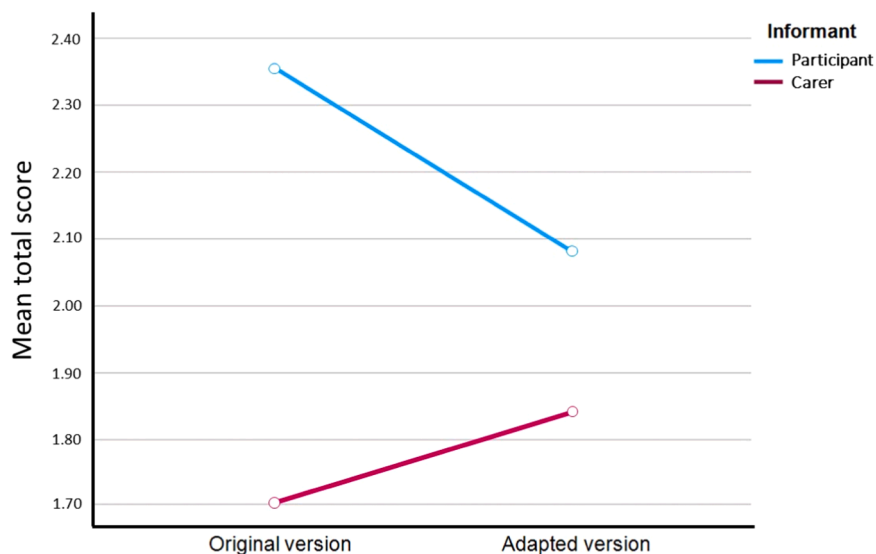

Fig. 1. Profile Plot Interaction Version vs Informant.

Table 7
Correlation matrix: Spearman rank correlations between ABAS-3 scores.

Variable	1	2	3	4
Mean total ABAS−3 score self-report - original	-			
Mean total ABAS−3 score informant report – original	.64**	-		
Mean total ABAS−3 score self-report - adapted	.90***	.60*	-	
Mean total ABAS−3 score informant report – adapted	.61*	.62**	.61**	-

* $p < .05$ ** $p < .01$ *** $p < .001$

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.

4.2. 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).

4.3. 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.

4.4. 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.

5. 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.

CRediT authorship contribution statement

Myrte Van Langen: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Hille Voss:** Writing – review & editing, Conceptualization. **Enid Reichrath:** Writing – review & editing, Methodology, Conceptualization. **Roel Kooijmans:** Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jarymke Maljaars:** Writing – review & editing, Methodology. **Ruth Dalemans:** Writing – review & editing, Methodology, Conceptualization. **Peter Langdon:** Writing – review & editing, Supervision. **Xavier Moonen:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

The authors have no relevant financial or non-financial interests to disclose.

Data availability

Data will be made available on request.

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.

Appendix B. Example of a question from the adapted ABAS-3

Before I make an important decision, I think it through calmly*

* formulation of original item: 'I make important decisions only after careful consideration and without rushing'

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.

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⟩.

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 In. K. Miller, S. Willson, V. Chepp, & J. Padilla (Eds.), *Cognitive Interviewing Methodology* (pp. 35–50). John Wiley & Sons.

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

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 Disability, 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

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 Disability, 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 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