Validity of an instrument that assesses functional abilities in people with profound intellectual and multiple disabilities

Wessels, Marleen D.; Paap, Muirne C. S.; van der Putten, Annette A. J.

Published in:
Journal of intellectual & developmental disability

DOI:
10.3109/13668250.2020.1785851

Publication date:
2021

Journal of Intellectual & Developmental Disability

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjid20

Validity of an instrument that assesses functional abilities in people with profound intellectual and multiple disabilities: Look what I can do!

Marleen D. Wessels , Muirne C. S. Paap & Annette A. J. Van der Putten

To cite this article: Marleen D. Wessels , Muirne C. S. Paap & Annette A. J. Van der Putten (2020): Validity of an instrument that assesses functional abilities in people with profound intellectual and multiple disabilities: Look what I can do!, Journal of Intellectual & Developmental Disability, DOI: 10.3109/13668250.2020.1785851

To link to this article: https://doi.org/10.3109/13668250.2020.1785851

Published online: 10 Aug 2020.

JOURNAL OF INTELLECTUAL & DEVELOPMENTAL DISABILITY
https://doi.org/10.3109/13668250.2020.1785851

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Validity of an instrument that assesses functional abilities in people with profound intellectual and multiple disabilities: Look what I can do!

Marleen D. Wessels [a], Muirne C. S. Paap [a,b] and Annette A. J. Van der Putten[a]

[a]Department of Inclusive and Special Needs Education, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands; [b]Department of Child and Family Welfare, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

ABSTRACT

Background: Research about the psychometric properties of the Behavioural Appraisal Scales (BAS) in people with profound intellectual and multiple disabilities (PIMD) is limited. This study evaluates invariance in factor structure, item bias and convergent validity of the BAS.

Methods: Data on the BAS from two studies ($n = 25$; $n = 52$) were analysed using the oblique multiple group method. The scale structure and item ordering were compared in the two groups. Convergent validity was assessed by correlating scores on the BAS with scores on two other instruments.

Results: Of all items, 16–18% correlated stronger with other subscales of the BAS than the subscale they were originally assigned to. Scale structure and order of difficulty differed between groups. Correlations between the BAS and two other instruments varied from low to excellent ($r = .48–.85$).

Conclusions: The results support the construct validity of the BAS. Removing, reassigning and adapting items may enhance construct validity.

KEYWORDS

People with profound intellectual and multiple disabilities; functional abilities; assessment; validity; psychometric properties; Behavioural Appraisal Scales

People with profound intellectual and multiple disabilities (PIMD) have profound intellectual disabilities combined with severe or profound motor disabilities (Nakken & Vlaskamp, 2007). Moreover, they often have several additional impairments (e.g., visual and auditory) and medical problems (Van Timmeren et al., 2016). Because of the complex disabilities, people with PIMD are dependent on others in almost all aspects of daily life (Nakken & Vlaskamp, 2007). They therefore need relationships with others (e.g., with parents, teachers or direct support professionals) in order to develop skills, express their needs and manage their lives (Vlaskamp et al., 2015). Support professionals cannot tune their support to the needs and abilities of people with PIMD without knowing what those needs and abilities are. Assessment is therefore crucial. Defined as the process of gathering information about the individual and their situations in order to understand someone better (Kendall & Norton-Ford, 1982), assessment can provide highly important information. The assessment of people with PIMD, however, is complicated by several factors (Vlaskamp, 2005b). First, people with PIMD have limited, if any, ability to use spoken language (Bellamy et al., 2010). They communicate primarily through sounds, gestures, facial expressions, or physiological signals. Assessment is therefore often based on observation or information from proxies and thus dependent on context and interpretation (Lyons et al., 2016). Furthermore, assessments that ignore the complex, interrelated patterns of disabilities characteristic of this population are likely to provide invalid and unreliable estimates of functioning and developmental patterns (Van der Putten et al., 2017; Vlaskamp, 2005b). Many instruments rely on visual or motor abilities and thus do not generate valid, reliable representations of the construct they aim to measure in people with PIMD. Instruments that were developed to measure functional abilities but rely on visual or motor abilities are possibly measuring a different construct in people with PIMD. An example are items about looking at faces to make contact. People with PIMD may not endorse these items because of a visual impairment, even though they may be able to initiate contact in a different way. In addition, many instruments assume linear developmental patterns characteristic of people without disabilities, despite evidence that, in people with PIMD, development might be both delayed and follow a different pattern (Visser et al., 2017; Vlaskamp, 2005a). Many assessment

CONTACT Marleen D. Wessels m.d.wessels@rug.nl
Supplemental data for this article can be accessed at https://doi.org/10.3109/13668250.2020.1785851

instruments that were not developed specifically to assess characteristics or abilities of people with PIMD assume an order in difficulty and include certain rules for when to stop filling in an assessment, which are based on the assumption that the other items are more difficult and will not be endorsed by the person. This assumption is based on a linear developmental model, which may not hold in the group of people with PIMD. Standardised tests suitable for people with PIMD are scarce (Carnaby, 2007). The Behavioural Appraisal Scales (BAS) were developed specifically for people with PIMD. This instrument was developed to measure functional abilities, which are of crucial importance to the ability of people to manage their own lives (Vlaskamp et al., 1999). In practice, the BAS is often combined with other assessment instruments, in order to construct personal profiles of people with PIMD (including information about abilities, preferences and needs) and to adapt support accordingly (Vlaskamp et al., 1999; Wessels & Van der Putten, 2017). The BAS covers several important aspects involved in assessing people with PIMD. The BAS has no age limit, as it was developed for children, adolescents and adults. This way, the possibly different order of development and considerable amount of variety in development between persons is taken into account, as the assumption that specific abilities are mastered or not relevant anymore at a specific age is possibly not applicable in this group. Furthermore, the BAS does not rely on verbal communication abilities and it considers visual and motor impairments as items were formulated in such a way that they do not rely on motor or sensory abilities. Items are formulated in a broad way to allow alternative ways of completing a task, for example grabbing objects with elbows or feet if a person cannot use his or her hands. Moreover, instructions allow flexibility in the scoring and testing procedure. For example, there is no maximum duration of the time a person has to respond, which takes into account the possible longer reaction time of people with PIMD. There are instructions for all items and the material needed, but a tester can use preferred material of the person with PIMD. Moreover, the BAS does not assume a linear developmental model, as there is no order of difficulty and the complete instrument is filled in (Vlaskamp et al., 2002). Finally, it combines information from proxies, observation and test situations to enhance the validity of its outcomes. Although the BAS can provide information that can be useful in providing support to people with PIMD, several aspects of the instrument require further analysis. Only one study has addressed its psychometric quality (Vlaskamp et al., 2002). It may therefore be advisable to study validity based on a variety of methods. In this study, construct validity is evaluated by analysing the scale structure and possible invariance in the scale structure and item ordering based on two different groups. The applicability of the BAS in specific subgroups of people with PIMD has not been studied yet. Even though the BAS was specifically developed for the entire population of people with PIMD, taking into account their disabilities, this group is heterogeneous (Nakken & Vlaskamp, 2007). Further inspection of the items suggests that some items seem to rely on visual behaviour or motor abilities, while they were developed to measure another construct. An example is the item "a person shows what he or she wants by pointing at it" to measure communication. As it is uncertain whether the BAS is applicable in specific subgroups, it is of paramount importance to examine possible invariance in factor structure and item bias across subgroups. In addition, in this study, we assess convergent validity by comparing scores on the BAS to scores on other instruments intended to measure related constructs.

Methods

Participants

People meeting the following criteria were identified as having PIMD (Nakken & Vlaskamp, 2007):

- Severe to profound intellectual disability: Adults (>18 years) were included if their developmental age was 36 months or lower. Children were included if their functioning was estimated at the level of half, or below half, of their chronological age (Vig & Sanders, 2007).
- Moderate to profound motor disability: Level III (moderate; walking with assistive technology, young children may creep, crawl or walk short distances indoors with assistive technology and/ or adult support), IV (severe; self-mobility with limitations, for example using adaptive technology for sitting or standing, or young children using adaptive seating for maintaining balance), or V (profound; no means of independent mobility and limited voluntary control of movement) on the Gross Motor Function Classification System (Palisano et al., 1997).

This study was based on secondary data analysis using datasets from two studies including children and adults with PIMD (see Table 1). The first study (Group 1) was a longitudinal study, called the OJKO-project, which aimed to study the development of young children with a significant cognitive and motor development delay (i.e., Van keer et al., 2019). In this study, 52 children participated. The second group consisted of 26

Table 1. Participant characteristics by group.

Characteristics		Group 1 ($n = 52$)	2 ($n = 26$)
Age	Mdn (sd)	3.1 (0.9)	34.2 (13.4)
Gender	Male (%)	25 (48.1)	12 (46.2)
	Female (%)	27 (51.9)	14 (53.8)
GMFCS level	III (%)	6 (11.5)	0 (0)
	IV (%)	12 (23.1)	2 (7.7)
	V (%)	34 (65.4)	24 (92.3)
Visual impairment	Blind (%)	3 (5.8)	15 (57.7)
	Severe (%)	16 (30.8)	11 (42.3)
	None (%)	24 (46.2)	0 (0)
	Unknown (%)	9 (17.3)	0 (0)
Seizure disorder	Yes (%)	30 (57.7)	17 (65.4)
	None (%)	21 (40.4)	9 (34.6)
	Unknown (%)	1 (1.9)	0 (0)

GMFCS = Gross Motor Function Classification System (Palisano et al., 1997).

adults and adolescents from three residential facilities, all of which provided support to people with visual acuity of less than 0.3 (a person has a vision of 30% or less compared to persons without visual impairments). This study was an intervention study, which looked into the effects of a motor activation program (Van Alphen et al., 2018). For both studies, the first measurement data were used.

Instruments

Behavioural Appraisal Scales

The BAS instrument is an adaptation of the Behaviour Assessment Battery (BAB) (Kiernan & Jones, 1982). The BAB was translated into Dutch and studied for feasibility (Vlaskamp et al., 2002). Content validity of the BAS was studied by adapting the items of the BAB, based on the feedback of professionals. Items of the BAB that relied on visual or motor abilities were adapted or removed. Moreover, instructions were adapted to make them applicable in the group of adults. These adaptations resulted in the BAS. The BAS encompasses a total of 100 items, distributed over five factors (i.e., subscales), which were defined in the study of Vlaskamp et al. (2002). The BAS consists of questions that are directed to a proxy, observation items in a daily situation and observation items in a test situation (in which the person with PIMD is stimulated to exhibit behaviours related to the subconstruct a subscale aims to measure), as well as a description of additional information. The subscales, a description of the subscales and example items with their instructions are provided in Table 2. Per subscale a description of suggested material is provided, such as "an object that makes sounds". However, as the BAS is a flexible test, the tester could choose the exact object, depending on the preferences of the person with PIMD.

Response categories are dichotomous (yes/no). The total scores on each subscale were used to calculate quartile scores, which were then used to establish a profile of strengths and weaknesses (Vlaskamp et al., 1999). In a study on the psychometric quality of the BAS in 96 persons with PIMD, Vlaskamp and colleagues (2002) reported high reliability, in terms of both Cronbach's alpha values (.96–.98) and interrater reliability coefficients. For four of the five subscales, interrater reliability coefficients of raw scores were high (.93–.96), although this was much lower for emotional communicative behaviour (.47). Although, to the best of our knowledge, there are no instruments that assess a similar construct as the BAS in people with PIMD, Vlaskamp and colleagues (2002) performed an exploratory study by correlating scores on the BAS with scores on a subsection of another instrument, the Inventory of the Personal Profile (IPP) to explore convergent validity. The subsection of the IPP that was used measures the ability of a support person to signal, interpret and respond to the behaviour of a person with PIMD. It was expected that this subsection would be positively related with the functional abilities of persons with PIMD, because people with PIMD can only develop their skills in relation with significant others, for which it is crucial that support persons can adequately interpret the meaning of the behaviour of the person with PIMD. The expectation was that if a person has more functional abilities, interpreting and responding to the behaviour of a person

Table 2. Subscales of the Behavioural Appraisal Scales.

Subscale (abbreviation)	Description (# items)	Example item	Instruction
Emotional communicative behaviour (Em)	Skills in communicating emotions or feelings (3)	The person makes sounds to indicate whether they are happy or sad.	Observation or ask a support person.
Receptive language behaviour (Rec)	Ability to respond to communications by another person (9)	The person understands simple words that relate to their needs.	Observation or ask a support person.
General communicative behaviour (Com)	Ability to express self in a social environment (15)	The person can say yes and/or no.	Observation or ask a support person.
Visual behaviour (Vis)	Reactions of the person to visual stimuli (36)	The person can follow an object, which moves in a horizontal direction, with their eyes.	An object is slowly moved in a horizontal direction.
Exploratory behaviour (Ex)	Exploring the environment (37)	The person grabs an object.	An object is placed in front of a person and the attention of the person is drawn to it by the tester.

becomes less complex. However, in the study of Vlaskamp et al. (2002), all correlations were .29 or lower, with the exception of the emotional communicative behaviour subscale (.47).

Communication and Symbolic Behaviour Scales-Developmental Profile (CSBS-DP)

The Communication and Symbolic Behaviour Scales Developmental Profile (CSBS-DP) was developed to measure early communication and symbolic skills in infants and young children (Eadie et al., 2010). The CSBS-DP consists of a short checklist for screening, a caregiver questionnaire (CQ) and a Behaviour Sample. In the current study, convergent validity was assessed only according to the CSBS-DP Caregiver Questionnaire, which was developed for children whose scores on the short CSBS-DP screening checklist fall outside the ranges expected for their age and for children at-risk for social-communication problems (Eadie et al., 2010). The CSBS-DP Caregiver Questionnaire is completed by someone very familiar with the person being assessed (e.g., a parent). The CSBS-DP Caregiver Questionnaire consists of seven clusters, which are part of an overarching set of scales (see Table 3). In this study, items were scored according to the CSBS-DP manual (Wetherby & Prizant, 2002). Although the CSBS was not studied in the specific group of children with PIMD, good findings from psychometric analyses of the CSBS-DP have been reported in related groups of children (young children with and without a risk of a developmental delay) (Wetherby et al., 2002). In addition, the CSBS has previously been used to assess communicative behaviour in this specific target group (Dhondt et al., 2020) and was chosen after thoughtful consideration, based on the applicability in the target group and the informant who completes the questionnaire, as it is filled in by someone who knows the child with PIMD well.

Table 3. Overarching scales and clusters of the CSBS-DP Caregiver Questionnaire and the BAS subscales they were correlated with.

Overarching scale	Cluster (# items)	BAS subscale it was correlated with
Social	Emotion and eye gaze (8)	Emotional communicative behaviour
	Communication (10)	General communicative behaviour
	Gestures (2)	General communicative behaviour
Speech	Sounds (4)	General communicative behaviour
	Words (4)	General communicative behaviour
Symbolic	Understanding (4)	Receptive language behaviour
	Object use (9)	–

CSBS-DP = Communicative and Symbolic Behaviour Scales Developmental Profile.
BAS = Behavioural Appraisal Scales.

Motor Development List (MDL)

The Motor Development List (MDL) is an adaptation of the "motor development" subscale of the Portage Program Developmental Checklist (Hoekstra et al., 2011a; Hoekstra et al., 2011b). The MDL consists of one scale with 145 items focussing on gross and fine motor skills (range scores 0–290), which can be completed by a direct support professional or parent. The same three response categories are used for all items: the child has not acquired the skill (0); the child has acquired the skill partly, sometimes, or with help (1); or the child has acquired the skill (2). Results of a Rasch analysis based on 736 children without disabilities (0–5 years) indicate that all of the subscales of the Developmental Checklist constitute reasonably fitting scales (Hoekstra et al., 2010). The MDL has been used to assess motor abilities in young children with PIMD and preliminary analyses about the psychometric properties of the MDL are promising and have been presented at both national and international conferences (e.g., Colla et al., 2015). Furthermore, the instrument was chosen after thoughtful consideration, as it is completed by someone who knows the child well and it allows for fine-grained measurement: the items are ordered according to their difficulty level, and the increments in difficulty are relatively small. This allows for high discriminative power when it comes to differentiating score groups.

Procedure

Data were collected in two studies. In the first study (Group 1), hospitals, diagnostic centres, health care organisations and specialised day care centres were contacted by email, telephone, or mail (with a flyer about the study) and asked to bring the study to attention to potential participants. This study was approved by the Social and Societal Ethics Committee of the University of Leuven and the Ethics Committee of the University of Groningen, Pedagogical and Educational Sciences. Written informed consent was obtained from all parents. The second study (Group 2) examined the effects of a motor activity program. Participants were living in three different residential facilities offering 24-hour support. The intervention study was registered in The Netherlands Trial Register (NTR), number 6627. The study was approved by the Ethics Committee of the University of Groningen, Pedagogical and Educational Sciences. Informed consent was obtained from all legal representatives.

Analysis

Missing data

Participants with more than 10% missing values on one of the instruments were excluded from the analysis. For the BAS, one participant was excluded for this reason, for the other participants missingness was 1.1% (Group 1) and 0.4% (Group 2). We addressed missing data by applying two-way imputation for each subscale (Van Ginkel et al., 2007). For the MDL, missingness was 3.4%, after excluding one child (100% missingness). For the CSBS-DP, missingness was 0.06%, after excluding one child (100% missingness).

Item analysis

The mean score, standard deviation and minimum and maximum scores were computed for each subscale. It was expected that participants of Group 2 would score lower on the subscale visual behaviour, as this scale specifically measures visual abilities. Mean scores on the exploratory behaviour subscale was expected to be somewhat lower in Group 2, as this was expected to be related to motor abilities, which were more impaired in Group 2. Scores on the other subscales were expected to be similar, as these items were developed to not rely on other abilities than the construct the items aim to measure. As a measure of reliability, Cronbach's alpha was computed for the separate subscales (Cronbach, 1951). Alpha values of less than .70 are considered inadequate, with values of .70–.80 being adequate, .80–.90 being good and .90 or higher being excellent (Evers et al., 2013). Item analysis was performed using tools from classical test theory (Nunnally & Bernstein, 1978). The difficulty of each item was assessed according to the proportion correct (Nunnally & Bernstein, 1978, p. 262). In this study, items with proportions higher than 0.90 or less than 0.10 were defined as having low discrimination.

BAS factors

The oblique multiple group (OMG) method was used to evaluate the dimensional structure of the BAS for both groups (Stuive et al., 2008). The OMG-method is a form of confirmatory factor analysis based on the correlation of items with the subscales to which they are assigned, as well as with other subscales. We expected the highest correlation of a given item to be with the subscale to which it is assigned. Therefore, items that correlated higher with another subscale than the one they were assigned to, were defined as incorrectly assigned and suggest contradicting evidence for the factor structure. Moreover, correlations between subscales were computed as well. We expected moderate to high correlations between subscales, with low correlations possibly indicating that subscales do not share an underlying dimension and high correlations indicating that subscales have a great deal in common and share an underlying dimension (Stuive, 2007).

Item ordering

The proportion correct analysis was studied to analyse whether the BAS yields a score that can be interpreted similarly in different subgroups. If there are items that are relatively easier or more difficult in a subgroup, this can indicate that the item has a different meaning for that subgroup. In each subscale of the BAS, the items were ordered from high to low based on the proportion correct score. With proportion correct score the item popularity is meant, which is the proportion of persons who score a "1". The easiest item was the item with the highest proportion correct score and was labelled with rank 1. The ranks, based on proportion correct scores, were compared across Groups 1 and 2.

Convergent validity

Convergent validity was studied by correlating the BAS subscale scores to those of other instruments measuring similar – but not identical – constructs. The subscales emotional communicative behaviour, receptive language behaviour and general communicative behaviour were correlated with clusters of the CSBS. The clusters of the CSBS and the subscales of the BAS they were correlated with are shown in Table 3. As the CSBS-DP aims to measure a similar construct as the communication subscales of the BAS, the expectation was that there would be a moderate to high correlation between the communication subscales of the BAS and the CSBS-DP.

The subscale exploratory behaviour was correlated with MDL. As being able to show exploratory behaviour partly relies on motor abilities, such as being able to move the arms, use the hands and having the ability to hold an object, a positive correlation was expected between motor abilities and exploratory behaviour (i.e Adolph & Franchak, 2017). These analyses were performed only for Group 1, as data on the other instruments were not available for the other group. Following the criteria of the European Federation of Psychologists' Associations, correlation coefficients of 0–.55 are considered inadequate; with correlations of .55–.65 considered adequate, .65–.75 considered good and >.75 considered excellent (Evers et al., 2013).

Software

All analyses were performed in the software package R, version 3.5.2 (R Core Team, 2013), using the following packages: haven (Wickham & Miller, 2018), psych (Revelle, 2018), readxl (Wickham & Bryan, 2018), TestDataImputation, (Dai et al., 2016), dplyr, (Wickham et al., 2018) and tidyverse (Wickham, 2017).

Results

Item analysis

An overview of the mean scores, standard deviations, range of the subscales and Cronbach's alpha values for both groups is presented in Table 4. The alpha values are .83 or higher for all subscales in both studies, except for the emotional communicative behaviour subscale in Group 2. The mean scores on the following subscales are comparable for both groups: emotional communicative behaviour, receptive language behaviour, general communicative behaviour and exploratory behaviour. For visual behaviour, scores for Group 1 were higher than those for Group 2.

BAS factors

Correlations among the subscales are shown in Table 5. In general, these correlations are of low to medium size. These results can be seen as evidence that the subscales measure distinct aspects of functioning. The correlations for the subscale pair, receptive language behaviour and general communicative behaviour, are moderate. In Group 1, scores on the exploratory behaviour subscale were moderately correlated with scores on the general communication and visual behaviour subscales. For Group 2, the exploratory behaviour subscale was moderately correlated with the receptive language behaviour subscale.

The items that are more strongly correlated with subscales other than with their own are displayed in Table 6. In Group 1, 16 items (16%) are more strongly correlated with other subscales. In Group 2, 18 items (18%) are more strongly correlated with other subscales. When comparing the 16 identified items in Group 1 to the 18 identified in Group 2, 7 items overlap. The items that were more strongly correlated with other subscales are approximately equally distributed among the subscales in both groups, except for the emotional communicative behaviour subscale. The correlations of all items with both the subscales to which they are assigned and the other subscales are presented in Appendices A1.1–A2.5 of the Online Supplement.

Item ordering

The proportions of correct scores on each subscale for both groups are presented in Appendices B 1.1–1.5 of the Online Supplement. For the emotional behaviour subscale, the proportion correct is very high for all items in both groups. The proportion correct for the items Ex23 ("the person deliberately rolls an object from a slope"); Com3 ("the person indicates what he or she wants by pointing at it"); Com7 ("the person names simple images using words or gestures") and

Table 4. Mean, standard deviation and minimum and maximum scores for each subscale, by group.

	Group							
	1 ($n = 52$)				2 ($n = 25$)			
Subscale (# items)	Mean	SD	Range	Cronbach's alpha	Mean	SD	Range	Cronbach's alpha
Emotional communicative behaviour (3)	2.85	0.57	0–3	.84	2.88	0.33	2–3	−.14
Receptive language behaviour (9)	2.87	2.43	0–9	.83	4.08	2.97	0–9	.87
General communication behaviour (15)	2.94	3.27	0–14	.88	3.32	3.31	0–11	.86
Exploratory behaviour (37)	14.52	10.87	0–35	.96	12.32	8.88	0–30	.94
Visual behaviour (36)	21.21	10.23	0–35	.96	9.20	8.80	0–30	.95

Table 5. Correlation between subscales of the BAS for Group 1 (above the diagonal) and Group 2 (below the diagonal).

	Subscale				
Subscale	Emotional communicative behaviour	Receptive language behaviour	General communicative behaviour	Exploratory behaviour	Visual behaviour
Emotional communicative behaviour	1	.28*	.15	.26	.32*
Receptive language behaviour	.14	1	.63*	.48*	.31*
General communication	−.08	.59*	1	.64*	.32*
Exploratory behaviour	.17	.55*	.48*	1	.64*
Visual behaviour	.11	.32	.13	.13	1

*$p<.05$.

Table 6. Items that are more strongly correlated with subscales other than their own, by group.

Group 1 ($n = 52$)	Item	Subscale with which the item is more strongly correlated than it is with the original subscale
	Rec1∼	Emotional communicative behaviour
	Rec2, Rec9, Ex29∼, Ex30, Ex36∼, Vis34	General communicative behaviour
	Com2∼*, Com3∼, Com7∼, Ex25	Receptive language behaviour
	–	Visual behaviour
	Com1∼, Com2∼*, Com15, Vis22, Vis27, Vis28	Exploratory behaviour
Group 2 ($n = 25$)	**Item**	**Subscale with which the item is more strongly correlated than it is with the original subscale**
	Com2∼, Ex22	Emotional communicative behaviour
	Rec8, Ex29∼, Vis24	General communicative behaviour
	Em3, Com1∼, Com4, Com7∼, Ex36∼	Receptive language behaviour
	Em1, Com3∼, Ex23, Ex26	Visual behaviour
	Em2, Rec1∼, Com8, Vis35	Exploratory behaviour

Note: Em: Emotional communicative behaviour; Rec: Receptive language behaviour; Com: General communicative behaviour; Ex: Exploratory behaviour; Vis: Visual behaviour. *Item Com2 is equally strongly correlated with the receptive language behaviour subscale and the exploratory behaviour subscale in Group 1; ∼: Item is more strongly correlated with a subscale other than the original in both groups.

Com12 ("the person imitates new combinations of syllables") are very low in both groups.

The order of difficulty is similar for the emotional communicative behaviour subscale and the receptive language behaviour subscales (except for one item) in both groups, but the order differs between the groups for the other subscales, with some items ranking lower in Group 1 and others ranking lower in Group 2. Items with a ranking difference of 4 or more for the general communication behaviour scale or 10 or more for the exploratory behaviour and visual behaviour scales (taking into account the difference in the length of the scales) are presented in Table 7. Lower rankings indicate lower proportions of participants scoring positively on the items.

Convergent validity

For Group 1, the correlation between the emotional communicative behaviour subscale and the CSBS-DP emotions and eye gaze is low (.48). The correlation between the general communicative behaviour subscale and the CSBS-DP cluster gestures was adequate (.56), with the cluster sounds correlation was excellent (.81), with the cluster words correlation was excellent (.78) and correlation with the cluster communication was excellent (.85). The correlation between the receptive language behaviour subscale and the CSBS-DP cluster understanding was excellent (.76). The correlation between the BAS exploratory behaviour subscale and the score on the MDL was excellent (.82).

Discussion and conclusion

The current study evaluated the psychometric properties of the BAS using several approaches. Item analysis indicates that the instrument's reliability (Cronbach's alpha) is good for all subscales except for emotional communicative behaviour. In addition, the items "the person deliberately rolls an object on a slope," "the person indicates what he or she wants by pointing at it", "the person names simple images using words or gestures," and "the person imitates new combinations of syllables" are apparently too difficult for participants with PIMD, while all three items on the emotional communicative behaviour scale are apparently too easy. These items should be removed or adapted to make them more applicable to the population of people with PIMD. Our results support the factorial structure of the BAS, as most items correlated most strongly with the subscales to which they were assigned, with only a minority of items being more strongly correlated with other subscales. We found that the items that are more strongly correlated with subscales other than their own differ between the two groups. The order of difficulty is similar for the receptive language behaviour and emotional communicative behaviour subscales, but different for the visual behaviour, exploratory behaviour and general communicative behaviour subscales. Finally, convergent validity in our study was adequate to excellent for the general communicative behaviour subscales and excellent for receptive language behaviour and exploratory behaviour subscale, but inadequate for emotional communicative behaviour. Our results might have been influenced by several aspects. Although the items and instructions might have been interpreted and tested in slightly different ways by the different assessors in the two studies, the differences are likely limited because of the high interrater reliability of the BAS (Vlaskamp et al., 2002). Moreover, in this study, sample sizes were small. More advanced techniques, such as Item Response Theory, would allow for stronger conclusions. However, for different reasons small sample sizes are common in this field, for example because of the health problems of our target group, the high care load of support persons and the total size of the population, as the group of people with PIMD is a small group (Vugteveen et al., 2014). We selected our methods with a small sample size in mind and reflected on this, by carefully formulating our conclusions. This exploratory study was a first

Table 7. Items for which the order of difficulty differs between the two groups, with lower rankings indicating fewer participants scoring positively on the items.

Lowest rankings for Group 1 Item	Difference in ranking (maximum possible difference in ranking)	Description item
Com4	6 (14)	The person uses words or gestures to name objects or people on request.
Com5	6.5 (14)	The person uses words or gestures to name objects or people spontaneously.
Com6	4.5 (14)	The person can say "yes" or "no."
Vis34	30 (35)	The person can identify familiar people.
Ex12	16 (36)	The person grasps an object with one or two hands quickly and without difficulty.
Ex22	10 (36)	The person deliberately pushes an object on a surface.
Ex28	12.5 (36)	The person drops an object and is interested in the dropping of an object.
Ex34	10.5 (36)	The person removes a screen to grab an object or makes a clear effort to do so.
Ex36	18 (36)	The person uses an object or gesture to clarify their needs.
Lowest rankings for Group 2 Item	**Difference in ranking (maximum possible difference in ranking)**	**Description item**
Com15	10 (14)	The person makes faces in the mirror.
Vis8	10 (35)	The person looks at pictures in a book for at least 5 s.
Vis9	15 (35)	The person stays interested in a book for at least 1 min in total.
Vis13	10 (35)	The person's eyes follow an object in the largest part of a vertical movement.
Vis14	10 (35)	The person's eyes follow an object that is moved in a circle for the largest part of the circle.
Vis17	12 (35)	The person's eyes and head follow a moving object in the largest part of the vertical movement.
Ex3	14.5 (36)	The person seems visually interested in their own hand movements.
Ex8	10.5 (36)	The person transfers an object from one hand to the other hand at least 2 times.
Ex19	12 (36)	The person deliberately makes movements with an object in order to make different sounds.
Ex37	15 (36)	The person touches their reflection in the mirror or makes a clear effort to do so.

Note: Ranking was established by numbering the items from easy to difficult, based on the proportion of correct scores. Items displayed in the table have a ranking difference similar to or greater than 4 (Com scale) or 10 (Vis scale, Ex scale); Com: General communicative behaviour; Ex: Exploratory behaviour; Vis: Visual behaviour.

step in developing knowledge and supporting the improvement of assessment practices in an area where knowledge is scarce. The high interrater reliability for all subscales, except for emotional communicative behaviour, reported by Vlaskamp and colleagues (2002), was in accordance with our results of reliability. Our results considering the factor structure are based on the OMG-method. Stuive (2007) suggests that this relatively simple, descriptive technique, performs equally well or even better in some situations than a more complex technique such as common factor analysis. Although the factor structure identified by Vlaskamp and colleagues (2002) was not completely replicated in our study, the methods used to analyse the factorial structure were different from those applied by Vlaskamp and colleagues. The results are thus not directly comparable. The method that we used is arguably more suitable than the principal component analysis is for purposes of evaluating latent variables rather than for performing data reduction (see Borsboom, 2006). An important strength of this study is that it considers possible differences between subgroups by comparing two groups of people with PIMD, making it possible to identify differences in the assignment of items to subtests (Stuive, 2007). One possible explanation for the low correlation between the emotional communicative behaviour subscale and the CSBS-DP is that the length and range of the emotional communicative behaviour subscale is small, which limits the reliability of the scores for this subscale (Nunnally & Bernstein, 1978). A limitation of our study concerning the MDL was that the stopping rule recommended in the manual of the Portage Program Developmental Checklist was used for five children that were enrolled in the study whereas the complete instrument was administered to the children who enrolled at a later stage. Using the stopping rule implies that items that are not administered are scored as 0. We ran a sensitivity analysis to explore whether omitting the scores of these five children would influence our results and found that it did not. Another important observation concerns the psychometric properties of the CSBS-DP and the MDL for the subpopulation of people with PIMD. The CSBS-DP was assessed among both typically developing children and children with developmental disabilities and the Developmental Checklist, on which the MDL was based, was assessed among typically

developing children (Hoekstra et al., 2010; Wetherby et al., 2002). Although several studies have reported good psychometric properties in other groups, neither the CSBS-DP nor the MDL have yet been extensively studied for psychometric properties in people with PIMD. As assessment is significant in providing appropriate support to people with PIMD, it is highly important to start obtaining knowledge in this field, which is currently only possible using analysis techniques that contain certain disadvantages. As to our knowledge, there are no instruments studied for their psychometric properties in the group of people with PIMD, which measure a related construct as the BAS. It is not yet possible to determine convergent validity using instruments that are studied in the specific group of people with PIMD. Our results are therefore exploratory and should be interpreted with caution. However, this exploratory study was a first step in generating evidence for developing instruments for people with PIMD with sound psychometric properties. Given the lack of assessment instruments for people with PIMD, this is an important step. It is advisable that future studies focus on further exploring the convergent validity of the BAS using other instruments that measure a related construct as the BAS, ideally with instruments that are studied for psychometric properties in the group of people with PIMD when such instruments become available.

According to Vlaskamp and colleagues (2002), several principles of the BAS are beneficial to the population of people with PIMD, including the limited influence of motor and sensory impairments on the scores. According to our results, however, the BAS in its current form is not neutral with regard to visual and motor skills. Given that the BAS was developed for the population of people with PIMD in general, without defining subgroups, we did not expect to find any differences in the factorial structure and order of difficulty of the items. Contrary to expectations, we identified several differences that may be related to the characteristics of the subgroups. The groups differed in terms of abilities, with Group 2 having more severe impairments in general. This is most clearly reflected with regard to visual impairments. Some of the items that are not part of the visual behaviour scale nevertheless rely on a visual component. For example, the item "the person makes faces in the mirror," which is part of the communicative behaviour subscale, is more difficult for or inapplicable to people with severe visual impairments. Moreover, some items in the visual behaviour subscale rely on abilities with an auditory component (e.g., "the person moves their eyes and head towards a sound"). Finally, some items that are not part of the exploratory behaviour subscale rely on motor skills. In conclusion, the results of this study support the subscale structure of the BAS and suggest that the convergent validity of the instrument is good. However, the results suggest that scale properties are not invariant across subgroups of people with PIMD; possibly due to interaction between the content of particular items and characteristics of the person being assessed. The next step could be evaluating the content of the items. The items that correlated more strongly with another subscale than the one they were assigned to, should be analysed for their content to see if they would better fit in a different subscale. Items that may rely on motor abilities or visual abilities should be adapted (for example formulated in a broader way) or removed. This could improve the BAS by building on the principles needed to assess people with PIMD. Such adaptations could also enhance content validity if this is specifically addressed in a revision. Moreover, if content validity is evaluated, the difference in length between the subscales should be taken into account, by analysing whether a more equal length of subscales could be established. Furthermore, in next studies, convergent could be further studied by comparing the BAS with different instruments. In addition, the BAS is frequently used to evaluate the effects of interventions or the development of functional skills over time. Further studies should involve a deeper examination of its responsiveness for measuring change in functional abilities. Assessment instruments that have been developed especially for people with PIMD and whose psychometric properties have been examined are scarce (Carnaby, 2007). To our knowledge, the BAS is one of the few instruments that was specifically developed for the group of people with PIMD and studied for psychometric properties and the first exploratory results of this study support the construct validity of the BAS. Assessors should nevertheless exercise critical reflection in interpreting the outcomes, particularly with regard to how they might have been influenced by the person's disabilities. Although its psychometric properties could be improved, the BAS can provide useful information about the functional abilities of a person with PIMD, thereby providing a starting point for adapting support services to the individual strengths and weaknesses of those being assessed (Vlaskamp et al., 1999).

Acknowledgements

We would like to thank the research groups that made their data available for our analyses, as well as the adults and children with PIMD and their families who took part in these studies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Marleen D. Wessels http://orcid.org/0000-0001-6331-577X
Muirne C. S. Paap http://orcid.org/0000-0002-1173-7070

References

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2), 1–18. https://doi.org/10.1002/wcs.1430

Bellamy, G., Croot, L., Bush, A., Berry, H., & Smith, A. (2010). A study to define: Profound and multiple learning disabilities (PMLD). *Journal of Intellectual Disabilities*, *14*(3), 221–235. https://doi.org/10.1177/1744629510386290

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Carnaby, S. (2007). Developing good practice in the clinical assessment of people with profound intellectual disabilities and multiple impairment. *Journal of Policy and Practice in Intellectual Disabilities*, *4*(2), 88–96. https://doi.org/10.1111/j.1741-1130.2007.00105.x

Colla, S., Van keer, I., Schalen, G. H., van der Putten, A., Visser, L., Maes, B., Vlaskamp, C., & Van der Meulen, B. (2015, October 30). *OJKO-project: Longitudinal study on the development of young children with a serious cognitive and motor developmental delay*. [Conference presentation] Eurlyaid conference, Paris, France.

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

Dai, S., Wang, X., & Svetina, D. (2016). TestDataImputation: Missing Item Response Imputation for Test and Assessment Data. R package version 1.0.0

Dhondt, A., Van keer, I., Van der Putten, A., & Maes, B. (2020). Communicative abilities in young children with a significant cognitive and motor developmental delay. *Journal of Applied Research in Intellectual Disabilities*, *33*(3), 529–541. https://doi.org/10.1111/jar.12695

Eadie, P. A., Ukoumunne, O., Skeat, J., Prior, M. R., Bavin, E., Bretherton, L., & Reilly, S. (2010). Assessing early communication behaviours: Structure and validity of the communication and symbolic behaviour scales—developmental profile (CSBS-DP) in 12-month-old infants. *International Journal of Language & Communication Disorders*, *45*(5), 572–585. https://doi.org/10.3109/13682820903277944

Evers, A., Hagemeister, C., Hostmaelingen, A., Lindley, P., Muniz, J., & Sjöberg, A. (2013). EFPA review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers. http://www.efpa.eu/professional-development/assessment

Hoekstra, A., Jansen, G., Van der Meulen, B., Oenema-Mostert, C., & Ruijssenaars, A. (2010). Home intervention: Validating the item order of a developmental Checklist. *Infants & Young Children*, *23*(4), 303–311. https://doi.org/10.1097/IYC.0b013e3181e9035d

Hoekstra, A. T., Oenema-Mostert, C. E., & Van der Meulen, B. F. (2011a). *Portage Programma. Theoretische en praktische handleiding*. [Portage Program. Theoretical and practical manual] (11/3849.03 ed.). Pearson.

Hoekstra, A. T., Van der Meulen, B. F., Oenema-Mostert, C. E., Jansen, G. G. H., & Smrkovsky, M. (2011b). *Vaardighedenlijst-Revised*. [Skillslist- Revised] (11/3849.03 ed.). Pearson.

Kendall, P. C., & Norton-Ford, J. D. (1982). *Clinical psychology: Scientific and professional dimensions*. Wiley.

Kiernan, C., & Jones, M. (1982). *Behaviour assessment Battery: Assessment of the cognitive, communicative and self-help skills of the severely mentally handicapped*. Nfer-Nelson.

Lyons, G., De Bortoli, T., & Arthur-Kelly, M. (2016). Triangulated proxy Reporting: A technique for improving how communication partners come to know people with severe cognitive impairment. *Disability and Rehabilitation*, *39*(18), 1814–1820. https://doi.org/10.1080/09638288.2016.1211759

Nakken, H., & Vlaskamp, C. (2007). A need for a taxonomy for profound intellectual and multiple disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, *4*(2), 83–87. https://doi.org/10.1111/j.1741-1130.2007.00104.x

Nunnally, J., & Bernstein, I. (1978). *Psychometric theory*. McGraw-Hill.

Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology*, *39*(4), 214-223. https://doi.org/10.1111/j.1469-8749.1997.tb07414.x

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Revelle, W. (2018). Psych: Procedures for Personality and Psychological Research. R package version 1.8.4.

Stuive, I. (2007). A comparison of confirmatory factor analysis methods: Oblique multiple group method versus confirmatory common factor method. [Doctoral dissertation, University of Groningen]. Groningen. https://www.rug.nl/research/portal/files/2786144/proefschrift_Ilse_Stuive.pdf

Stuive, I., Kiers, H. A., Timmerman, M. E., & Ten Berge, J. M. (2008). The empirical verification of an assignment of items to subtests: The oblique multiple group method versus the confirmatory common factor method. *Educational and Psychological Measurement*, *68*(6), 923–939. https://doi.org/10.1177/0013164408315264

Van Alphen, H. J. M., Waninge, A., & Van der Putten, A. A. J. (2018, July 18-20). *Evaluation of implementation of a motor activation program for people with profound intellectual and multiple disabilities*. [Conference presentation] 5th IASSIDD Europe Congress, Athens, Greece.

Van der Putten, A. A. J., Vlaskamp, C., Luijkx, J., & Poppes, P. (2017). *Kinderen en volwassenen met zeer ernstige verstandelijke en meervoudige beperkingen: tijd voor een nieuw perspectief*. [Children and adults with profound intellectual and multiple disabilities: time for a new perspective] [Position Paper]. https://www.rug.nl/gmw/pedagogical-and-educational-sciences/research/researchpmd/mensen-met-zeer-ernstige-verstandelijke-en-meervoudige-beperkingen.pdf

Van Ginkel, J. R., Van der Ark, L. A., Sijtsma, K., & Vermunt, J. K. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics and Data Analysis*, *51*(8), 4013–4027. https://doi.org/10.1016/j.csda.2006.12.022

Van keer, I., Ceulemans, E., Bodner, N., Vandesande, S., Van Leeuwen, K., & Maes, B. (2019). Parent-child interaction: A micro-level sequential approach in children with a significant cognitive and motor developmental delay. *Research in Developmental Disabilities*, *85*, 172–186. https://doi.org/10.1016/j.ridd.2018.11.008

Van Timmeren, E., Van der Putten, A. A. J., Van Schrojenstein Lantman-de Valk, H. M. J., Van der Schans, C., & Waninge, A. (2016). Prevalence of reported physical health problems in people with severe or profound intellectual and motor disabilities: A cross-sectional study of medical records and care plans. *Journal of Intellectual Disability Research*, *60* (11), 1109–1118. https://doi.org/10.1111/jir.12298

Vig, S., & Sanders, M. (2007). Assessment of Mental Retardation. In M. R. Brassard & A. E. Boehm (Eds.), *Preschool assessment: Principles and practices* (pp. 420–446). Guilford Press.

Visser, L., Vlaskamp, C., Emde, C., Ruiter, S. A., & Timmerman, M. E. (2017). Difference or delay? A comparison of Bayley-III Cognition item scores of young children with and without developmental disabilities. *Research in Developmental Disabilities*, *71*, 109–119. https://doi.org/10.1016/j.ridd.2017.09.022

Vlaskamp, C. (2005a). Assessing people with profound intellectual and multiple disabilities. In J. Hogg, & A. Langa (Eds.), *Assessing adults with intellectual disabilities: A service providers' guide* (pp. 152–163). Blackwell Publishers.

Vlaskamp, C. (2005b). Interdisciplinary assessment of people with profound intellectual and multiple disabilities. In J. Hogg, & A. Langa (Eds.), *Assessing adults with intellectual disabilities: A service providers' guide* (pp. 39–51). Blackwell Publishers.

Vlaskamp, C., Poppes, P., & Van der Putten, A. A. J. (2015). *Databank interventies langdurende zorg: beschrijving Opvoedings-/ondersteuningsprogramma voor mensen met (zeer) ernstige verstandelijke en meervoudige beperkingen.* [Database interventions of longterm care: description of the Support Program for people with profound/ severe intellectual and multiple disabilties]. Vilans.

Vlaskamp, C., Van der Meulen, B. F., & Smrkovsky, M. (1999). *Het Gedragstaxatie- Instrument* [The Behavioural Appraisal Scales]. Stichting Kinderstudies.

Vlaskamp, C., Van der Meulen, B. F., & Zijlstra, H. P. (2002). De instrumentele realisering van het Gedragstaxatie-Instrument [The instrumental realisation of the Behavioural Appraisal Scales]. *Tijdschrift Voor Orthopedagogiek*, *41*(1), 22–31.

Vugteveen, J., Van der Putten, A. A. J., & Vlaskamp, C. (2014). *Inventarisatieonderzoek personen met ernstige meervoudige beperkingen: prevalentie en karakteristieken [Inventory study about persons with severe multiple disabilities: Prevalence and characteristics]*. Stichting Kinderstudies.

Wessels, M. D., & Van der Putten, A. A. J. (2017). Assessment in people with PIMD: Pilot study into the usability and content validity of the Inventory of the personal Profile and support. *Cogent Psychology*, *4*(1), 1–10. https://doi.org/10.1080/23311908.2017.1340082

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1202–1218. https://doi.org/10.1044/1092-4388(2002/097)

Wetherby, A., & Prizant, B. (2002). *Communication and symbolic behavior scales: Developmental profile – first normed edition*. Paul H. Brookes.

Wickham, H. (2017). Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.

Wickham, H., & Bryan, J. (2018). readxl: Read Excel Files. R package version 1.1.0.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.5.

Wickham, H., & Miller, E. (2018). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 1.1.1.