ORIGINAL ARTICLE

Low Motor Assessment: A Comparative Pilot Study with Young Children With and Without Motor Impairment

Selma Anne José Ruiter · Han Nakken · Bieuwe F. van der Meulen · Carolien B. Lunenborg

Abstract Most of the developmental instruments that measure cognitive development in children rely heavily on fine motor skills, especially for young children whose language skills are not yet well developed. This is problematic when evaluating the cognitive development of young children with motor impairment. The purpose of this study is to assess the need for a Low Motor adapation of a standardized instrument when testing children with motor impairment. To accomplish this, we have adapted the procedures, item instructions and play material of a widely used and standardized instrument, the Bayley Scales of Infant Development-Second Edition (BSID-II, Bayley 1993). The Original and the Low Motor versions were administered to 20 children experiencing typical development and 19 children with motor impairment within a period of two weeks. Results showed that children with motor impairments scored significantly higher on the Low Motor version of the Bayley Mental Scale than on the Original version: a difference of between 5 and 10 points when the score is expressed in terms of a developmental index score. Results from children with typical development support the assumption that item content and difficulty remain unchanged in the Low Motor version.

Keywords Adaptive assessment · Motor impairment · Bayley scales · Young children · Cognitive assessment

S. A. J. Ruiter (✉) · H. Nakken · B. F. van der Meulen · C. B. Lunenborg
Department of Special Needs, Education and Child Care, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Rozenstraat 38, 9712 TJ Groningen, The Netherlands
e-mail: s.a.j.ruiter@rug.nl

Adaptive Assessment of Young Children with Motor Impairment

To facilitate satisfactory decision-making (e.g., for school type) and effective intervention measures for children at risk for developmental problems, it is essential to make sure that appropriate and fair assessment is possible. This becomes even more critical when a child has a specific disability, such as motor impairment. However, when evaluating young children with motor impairment, most standardized tests are of limited use because of their strong motor component (e.g., Becher et al. 2003; Bagnato et al. 2007; Snow and Van Hemel 2008). Especially in the case of young children, whose language skills are not yet well developed, performance in cognitive tasks relies heavily on their ability to explore and manipulate play material.

Worldwide, only a few diagnostic developmental tests are available that measure young children's cognitive development without relying heavily on their motor abilities. The best known instruments are the Kent Infant Developmental Scale (KIDS, Reuter and Gruber 2000), the Fagan Test of Intelligence (FTII, Fagan and Shepherd 1991), the Snijders-Oomen Non-Verbal Intelligence Test 2.5 to 7 Years (SON-R 2½–7, Tellegen et al. 2007) and the Kaufman Assessment Battery for Children (K-ABC (-II), Kaufman and Kaufman 1983, 2004). These scales all cover a large proportion of the aspects of a child's cognitive development, but do not specifically take motor disabilities into account. The KIDS evaluates psychomotor development based on information from parents on 252 behavioral items grouped into five domains: cognitive, language, motor, self-help and social. Although it is not a traditional instrument where a child performs tasks to demonstrate his cognitive skills, developmental age estimates from the KIDS are strongly correlated and similar to results from a traditional test like the Bayley Scales (BSID) (Reuter and Gruber 2000; Schneider et al. 1990). The major drawbacks of the test are, however, its limited age range (1–14 months) and the fact that the data are entirely dependent on informant reporting; this increases the possibility of bias. The Fagan Test of Infant Intelligence (FTII) is a screening device for testing visual recognition (novelty preference). In their examination of the validity of the FTII, Tasbihsazan et al. (2003) concluded that correlations between the BSID-II and the FTII were low and non-significant, and that its clinical utility was limited as a result of its low sensitivity. Meta-analyses have found average correlations between results from the FTII and later scores on intelligence tests (e.g., Bornstein et al. 1997). The SON-R is a non-verbal intelligence test for young children, which does not demand much from the fine motor skills of a child, yet the manual does not provide any standard Low Motor adaptations for the test procedure and item instruction. It is left to the clinician to determine how and to what degree the adapted procedures and/or item instructions are necessary on a case-by-case basis (Tellegen et al. 2007). Research conducted with the K-ABC has shown that this test is suitable for children with motor impairment (Jurjens and Hendriksen 1991). Although many items have an important motor component, one important advantage is that this test differentiates between simultaneous and sequential information processing. This dovetails well with clinical (rehabilitation) practice, because it provides information about the information processing possibilities in the right and left hemispheres. A number of other neuropsychological tests is also suitable for this target group, including the Motor-Free Visual Perception Test-Third Edition (MVPT-3, Hammill and Colarusso 2003)

and the Colored Progressive Matrices (Raven and Raven 2003). However, these tests only measure a specific area of cognitive functioning and are not suitable for very young children.

Based on these findings, it can be stated that there are very few useful tests for the target group of young children with motor impairment that provide an appropriate and fair picture of their cognitive development and possibilities. Although the tests used are those that offer some advantages for children with motor impairment, they remain for the most part unsuitable due to the many items containing a significant motor or fine motor component (K-ABC), or because of a very limited age span (KIDS), or due to the fact that they only measure one aspect of cognitive functioning (SON-R and FTII). Although the literature frequently mentions the need for diagnostic tools adapted for children with motor impairment, there is little empirical research that provides data on differences in test performance on an original and adapted developmental instrument for these young children. The purpose of this study is to experimentally assess the need for Low Motor adaptations when a standardized developmental instrument is administered to children with motor impairment.

With this end in mind, we have adapted the procedures, item instructions and play material of a widely used and standardized instrument, the Bayley Scales of Infant Development-Second Edition (BSID-II) for assessing young children's general cognitive and motor development, also known as the "gold standard" (Aylward 2002; Gauthier et al. 1999). The present study was based on the Mental Scale of the Dutch version of this test (BSID-II-NL; Van der Meulen et al. 2002). The Dutch version consists of a translation of the BSID-II, a Dutch standardization, along with a discussion of validity and reliability studies for the BSID-II-NL (Ruiter et al. 2005). The test procedure, item instructions and play material were adapted so as to minimize the impact of motor skills impairment. Although each adaptation of the test procedure, item instruction and/or play material relevant to the Standard instructions implies that the Original norm tables no longer apply, the adaptations we propose are only meant to compensate for a child's impairment. By leaving the item content and degree of difficulty essentially unchanged, the Original norm tables can apply as before. Batshaw Clair et al. (2002) describe such changes in format, response, environment, timing or scheduling, that do not significantly alter what a test measures, as accommodations. In contrast, when changes in assessment alter what a test measures or the comparability of its scores, the term modification is used. If there is sufficient evidence to show that the adapted version of the BSID-II involves only accommodations, it will not be necessary to conduct large-scale and time-consuming standardization research for this specific group of children. The use of a standardized Low Motor version, combined with the Original norm tables, will permit a comparison of the cognitive development of children with motor impairment with that of pears without.

To determine whether test results for the Low Motor version of the BSID-II better reflect a child's true score than the Original version does, we hypothesized that:

(1) Item content and difficulty in the Low Motor version will remain unchanged in comparison with the Original version

(2) Test results for the Low Motor version of the BSID-II will be significantly
      higher than test results for the Original version for children with motor
      impairments.

Method

Subjects

Seventy-one children participated in this study. Twenty children represented the
Standard population. These latter children experience typical development without
any specific physical or mental problems (the "Standard Group"). All the children in
this group are aged 12 to 42 months, six male and fourteen female. All come from
the city of Groningen. This group had already participated in the Dutch
standardization study for the BSID-II. Their parents agreed to their also participating
in this study.

   Sixty-six children with motor impairments participated in this study. They were
all known to one of the centers for assessment and intervention for motor-impaired
children in the Netherlands, or to one of the day-care centers for children with severe
profound developmental impairment. We approached, in total, 30 centers via their
educational psychologists with the request that they participate in this research,
testing children from the target group twice within a period of two weeks, one test a
BSID-II-NL Standard version and the other a Low Motor version of the test. Since
the BSID-II is a commonly used instrument in the diagnosis and intervention
planning for children with motor impairment in the Netherlands, the test results were
not for research purposes alone. Results could also be integrated into the regular
treatment structures of the centers. While this was an important advantage and all
centers applauded the initiative of addressing the need for an accurate cognitive
measurement for children with motor impairment, only 18 centers volunteered to
take part. The most important reasons given for not participating were the short
timeframe in which the tests had to be administered, the overly heavy workload of
the educational psychologists working at the centers, and the lack of availability of
children who fully met the inclusion criteria.

   The educational psychologists working at the centers recruited the children on the
basis of formulated inclusion criteria. These criteria describe the minimum abilities
required of a child in order to perform the test items.

   The children in the Clinical Group met the following inclusion criteria:

- The child was diagnosed with mild to severe motor impairment
- The child's developmental age ranged between 12 and 42 months
- The child was able to sit up in a chair or wheelchair to work at a table
- The child had sufficient visual and auditory abilities
- The child could use at least one hand.

   Their motor impairments were the result of various syndromes, anomalies or
diseases. Inclusion criteria were not based on diagnosis, because even children with
identical diagnoses could experience great differences in the degree of limitation to

their motor abilities. Children were not included in the study when parents did not agree to participation in the project, when the motor impairment did not affect hand function, when the child had been diagnosed with a hearing problem, or if it was thought that carrying out the test would be too stressful for the child.

The research started off with 66 children with Low Motor, but only nineteen children were finally able to be included in the analyses. Reasons for this loss of data were the following. First, the Standard and Low Motor versions of the test could not be administered to a significant group of children ($n=32$) within a given period of time (only one test result was available), mostly because of time constraints affecting the educational psychologist (test administrator) or because of the children's health problems. Second, twelve children did not reach the basal level within the item set at twelve months. Because the Low Motor version was designed for ages ranging between 12 and 42 months, a test result for the Low Motor version could not be determined. And, finally, three test administrations were excluded because the child refused to perform on any item of the test and the educational psychologist decided not to organize a second trial because the child reacted in an extremely tense way to the test situation ($n=1$). One test administration was excluded because the parents withdrew their initial agreement concerning use of the data for research purposes (for reasons unknown), and one test administration was discontinued because the adapted Low Motor material appeared to be too heavy and large.

Eventually, 19 children met the inclusion criteria and were able to be included in the data analysis. Their chronological age varied between 14 months and 146 months (12 years), and they numbered 12 males and 7 females.

Instruments

The BSID-II The BSID-II (Bayley 1993) is a well developed, standardized and norm-referenced instrument for evaluating the general development of infants aged one to 42 months. The Bayley Scales (Bayley 1969, 1993) are known as the "gold standard" of psychometric excellence for infant and preschool tests (Koseck 1999; Flanagan and Alfonso 1995). The instrument consists of three scales: the Mental, Motor, and Behavior Rating Scales. The Mental Scale consists of 178 items that measure the child's cognitive skills. The Mental Scale items relate to the processing of visual and auditory information, hand-eye coordination, imitation, language development, memory and problem solving. The Motor Scale consists of 111 items that measure skills related to gross and fine motor control, including movements such as rolling, crawling, standing, walking, running, and jumping. This scale also tests fine motor manipulations, such as hand-eye coordination, adjusted use of writing materials and imitation of hand gestures. Most of the items in the Mental and Motor Scale are task items, dichotomously scored. Depending on the age and developmental level of the child, an age-appropriate item set with a specific start and stop item was administered for both the Mental and Motor scales. Raw scores were converted into a mental developmental index (MDI) score and a motor developmental index score (PDI), with a mean of 100 (SD 15).

The Behavior Rating Scale contains questions answered by the administrator after administering the test. These questions allow the administrator to judge the behavior of the child during the test.

The BSID-II Low Motor The Low Motor version of the Original BSID-II offers adapted test procedures, item instructions and play material for administering the test to children with motor impairments.

The Low Motor adaptations were designed for the Mental Scale, and for the age range of 12 to 42 months. For age groups under 12 months old, internal consistency was found to be very low (.50 ≤ lambda.2≤.65). This led to very broad reliability intervals and, therefore, a relatively low reliability for the test result. For this reason, the lower age boundary was set at 12 months. The Low Motor adaptations were developed in close cooperation with practitioner experts—qualified educational psychologists with extensive experience in working with children with motor impairment. This project began with a literature survey and interviews with experts in order to formulate recommendations for developing an adapted version of the BSID-II for use with children with motor impairment. The important outcomes were, first, that due to an often delayed information processing system, time limits needed to be eliminated. Second, play material (e.g., pegs and pegboard, puzzle board, and cubes and dolls) needed to be enlarged so that these items required less from the disabled hand function of a child with motor impairment. Third, children needed to be allowed to use eye pointing (glance direction) instead of finger pointing, in combination with the presentation of pictures from the stimulus booklet on a transparent frame (a Perspex frame standing on the table between the child and the administrator, on which pictures can be attached, and spaced well apart) so that the eye pointing of the child could be clearly observed. Finally, when it was clear that the intention was to carry out the item properly, only problems in motor and fine motor control hindered a child in executing the item; light motor support was permitted to be given by the tester. For instance, where a child brought a peg close to a hole in the pegboard, but repeatedly could not make the final move to get the peg into the hole in the pegboard, supporting the arm of the child by the elbow was allowed in order to enable the child to control his arm/hand. An experimental version of the BSID-II Low Motor was designed and piloted (Geerds 2004). The experimental Low Motor version of the BSID-II was administered by practitioners to 21 children from the target group. The practitioners were supplied with the materials and trained in applying the adapted test procedure and play material. Practitioners were all experienced BSID-II users. They were asked to administer the Low Motor version of the BSID-II and to evaluate, using standardized forms, the applicability and usability of the adapted test procedure in general, and item instruction and adapted play material per item specifically. Positive results confirmed its applicability and usability for children with motor impairment. Most of the adapted testing procedure, item instruction and play material met the applicability requirements for this target group. Suggested improvements and changes to item procedures, instructions and/or play material were evaluated and, when judged to be clear and usable improvements, they were assimilated into the final version of the BSID-II Low Motor. A major standardizing factor was that the detailed test procedure and item instructions were given in the manual. How the item should be administered and scored was exactly described item by item. This was of course the case for the Standard version of the BSID-II, but for the Low Motor version as well. Following the example of the Standard instructions, the test procedure for the Low Motor version was described and detailed instructions were given for each item.

The adaptations in test procedure, item instruction and play material were an addition to the Standard material of the BSID-II. The final instrument consisted of 108 items: items requiring only adaptation of their material (13), items requiring only adaptation of their instructions (9), items requiring both instructional and material adaptation (47), items that were acceptable as they were (29), and biased items that could not be changed so as to be acceptable (10).

Procedure

Children in the Standard Group were examined by Master's students in the playroom of the Department of Special Needs, Education and Child Care in Groningen. Students were all very experienced BSID-II administrators and they received intensive training in administering the Low Motor version. Children in the Clinical Group were examined at the centers by qualified educational psychologists with extensive experience in the assessment of children with motor impairment and in administering the BSID-II. They also received training in the Low Motor version.

The tests for children with motor impairment were administered by practitioners who were familiar with the children and their specific limitations. Despite their extensive experience with the BSID-II, the Master's students lacked the necessary knowledge and experience with children with motor impairment so as to be able to administer the test well. As for the practitioners, it was virtually impossible for them to test the control group children (Standard Group).

To test the two hypotheses, content validity was examined in two ways. First, performance on the BSID-II (Original version) and the BSID-II Low Motor (Low Motor version) were compared. Children from the Standard and Clinical Groups were tested twice within two weeks, once with the Original version and once with the Low Motor version in counterbalanced order and under identical conditions. We expected that the Standard Group's test results for the Original and Low Motor versions would not differ significantly (Hypothesis 1). For the children with motor impairment, we expected that applying the Low Motor version would maximize the test results, and that the test results for the Low Motor version would exceed those for the Original version. As a result of the adaptation, the validity of the test for the group of children with motor impairment should increase (Hypothesis 2).

Second, test performance at the item-category level was compared. Test results for the Low Motor and Original versions of the BSID-II for both groups of children were compared at the item-category level. Items were assigned to five categories. One category was made up of the non-verbal items (Category 0); the other categories (1–5) were based on the type of adaptation (Table 1). Categories 4 and 5 were left out of the analysis, because these items remained unchanged in the Low Motor version.

Test results for typically developing children should not vary significantly between the Low Motor and Original versions (Hypothesis 1). We expected that children with low motor would obtain significantly higher test scores for the items in Categories 1, 2, and 3 (Hypothesis 2).

Table 1 Overview of the number of adapted items per category in the Mental Scale between 12 months and 42 months (item no. 71–178)

Category	Adaptation	Number of items
1	Item requires adaptation of its material	13
2	Item requires adaptation of its instructions	9
3	Item requires instructional and material adaptation	47
4	Item is acceptable as it is	29
5	Item requires adaptation, but cannot be changed to be acceptable	10

Results

Test performances at the scale-category and item-category levels were compared. Results were based on raw scores. Developmental indices could not be obtained for most of the children in the Clinical Group; either they were older than 42 months or their raw scores were too low to convert to a developmental index. The concrete difference in raw score points is presented in Fig. 1. In both sets of plots, a positive score (>0.00) means that the raw score for the Low Motor version is higher than the raw score for the Original.

As can be concluded from Fig. 1, discrepancies in the Clinical Group are clearly to the advantage of the results from the Low Motor version of the BSID-II. Almost 75% of the children achieved higher results with the Low Motor version. On average, the children with motor impairments achieved a score 3.5 raw-score points higher with the Low Motor version than with the Original version of the BSID-II. Converted to an index score, this can mean a difference of about 5 to 10 developmental index points. This difference was confirmed by an analysis at group level using the Wilcoxon signed-rank test ($z=-2.85$, $p<0.01$, $r=-.46$).

In the Standard Group, children achieved higher scores both with the Low Motor and Original versions, equally. As expected, analysis showed no significant difference for the Standard Group ($z=-1.23$, $p=.22$).
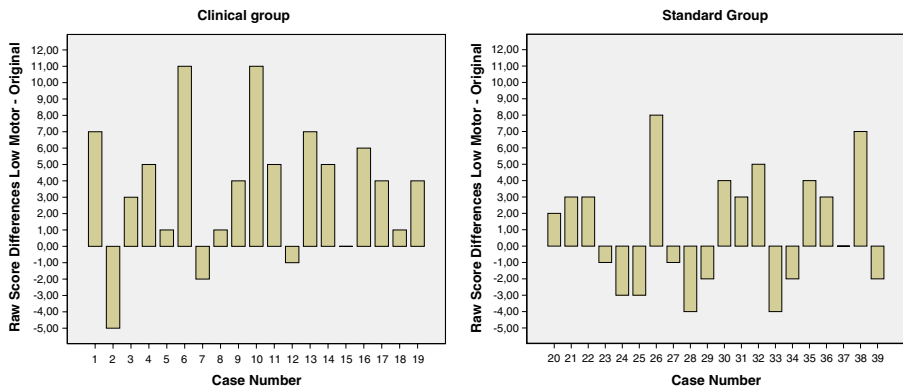
Fig. 1 Differences between the raw scores for the Low Motor and the Original versions of the BSID-II when administered to the Clinical and Standard Groups

To analyze differences at the individual level, a minimally significant difference was calculated between the individual scores for the Original and Low Motor versions, using $d_\alpha = z_\alpha.s.\sqrt{(2 - r_{xx} - r_{yy})}$ (based on reliability and standard deviations of the standardization sample; see Ruiter et al. 2005, pages 12 and 18). An individually significant difference was found when the difference in test scores ranged over a minimum of six raw points. From the analysis, it would appear that five participants from the Clinical Group had deviant scores, all in favor of the Low Motor version. From the Standard Group, only two out of 20 children showed significantly different scores between the Original and Low Motor versions of the BSID-II.

To determine whether performance on the Mental scale was significantly different when the items were assigned to categories, we designed item categories based on the nature of the adaptation: items with material adaptation (Category 1), items with instructional adaptation (Category 2), and items with both material and instructional adaptation (Category 3). Figure 2 shows the individual differences between the raw scores for the Low Motor and Original versions within item categories that reveal a difference between the two versions because of material and/or instructional adaptations.

In all three categories, the vast majority of the Clinical Group performed better on the Low Motor version. The mean total raw-score difference was 3.5 points. Children with motor impairment on the average scored .95 raw points higher on items in Category 1, .21 points higher in Category 2, and 1.8 points higher on the Low Motor version of the items in Category 3.

Analysis with Wilcoxon signed-rank tests for each item category revealed significant differences between the test results in favor of the Low Motor version for two of the item categories in the Clinical Group. As expected, children profited significantly from the material adaptation alone (Category 1; $z = -2.05$, $p = .04$, $r = -.33$), as well as from the material adaptation in combination with the instructional adaptation (Category 3;
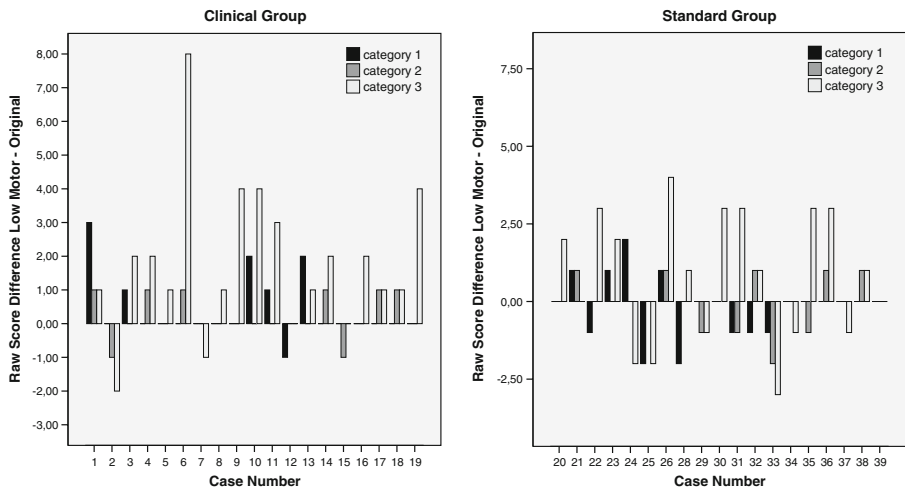

Fig. 2 Raw score differences within three item categories between the Low Motor and the Original versions of the BSID-II for the Clinical and Standard Groups. Note: Case 27 was excluded because no data at the item level was available to determine raw scores within categories

$z=-2.90$, $p=.00$, $r=-.47$). For the Standard Group, the mean difference between the Low Motor and Original versions was 1.1 raw score points. Within the categories this was $-.16$ in Category 1, .00 in Category 2, and .84 in Category 3. An analysis confirmed the expectation: No significant differences were found between the results for the Low Motor and the Original versions of the BSID-II for the group of children with typical development.

For the Clinical Group, the results were also calculated for the non-verbal items of the Mental scale. Children with motor impairment achieved significantly higher scores ($z=2.61$, $p=.00$, $r=.42$) for the non-verbal items with the Low Motor version than with the Original version. This was probably due to the large number of items in which the child needed to manipulate material. In addition, the Clinical Group mostly profited from those item adaptations that concerned material and instruction.

Discussion

To date little work has been done to develop accurate and fair tests for children with motor impairment. This has led to use of "home-made" adaptations to commonly used instruments—changes in the test process, the item instructions, the response format or the standardized play material—in order to more accurately portray the performance of children with motor impairment. In this study, we examined whether there was a need for an adapted diagnostic tool to compensate for motor impairment bias when a standardized instrument was administered to young children with motor impairment. The BSID-II provides such standardized findings. The Standard test material depends heavily on the visual and motor skills of a child. It was expected that using the Original BSID-II with children with motor impairment would affect individual performance and, in general, the efficacy and validity of the instrument. In close cooperation with practitioners, a BSID-II Low Motor version was designed to minimize the impact of motor impairment on a child's Mental Scale performance.

The hypotheses were confirmed. Test performance on the BSID-II was not affected by using adapted Low Motor material (Hypothesis 1). The scale scores for the Low Motor version for the group of children experiencing typical development did not differ significantly from the scores achieved with the Original version. When items were grouped based on how they were adapted, the results for both versions continued not to differ significantly. Item content and difficulty remained essentially unchanged. These results support the supposition that the administrator is in fact free to make any necessary changes for the specific needs of a child without threatening the validity of the test and the applicability of the Standard norm tables. However, to demonstrate conclusively that the Standard norm tables still apply to the Low Motor version of the BSID-II, a study on a larger scale should be conducted. This is necessary in order to be able to analyze the item-difficulty ratings more robustly, not only on the scale level but also on the item level.

The BSID-II Low Motor compensates for motor impairment bias when a standardized instrument is administered to children with motor impairments (Hypothesis 2). The results indicated that children with motor impairments profited from the use of the adapted test materials. Their scale scores for the Low Motor version of the test were, on average, significantly higher by 3.5 raw score points.

When converted to a developmental index score, this can mean a difference of between 5 and 10 points. This can tip the balance when judging eligibility for intervention programs or special education. When items were grouped based on how they were adapted, children with motor impairment proved to benefit most from adaptations that concerned item instruction in combination with adapted play material. With respect to Categories 1 and 2, obtaining a significant difference was difficult because of the low number of items in these categories. The significantly higher scores for Low Motor children on the non-verbal items of the Low Motor version showed that Low Motor children profited from the expanded response possibilities for those items that generally demand the most from the fine motor skills of a child. For typically developing children, no such difference was found.

Although expert opinion was not systematically investigated in this study, in terms of personal communication, expert administrators noted that the adapted test materials increased the response possibilities of a child and therefore increased the child's motivation to show his abilities. The optional use of personal aids and gestures was considered particularly helpful. Another remark often made concerned the favorable effect of being able to decide at the item level whether an item should be administered according to the Original or Low Motor version of the BSID-II. Due to the interchangeability of the items (instruction and material) an optimal test situation can be created for a child. Several experts emphasized that the publication of the BSID-II Low Motor would finally enable them to administer the test in a standardized manner. Previously, the test had been administered using "home-made" adaptations.

Although relevant and valuable, the findings in this article are subject to at least four important limitations. First, in the present study we addressed the validity issue of the Low Motor adaptations of the BSID-II. Information on the reliability is limited to a measure of inter-rater reliability from a previous pilot study that applied an experimental version of the BSID-II Motor Low ($n=15$). This yielded a median kappa value of .87. This can be described as "almost perfect."

Second, though all the test administrators involved in this study were experienced users of the BSID-II and were trained in administering the Low Motor version, they had not received any additional guidance after the one-off training. It became apparent, while processing the data provided by clinicians, that standard procedures and item instructions were regularly applied incorrectly. From video recordings and notes and remarks on the score forms, we deduced that mistakes were mostly made as to how the items were presented and the amount of assistance that should be given. In some cases, the test administrator was forced to deviate from the instructions because of the specific limitations of the child being tested, while in other cases the administrator clearly had not sufficiently understood the instructions as described in the manual. When processing the results, it was usually possible to correct for these mistakes. Where this was not possible, the test was excluded from the analysis. Video recordings and evaluation forms made it possible to examine carefully whether the items were administered properly. In particular, the evaluation form on which the test leader (administrator) evaluated each item individually made it possible to obtain additional information about the administration method. The training emphasized that in processing the scores the manual should be used as a control. Immediately after administering, each item in the instruction manual was

read again to assess the accuracy of the administration and the scoring of each item. The test leaders were invited to share this evaluation on an evaluation form. The test leader was thus able to describe the difficulties related to that item in general or specifically in relation to an individual child, and to offer possible solutions. One compensating factor was the fact that a large group of test leaders was involved in implementing the test, which means that the influence of administrator bias was limited. Third, 10 of the 108 items in the Low Motor version were biased, but could not be changed to be acceptable. For example Items 169–171, where the child has to push a little car on a map, and Item 132, where a child has to put a string of beads in a tube, require fine motor control. These ten items are impossible to implement for children with serious motor impairments and will in many cases yield an unjustifiably negative score. Fortunately, the items were broadly distributed across the item sets, and negative effects on the test result were thus limited. Nevertheless, all test items biased for children with motor impairments should be adapted in future research. Items incapable of change should be replaced by new items that measure the same ability as the Original item, but that are applicable for children with motor impairment. Finally, results from this validity study support the assumption that the Low Motor version is a more valid instrument for examining cognitive abilities in young children with motor impairment than the Original version of the BSID-II; however, to answer the research question more conclusively, an extended study to supplement the sample would be required. Larger sample sizes are necessary to ensure that the findings are robust. For this target group, it is known that test results can be strongly influenced by the time of day, the health of the children and the environment where the test is administered. Obtaining a large number of test scores that are adequate and useful for research purposes is therefore a difficult process. Add to this the fact that the target group of children with Low Motor is small and heterogeneous, and the result is that obtaining a large homogeneous sample would be extremely ambitious. However, a larger sample of children with typical development using both versions will enable us to determine more conclusively whether item content and difficulty remain unchanged using Low Motor adaptations. A more thorough analysis at the item level could then be implemented. With respect to the second hypothesis, an extended sample would enable us to classify the heterogeneous sample of children with motor impairment into more homogeneous groups. For example, according to Mittler (1974), the sample could be divided into two main groups: children with motor impairment associated with or as a result of cerebral palsy, and children with motor impairment not related to cerebral palsy. Children with cerebral palsy often experience more complex disabilities that affect their intellectual, perceptual, sensory and/or speech and language abilities. An extended study could also address the following questions: Is there a tipping point in the degree of Low Motor at which a child starts to profit from Low Motor adaptations of the test material? Should Low Motor adaptations be more rigorous in order to draw an optimal distinction, for example, by using a console or developing a shortened version? For children with multiple disabilities, should the Low Motor adaptations be combined with low-vision and non-verbal adaptations? Finally, continuing research should include similar instruments, such as the third revision of the Bayley Scales (Bayley 2006), but also the Wechsler scales for young children (WPPSI-III, Wechsler 2002), or the Mullen Scales of Early Learning (Mullen 1995).

References

Aylward, G. P. (2002). Cognitive and neurological outcomes: more than IQ scores. *Mental Retardation and Developmental Disabilities Research Reviews, 8*, 234–240.

Bagnato, S. J., Macey, M., Salaway, J., & Lehman, C. (2007). *Research foundations for authentic assessments to ensure accurate and representative early intervention eligibility*. Washington: U.S. Department of Education, Office of Special Education Programs, TRACE Center for Excellence.

Batshaw Clair, E., Church, R. P., & Batshaw, M. L. (2002). Special education services. In M. L. Batshaw (Ed.), *Children with disabilities*. Baltimore: Paul H. Brookes Publishing Co.

Bayley, N. (1969). *Manual for the Bayley scales of infant development*. New York: Psychological Corporation.

Bayley, N. (1993). *Manual for the Bayley scales of infant development* (2nd ed.). San Antonio: The Psychological Corporation.

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio: Harcourt Assessment.

Becher, J. G., Smit, L. M. E., & Gorter, J. W. (2003). Cerebrale parese. In M. J. Meihuizen-de Regt, J. M. H. de Moor & A. H. M. Mulders (Eds.), *Kinderrevalidatie, 3^e geheel gewijzigde druk*. Assen: Koninklijke Van Gorcum.

Bornstein, M. H., Slater, A., Brown, E., Roberts, E., & Barrett, J. (1997). Stability of mental development from infancy to later childhood: Three "waves" of research. In G. Bremner, A. Slater & G. Butterworth (Eds.), *Infant development: Recent advances* (pp. 191–215). Hove: Psychology.

Fagan, J., & Shepherd, P. (1991). *Manual: The Fagan test of infant intelligence*. Cleveland: Infantest Corp.

Flanagan, D. P., & Alfonso, V. C. (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. *Journal of Psychoeducational Assessment, 13*, 66.

Gauthier, S. M., Bauer, C. R., Messinger, D. S., & Closius, J. M. (1999). The Bayley scales of infant development II: where to start? *Developmental and Behavioral Pediatrics, 20*, 75–79.

Geerds, F. G. (2004). *The usability of the experimental low motor version of the BSID-II. A pilot study.* Groningen: University of Groningen, (unpublished master thesis).

Hammill, D. D., & Colarusso, R. P. (2003). *Motor-free visual perception test-third edition (MVPT-3)*. Novato: Academic Therapy.

Jurjens, H. B., & Hendriksen, J. G. M. (1991). De Kaufman assessment battery for children: Psychometrische kenmerken en toepasbaarheid. *Proceedings van de najaarsconferentie Sectie Revalidatiepsychologen.* Amsterdam: Nederlands Instituut van Psychologen.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman assessment battery for children*. Circle Pines: AGS.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines: AGS.

Koseck, K. (1999). Review and evaluation of psychometric properties of the revised Bayley scales of infant development. *Pediatric Physical Therapy, 11*, 198–204.

Meulen, B. F. van der, Ruiter, S. A. J., Lutje Spelberg, H. C., & Smrkovsky, M. (2002). *Bayley scales of infant development* (2nd ed., Nederlandse versie). Amsterdam: Harcourt Test. Nederlandse versie.

Mittler, P. (1974). *The psychological assessment of mental and physical handicaps*. London: Tavistock Publications in association with Methuen & Co Ltd.

Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines: American Guidance Services Inc.

Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–240). New York: Kluwer Academic/Plenum.

Reuter, J., & Gruber, C. (2000). *KIDS. Kent inventory of developmental skills. Manual*. Los Angeles: Western Psychological Services.

Ruiter, S. A. J., Lutje Spelberg, H. C., & Meulen, B. F. van der (2005). *BSID-II-NL, deel II: Normering en psychometrische kenmerken*. Amsterdam: Harcourt Test.

Schneider, M. J., Loots, G. M. P., & Reuter, J. (1990). *Kent infant developmental scale. Nederlandse bewerking. Handleiding*. Lisse: Swets & Zeitlinger.

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what and how*. Washington DC: National Academy Press.

Tasbihsazan, R., Nettelbeck, T., & Kirby, N. (2003). Predictive validity of the Fagan test of infant intelligence. *British Journal of Developmental Psychology, 21*, 585–597.

Tellegen, P. J., Winkel, M., & Laros, J. A. (2007). *Snijders-Oomen Non-Verbal Intelligence Test (Revised), SON-R 2.5–7 years*. UK Adaptation of the third edition. Amsterdam: Hogrefe.

Wechsler, D. (2002). *WPPSI-III Administration and scoring manual*. San Antonio: The Psychological Corporation.