RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

Reliability of Movement Assessment Battery for Children, Age Band 2 (Second Edition): Exploratory Study

Eryk Przysucha[1], Alexandra Boyle[2], Carlos Zerpa[3]
[1,2,3,] Lakehead University

Abstract:
Objective: The purpose was to explore the reliability of the Total Test Score (TTS) and three subsections of age band 2 of MABC-2 using test re-test and internal consistency measures.
Patients and Methods: Forty typically functioning children (18 boys, 22 girls) (M = 9 years, 2 months, SD = 1 year, 3 months) were assessed twice, two weeks apart, by the same researcher in the same laboratory setting. Intra-class correlation (ICC) coefficient examined the test-retest stability whereas Cronbach's alpha was used to examine the internal consistency of the items.
Results: The normality, skewness and kurtosis of standard, component and percentile scores were examined first. Among the three scoring systems, the standard scores met the majority of the assumptions tested. Based on these scores, the test-retest reliability revealed questionable stability for the TTS (ICC = .67), with balance exhibiting the poorest reliability across the three subsections (ICC = .56). The analysis of the internal consistency revealed a similar trend with aiming and catching exhibiting the lowest Cronbach alpha (.49)
Conclusion: From the clinical perspective, caution is warranted when component and percentile scores are used. In terms of reliability, while the TTS was most stable, the analyses confirmed previous findings suggesting that overall the MABC-2 is not a reliable assessment tool for children between the ages of 7 to 10. Due to these findings, when using this age band the child should be retested and/or the inferences emerging from this test should be triangulated with other formal assessment tests.

Key words: reliability; measurement error; assessment; MABC

INTRODUCTION

One important role of an adapted professional is to carry out reliable and valid assessments. Since the latter cannot be accomplished without the former, reliability represents a key concept for persons working with atypically functioning individuals. Measurement error, hence lack of reliability, is inevitable when assessing social, psychological and physical attributes of persons. However, when the amount of measurement error (systematic or random) is substantial, the conclusions drawn from the tests and coinciding scores have to be treated with caution or disregarded altogether. They may lead to false positive or negative inferences, which in turn may have detrimental consequences for the client [1].

The Classical True Score Theory (CTT), which is one of many theories of measurement (e.g. Generalizability Theory, Item-Response Theory), is able to estimate how much error is within the measurement [2, 3]. The hypothetical true ability (score) of the person is difficult if not impossible to capture via any assessment tool. One way of estimating this (true score) would be by taking the mean of many trials performed by the same individual, under the same circumstances, and on the same test. Unfortunately, this method is impractical. Thus, according to CTT, if the amount of measurement error is small one can be confident that the emerging observed score approximates the true ability of the person. The degree of this confidence can be wagered based on the magnitude of the reliability coefficient which is often derived via test-retest or internal consistency approaches. The former infers the consistency/stability of repeated performances that are separated in time and measured by the same examiner under the same conditions [4]. When the emerging correlation coefficient is around 0.8 generally this would indicate a good reliability, any value between 0.7 and 0.8 is deemed as moderate, and a value below .7 is considered as questionable or week. The internal consistency, on the other hand, assesses how well the items of a test or instrument measure a specific construct [4]. If the individual items within a set of tasks are highly correlated that would indicate that the items represent the same construct. This form of reliability only requires one test administration and the estimations are often inferred from the magnitude of coefficient alpha. One potential issue which directly affects the degree of internal consistency estimate is the length of a test and number of items within the test. A longer test will generally provide higher values of Cronbach's alpha than a shorter test due to the higher degree of variance which is indispensable when correlations are considered [4, 5].

In relation to the assessment, the choice of a particular test depends on many factors such as its purpose, population of

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

interest, theoretical framework and its psychometric properties [1]. For decades Movement Assessment Battery for Children (MABC) [6, 7] has been considered as a gold standard in the area of adapted physical activity as related to the assessment of children with non-congenital, developmental coordination problems [6]. The MABC is a standardized, norm-referenced test which is used for the purpose of screening, identifying and diagnosing mild to moderate movement problems in children [7]. The test involves completion of three subsections involving performance of manual dexterity, ball skills, and balance tasks. These scores when accumulated constitute the Total Impairment Score (TIS), which reflects the overall motor performance of a child. Recently, a revised version of MABC has been released (Movement Assessment Battery for Children – Second Edition) [8]. The "kit is easier to carry, the performance test items are more engaging for children, and the scoring system for both the performance test and checklist are more user-friendly" (p. 92) [7]. Also, the new version encompasses a broader age range (3 to 17 years of age) than the previous test (4 to 12 years of age), and the number of age bands has been reduced from four to three (3 to 7; 7 to 10 and 11 to 17), resulting in different number of items and new equipment. In regards to age band 2, which is off primary interest here, in the manual dexterity subsection, the placing pegs task now has a new starting position and layout. The lacing board is longer for the threading lace activity, and the shape of the drawing trail has changed. For the aiming and catching section, the beanbag task a box was replaced with a target, whereas in the balance subsection floor mats were added for the one leg hopping activity. Despite these numerous changes, the authors maintained that the research pertaining to the reliability and validity of the original tool applies to the new version. Some researchers [7, 8] in the field remained skeptical in regards to this assumption and called for more investigations examining the psychometric properties of the new version.

Most of the existing research involving MABC-2 pertains to age band 1 (3 to 7 years old) [9, 10, 11, 12]. Collectively, these studies showed that the reliability of the TTS and the individual subsections was satisfactory as inferred from test-retest approach. In contrast, the internal consistency, as inferred from Cronbach's alpha, was substantially lower across the studies. Among the items, drawing trail, walking heels raised, and balance appeared to have the lowest reliability. In terms of age-band 2, only one study examined this age group explicitly [13], although Wuang and colleagues [14] examined the reliability of this and other age-groups combined. The study by Holme and colleagues [12] examined reliability using intra and inter-raterreliability coefficients. Forty-five typically functioning children, 23 girls and 22 boys, with a mean age of 8.7 years were recruited. When the children attended the first testing session, they were tested twice by two physiotherapists, who scored them independently (inter-rater). On the second testing session, one to two weeks later, the children were re-assessed by one of the examiners (intra-rater). The analysis of the component scores showed that there was a lack of reliability for the sub-sections as well as the TTS. Intra-rater reliability had ICC values of 0.62, 0.49, and 0.49 for manual dexterity, aiming and catching, and balance respectively. The ICC for the TTS was also low (0.68), and the SEM for the different scores were high (2.4 to 4.9). The analysis of inter-ratter reliability, also calculated based on the component scores, had ICC values of 0.63, 0.77 and 0.29 for manual dexterity, aiming and catching, and balance. As well, the ICC for the TTS was 0.62, indicating a questionable degree of reliability. Among the individual items, the threading lace and one board balance tasks were the least reliable. Overall, this study showed that across the different scores the reliability of the test was moderate to questionable. Also, it revealed that certain tasks within the MABC-2 (e.g. tasks within the balance sub-component) maybe problematic as they are too difficult or not challenging enough. The important limitation of this research was that the researchers only used inter and intra-rater reliability, thus the internal consistency of the MABC-2 was not examined. As well, the analysis involved the component scores alone and did not address the reliability of the other scores (standard & percentile), which are often used in research and clinical settings. Thus the purpose of this study was to examine different facets of reliability (test-retest & internal consistency) of the age band 2 (7 to 10 years old) across standard, component, and percentile scores for the three subsection and TTS scores. It was hypothesized that test re-test reliability (ICC) for the TTS would be moderate to high (> 0.70), and it would be higher when compared to the three sub-sections (manual dexterity, aiming and catching, and balance). Also, it was expected that the internal consistency (Cronbach's alpha) would be moderate (> 0.70) for manual dexterity and aiming and catching sub-sections, but it would be questionable (< 0.60) for the balance sub-section.

PARTICIPANTS AND METHODS

Participants

Forty participants (18 males and 22 females) between 7 and 10 years of age (ME = 9 years, 0 months and 5 days) were

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

recruited for this study. In order to be included in this study, children were required to be typically functioning in terms of their motor and cognitive status, as reported by the parents/guardians. Atypically functioning children with an official diagnosis for any developmental disabilities in the cognitive or motor domains were excluded from the study.

Purposive sampling was implemented. The children were recruited from local elementary schools with the permission of the respective school boards in the region. The information packages were delivered to the teachers of the appropriate grades. The recruitment package contained a recruitment letter, consent form, and an ExPARA (Exercise and Physical Activity Readiness Assessment). Consent forms required the parents' signature as the children were all under the age of 18. When the forms were returned, parents were contacted via phone or email to set up the testing sessions. Prior to the data collection, participants were given a brief description of the study. The children and parents were all informed that participation was voluntary and that all data would remain confidential.

DESIGN AND PROCEDURE

Participants were asked to commit to two sessions, one to two weeks apart. Children were assessed individually and each session took approximately 45 minutes to one hour. At the second testing session, the child was re-assessed at the same location, under the same conditions, and by the same examiner.

There are 8 different tasks in the MABC-2 test for age band [8] which are divided into three sub-sections; manual dexterity, ball skills and balance. The tasks within each section are safe, relatively simple and resemble activities that a child performs on a daily basis. There are 3 manual dexterity tasks (placing pegs, threading lace, and the drawing trail), 2 ball skills tasks (catching with two hands and throwing beanbags onto a mat), and 3 balance tasks (one board balance, walking heel-to-toe, and hopping on mats). For each child, a Total Test Score (TTS) and three sub-section scores were derived as raw scores and were converted to standard, component, and percentile scores, respectively.

MEASURES, DESIGN AND ANALYSIS

The reliability of the MABC-2 was analyzed using ICC correlation coefficient [4, 16] for the test-retest approach, whereas Cronbach's alpha [4, 17] was used to infer internal consistency for the three types of scores (standard, component, and percentile). A two-way mixed model with absolute agreement was implemented for the calculation of ICC. An analysis of variance (dependent samples t-tests) was also implemented to determine if there were statistically significant differences between the group means across the two sessions. Cronbach alpha was used to infer the internal consistency of the items across manual dexterity, aiming and catching, and balance subsections. The scores from time one were used. Also, the Cronbach's alpha with items deleted was implemented [17]. This allows inferring if the internal consistency of the sub-component improves when a certain item is removed. If so, this would indicate that that specific task is not measuring the same domain as the remaining items. If the Cronbach's alpha stays the same or goes down, while the item is deleted, this indicates that the item enhances the internal consistency of the sub-component. The aiming and catching sub-component is composed of only two tasks therefore the data was not further analyzed with items deleted approach.

Prior to the analyses of reliability, the standard, component, and percentile data sets were examined to verify if they met the necessary assumptions. To test if the data sets were normally distributed, a Shapiro-Wilks test was implemented. The skewness and kurtosis of each data set were also examined. The homoscedasticity of the data was inferred from the respective scatterplots for each coefficient.

RESULTS

Tests of Assumptions

The results from the tests of normality across all three types of scores were mixed and somewhat equivocal (Table 1). The standard scores exhibited the desired characteristics of normally distributed data. In fact, all 8 demonstrated minimal skewness and kurtosis, and 5 out of the 8 data sets were normally distributed, according to the Shapiro-Wilks test. On the other hand, the component scores demonstrated skewness for 6 out of the 8 areas, thus indicating that these sub-components were not normally distributed. The percentile scores met the normality assumptions based on the skewness and kurtosis tests; however only 2 out of the 8 areas were normally distributed according to Shapiro-Wilks test. When the data is not normally distributed, the inferences need to be treated with caution. In terms of homo/heteroscedasticity, when the data is heteroscedastic "individuals who score the highest values on a test also show the greatest amount of measurement error … and smallest changes in responses" [4]. Therefore it may be difficult to identify these small changes in participants that

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

are performing the best, even though these changes might allow for the detection of measurement error. Collectively, the analysis for the TTS and three sub-components revealed that the standard scores exhibited the characteristics of normally distributed data most frequently among the three other scores. Consequently, the results and discussion sections focused on the data pertaining to the standard scores alone.

Table 1.

Number of Domains (out of 8) That Met The Three Normality Assumptions Across First (T1) and Second (T2) Testing Sessions.

	Type of Score		
	Standard Scores	Component Scores	Percentile Scores
Normality	5/8	5/8	2/8
Skewness	8/8	2/8	7/8
Kurtosis	8/8	7/8	8/8

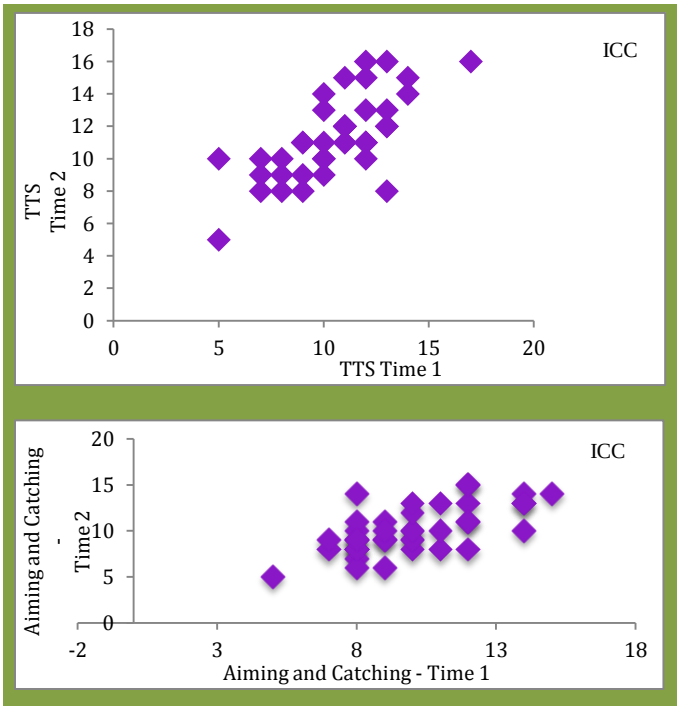
TEST RE-TEST.

Total Test Score.

The analysis of the standard scores did not support the hypothesis regarding the TTS. The ICC coefficient (0.67) indicated that this score only approached the expected degree of reliability. The additional analysis of variance (t-test) confirmed these inconsistencies as scores at time one (M =10.55, SD = 2.49) were significantly lower as compared to time two (M = 11.53, SD = 2.53) (t (39) = -2.53, p < 0.05). As evident from the scatterplot capturing the correlation coefficient for the TTS (Figure 1, top graph), the data set was homoscedastic, but it did demonstrate a restricted range.

SUB-COMPONENT SCORES.

In line with TTS the correlation coefficients pertaining to the sub-scores revealed a moderate degree of consistency. The analysis of the manual dexterity revealed an ICC = .68, which was supported by the t-tests revealing a statistically significant difference between time one (*M = 9.85, SD = 3.10*) and time two (*M = 10.70, SD = 3.32*) (*t* (39) = -2.14, p < 0.05). The scatterplot (Figure 1, second graph from the top) showed the data had an expected amount of variance around the hypothetical line of best fit. The aiming and catching sub-component also revealed a moderate degree of reliability (ICC = .65), however no differences were found between time one (*M = 9.95, SD = 2.31*) and two (*M = 10.13, SD = 2.54*) (*t* (39) = -0.54, p = 0.59). The respective scatterplot (Figure 1, third graph from the top) showed that the data points were more homogeneous compared to the manual dexterity domain. Similarly to aiming and catching, also the balance domain revealed a moderate degree of consistency (ICC = .56), however this was accompanied by significant differences between the two sessions (*M = 11.55, SD = 2.57* vs. *M = 14.10, SD = 2.18*) (*t* (39) = -2.09, p < 0.05). The review of the corresponding scatterplot (Figure 1, bottom graph) demonstrated a lack of variability within the data set for this domain, which may have contributed to the relatively moderate correlation coefficient. It is important to note that all the scatterplots (Figure 1) showed less than forty data points. This indicates that some children achieved the same scores across the sessions and this could be attributed to floor or ceiling effects.



RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

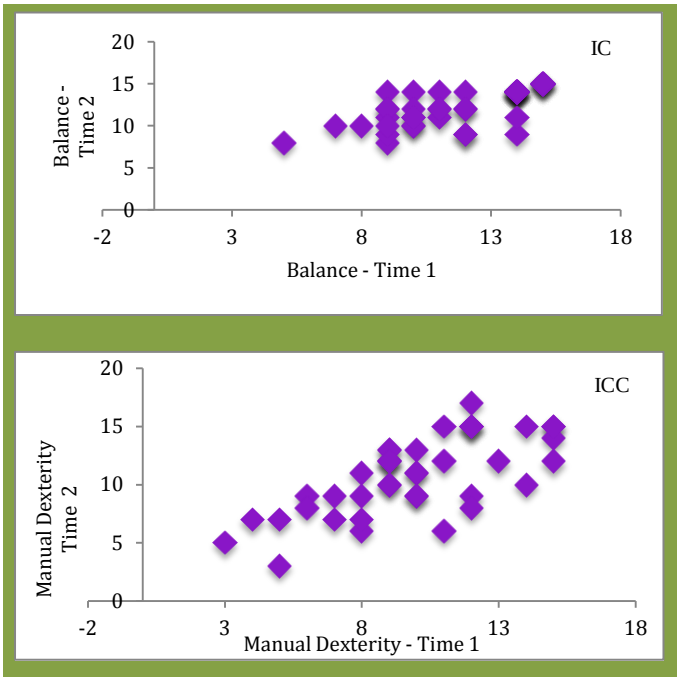


Figure 1.Intraclass correlation coefficients and respective scatterplots for standard scores for total test score (top), manual dexterity (second from the top), aiming and catching (third from the top), and balance sub-components (bottom).

Internal Consistency.

There were two different testing sessions (T1 and T2) that could be used for the analysis of internal consistency. As the analysis of normality assumptions provided similar results for both times, the date from time one was chosen to ensure that the memory and/or practice effect did not affect the inferences. The analysis of the internal consistency of manual dexterity, aiming and catching, and balance was carried out on the item standard scores, which were derived from the raw scores from the MABC-2.

MANUAL DEXTERITY.

The results did not support the hypothesis, as Cronbach's alpha was 0.61, representing questionable internal consistency for this sub-component. Further analysis was implemented using Cronbach's alpha with items deleted. As the coefficient did not increase when any of the three items were deleted, this indicated that not one specific task jeopardized the internal consistency of this set of items (Table 2).

AIMING AND CATCHING.

Based on previous results, it was hypothesized that this sub-component would have the highest internal consistency, as compared to manual dexterity and balance. The data did not support this hypothesis as the aiming and catching domain had a Cronbach's alpha of 0.49, indicating a questionable internal consistency. Cronbach's alpha with items deleted could not be calculated as there are only two tasks within this sub-section.

BALANCE.

As it was the case with the two previous sub-scores, the findings failed to support the hypothesis as the Cronbach alpha was 0.53. When items were deleted (Table 2), the Cronbach's alpha value did not increase for any of the three items, thus once again indicating that not one specific item caused the questionable internal consistency.

Table 2
Cronbach's Alpha with Items Deleted for the Item Standard Scores for Time One.

Item	Cronbach's Alpha if Deleted
Manual Dexterity – Cronbach's Alpha = 0.61	
Manual Dexterity 1	0.54
Manual Dexterity 2	0.54
Manual Dexterity 3	0.41
Aiming and Catching – Cronbach's Alpha = 0.49	
Aiming and Catching 1	-
Aiming and Catching 2	-
Balance – Cronbach's Alpha = 0.53	
Balance 1	0.51
Balance 2	0.41
Balance 3	0.41

DISCUSSION

The research regarding the psychometric properties of the revised version of MABC-2 is limited, particularly in regards to age band 2. As a result, the aim of this study was to examine more comprehensively the reliability of this age

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

band (7 to 10 years old) for TTS and the three subsections. As the statistical assumptions pertaining to component and percentile scores were not met, the subsequent discussion pertains to standard scores alone.

TEST RE-TEST

Total Test Score. The analysis of the TTS revealed an ICC of 0.67 when test re-test reliability was examined, thus indicating a questionable degree of reliability. The analysis of the scatterplots showed that the weak relationship could be due to the restricted range of the data, which may have caused the data to be homoscedastic. As the use of correlations alone can be problematic to assess the degree of systematic bias, particularly when the inter-individual variability is low, the analysis of variance was also carried out. In line with the correlation, the data showed significant differences between time one and time two. In fact only 9 out of the 40 participants achieved the same standard score across the two trials (Figure 1, top graph). This indicated that there is a lack of consistency across performances in more than three quarters of the individuals. To date, there has been only one study which examined the reliability of the age band 2, even if it did involve the component scores. Holm and colleagues [14] examined intra- and inter-rater reliability based on component scores and reported ICC values of 0.68 and 0.62, respectively. These findings are similar to those reported here. The sample size and characteristics of the participants were also comparable between the studies. Thus, the emerging results appear to be robust even though different reliability coefficients were examined. In regards to studies examining this issue, but with a different age band, Ellinoudis and colleagues [10] (ICC = 0.85) and Wuang and colleagues [15] (ICC = .97) both reported substantially higher coefficients for the TTS of age band 1. The discrepancies could be attributed to many factors, aside from the plausibility that this age band is less reliable as compared to the others. The former study [10] incorporated 183 children, whereas the study by Wuang et al., [15] involved a sample of 144 atypically functioning individuals. In both cases the inter-individual variability present in the sample may have contributed to larger ICC values. In terms of the research examining the test re-test reliability based on the original version of MABC, Chow and Henderson [18] found that the TTS had a moderate reliability (0.77). However, their sample size was also larger (75 participants), and the study was conducted on age band 1 (4 to 6 years). Also, the authors did not indicate which type of score was used (standard, component, or percentile) for the calculation of ICC, thus a caution is warranted when comparing the results from the two studies.

Sub-Components. The analysis of manual dexterity revealed the ICC of 0.68, which once again indicated poor reliability. The analysis of the scatterplot (Figure 1) showed that the data appeared to be normally distributed around the hypothetical line of best fit, thus suggesting that factors other than restricted range or outliers contributed to low correlation value. The lack of consistency was also supported by the analysis of variance, which showed that there were statistically significant differences between the two testing sessions. The analysis of the individual data across both sessions two also showed that 25 of the 40 participants scored higher on time two suggesting that some systematic and/or random bias emerged. The standard deviation also increased between the sessions indicating that as the group the sample was more variable despite that they were assessed with the same tool for the second time. In terms of previous research, the present results are higher than those reported in the past studies. Holm and colleagues (2013) showed that age band 2 had ICC values of 0.62 (intra-rater) and 0.63 (inter-rater) for the manual dexterity sub-component. However, once again when compared to the other age bands the results reported here and in the past studies were substantially lower. Ellinoudis et al., [10] revealed an ICC of 0.82 and Wuang and colleagues [15] reported an ICC of 0.97 for the test re-test reliability, for age band 1. No studies were conducted on age band 3, or the original MABC, that examined the test re-test reliability for manual dexterity.

In terms of aiming and catching, the ICC for the test re-test reliability was 0.65, thus indicating a questionable consistency. The analysis of variance showed that there were no statistically significant differences between time and time two. Nevertheless, the individual data showed that only 9 of the 40 participants had the same standard score across both testing occasions, thus from the standpoint of absolute reliability little consistency was evident. The analysis of the scatterplot also showed that the low degree of reliability could be due to the restricted range of the data, as the tasks were too easy for the individuals resulting in a ceiling effect. The present results are to some extend comparable to those reported by Holm and colleagues [14] who showed that the ICC values ranged between 0.49 and 0.77 for intra- and inter-rater reliability, respectively. In regards to previous research, with other age bands, similar results (ICC = .61) were reported by Ellinoudis and colleagues [10] despite the fact that their sample was much

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

larger than the current data set. This value was much lower than the ICC findings for the TTS and the other sub-components. This was in line with the present study, where the aiming and catching sub-component also had the lowest coefficient as compared to the other domains. Another potential reason for the lower reliability coefficient may be the fact that the aiming and catching is comprised of only two tasks (two handed catch &throwing a bean bag on to a target).

The analysis of the balance sub-component also revealed a questionable degree of reliability. The analysis of the scatterplot (Figure 1, bottom graph) demonstrated a restricted range as there were less than forty data points. This could be due to the fact that tasks such as one-board balance were too challenging to most of the children. In fact, 50% of participants achieved the same score across trials, in comparison to the other two sub-components and the TTS. Thus, the individuals were scoring consistently but poorly across the both sessions, leading to small intra-group variability. The low reliability found here is consistent with the coefficients (.49 & .29) reported in previous research, for age band 2[14]. The authors of that research also supported the notion that low ICC values for this subsection may be due to the one board balance task as evident from high SEM. In regards to other research, Ellinoudis and colleagues [10] reported an ICC value of 0.75 for age band 1. However, it should be noted that the nature of the tasks included is different across the age bands. For example, in age band 1 the participants are required to complete 3 tasks including one leg balance, walking with heels raised and jumping on mats. Although the latter two items are comparable to those in age band 2, the one leg balance task would be considerably easier than the one-board balance task. Therefore, this task likely contributed to a restricted range in age band 2. Furthermore, the difference in the strength of the reliability coefficient, as compared to the present study, could be due to the size and nature of the sample. Ellinoudis and colleagues [10] assessed 183 children, between 3 and 6 years of age, which is a substantially larger and younger sample as compared to the one used in this study. Also, Wuang and colleagues [15] reported an ICC of 0.97, while testing atypically functioning individuals. Thus, a larger and more heterogeneous sample size likely contributes to stronger reliability coefficients, regardless if the tasks/or performances are actually more stable/reliable [15].

INTERNAL CONSISTENCY

Manual Dexterity.The Cronbach's alpha for manual dexterity was 0.61 indicating a questionable reliability. The low internal consistency could be due to the fact that one of the three items is not measuring the construct of interest, therefore lowering alpha. For example, the placing pegs and threading lace task are very similar in that they are both timed tasks. However, the drawing trail-2 task is self-paced and it requires effective use of a pen/pencil. Thus, although similar, the additional constraint of time and use of an implement may require different aspects of motor functioning. A low alpha could also be attributed to the fact that there are only three tasks within the manual dexterity sub-component, as Cronbach's alpha is higher when there are more test items. As Holm and colleagues [14] did not address the internal consistency of age band 2, the existing literature pertains to other age bands. Wuang and colleagues [15] examined the internal consistency of age band 1 and reported a much higher Cronbach's alpha (0.81). However, in that study the age range was larger (6 to 12 years of age), as was the number of the items examined. The manual dexterity subsection examined here had three items whereas the previous research combined items from each age group resulting in 9 items.

Aiming and Catching. The present study revealed that this subsection had the lowest reliability as evident from Cronbach's alpha of 0.49. This is not surprising. There are only two tasks included in the section, and they differ considerably in terms of their characteristics. Catching with two hands involves interceptive skills whereas throwing a bean bag on to a mat is an accuracy type of the task, without external time demands. Thus, although both involve goal-directed manual actions they do not belong to the same domain such as ball skills. In regards to previous research, Ellinoudis and colleagues [10] found the internal consistency of age band 1 to be acceptable with an alpha of 0.70. The only other study that reported an internal consistency for aiming and catching had an alpha value of 0.84 [14].

Balance.The analysis of the internal consistency for the balance sub-component revealed a Cronbach's alpha of 0.53 indicating that the three items may not be measuring the same domain. When examining the results of Cronbach's alpha with items deleted it was evident that when each task was individually removed, the alpha decreased. The walking on a line and hopping on mats tasks, although also completed on one foot, revealed a much higher scores when compared to the one-board balance task. In fact, the individual data revealed that the one board balance task was

RA Journal of Applied Research
||Volume||2||Issue||09||Pages-611-619||Sept-2016|| ISSN (e): 2394-6709
www.rajournals.in

the most difficult, while the remaining two (walking on a line and hopping on mats) proved to be too easy as almost all children had a perfect score. In terms of the previous research, Ellinoudis and colleagues [10] also reported a relatively low Cronbach's alpha (0.66), for the balance sub-component of age band 1. In contrast, Wuang et al., [15] reported the highest internal consistency for balance with an alpha of 0.84. However, as previously mentioned, the findings from that study are difficult to compare to the present results due to differences in the age of participants and number of items.

CONCLUSIONS

The original version of MABC [6, 7] has been one of the most widely used assessment tools for screening, assessment and diagnosis of children with movement difficulties. The authors of the new version [8] hoped that this legacy would continue with MABC-2, following the improvements implemented in the tasks and scoring system. However, the present and previous studies examining this age band [14] collectively failed to show the desired characteristics in relation to the TTS. In other words, if the child was to be retested again by the clinician, his/her score may be drastically different as compared to the initial testing. This may lead to false positive or negative inferences in regards to the overall proficiency of the client. Hence, it may prevent the child from entering a particular program or resources, or vice-versa it may place the child in the clinical setting when no rehabilitation is in fact required. In regards to three sub-components a similar scenario emerged. The reliability of manual dexterity, even if did not meet the expected level, was the highest. On the other hand, the balance sub-component was uniformly least consistent as evident from the weakest ICC values and the lowest internal consistency. At the individual item level of analysis, this could be attributed to the fact that one board balance task proved to be too difficult whereas the hopping on mats and walking heel to toe were too easy. The resulting celling and/or floor effects have an impact on variability, thus the correlations. However, this also indicates that both tasks are not able to differentiate between "good" versus "bad" movement proficiency resulting once again in invalid inferences. Likely these items require further validation [17].

In conclusion, from the clinical perspective teachers, clinicians, and researchers should implement this age band with caution and either examine the child more than once, and/or rely on other sources of information (e.g. Bruininks; parental reports) when trying to infer a child's movement status. In terms of the limitations of this research, the primary issue was the characteristics and size of the sample. A relatively small and homogeneous sample possibly resulted in restricted range due to ceiling or floor effects. This contributed to the lack of variance within the data, which often deflates the magnitude of the correlation. Thus, without taking away from the merits of the inferences emerging this preliminary research, future studies should examine the reliability of this age band with larger and more heterogeneous sample.

REFERENCES

[1] Burton A, Miller D. Movement skill assessment. 1998; Champaign, IL: Human Kinetics.

[2] Nunnally JC, Bernstein IH. (1994). Psychometric theory – 3rd Edition. 1994; New York, NY: McGraw-Hill Inc.

[3] Crocker L, Algina J. Introduction to classical and modern test theory. 1986; Philadelphia: Harcourt Brace Jovanovich College Publishers.

[4] Atkinson G., Nevill, A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine. 1998; 26(4), 217-238.

[5] Furr R., Bacharach V. Psychometrics. 2008; Thousand Oaks, CA: Sage Publications

[6] Henderson S, Sugden D. Movement assessment battery for children. 1992; Kent, UK: The Psychological Corporation.

[7] Barnett A, Henderson S. An annotated bibliography of studies using the TOMI/Movement ABC. (1998). London, UK: The Psychological Corporation/Harcourt Brace & Company Publishers.

[8] Brown T, Lalor, A. The movement assessment battery for children – second edition (MABC-2): A review and critique. Physical and Occupational Therapy in Pediatrics. 2009; 29(1), 86-103.

[9] Chow S, Chan L, Chan C, Lau C. Reliability of the experimental version of the Movement ABC. British Journal of Therapy and Rehabilitation. 2002; 9, 404-407.

[10] Ellinoudis T, Evaggelinou C, Kourtessis T, Konstantinidou Z, Venetsanour F, Kambas A. Reliability

and validity of age band 1 of the movement assessment battery for children – second edition. Research in Developmental Disabilities. 2011; 32, 1046-1051.

[11] Hua J, Wu Z, Gu G, Meng W. Assessment on the application of 'Movement Assessment Battery' for Children. Zhonghua Liu Xing Bing XueZaZhi. 2012; 33(10), 1010-1015.

[12] Hua J, Gu G, Meng W, Wu Z. Age band 1 of the Movement Assessment Battery for Children – Second Edition: Exploring its usefulness in mainland China. Research in Developmental Disabilities. 2013; 34(2), 801-808.

[13] Smits-Engelsman B, Niemeijer A, van Waelvelde H. Is the movement assessment battery for children- 2nd edition a reliable instrument to measure motor performance in 3 year old children? Research in Developmental Disabilities. 2011; 32, 1370-1377.

[14] Holm I, Tveter A, Aulie V, Stuge B. High intra- and inter-rater chance variation of the movement assessment battery for children 2, age-band 2.Research in Developmental Disabilities. 2013; 34(2), 795-800.

[15] Wuang YP, Su JH, Su CY. Reliability and responsiveness of the Movement Assessment Battery for Children- Second Edition Test in children with developmental coordination disorder.Developmental Medicine and Child Neurology. 2012; 54;160-165

[16] McGraw K, Wong S. (1996). Forming inferences about some intraclass correlation coefficients.Psychological Methods. 1996; 1(1), 30-46.

[17] Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education. 2011; 2, 53-55.

[18] Chow S, Henderson S. Interrater and test-retest reliability of the Movement Assessment Battery for Chinese preschool children. American Journal of Occupational Therapy. 2003: 57(5), 574-577.