

Psychometric properties of the qualitative assessment of manual dexterity in the MABC-2 test

Rudolf Psotta^{1,*}, Dominika Pechová¹, and Josef Kraus²

¹Faculty of Physical Culture, Palacký University Olomouc, Olomouc, Czech Republic; and ²Second Faculty of Medicine, Charles University, Prague, Czech Republic

Copyright: © 2018 R. Psotta et al. This is an open access article licensed under the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

Background: The Movement Assessment Battery for Children Test, 2nd edition (MABC-2 test) is a widely accepted standardized method for the assessment of motor competency and identification of developmental coordination disorder in children. However, there are no information on psychometric qualities of the qualitative part of the test. **Objective:** To reveal possible limitations of the qualitative assessment of the MABC-2 test tasks of the age band 2 (7–10 years), we examined the inter- and intra-rater reliability and construct (known-groups) validity, specifically for observation of the manual dexterity (MD) tasks. **Methods:** The MD tasks of the MABC-2 executed by children with and without motor difficulties in manual dexterity (MD⁻ and MD⁺; $N = 6$) were recorded by a video camera. A group of investigators ($N = 9$) performed qualitative observational assessment of children's movements from video recordings, in two observation sessions apart by four weeks (time 1, time 2). **Results:** The statistical tests rejected the hypothesis on significant inter-rater reliability in the number of marked signs in the MD tasks. Regarding the intra-rater reliability, the number of marked signs in the MD tasks was not significantly different between observation in time 1 and time 2 in all investigators. 95% agreement limits for the number of marked movement signs in the two repeated observations ranged from $M_{\text{diff}} \pm 0.73$ up to $M_{\text{diff}} \pm 2.63$ signs. A good construct validity of the qualitative assessment of movements in the MD tasks was indicated by significantly higher number of movement signs marked in the MD⁻ children as compared to the MD⁺ children, with exception of the graphomotor task. **Conclusions:** The qualitative assessment within the MABC-2 test is a useful tool for the identification of manual dexterity impairment and weaker aspects of hand-eye coordination. However, the qualitative observational assessment of the MABC-2 tasks demands specific training of users.

Keywords: motor test, developmental coordination disorder, validity, reliability, children

Introduction

A smaller proportion of the motor tests developed for children used nowadays involve the assessment of both the quantitative and qualitative aspects of movements. The Movement Assessment Battery for Children Test, 2nd edition (MABC-2 test) (Henderson, Sugden, & Barnett, 2007) is a widely accepted standardized method for the assessment of motor competency in children and identification of motor difficulties that significantly contributes to the decision for a diagnosis of developmental coordination disorder (DCD).

The test was developed for use in physiotherapeutic, psychological and clinical practice, and contains three age versions: age band 1 (AB1, 3–6 years), AB2 (7–10 years) and AB3 (11–16 years). The quantitative assessment of the motor competency of children is derived from the results of performance in eight test tasks that are grouped into three motor domains: manual dexterity (MD), aiming & catching (AC), and balance (Bal). Based on quantitative, norm-referenced results of the test, a decision about criterion A of the diagnosis of DCD is made. According to DSM-5 (American Psychiatric Association, 2013), the definition of criterion A is related to the execution and acquisition of coordinated motor skills that are substantially below that expected, given the individual's chronological age and opportunity for skill learning and use.

* Address for correspondence: Rudolf Psotta, Department of Natural Sciences in Kinanthropology, Faculty of Physical Culture, Palacký University Olomouc, třída Míru 117, 771 11 Olomouc, Czech Republic. E-mail: rudolf.psotta@upol.cz

The qualitative assessment with the MABC-2 test consists in observation and judgement whether, during execution of a given task, a child manifests some of listed critical movement signs that are considered to be deviations from typical movement patterns (see Appendix). This observational assessment brings information on quality in posture, control of movement of body segments, spatial accuracy and timing of movements, and control in muscle effort (Henderson et al., 2007).

Quantitative and qualitative results of the MABC-2 test are judged in their mutual context, and together with other information such as possible psychological and physical factors entering into the motor test performance of a child, and other information obtained from his/her teachers and parents, the final diagnostical decision about motor normality or motor difficulties of the child is completed. Results of the qualitative observational assessment in the MABC-2 test in particular serve psychologists and physiotherapists to specify weaker aspects of motor coordination/control and thus to plan appropriate interventions, and to evaluate their effects (Henderson et al., 2007; Sugden & Henderson, 2007).

The current study is focused on the qualitative observational assessment of the MD tasks of the MABC-2 test – AB2 (7–10 years). To reach valid diagnostical conclusions in both clinical practice and research, several issues of the reliability and validity of the qualitative assessment of movement patterns in the test tasks should be considered. Previous studies confirmed the original assumption that AB2 of the MABC-2 test captures three specific latent motor factors represented by the MD, AC and Bal components, and further, the particular motor tasks in the test are valid for the assessment of the corresponding motor component (Schulz, Henderson, Sugden, & Barnett, 2011; Wagner, Kastner, Petermann, & Bös, 2011). However, there is no information how a method of qualitative assessment in AB2 is valid and reliable. Only one study (Gard & Rösblad, 2009) has verified the psychometric qualities of the qualitative observational assessment of two tasks, Threading beads, and Jumping over a cord in 4–6-year-old children that are involved in the first edition of the MABC test (Henderson & Sugden, 1992).

According to the concept of construct validity (Smith, 2005) and validity of measurement in physiotherapy (Sim & Arnell, 1993), the first issue of qualitative observation in the MABC-2 test could be related to whether the critical movement signs listed for the examiner's observation of particular test motor tasks (see Appendix) are representative of neurological soft signs of developmental coordination disorder in children. The second problem of construct validity could concern to some degree the difficulty to verbally

describe movement signs (Gard & Rösblad, 2009; Henderson et al., 2007).

Thirdly, the Examiner's Manual for the MABC-2 test (Henderson et al., 2007) provides users with a general recommendation how to observe a child's movements during execution of the motor tasks. More specific criteria or rules are not available on how to judge the incidence of individual movement signs. For instance, there is the question of whether or not a given sign was typically manifested during execution of a task when an observer noticed this sign in half or in one third of the trials. Similarly, it is difficult for an observer to decide whether a child actually used excessive force while inserting pegs into a desk in the MD 1 task, or if he/she manifested poor sitting posture, in each of the MD tasks (Appendix). These problems might lead both to wider variations in the identification of the critical movement signs among observers (a problem of inter-observer reliability), and to the disruption of self-reproducibility of the tester's judgements over time, i.e. intra-observer reliability.

To reveal possible limitations of the qualitative assessment in AB2 of the MABC-2 test in psychological, physiotherapeutic and clinical practice, we examined the inter- and intra-rater reliability and construct (known-groups) validity, specifically for observation of the manual dexterity tasks.

Methods

Participants

Three children with difficulties in manual dexterity (MD⁻) and three children without difficulties in manual dexterity (MD⁺), two boys and one girl in each group, recruited from one primary school, participated in the study (mean age 9.3 ± 0.5 years, range 8.0–10.9 years). The following exclusion criteria were used: presence of intellectual, physical and sensory disabilities, medical and other neurological impairments, and current diagnosis of ADHD. School psychologists were responsible for recruitment of children into the study according to the exclusion criteria. In addition, attentional impairment was checked by the d2 test of attention (Brickenkamp & Zillmer, 2000). Children with a score $\leq 5^{\text{th}}$ percentile were not included in the study.

The study was approved by the Ethics Committee of the Faculty of Physical Culture, Palacký University Olomouc (No. 65/2016). A written informed consent was obtained from the parents or legal guardians of the children, and the children gave their oral assent to participate in the study. They were not aware of the specific purpose of the study.

Manual dexterity assessment

To gain a sample of MD⁻ and MD⁺ children, the selected children were tested with three manual dexterity (MD) tasks of the MABC-2 test – AB2: Placing pegs (s) (MD 1), Threading lace (s) (MD 2), and Drawing trial (number of errors) (MD 3). The MD tasks were administrated according to the Examiner's Manual for MABC-2 (Henderson et al., 2007). Test scores were converted according to the norms for the population of Czech children (Psotta, 2014). MD⁻ children were those who achieved a component score ≤ 7 ($< 16^{\text{th}}$ percentile), with classification of mild difficulties (score 6–7) or significant difficulties (score ≤ 5 ; $\leq 5^{\text{th}}$ percentile) in manual dexterity. Children with a component score > 7 ($\geq 16^{\text{th}}$ percentile) were classified as having no difficulties in manual dexterity (MD⁺ children) (Table 1).

Video records of the motor tasks

Execution of the MD tasks by the MD⁻ and MD⁺ children was recorded with a Panasonic HDC-TM 900 digital camera (Panasonic, Kadoma, Japan; 50 Hz, resolution 14.2 Mpx). The camera was placed on a tripod and positioned at a distance so that movement patterns could be well seen, without distracting the child. The height of the location of the camera on the tripod was 160 cm, corresponding approximately at the eye level of a typical observer. The location of the camera enabled a frontal view at a child so that an observer could look at the child's sitting posture, head position, direction of his/her gaze, reaching, grasping, holding and manipulation with an object (peg, board and lace, pen).

The Examiner's Manual for MABC-2 (Henderson et al., 2007) describes two approaches how to carry out the qualitative assessment of the tasks. Firstly, observation can be performed by a qualified observer from an appropriate position while another examiner as a tester concurrently deals with the quantitative administration

of the test tasks with the child. A second approach is that a single tester administrates the formal quantitative test and concurrently makes a qualitative assessment. According to our experience, the latter approach is very difficult for a single tester to complete both types of motor assessment. Therefore, we prefer the first approach, and therefore we examined the reliability and validity of this approach of qualitative observational assessment.

Eight video records of each child were put in the individual set of video records in following order: two trials of the MD 1 task performed preferred and non-preferred hand (in total 4 trials), and two trials of the MD 2 and MD 3 tasks. These individual sets of video records from all children were randomly ordered so that a final set of video records for investigation included 48 video records (6 children \times 8 video records).

Investigators

Five physiotherapists, one paediatric neurologist and three academic workers in human movement science ($N = 9$) participated in the study as investigators. Their mean age was 42.7 ± 15.2 years, and experience in their profession 20.7 ± 17.1 years (range 4–48 years). The inclusive criteria for involvement in the study were: minimum four years in professional work with children, and experience in administration of the quantitative form of the MABC-2 test (minimum 15 children tested). With the exception of one, all investigators reported some experience with qualitative motor observation of the MABC-2 tasks.

Observation by the investigators

The potential investigators were informed by e-mail and phone of the aim of the study, and the course of their participation. After agreement with their involvement in the study, the co-author met each investigator to provide him/her with both verbal and written detailed information on the schedule of investigation, and instructions for qualitative observation of the MD tasks according to the Examiner's Manual for the MABC-2 test (Henderson et al., 2007). Then, video recording by video recording, the investigator practiced his/her observation and judgement whether or not the critical movement signs listed (Appendix) were involved in the execution of a given task. In the case of questionable judgement, critical moments were re-observed accompanied with the consensus of the co-author and the investigator. This practicing involved the observation of eight video records in the three MD tasks performed by one child with significant difficulties in MD (standard score 5, i.e., 5^{th} percentile). These video records were not included in the set of 48 video records determined for the investigation.

Table 1

The results of the formal quantitative assessment of manual dexterity with the tasks of the MABC-2 test (mean and SD of standard scores)

	MD ⁻	MD ⁺
MD 1	2.0 \pm 0.8	10.3 \pm 0.5
MD 2	5.0 \pm 3.7	10.7 \pm 0.5
MD 3	3.7 \pm 2.1	10.0 \pm 0.0
MD	2.3 \pm 1.2	11.0 \pm 0.0

Note. MD⁻ = children with difficulties in manual dexterity; MD⁺ = children without difficulties in manual dexterity; MD 1, MD 2, MD 3 = manual dexterity tasks (see Appendix); MD = manual dexterity component.

After the practice, each investigator received a DVD with the set of 48 video recordings and was asked to carry out qualitative observation and to mark the observed critical movement signs into the record form immediately after a child finished the last trial of a given task. They were also asked to carry out observation of all children within one or two consecutive days, without repeated observation. The investigators were informed of the age of each child, however blinded to the quantitative scores achieved in the MD tasks.

Four weeks after completing the observation of the set of video records in time 1, the investigators received a new DVD with the same 48 video records as on the first DVD, but in a different random order, to perform a re-observation (time 2). No specific instruction and training was provided the investigators before their observational assessment in the time 2.

Statistical analysis

Inter-observer reliability was tested for two variables: (i) the number of movement signs marked for each child, for each task separately, in time 1 and time 2; (ii) the total number of movement signs for each child marked in all MD tasks, in time 1 and time 2. The Friedman test and Kendall's W test were used for this purpose. As the statistical task was too large for exact tests, Monte Carlo simulation was also calculated, with a random sample size of 100,000.

To assess the intra-observer reliability of each investigator, the Wilcoxon matched-pairs signed-rank test (two-tailed) and 95% agreement limits (Bland & Altman, 1999) in time 1 and time 2 were calculated for the following variables: (i) the number of movement signs marked for each child, for each task separately; (ii) the total number of movement signs marked for each child in all MD tasks.

To assess construct (known-groups) validity of the qualitative assessment of the MD tasks, the Mann-Whitney U test (two-tailed) was used to test the difference of the following variables achieved in children

with and without difficulties in MD: (i) the number of movement signs marked for each child by each investigator, for each MD task separately; (ii) the total number of movement signs in all MD tasks for each child by each investigator. The effect size for the Mann-Whitney U test (r) was calculated according to the formula by Fritz, Morris, and Richler (2012), with interpretation as a small, medium and large effect size for .1, .3 and .5, respectively (Cohen, 1988). The analyses of construct validity were made separately for the variables obtained in time 1 and time 2.

A level of $\alpha = .05$ was set for all tests. Data analyses were conducted with IBM SPSS Statistics (Version 21 for Windows; IBM, Armonk, NY, USA).

Results

Table 1 shows the different level of the manual dexterity in the MD⁻ and MD⁺ children assessed by the formal quantitative part of the MABC-2 test. All statistical tests rejected the hypothesis on significant inter-rater reliability in the number of marked signs in both the particular MD tasks and totally in all MD tasks. The variation range of movement signs marked in the MD tasks by the nine investigators varied from 0–1 sign up to 0–6 signs in particular children (Table 2).

Regarding the intra-rater reliability, the number of marked signs in the particular MD tasks and in all MD tasks, there were not significant differences between observation in time 1 and time 2 in all investigators (Table 3). The only exception was the significant difference in the total number of signs in the MD tasks marked by the investigator I4 (Table 3). Mean difference (M_{diff}) of the number of movement signs marked in time 1 and time 2 in the individual investigators ranged between -0.50 and 0.83 in the MD 1 task, -0.83 to 0.67 in the MD 2 task, -0.17 to 1.00 in the MD 3 task, and -0.50 and 0.78 in overall MD tasks (Table 3). 95% agreement limits for the number of marked movement

Table 2
The minimum and maximum range, and mean interquartile range of the movement signs per child marked by a group of investigators (N = 9)

	Time 1			Time 2		
	Minimum	Maximum	IQR-M	Minimum	Maximum	IQR-M
MD 1	0–2	0–6	1.5	0–1	0–6	1.3
MD 2	0–2	1–6	1.3	0–1	1–6	1.2
MD 3	0–3	0–5	2.8	0–3	0–5	2.5
MD	0–5	0–11	3.0	0–6	0–11	4.3

Note. Minimum, maximum = minimum and maximum range; IQR-M = mean interquartile range; MD 1, MD 2, MD 3 = manual dexterity tasks (see Appendix); MD = manual dexterity component.

Table 3

The intra-observer reliability of the qualitative observational assessment of movements in the manual dexterity tasks: A significance of difference (Wilcoxon matched-pairs signed-rank test) and the 95% limits of agreement in the number of movement signs marked in repeated two observations

Investigator	MD 1			MD 2			MD 3			MD		
	Z	p	95% LoA	Z	p	95% LoA	Z	p	95% LoA	Z	p	95% LoA
I1	1.342	.180	0.33 ± 0.92	1.342	.180	-0.67 ± 2.18	1.342	.180	0.33 ± 0.92	0.943	.345	-0.22 ± 1.67
I2	0.000	> .99	0	1.604	.109	-0.83 ± 1.76	0.447	.655	0.17 ± 1.76	0.944	.345	-0.22 ± 1.67
I3	1.095	.273	-0.50 ± 1.88	1.604	.109	-0.83 ± 2.10	0.270	.787	-0.17 ± 2.63	1.569	.117	-0.50 ± 2.29
I4	1.604	.108	0.83 ± 1.76	1.604	.109	0.67 ± 1.47	1.483	.138	0.83 ± 2.10	2.623	.009	0.78 ± 1.80
I5	1.095	.273	-0.50 ± 1.88	0.000	> .99	0.00 ± 1.14	0.913	.361	0.33 ± 1.47	0.255	.799	-0.06 ± 1.67
I6	0.000	> .99	0	0.802	.423	-0.30 ± 1.84	0.270	.787	-0.17 ± 2.63	0.628	.530	-0.17 ± 2.10
I7	1.826	.068	0.67 ± 0.92	0.000	> .99	0.00 ± 1.61	0.270	.787	-0.17 ± 2.63	0.889	.374	0.17 ± 1.35
I8	0.816	.421	0.17 ± 0.73	1.095	.273	0.50 ± 1.88	0.000	> .99	0.00 ± 1.96	0.980	.327	0.22 ± 1.67
I9	0.000	> .99	0	0.000	> .99	0	0.000	> .99	1.00 ± 2.10	0.000	> .99	0.00 ± 1.31

Note. MD 1, MD 2, MD 3 = manual dexterity tasks (see Appendix); MD = manual dexterity component; Z = standardized value of the *W* statistic; *p* = *p*-value (two-tailed); 95% LoA = 95% limits of agreement (presented as mean difference ± SD of differences).

signs in the two repeated observations ranged from $M_{diff} \pm 0.73$ up to $M_{diff} \pm 2.63$ signs (without involvement of absolute agreement of signs of the MD 1 task marked by three investigators) (Table 3).

As compared to the MD⁺ children, the number of movement signs marked in the MD⁻ children was significantly higher in the MD 1 and MD 2 tasks, as well as the total number of marked signs in all MD tasks (Figure 1). This type of significant difference was not found in observation of the execution of the MD 3 task. These findings were the same for observation in time 1 and time 2.

Discussion

The clinical meaning of inter-observer reliability consists in the interchangeability of observers (Gwet, 2012). In the case of the qualitative observational assessment of an execution of MABC-2 test tasks, the inter-observer reliability concerns the extent to which observers agree with each other in the nominal coding of movement signs (manifested – not manifested). The results of the study showed that observers can differ reasonably in their judgement of whether or not the particular critical movement signs occurred during execution of the MD tasks. Specifically, the investigators differ in several children by even five or six signs marked per child in particular MD tasks and by eleven signs in total across all MD tasks (see variation ranges in Table 2). Wider differences between the investigators were found in the MD⁻ children rather than in MD⁺ children.

After the finding of non-significant inter-rater reliability of observational assessment, we performed a supplementary analysis to examine whether this finding might be affected by the age and the length of professional experience of the investigators. For the purpose, we compared the mean total number of movement signs marked in all MD tasks in each child between: (i) the groups of four youngest and four oldest investigators; (ii) the groups of four investigators with the shortest and four investigators with longest length of professional experience. The small effects of both the age and length of professional experience on observational assessment were found (Cohen's *d* = 0.11–0.45 and 0.14–0.46, respectively). To increase validity of the observational assessment in clinical practice, two investigators could be recommended to observe a child during performing of the MD tasks.

In contrast to inter-observer reliability, intra-observer reliability of the qualitative assessment of the MD tasks showed itself to be very acceptable. Individual investigators observed the tasks with non-significant

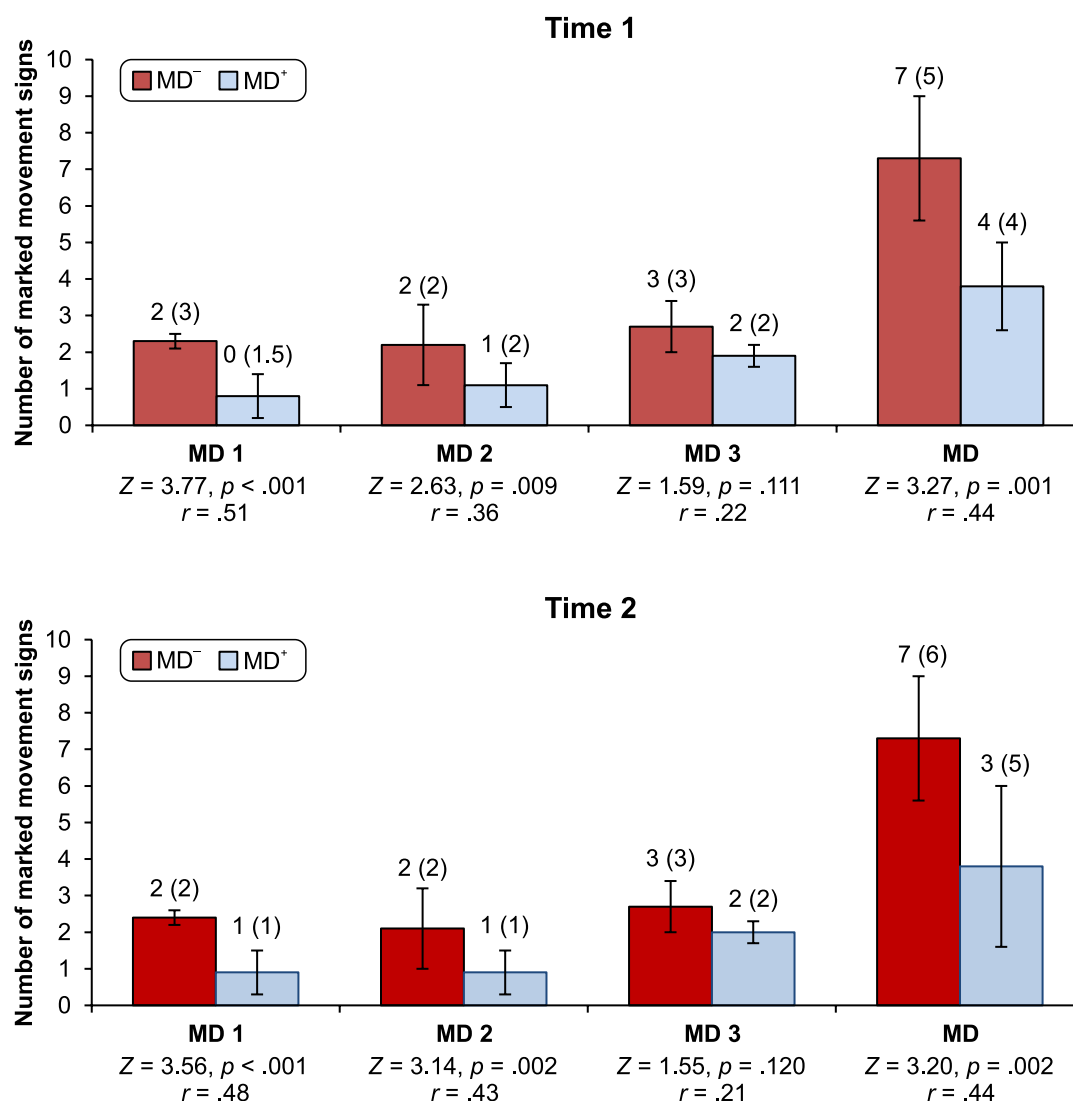


Figure 1. The number of movement signs marked by investigators in children with and without difficulties (MD⁻ and MD⁺) in manual dexterity, separately in Time 1 (observation) and Time 2 (re-observation). Values above the bars – median (IQR); MD 1, MD 2, MD 3 = the manual dexterity tasks; MD = sum of the MD tasks; Z = standardized value of U statistic; p = p-value (two-tailed, $\alpha = .05$); r = effect size.

differences in the number of marked movement signs in the repeated observations, i.e. in time 1 and time 2. For practical purposes, intra-observer reliability of qualitative observation of body movements should be quantified as a degree of agreement. Therefore, we calculated the 95% agreement limits according to the equation by Bland and Altman (1999). This measure expresses test-retest repeatability that is calculated as $M_{diff} \pm SD_{diff} \cdot .96$, where M_{diff} is the mean of differences between the test and retest scores, and SD_{diff} is the standard deviation of these differences. Most of the 95% agreement limits (in 75% of calculations of 95% agreement limits; Table 3) showed an error of repeated observations in the individual investigators ranged from 0 to ± 2.00 signs around the M_{diff} that ranged from

0 to 0.83 (with the exception of a 1.00 sign in one case; Table 3). The higher intra-rater than inter-rater reliability was also reported for qualitative assessment of manual dexterity and diadochokinesis of Maastricht's Motor Test in 5–6-year-old children (Kroes et al., 2004).

Observation of the critical movement signs in the graphomotor MD 3 task seemed to be more difficult for investigators compared to the MD 1 and MD 2 tasks. This tendency was obvious with the SD_{diff} of repeated observations of the MD 3 task being higher than the ± 2.00 signs in five investigators (Table 3). Also, in observation of the MD 3 task, the investigators differed between each other almost twice as much as in observation of the MD 2 and MD 3 tasks, as the difference of the mean IQR of movement signs marked for

the MD 3 task across all children shows (see IQR-M in Table 2). One possible explanation may be that slight movements of the drawing hand of a child might not be easily recognized from a distance of 1.5 m for an observer, in contrast to the MD 1 and MD 2 tasks with more noticeable, discrete movements of the arm, hand and fingers.

The investigators identified almost twice as many critical movement signs in the MD 1 and MD 2 tasks, and totally in all MD tasks in children with poor manual dexterity as compared to children with normal manual dexterity, with significant difference and medium effect (Figure 1). These findings support the construct (known-groups) validity of the method of qualitative assessment in the MABC-2 test. The critical movement signs listed for observation of the MD tasks are related to perceptual-motor aspects such as body control/posture, functioning of limbs, spatial accuracy, control of force/effort, and timing of actions (see Henderson et al., 2007). A higher number of observed movement signs in the test tasks allow the examiner to reach a more valid conclusion about perceptual-motor deficits of a child.

In contrast to the above-mentioned results, observation of the graphomotor task MD 3 did not prove appropriate construct validity. What could be a possible explanation? Firstly, lower intra-observer reliability of observation of this task (see above) may constrain its validity. It is known that the assessment tool that is not reliable enough cannot be valid (Burton & Miller, 1998). Secondly, a mean number of movement signs in the MD 3 task marked in MD⁺ children was obviously higher than in the MD 1 and MD 2 tasks and was a little closer to the number of marked signs observed in MD⁻ children. For school-age children, handwriting and drawing is a daily activity and an important part of learning (Sugden & Wade, 2013). One can consider that these sensorimotor actions are more real-life-related activities for children than the MD 1 and MD 2 tasks. As a consequence, the investigators could adopt a more critical view on the qualitative assessment of the MD 3 task.

The limitation of the study

We see three major limitations of the study. Firstly, there is not a method how to check whether observational assessments by the investigators were made according the rules instructed by the authors of the study (see Methods). Secondly, the finding of construct known-groups validity of the qualitative observational assessment of the MD tasks could be affected by a small sample size of children. Thirdly, conclusions on validity and reliability of the qualitative assessment are derived from the observation of video recordings.

Video recordings are two-dimensional representation of a three-dimensional real scene. Therefore, validity and reliability of observation of the MD tasks could be also verified in real situations in the next study.

Conclusions

This study is the first to provide information on the psychometric qualities of the qualitative observational assessment of movement patterns in the MABC-2 test. The study showed good construct validity of the qualitative assessment of manual dexterity in younger school-age children. Thus, the critical movement signs used for observation seem to be valid clinical symptoms of manual dexterity impairment. Intra-observer reliability of the qualitative assessment of manual dexterity in children was good in contrast to non-significant agreement among observers. These findings suggest that qualitative diagnosis of manual dexterity impairment in the MABC-2 test demands specific training of examiners in their observation. The qualitative assessment within the MABC-2 test is a useful tool for the identification of manual dexterity impairment and weaker aspects of hand-eye coordination, on the basis of which an intervention can be planned and its effects evaluated.

Acknowledgments

The study was funded by the Czech Science Foundation (GAČR) under the project No. 16-17945S.

Conflict of interest

There were no conflicts of interest.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135–160.
- Brickenkamp, R., & Zillmer, E. (2000). *d2 test of attention*. (Czech version by P. Balcar). Prague, Czech Republic: Hogrefe-Testcentrum.
- Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Champaign, IL: Human Kinetics.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18.
- Gard, L., & Rösblad, B. (2009). The qualitative motor observation in Movement ABC: Aspects of reliability and validity. *Advances in Physiotherapy*, 11, 51–57.
- Gwet, K. L. (2012). *Handbook of inter-rater reliability* (3rd ed.). Gaithersburg, MD: Advanced Analytics.
- Henderson, S. E., & Sugden, D. A. (1992). *Movement Assessment Battery for Children*. San Antonio, TX: The Psychological Corporation.
- Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2007). *Movement Assessment Battery for Children-2*. London, United Kingdom: Pearson Education.
- Kroes, M., Vissers, Y. L. J., Sleijpen, F. A. M., Feron, F. J. M., Kessels, A. G. H., Bakker, E., ... Vles, J. S. H. (2004). Reliability and validity of a qualitative and quantitative motor test for 5- to 6-year-old children. *European Journal of Paediatric Neurology*, 8, 135–143.
- Psotta, R. (2014). *MABC-2. Test motoriky pro děti MABC-2* (1. české vydání) [The Movement Assessment Battery for Children-2 Test (1st Czech Edition)]. Prague, Czech Republic: Hogrefe Testcentrum.
- Schulz, J., Henderson, S. E., Sugden, D. A., & Barnett, A. (2011). Structural validity of the Movement ABC-2 test: Factor structure comparisons across three age groups. *Research in Developmental Disabilities*, 32, 1361–1369.
- Sim, J., & Arnell, P. (1993). Measurement validity in physical therapy research. *Physical Therapy*, 73, 102–110.
- Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, 17, 396–408.
- Sugden, D. A., & Henderson, S. E. (2007). *Ecological intervention for children with movement difficulties*. London, United Kingdom: Pearson Education.
- Sugden, D., & Wade, M. (2013). *Typical and atypical motor development*. London, United Kingdom: Mac Keith Press.
- Wagner, M. O., Kastner, J., Petermann, F., & Bös, K. (2011). Factorial validity of the Movement Assessment Battery for Children-2 (age band 2). *Research in Developmental Disabilities*, 32, 674–680.

Appendix: The critical movement signs for the qualitative observational assessment

Manual dexterity task 1 (MD 1): Placing pegs

Posture/body control

- Sitting posture is poor
- Holds head too close to task
- Holds head at an odd angle
- Does not look at board while inserting pegs
- Does not use pincer grip to pick up pegs
- Exaggerates finger movements in releasing pegs
- Does not use the supporting hand to hold board steady
- Does extremely poorly with one hand (asymmetry striking)
- Changes hands or uses both hands during a trial
- Hand movements are jerky
- Moves constantly/fidgets

Adjustment to task requirements

- Misaligns pegs with respect to holes
- Uses excessive force when inserting pegs
- Is exceptionally slow/does not change speed from trial to trial
- Goes too fast for accuracy

Manual dexterity task 2 (MD 2): Threading lace

Posture/body control

- Sitting posture is poor
- Holds materials too close to face
- Holds head at an odd angle
- Does not look at board while inserting tip of lace
- Does not use pincer grip to hold lace
- Holds lace too far from tip
- Holds lace too near tip
- Finds it difficult to push tip with one hand and pull it through with the other
- Changes threading hands during a trial
- Hand movements are jerky
- Moves constantly/fidgets

Adjustment to task requirements

- Sometimes misses hole with tip of lace
- Gets muddled in the threading sequence
- Is exceptionally slow/does not change speed from trial to trial
- Goes too fast for accuracy

Manual dexterity task 3 (MD 3): Drawing trial 2

Posture/body control

- Sitting posture is poor
- Holds head to near paper
- Holds head at an odd angle
- Does not look at trail
- Holds pen with an odd/immature grip
- Holds pen too far from point
- Holds pen too close to point
- Does not hold paper still
- Changes hands during a trial
- Moves constantly/fidgets

Adjustment to task requirements

- Progresses in short jerky movements
- Uses excessive force, presses very hard on paper
- Is exceptionally slow
- Goes too fast for accuracy