



UNIVERSITY *of* LIMERICK

OLLSCOIL LUIMNIGH

Conor Martin

0725927

BSc (Physiotherapy)

2011

**The inter and intra-rater reliability of age band 2 of the Movement
Assessment Battery for Children - Second edition (MABC2)**

Conor Martin

0725927

Supervisor: Dr. Amanda Connell

PY4007 and PY4008: Final Year Project

Word Count: 4912



AUTHORS DECLARATION

I, the undersigned declare that this project which I am submitting is all my own work and that the data presented is authentic.

_____ (Printed Name)

_____ (Signature)

Date / /

Acknowledgements

I would like to thank Dr. Amanda Connell, my project supervisor. Her experience and guidance helped to make the project a hassle free and enjoyable experience.

I would like to thank Eanna Clifford, my research partner on this project. He made the research and testing process much easier and fun.

I would like to thank Patrick Hegarty for his kind words of wisdom throughout the project. Without which it would not have been possible.

Dr. Susan Coote, the module leader was extremely helpful in giving advice and direction which has helped to make this project possible.

I would like to thank all the children and their parents for their participation and co-operation with the testing procedure, without which this project would not have been possible.

I would like to thank the Principal of Scoil Iosagain for his co-operation and support with the testing procedure.

To Timon and Pumbaa, I want to say thank you for your wise words of guidance and helping me to broaden my outlook on life over the last number of years: *Hakuna Matata*

Table of Contents

1. Introduction	7
2. Aims and Objectives	9
3. Methodology	9
3.1 Study Design	9
3.2 Sample Recruitment/consent.....	9
3.3 Inclusion Criteria.....	10
3.4 Exclusion Criteria:	10
3.5 Ethical Approval	11
3.6 Outcome Measure	11
3.7 Data Analysis	12
4. Results	13
4.1 Subjects Information	13
4.2 Inter-rater Reliability.....	13
4.3 Intra-rater Reliability.....	19
5. Discussion	25
6. Conclusion.....	28
7. References	30
Appendix 1: Letter to the principal	33
Appendix 2: Subject consent form	34
Appendix 3: Information Leaflet	35
Appendix 4: Raw Data	38

Abstract

Authors: Conor Martin, Dr. Amanda Clifford (MISCP) (MCSP)

Background: The MABC-2 is one of the most widely used outcome measures as means of measuring developmental delay. (Van Waelvelde et al 2006). Despite this, there is little literature surrounding the reliability of the outcome measure.

Objectives: To determine the inter-rater and intra-rater reliability of the second age band of the MABC-2

Methods: 11 children (mean age 7.55 years) enrolled in second class of primary education completed the Movement Assessment Battery for Children 2 on two separate occasions. The results of these two testing sessions were used to calculate inter and intra-rater reliability.

Results:

Inter-rater reliability

The ICC values for the inter-rater reliability were very good with only one task falling below 0.9. One task received an ICC .892, which is still considered excellent reliability. (Portney & Watkins 1993).

Intra-rater reliability

The ICC values for the intra-rater reliability do not appear as strong as the results for inter-rater reliability. Only two tasks had a reliability that can be considered excellent (>0.9). Three other tasks appear to have moderate intra-rater reliability(.5-.75). All other tasks fell below and ICC of 0.5 indicating poor intra-rater reliability.

Conclusions: The MABC-2 appears to have good inter-rater reliability but much poorer intra-rater reliability. Although there were some limitations to the methodology, questions must arise to the quality of an outcome measure if the intra-rater reliability is not seen to be of a high quality.

Keywords: MABC-2, reliability, Developmental Co-ordination Delay (DCD)

1. Introduction

Many children experience difficulties with everyday tasks that require motor coordination. When these children are assessed objectively, they fall substantially below the level of motor coordination expected given their chronological age (Van Waelvelde et al 2006). However, many of these children are not diagnosed with any physical disorder and this deficit is not being attributed to any known medical condition (Smits-Engelsmen et al 2008). Many of these children are simply labelled as “clumsy”, without any known cause, while others are given a diagnosis of developmental coordination disorder (DCD), (Barnhart et al 2003). In a paper by Lingman et al (2009), the author reports that the incidence of developmental co-ordination disorder (DCD) may range anywhere from 1.4%, to as high as 19% of children. The reason for such a large range includes problems surrounding the diagnosis of DCD (Lingman et al 2009).

Developmental Coordination Disorder is strongly associated with attention deficits/ADHD, Asperger syndrome, and other autism spectrum disorder symptoms (Gillberg 2003). Identification of these conditions and mild motor impairments is imperative in helping to reduce or limit the child’s development of any associated academic or social disorders (Jongmans et al 2005).

Valid outcome measures and motor coordination tests are valuable and essential tools for therapists and school teachers. Outcome measures help not only in identifying children with motor impairment, but also in evaluating motor development and helping to determine the efficacy of any interventions used (Tieman et al 2005). The area of assessing movement skills and coordination is still a relatively young topic of research with very few tests fulfilling the criteria of standardisation, reliability and validity. Outcome measures such as the Movement Assessment Battery for Children (MABC), which was only established in 1992, have been used for a number of years in several different countries to determine mild to moderate motor impairments in children, particularly in European primary health care systems (Van Waelvelde et al 2006). As a result, the outcome measures being used are constantly being revised and altered to improve their reliability, validity and efficacy.

The Movement Assessment Battery for Children 2 (MABC-2) is a revised version of the original MABC. The reason for revising the MABC was to generate a “reliable, easily administered and valid measure of competence in three broad and carefully selected areas of motor performance” (Henderson et al 2007). Changes undertaken to create the second edition involved revising existing items and introducing new ones. Changes were made in the testing section and concern materials, tasks, and instructions. Structural changes to the MABC comprised of adding a 3–6-year old age band, and joining the 7–8 year-old and 9–10-year old age bands. By changing the age brackets, the MABC-2 can now be administered to a larger range of children. Crucially, it now allows for younger children to do the MABC. As a result, motor impairments may be detected and intervened with earlier, minimising any of the associated problems mentioned above.

From the available literature there have been only three studies that have assessed the reliability of the MABC-2. In the first study, intra-rater reliability of the test was calculated using all three age bands with 20 children in each band. The results of this study concluded that the MABC-2 had a good reliability with Pearson’s Coefficients of .77, .84, .73 and .80 for manual dexterity, aiming and catching, balance and total test score respectively (Henderson et al 2007). In the second study, the author looks at the intra-rater reliability of an experimental version of the MABC-2 (Chow et al 2002). The authors of this paper conclude the reliability of the MABC-2 to have an intraclass correlation coefficient of 0.62, indicating moderate reliability (Portney & Watkins 1993). The most recent study, which looks at the reliability of the first age band of the MABC-2, shows the outcome measure to have moderate reliability with alpha coefficients ranging from .51 to .7 (Ellinoudis et al 2011). From the available evidence, it is clear to see there have been very few studies surrounding the MABC-2 and its reliability. As a result, only these studies and studies assessing the reliability of the MABC can be used as evidence for evaluating the reliability of the MABC-2 (Van Waelvelde et al 2007)

As there have been many changes between the original MABC and the MABC-2, one cannot simply assume that the evidence surrounding the reliability of the MABC can be used for the MABC-2 (Brown and Lalor 2009). The changing of age brackets and

addition of new tasks being assessed means that new studies are required to determine the reliability of the MABC-2. Questions must surely arise as to the validity of any outcome measure if the reliability of its use is not of a high quality. Thus, the need for new studies/evidence on the reliability of the MABC-2 is considerable and of paramount importance in establishing the MABC-2 as a valid outcome measure for assessing motor development.

The purpose of this study was to assess the inter-rater and intra-rater reliability of the second age bracket (7-10 years) of the revised MABC-2. The focus of this study was on children falling into the second age bracket (7-10 years) as this had not yet been assessed.

2. Aims and Objectives

- (1) To determine whether the M-ABC 2 is a reliable instrument when tested by one rater on two separate occasions
- (2) To determine whether the M-ABC 2 is a reliable instrument when tested between two raters on the same child.

3. Methodology

3.1 Study Design

The study took the form of a quantitative reliability study. Two sessions were conducted in this study, based two weeks apart. The first testing session took place in the children's primary school, while the second testing session took place two weeks later in a local health centre as the testing room within the school was not available on the second date. This time interval aimed to improve the accuracy of the reliability of the measure by decreasing the "practice effect" (Smits-Engelsman 2007)

3.2 Sample Recruitment/consent

Approval was gained from the Principal of a local Primary School by means of a letter (See Appendix 1) which explained the outline of the study and what would be required of his/her school. After securing the primary school as a recruitment centre, children

who fell into the second age bracket were recruited. Children in second class (n=14) of the primary school were asked to partake in the reliability study of the MABC-2. Information sheets and consent forms (See Appendix 2&3), which described the study were given to each of the 14 children in second class to be read and signed by their parents/guardians. Of the 14 children, 11 consent forms were returned and signed by the parents. These 11 children were included in the study.

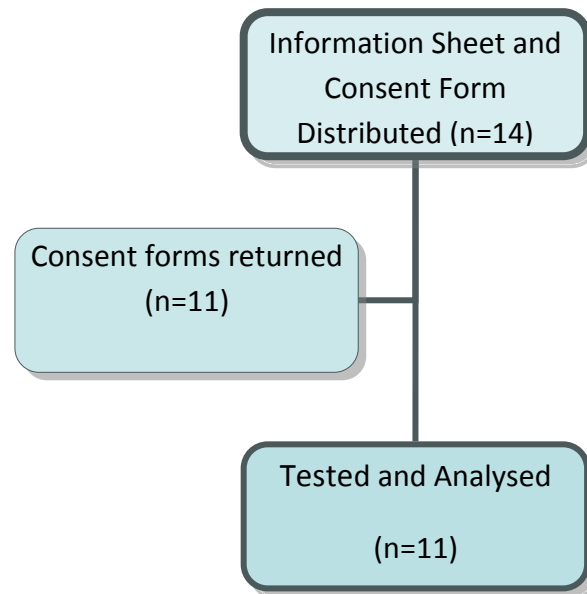


Figure 1. Flow Diagram of Recruitment Course

3.3 Inclusion Criteria

- Children between the ages of 7 and 10 enrolled in primary education

3.4 Exclusion Criteria:

- Non English speaking children
- Children suffering from any injury at the time of testing
- Children diagnosed with severe neurological conditions resulting in cognitive impairments (Van Waelvelde et al 2007)

3.5 Ethical Approval

Ethical approval was applied for and granted by the University of Limerick Research Ethics Committee. The University of Limerick Child Protection Guidelines (2006) were adhered to at all times.

3.6 Outcome Measure

The outcome measure being assessed was the MABC-2. (The testing kit can be seen in Figure 1.) In the MABC-2 the child is asked to do 11 different tasks which are then marked based on the performance of the child. These markings are then given a standard score which are then used to calculate the percentile of the child for three different subsections. These subsections include; *Manual Dexterity*, *Aiming and Catching* and *Balance*. The tasks included for the MABC-1 are displayed in Table 1.

Movement Assessment Battery for Children 2	
Section 1	Manual Dexterity
Task 1	Placing pegs preferred hand
Task 2	Placing pegs non preferred hand
Task 3	Threading lace
Task 4	Drawing trail
Section 2	Aiming and Catching
Task 5	Catching
Task 6	Throwing beanbag
Section 3	Balance
Task 7	One board balance best leg
Task 8	One board balance other leg
Task 9	Walking heel-toe forwards
Task 10	Hopping best leg
Task 11	Hopping other leg

Table 1. Description of tasks for MABC-2



Figure 1. Testing Kit for the MABC-2.

3.7 Data Analysis

Data analysis was conducted using the Statistical Programme for the Social Sciences (SPSS) version 18.

For the purpose of this study, the standard scores which the children received were used to determine reliability. Recent papers surrounding the reliability of the MABC-2 (Ellinoudis et al 2011), as well as older papers assessing the reliability of the original MABC (Chow et al 2003) have used standard scores and percentiles opposed to raw data. While the standard score allows for less variance than the raw data, it may be more accurate in determining the overall reliability of the test as the standard scores are then used to calculate percentiles for each subsection.

Intra-class coefficients (ICC's) were calculated to evaluate the inter-rater and intra-rater reliability of the MABC-2. For ICC's, coefficients below .50 indicate poor reliability, those between .50 and .75 indicate moderate reliability, and those above .75 indicate good reliability (Portney & Watkins 1993). To further test the reliability 95% confidence intervals were used. While data may appear reliable from the ICC's it is impossible to tell how precise this measurement is. The use of a 95% confidence interval allows us to see the level of association between two measurements. The equations used for the inter-rater reliability portion of the project was two-way mixed and inter-rater reliability was one way random as recommended by Rankin and Stokes (1998).

Bland and Altman calculations were also used to measure inter and intra-rater reliability. With Bland and Altman calculations, the closer the Mean Difference (DIFF), the Standard Error (SE) and the Standard Deviation (SD) are to zero, the more reliable the measurement is considered (Altman 1991). Within the Bland and Altman calculations, the standard mean has also been calculated. This calculation shows how good or reliable the estimate of the mean difference is. Bland and Altman calculations complement the ICC calculations by assessing the level of agreement between measurements and should be done in conjunction with one another (Rankin and Stokes 1998).

4. Results

4.1 Subjects Information:

Mean Age: 7.55 years, Standard deviation: +/-0.5, Sex: All Male

4.2 Inter-rater Reliability

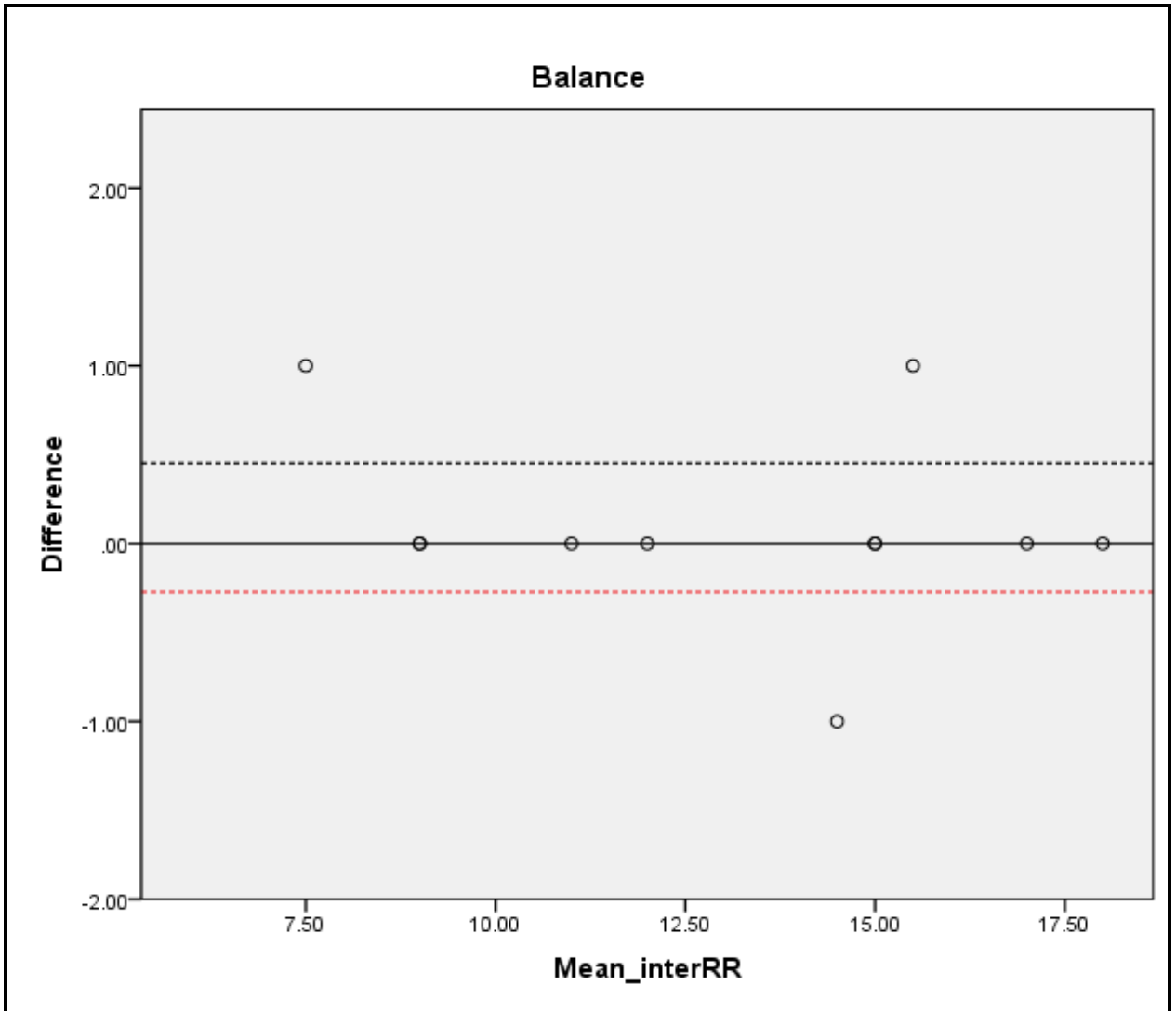
Inter-rater reliability is the degree of agreement among raters. It gives a score of how much homogeneity, or consensus there is in the ratings given by judges (Bland and Altman 1986). Table 2 summarises the statistical analysis for the inter-rater reliability portion of this paper. This table includes both the Intraclass Correlation Coefficients and Bland and Altman statistical analysis results for this study. For ICC's, a value of 1 implies perfect reliability and a value of zero signifies no reliability. The ICC values for the inter-rater reliability were very good with only one task falling below 0.9. Hopping on the worse leg received the lowest ICC of .892, which is still considered excellent reliability (Portney & Watkins 1993). The 95% confidence intervals for the ICC results also demonstrate very small amounts of variance, improving the standard of the ICC results.

The Bland and Altman results of each of the subsections, (*Manual Dexterity, Aiming and Catching and Balance*) and total test scores are also displayed on a scatter plot diagrams in Figure 2, 3, 4, and 5. The Bland and Altman Values for Mean Difference (DIFF), standard error of the mean difference (SE of DIFF), and the standard deviation

of the mean difference (SD of DIFF), all lie close to zero showing good reliability, similarly to the ICC results. The 95% confidence intervals also show a good range, going through zero which indicates that no systematic difference exists between the results (Rankin and Stokes, 1998)

Table 2: ICC and Bland and Altman results for Inter-rater reliability

	Intraclass Correlation Coefficients			Bland and Altman				
	ICC	95%CI		DIFF	SE of DIFF	95%CI of DIFF		SD
		Lower Bound	Upper Bound			Lower Bound	Upper Bound	
Placing pegs preferred	0.928	0.757	0.98	0	0.23355	-0.5204	0.5204	0.7746
Placing pegs non preferred	0.97	0.894	0.992	0	0.13484	-0.3004	3.004	0.44721
Threading lace	0.952	0.833	0.987	-0.909	0.16262	-0.4533	0.2714	0.53936
Drawing trail	0.977	0.918	0.994	0.0909	0.21125	-0.3798	0.5616	0.70065
Manual Dexterity	0.979	0.923	0.994	0.18182	-0.1818	-0.5869	0.2233	0.60302
Catching	0.976	0.914	0.993	-0.1818	0.18182	-0.5869	0.2233	0.60302
Throwing beanbag	0.956	0.845	0.988	0	0.26968	-0.6009	6.009	0.89443
Aiming and Catching	0.973	0.903	0.993	0.18	0.18182	-0.2233	0.5869	0.60302
One board balance best leg	0.941	0.798	0.984	0.2727	0.14084	-0.0411	0.5865	0.4671
One board balance other leg	0.984	0.943	0.996	0.2727	0.23706	-0.2555	0.8009	0.78625
Walking heel-toe forwards	0.981	0.932	0.995	-0.0909	0.16262	-0.4533	0.2714	0.53936
Hopping best leg	0.992	0.969	0.98	-0.0909	0.09091	-0.2935	0.1116	0.30151
Hopping other leg	0.892	0.649	0.997	0.1818	0.29598	-0.4777	0.8413	0.98165
Balance	0.989	0.959	0.997	0.0909	0.16262	-0.2714	0.4533	0.53936
Total test score	0.977	0.915	0.994	-1	0.21125	-0.5616	0.3798	0.70065



_____ Mean of differences

----- 95% confidence interval (upper bound)

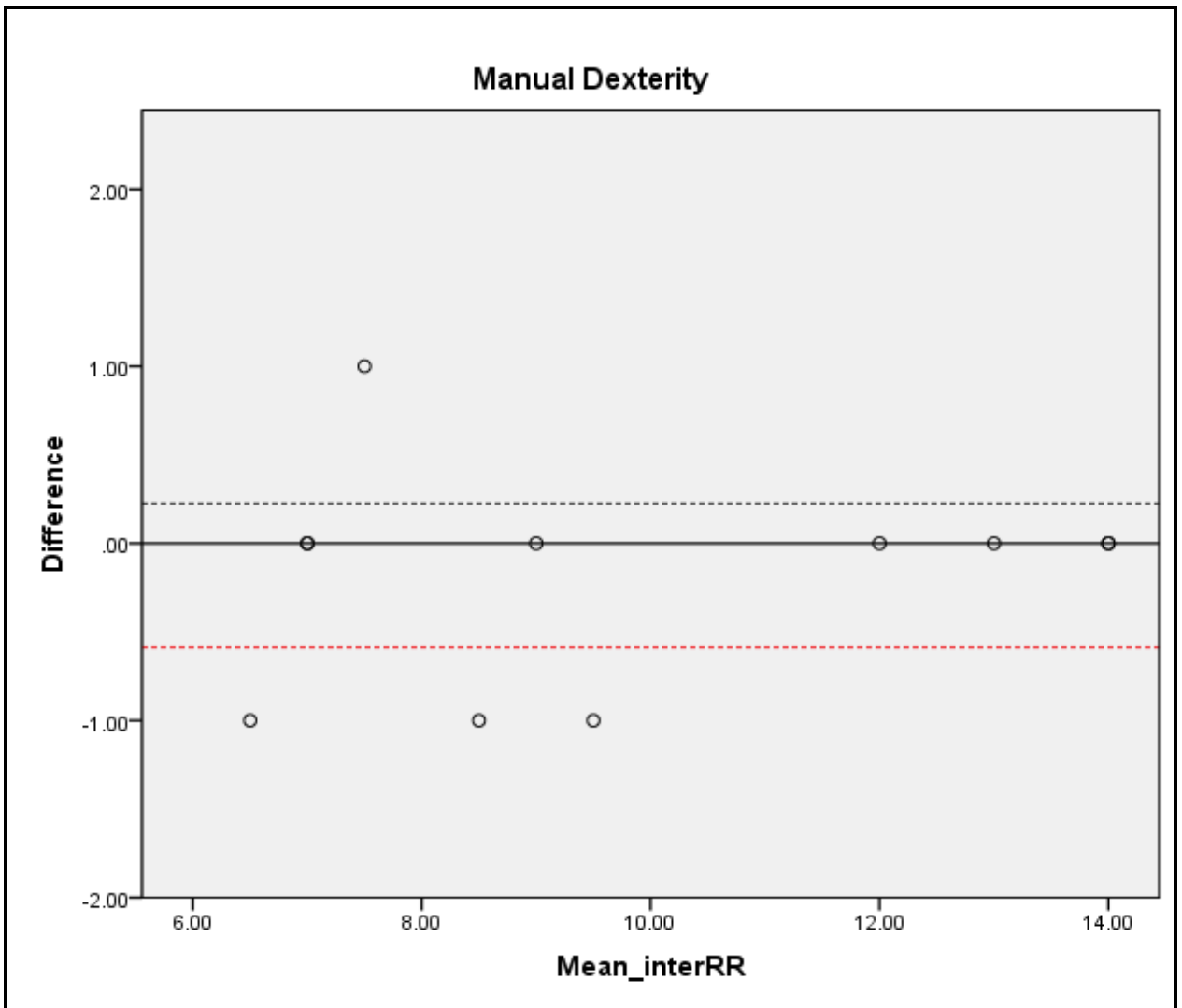
..... 95% confidence interval (lower bound)

X-axis: Mean inter-rater reliability for the balance subsection of the M-ABC 2.

Y-axis: Differences between the mean M-ABC 2 scores.

Figure 2.

Distribution Plot from Bland and Altman (1991) test showing the inter-rater reliability for the **Balance** subsection of the M-ABC 2.



_____ Mean of differences

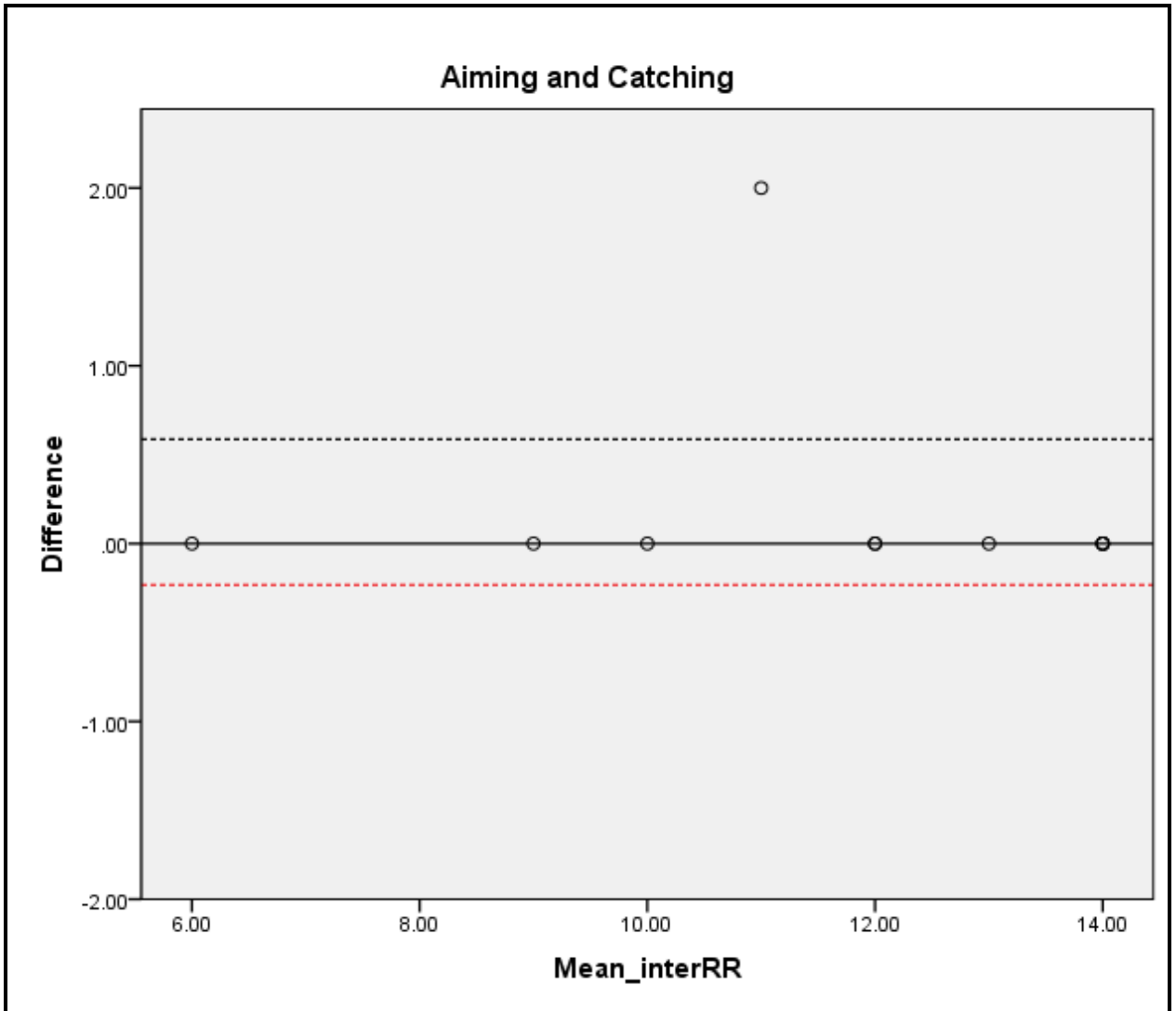
----- 95% confidence interval (upper bound)

----- 95% confidence interval (lower bound)

X-axis: Mean inter-rater reliability for the *Manual Dexterity* subsection of the M-ABC 2.
 Y-axis: Differences between the mean M-ABC 2 scores.

Figure 3.

Distribution Plot from Bland and Altman (1991) test showing the inter-rater reliability for the **Manual Dexterity** subsection of the M-ABC 2.



_____ Mean of differences

----- 95% confidence interval (upper bound)

----- 95% confidence interval (lower bound)

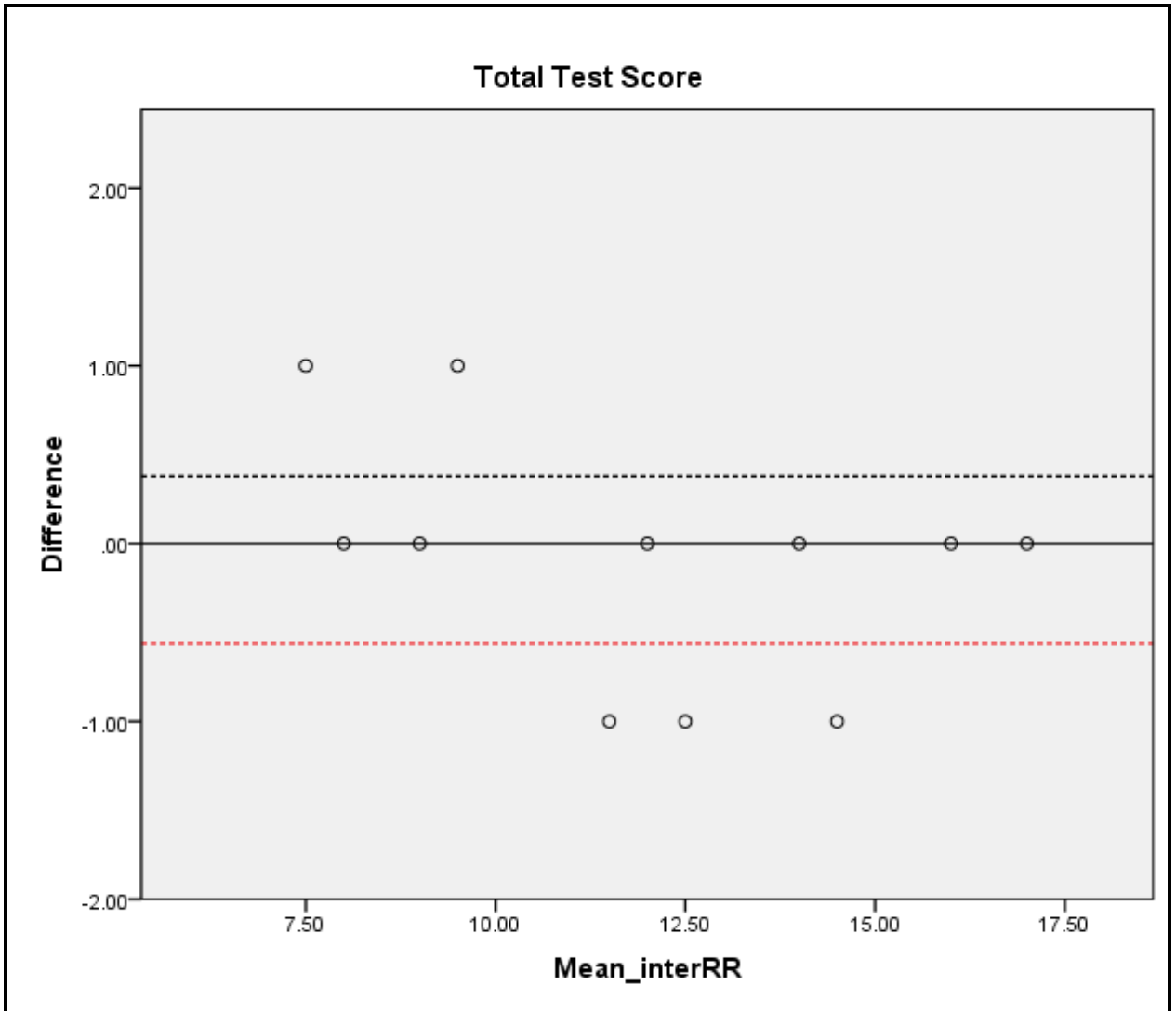
X-axis: Mean inter-rater reliability for the *Aiming and Catching* subsection of the M-ABC 2.

Y-axis: Differences between the

mean M-ABC 2 scores.

Figure 4.

Distribution Plot from Bland and Altman (1991) test showing the inter-rater reliability for the **Aiming and Catching** subsection of the M-ABC 2.



_____ Mean of differences

----- 95% confidence interval (upper bound)

..... 95% confidence interval (lower bound)

X-axis: Mean inter-rater reliability for the *Total Test Score* of the M-ABC 2.

Y-axis: Differences between the mean M-ABC 2 scores.

Figure 5.

Distribution Plot from Bland and Altman (1991) test showing the inter-rater reliability for the **Total Test Score** of the M-ABC 2.

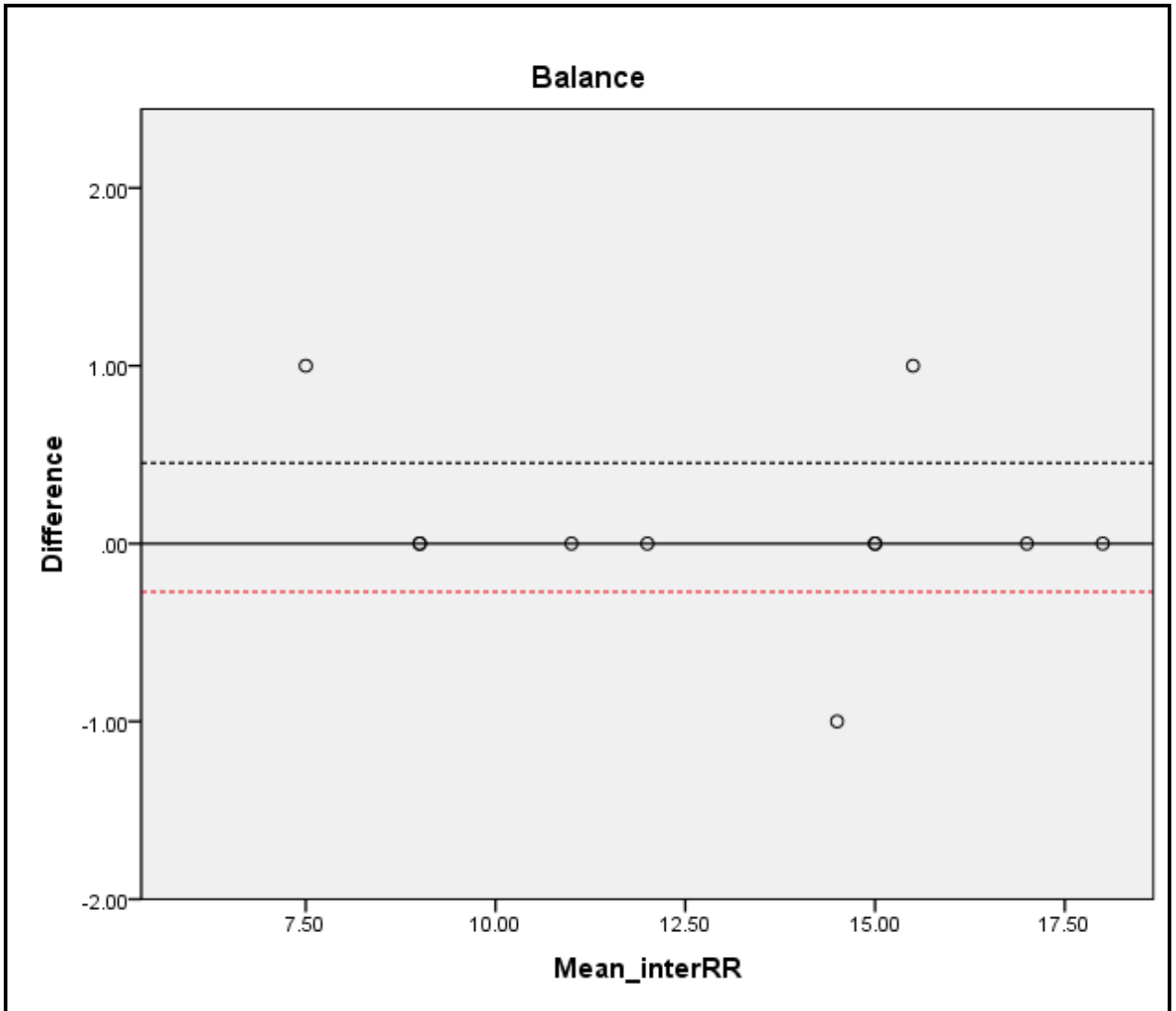
4.3 Intra-rater Reliability

Intra-rater reliability is the degree to which repeated measurements vary for individuals (Bruton et al 2000). Table 3 summarises the statistical analysis for the intra-rater reliability portion of this paper. This table includes both the Intraclass Correlation Coefficients and Bland and Altman statistical analysis results for this study. The ICC values for the intra-rater reliability do not appear as strong as the results for inter-rater reliability. The walking heel-toe and hopping on the best leg exercises are the only tasks with a reliability that can be considered excellent with ICC's of .971 and .976 respectively. Only three other tasks appear to have moderate intra-rater reliability with placing pegs, drawing a trail and one board balance all receiving ICC's of between .5 and 0.75. The remainder of the tasks fall below an ICC of 0.5 indicating poor reliability (Portney & Watkins 1993).

The Bland and Altman results of each of the subsections, (*Manual Dexterity, Aiming and Catching and Balance*) and total test scores are also displayed on a scatter plot diagrams in graphs 5, 6, 7 and 8. The Bland and Altman Values for Mean Difference (DIFF), standard error of the mean difference (SE of DIFF), and the standard deviation of the mean difference (SD of DIFF), are very varied between the different tasks. Certain tasks appear to have good reliability with the Bland and Altman results such as walking heel toe and hopping best leg. However, most other tasks display poor reliability. The 95% confidence intervals are again extremely varied ranging from -1.4825 →1.3007 for placing pegs with the preferred hand, to -0.479→13.154 for one board balance on the best leg (Rankin and Stokes, 1998)

Table 3 – ICC and Bland and Altman results for Intra-rater reliability

	Intraclass Correlation Coefficients			Bland and Altman				
	ICC	95%CI		DIFF	SE of DIFF	95%CI of DIFF		SD
		Lower Bound	Upper Bound			Lower Bound	Upper Bound	
Placing pegs preferred	0.481	-0.105	0.826	-0.0909	0.62457	-1.4825	1.3007	2.07145
Placing pegs non preferred	0.564	0.009	0.859	0.0909	0.56334	-1.1643	1.3461	1.8684
Threading lace	0.461	-0.13	0.817	0.10909	0.41461	-2.0147	-0.1671	1.3751
Drawing trail	0.55	-0.011	0.853	0.7273	0.95433	-0.1399	2.8537	3.16515
Manual Dexterity	0.804	0.445	0.942	-0.3636	0.66432	-1.1844	1.1166	2.2033
 								
Catching	0.383	0.139	0.889	-0.0909	0.7318	-1.7215	1.5397	2.42712
Throwing beanbag	0.447	-0.147	0.811	0.8182	0.92263	-2.8739	1.2376	3.06001
Aiming and Catching	0.806	0.45	0.943	0.0909	0.51265	-1.0513	1.2332	1.70027
 								
One board balance best leg	0.383	-0.222	0.783	0.2727	0.57352	-1.0052	1.5506	1.90215
One board balance other leg	0.762	0.354	0.929	4.1818	4.02677	-0.479	13.154	13.3553
Walking heel-toe forwards	0.971	0.903	0.992	0	0.19069	-0.4249	0.4249	0.63246
Hopping best leg	0.975	0.914	0.993	0	0.13484	-0.3004	0.3004	0.44721
Hopping other leg	0.367	-0.24	0.776	0.9091	0.53009	-0.272	0.899	1.7521
Balance	0.766	0.364	0.889	1.1818	0.73631	-0.4588	2.8224	2.44206
 								
Total test score	0.926	0.76	0.979	-0.0909	0.41461	-1.0147	0.8329	1.13751



_____ Mean of differences

----- 95% confidence interval (upper bound)

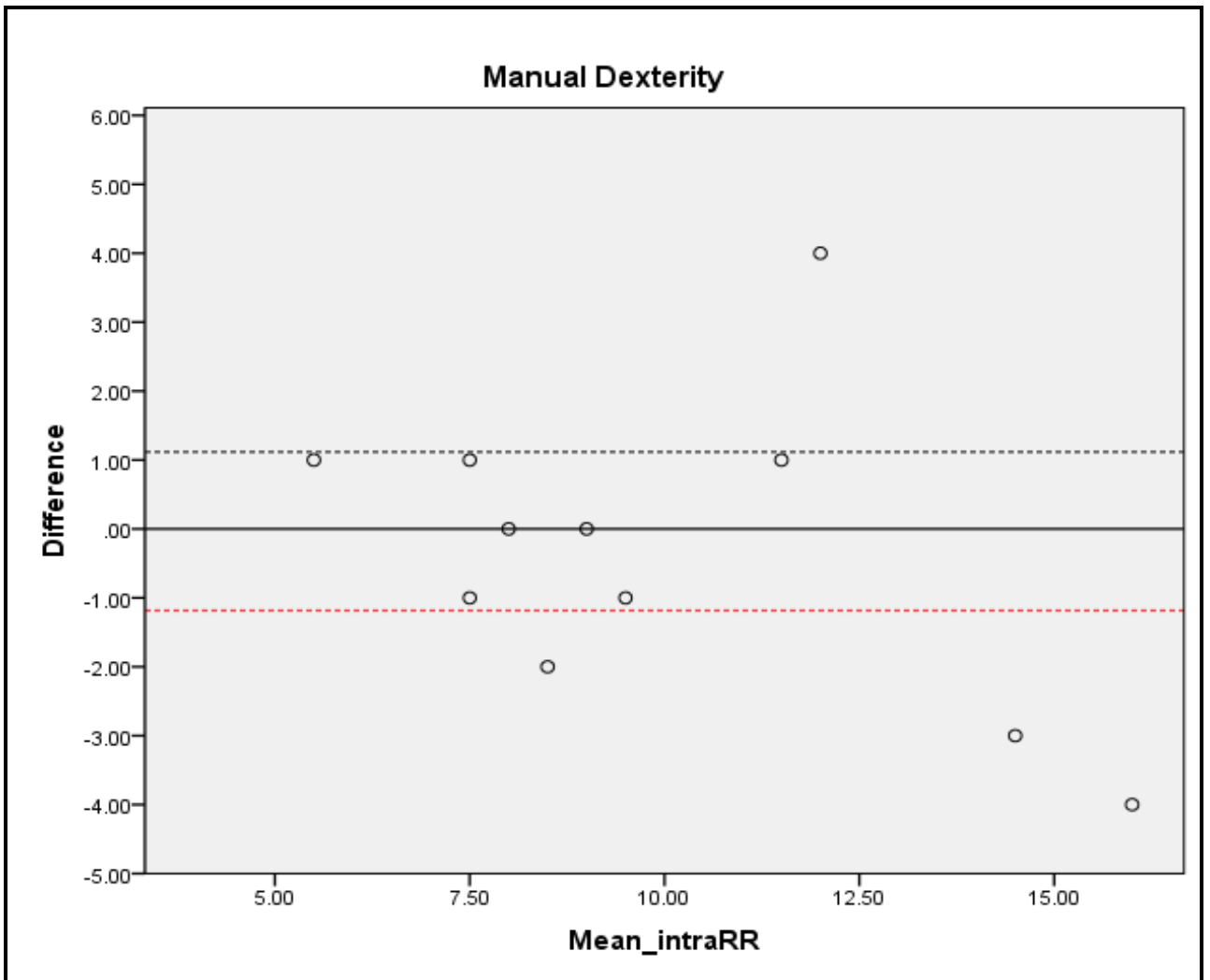
..... 95% confidence interval (lower bound)

X-axis: Mean intra-rater reliability for the balance subsection of the M-ABC 2.

Y-axis: Differences between the mean M-ABC 2 scores.

Figure 6.

Distribution Plot from Bland and Altman (1991) test showing the intra-rater reliability for the **Balance** subsection of the M-ABC 2.



_____ Mean of differences

----- 95% confidence interval (upper bound)

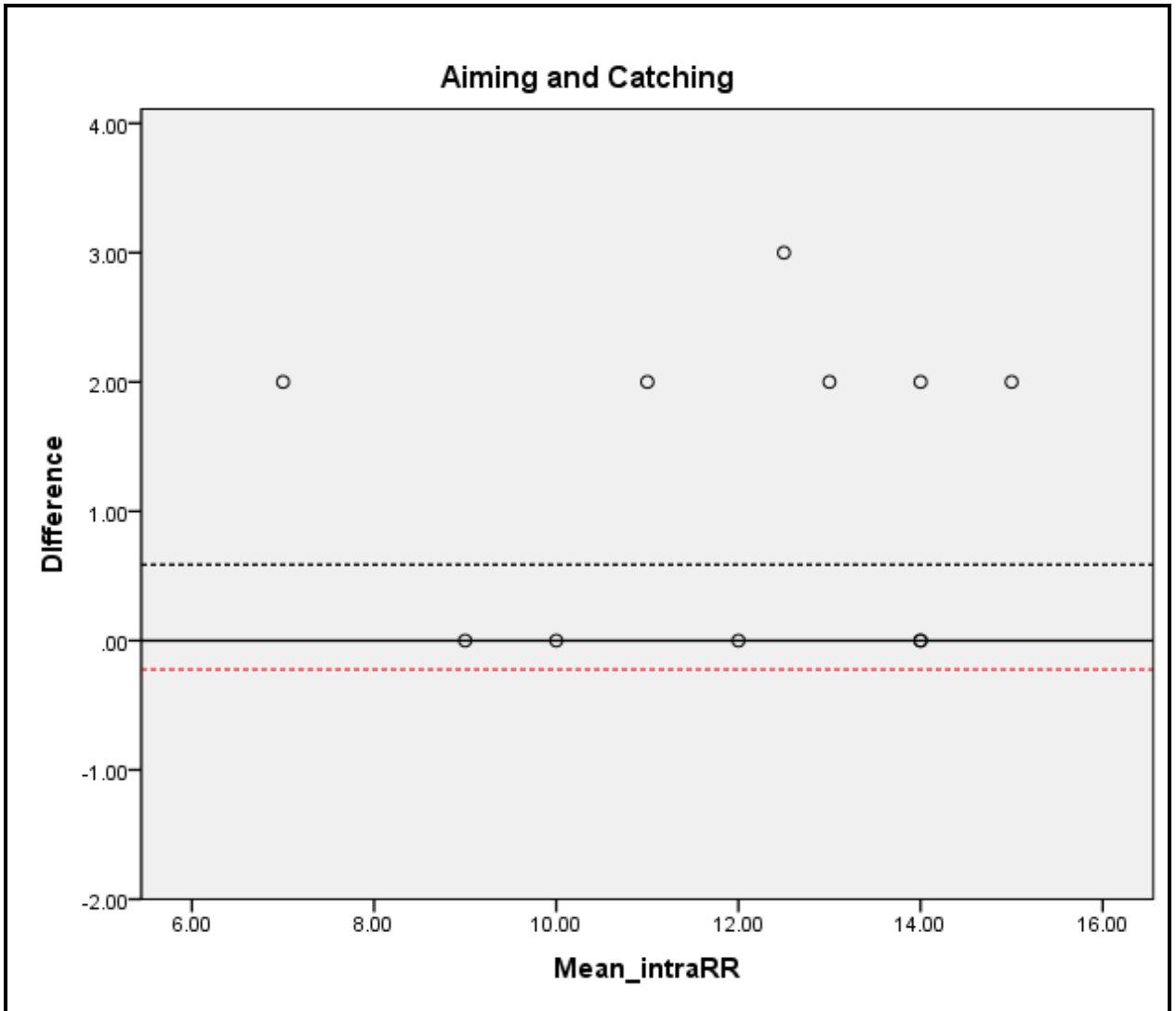
----- 95% confidence interval (lower bound)

X-axis: Mean intra-rater reliability for the *Manual Dexterity* subsection of the M-ABC

2. Y-axis: Differences between the mean M-ABC 2 scores.

Figure 7.

Distribution Plot from Bland and Altman (1991) test showing the intra-rater reliability for the **Manual Dexterity** subsection of the M-ABC 2.



_____ Mean of differences

----- 95% confidence interval (upper bound)

..... 95% confidence interval (lower bound)

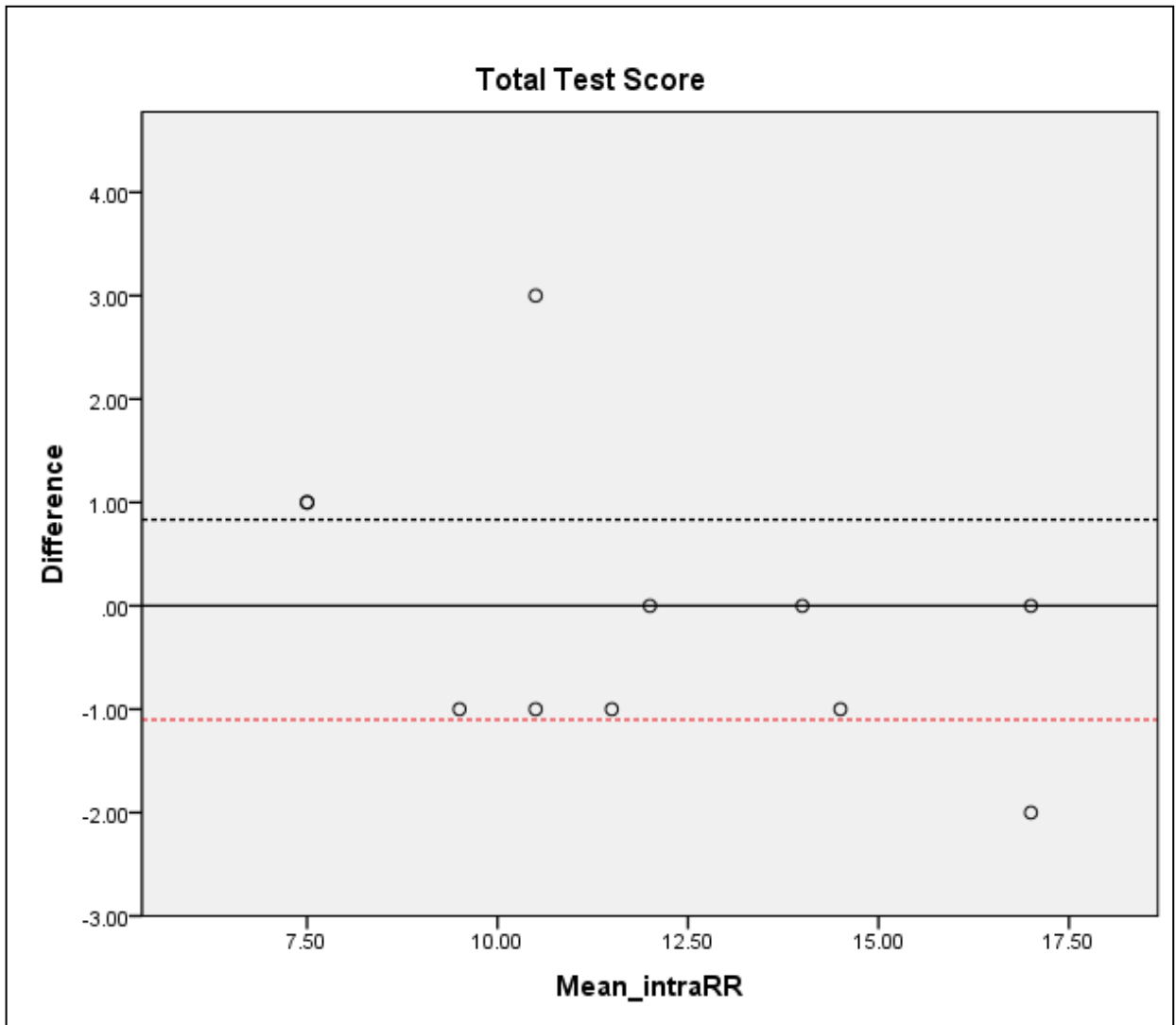
X-axis: Mean intra-rater reliability for the *Aiming and Catching* subsection of the M-ABC 2.

Y-axis: Differences between the mean M-ABC 2 scores.

_____ Mean of differences

Figure 8.

Distribution Plot from Bland and Altman (1991) test showing the intra-rater reliability for the **Aiming and Catching** subsection of the M-ABC 2.



_____ Mean of differences

----- 95% confidence interval (upper bound)

----- 95% confidence interval (lower bound)

X-axis: Mean intra-rater reliability for the *Total Test Score* of the M-ABC 2.

Y-axis: Differences between the mean M-ABC 2 scores.

Figure 9.

Distribution Plot from Bland and Altman (1991) test showing the intra-rater reliability for the **Total Test Score** of the M-ABC 2.

5. Discussion

The aim of this study was to assess both the inter-rater and intra-rater reliability of the second age bracket of the Movement Assessment Battery for Children – Second Edition. At the time of this study, there had been no studies evaluating the reliability of this age bracket of the MABC-2. As a result, the reproducibility of results when using this outcome measure on children aged between 7 and 10 was unknown until now. The only recently available published material concerns the reliability of the first age band of the MABC-2, which is meant for children aged 3-7 (Ellinoudis et al 2011). The results of the study showed the reliability of the test within this age bracket to be moderate to high with alpha coefficients ranging from .51 to .7. However, if the outcome measure is to be considered reliable as a whole, each age bracket would need to be evaluated for reliability and validity.

Within clinical practice, it is highly important that an outcome measure surrounding motor development is reliable as it allows for the analysis and efficacy of an intervention to be measured objectively (Goyen & Lui 2009). Within this study the MABC-2 appears to have an excellent level of inter-rater reliability. The only task which did not get an excellent ICC (<0.9) was the *hopping on the worse leg* task which scored an ICC of 0.892 which is still considered very high (Portney & Watkins 2000). One aspect of the methodology may account for the very high levels of inter-rater reliability seen for the MABC-2. The use of standard scores for data analysis, while widely used (Ellinoudis et al 2011; Van Waelvelde 2007), may make the reliability of the outcome measure appear better than it actually is. As there is a greater degree of variance in the raw data or the initial scores that the children receive, this would make the reliability of the outcome measure appear worse than the values for reliability demonstrated in this study. However, as these raw scores are then transformed into standard scores and further again into percentiles, if the child is still receiving the same standard scores, but different raw data, this would mean that the outcome measure is being reliable in what it is set out to measure. The positive results for inter-rater reliability are encouraging as it means that the MABC-2 can be administered by two separate raters/clinicians and still produce very similar results.

Intra-rater reliability is important in distinguishing whether or not an outcome measure is reliable when used on the same child on two separate occasions by the same rater (Bland & Altman 1986). This is imperative in a clinical setting whereby the same clinician will be performing the test on the same group of children within the clinical setting to assess any progress in motor development.

What is interesting to note about the results of this study is that while they indicate a very high level of inter-rater reliability, the results for intra-rater reliability are not as positive. The walking heel-toe and hopping on the best leg exercises are the only tasks with a reliability that can be considered excellent with ICC's of .971 and .976 respectively. Only three other tasks appear to have moderate intra-rater reliability with placing pegs, drawing a trail and one board balance all receiving ICC's of between .5 and 0.75 (Portney & Watkins). If the inter-rater reliability appeared so good, one must question why the results for intra-rater reliability are not so positive. When looking at the methodology there are a number of aspects that may have accounted for the poor results in intra-rater reliability. Between the first and second testing day the location was moved from the children's primary school to a nearby health centre. This change in location appears to have been the major difference between the two testing days. On the first testing day, the children were assessed in their school by two raters. Within the school, the children were in an environment which they were comfortable and familiar with. Many papers suggest that children tend to perform better and more honestly in an environment where they are comfortable, specifically school (Parkinson 2001; Macy et al 2003). Macy et al (2003), look at implementing interventions within the classroom as children tend to perform better as they feel safer and less anxious. On the second testing day, the children were moved to the unfamiliar setting of a nearby health centre. This change of location may have impacted on the children's performance, causing worse scores in the MABC-2 on this day. This in turn, will have impacted on the data received and the intra-rater reliability in this study.

Another limitation of the study that may have accounted for these results was the lack of familiarity of the raters with the outcome measure. Both of the raters who assessed the children were physiotherapy students in their final year of university study. As a result, the raters had little clinical practice with the MABC-2 prior to this study. This lack of familiarity meant that as the study progressed, the raters became more comfortable with

the outcome measure and by the second testing day, they may have been marking the outcome measure more effectively. This would have impacted on the intra-rater reliability as the results from the first testing session were compared to the results on the second testing session. This improved familiarity with the administration of the outcome measure may have affected the apparent intra-rater reliability of the MABC-2. Other studies assessing reliability have tended to be carried out by experienced therapists who are sufficiently familiar with the outcome measure prior to the study (Smits-Engelsman et al 2008; Van Waelvelde et al 2007). By ensuring that the raters were experienced in carrying out the MABC-2, this may have improved the accuracy of the results.

Another factor that may have impacted on the intra-rater reliability is the notion that children, while competitive in nature, often lack well-developed verbal skills, do not sit attentively for extended periods of time, get bored easily with repetitive activities, and are not very concerned with compliance to please the examiners (Payne & Isaacs, 1998). During the testing process, many of the children did display poor levels of compliance, requiring constant prompting to finish the tasks and getting distracted easily. This lack of compliance or concern with completing the tasks correctly may have accounted for the poor results for intra-rater reliability. If the children are not concerned with doing their best in the tasks, the outcome measure may not be effective in recording the actual level of motor development of the child. While instructions and practice trials were given to the child prior to each task, there were still issues with certain children not complying appropriately with the task procedure. This lack of compliance or concern with pleasing the raters would have impacted on the reliability as they may have been compliant with different tasks between the two testing sessions.

Despite the limitations within the methodology discussed above, questions must still arise as to how robust an outcome measure the MABC-2 is and whether the results of poor intra-rater reliability displayed in this study are due the study's limitations or due to the outcome measure itself. As there have been no previous studies concerning the intra-rater reliability of this age band of the MABC-2, only the results from studies concerning different age bands within the MABC-2, and results from the original MABC, can be used to draw comparison with. Visser and Jongmans (2004) reported

similar results for intra-rater reliability of the first age band of the MABC-2. The authors conclude that the MABC-2 displayed only moderate intra-rater reliability with Pearson's coefficients ranging from 0.49 to 0.7. Chow et al (2002) evaluated the inter-rater and intra-rater reliability of the third age band of an experimental version of the MABC-2. The results from Chow et al (2002) bare many similarities with the results from this study. Chow et al (2002) evaluated that the MABC-2 displayed excellent inter-rater reliability with ICC's ranging from 0.92 to 1.00. However, similarly to the results found in this study, the authors concluded that the intra-rater reliability was not as good with ICC's ranging from 0.62 to 0.92. While these ICC's are based on an experimental version of the MABC-2 and are still within the moderate to very good range (Portney & Watkins 1993), it is still very interesting to note the vast difference between inter-rater and intra-rater reliability. Questions must arise as to how appropriate the MABC-2 is as an outcome measure to assess motor-development if it does not display both good inter-rater and intra-rater reliability.

6. Conclusion

The test authors of the MABC-2 assume that the literature surrounding the reliability of the original MABC is generalizable to the updated MABC-2. "Confidence in the MABC-2 score interpretation can be derived not only from the UK standardisation study but also from the extensive validation data reported in this and earlier manuals" (Henderson et al 2007, p. 132) While the original MABC and the MABC-2 may be assessing the same motor skills, the MABC also has four added tasks from the first edition. These new tasks, along with the changing of the age bands mean that the reliability of the MABC-2 cannot simply be assumed from the evidence surrounding the MABC.

The MABC-2 is still a relatively new outcome measure (Henderson et al 2007), and there is a limited amount of literature surrounding its validity and reliability (Ellinoudis et al 2011). Brown and Lalor (2009) assessed the MABC-2 and the evidence surrounding its reliability and validity. The authors concluded that this lack of research surrounding the reliability and validity of the MABC-2 were the outcome measures

primary limitations. While there have been some studies published since Brown and Lalor's paper in 2009, (Ellinoudis et al 2011), the results have not been completely encouraging. With the results of this and other studies in mind, until further reliability and validity studies are completed, therapists should be cautious in making clinical decisions related to clients solely based on MABC-2 results.

In summation, further tests with larger sample sizes and improved methodology will be needed to verify the true reliability of the outcome measure as a whole. Until these studies are done, the reliability of the MABC-2 can only be drawn from the limited amounts of available literature (Chow et al 2002; Visser and Jongmans 2004; Henderson et al 2007; Ellinoudis et al 2011)

7. References

- Altman D.G (1991) 'Practical statistics for medical research' London: Chapman & Hall.
- Barnhart, R.C., Davenport, M.J., Epps, S.B. and Nordquist, V.M. (2003) 'Developmental coordination disorder', *Physical Therapy*, 83(8), 722-731.
- Bland, J. M., and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* i, pp. 307–310.
- Brown, T. and Lalor, A. (2009) 'The Movement Assessment Battery for Children—Second Edition (MABC-2): A Review and Critique', *Physical & Occupational Therapy in Pediatrics*, 29(1), 86-103.
- Bruton, A., Conway, J.H. and Holgate, S.T. (2000) 'Reliability: What is it, and how is it measured?', *Physiotherapy*, 86(2), 94-99.
- Chow, S. M. K., Chan, L. L., Chan, C. P. S., & Lau, C. H. Y. (2002). 'Reliability of the experimental version of the movement ABC', *British Journal of Therapy and Rehabilitation*, 9, 404–407.
- Chow, S.M. and Henderson, S.E. (2003) 'Interrater and test-retest reliability of the Movement Assessment Battery for Chinese preschool children', *AMERICAN JOURNAL OF OCCUPATIONAL THERAPY*, 57(5), 574.
- Ellinoudis T, Evaggelinou C, Kourtessis T, Konstantinidou Z. (2011). Reliability and validity of age band 1 of the Movement Assessment. *Research in Developmental Disability*. 32 (1), 1046-1051.
- Gillberg C, Kadesjo B (2003), Why bother about clumsiness? The implications of having developmental coordination disorder (DCD). *Neural Plast* 10:59–68

Goyen T, Lui K (2009). Developmental coordination disorder in “apparently normal” schoolchildren born extremely preterm. *Disabled Children*. 94 (4), 298-302.

Henderson, S. E., & Sugden, D. A. (1992). *Movement assessment battery for children manual*. Kent, UK: The Psychological Corporation.

Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2007). *Movement assessment battery for children-2 second edition manual* [Movement ABC-2]. London, UK: The Psychological Corporation.

Jongmans MJ. (2005) ‘Early identification of children with developmental coordination disorder.’ In Sugden D, Chambers M *Children with developmental coordination disorder*. 3rd ed. London: Whurr Publishers: 155–67

Lingam R, Hunt L, Golding J, Jongmans M, Emond A . (2009). Prevalence of Developmental Coordination Disorder Using the DSM-IV at 7 Years of Age: A UK Population–Based Study. *Pediatrics*. 123 (4), 693-700.

Macy, R.D., Mary, D.J., Gross, S.I. and Brighton, P. (2003) 'Healing in familiar settings: Support for children and youth in the classroom and community', *New Directions for Youth Development*, 2003(98), 51-79.

Payne, G., & Isaacs, L. (1998). *Human movement development: A lifespan approach*. California: Mayfield Publishing Company.

Portney, L. G., & Watkins, M. P. (1993). *Foundations of clinical research: Applications to practice*. Norwalk, Connecticut: Appleton and Lange.

Rankin, G. and Stokes, M. (1998) 'Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses', *Clinical rehabilitation*, 12(3), 187-199.

Smits-Engelsman, B., Fiers, M.J., Henderson, S.E. and Henderson, L. (2008) 'Interrater reliability of the Movement Assessment Battery for Children', *Physical Therapy*, 88(2), 286-294.

Tieman, B.L., Palisano, R.J. and Sutlive, A.C. (2005) 'Assessment of motor development and function in preschool children', *Mental retardation and developmental disabilities research reviews*, 11(3), 189-196.

Van Waelvelde, H., Peersman, W., Lenoir, M. and Smits Engelsman, B. (2007) 'The reliability of the Movement Assessment Battery for Children for preschool children with mild to moderate motor impairment', *Clinical rehabilitation*, 21(5), 465-470.

Visser, J., & Jongmans, M. (2004). 'Extending the movement assessment battery for children to be suitable for 3-year-olds in the Netherlands' Unpublished manuscript.

Appendix 1: Letter to the principal

To whom it may concern,

My name is Conor Martin and I am a fourth year physiotherapy student in the University of Limerick. As part of final year, students are required to undertake a project relating to an area of their choice. My chosen area is paediatrics and the title of my project is, 'To assess the reliability of the Movement Assessment Battery for Children-2 when used in different settings, the school and the clinic'. The aim of the study is to ascertain whether environment has a role to play in a child's motor performance, thus influencing the degree to which results obtained on a measure of motor skills are transferable between these settings. In order to conduct this project I am looking to recruit ten to twenty pupils from your school aged between 7 and 10 years. It would be greatly appreciated if you were prepared to facilitate the undertaking of this project in your school.

Children will be required to complete the following tasks: 1.) Placing pegs 2.) Threading lace 3.) Drawing a trail 4.) Catching with two hands 5.) Throwing a bean bag onto a mat 6.) One-board balance 7.) Walking heel-to-toe forward 8.) Hopping on mats. Children would be tested once in the school, and once in a clinical setting close to your school one week later. Testing should take no longer than 40 minutes per child on each occasion and children will be accompanied by myself and one other investigator to the clinical setting. At all times testing will be carried out under the supervision of Dr. Amanda Connell (PhD, MSc, MCSP, MISC), a lecturer from the Physiotherapy Department in the University of Limerick. The data collected will remain anonymous and be stored in a locked drawer in the Health Sciences Building in the University of Limerick when testing is completed. Participation is entirely voluntary and informed consent will be sought from each student and their respective parent/guardian(s) – please find enclosed the information sheet. If at any stage the student or their parent/guardian(s) wishes to withdraw they will be allowed to do so immediately. Ethical approval has been obtained from The University of Limerick Research Ethics Committee allowing the project to take place.

Should you wish your school to partake in this project, I would propose conducting testing sometime in December or at a date convenient for you. Please feel free to contact me or my supervisor at any stage with any queries you may have.

Yours sincerely,

Conor Martin

Conor Martin

Eanna Clifford

Amanda Connell

Email: 0753726@studentmail.ul.ie

0725927@studentmail.ul.ie

amanda.connell@ul.ie

Appendix 2: Subject consent form

Consent Form



UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

- I have read and understood the **subject information sheet**.
- I understand what the project is about, and what the results will be used for.
- I am fully aware of **all** of the procedures involving my child, and of any **risks and benefits** associated with the study.
- I know that my child's participation is voluntary and that he/she can withdraw from the project at any stage without giving any reason.
- I am aware that my child's results will be kept confidential.

Print name: _____

Signed: _____

Date: _____

Tester Signature: _____

Print Name: _____

Second Tester: _____

Print Name: _____

Date: _____

Appendix 3: Information Leaflet

What is the study about?

There are many measures currently being used to assess motor skills/co-ordination in children. This study aims to assess how reliable one of these measures, the Movement Assessment Battery for Children-2 is.

Where and when?

Testing will take place during school hours, once in the school, and once in a clinical setting 1 week later, with as little disruption to class time as possible. Your child will be accompanied to the clinical setting by the project investigators and will be supervised at all times.

What will I have to do?

If you wish for your child to participate in the study, please sign the attached consent form. You will then be notified when the testing sessions will take place. Your child will have to complete a series of simple tasks in the categories of dexterity, aiming and catching, and balance. Such tasks will include placing pegs, throwing bean bags onto a mat, and hopping. Your child's performance in these tasks will be recorded on the M-ABC-2 scale and you may request information about your child's performance when the study is completed. The tasks will include:

Manual Dexterity

- 1.) Placing pegs
- 2.) Threading lace
- 3.) Drawing a trail

Balance

- 4.) Catching with two hands
- 5.) Throwing a bean bag onto a mat

Aiming and catching

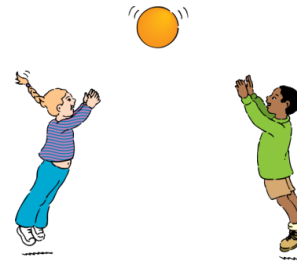
- 6.) One-board balance
- 7.) Walking heel-to-toe forwards
- 8.) Hopping on mats

What are the benefits?

Your child will be providing valuable information about the reliability of the M-ABC2.

Should your child display any difficulties with the tasks, these will be brought to your attention and can be dealt with.

What are the risks?



Testing involves simple tasks, no different from playtime activities in school and is non-invasive and should therefore have no serious risks associated with it.

Should your child wish at any point to discontinue the testing, he/she is fully entitled to withdraw without giving a reason.

What if I do not want to take part?

Your child may only take part in this project with your permission

Participation is completely voluntary and you are under no obligation to take part.

Parents/Children can withdraw from the study at any stage without giving a reason.

What happens to the information?

Any information gathered during the study will remain confidential and anonymous. Participants will not be referred to by name anywhere in the study. At the end of the study information gathered will be used as part of a Final Year Physiotherapy project in the University of Limerick. Upon completion of the study, the information will be stored in a locked cabinet in an office in the Health Sciences Building in the University of Limerick for 7 years in accordance with research ethics.

Parents may request information pertaining to their child from the researchers following completion of the study.

Who else is taking part?

Approximately 10-20 primary school children, aged 7 to 10 years will be recruited from the school.

What if I have more questions or do not understand something?

If you require any further information or do not understand something you can at any point contact the project investigators through the contact details below or ask the project investigators on the day of testing.

What happens if I change my mind during the study?

As participation is voluntary you are free to withdraw from the study at any time for any reason with no explanation required.



What to do?

If you would like your child to participate in the study, please fill out the adjoining consent form and return it to the school as soon as possible.

Thank you for taking the time to read
this leaflet

Contact name and number of Project Investigators.

Amanda Connell Amanda.Connell@ul.ie

Eanna Clifford 0753726@studentmail.ul.ie

Conor Martin 0725927@studentmail.ul.ie

'If you have concerns about this study and wish to contact someone independent, you may contact'

*The Chairman of the University of Limerick
Research Ethics Committee,*

*C/o Vice President Academic and Registrar's
Office,*

University of Limerick,

Limerick.

Tel: (061) 202022



Appendix 4: Raw Data

Rater 1: Day 1

	Pegs pref	Pegs non-pref	Lace	Trail	Manual Dexterity	Catch	Throw	Aim & Catch	OBB best	OBB other	Walking HT	Hop best	Hop other	Balance	Total test score
Subject	Raw score				Component score										
1	33(6)	36(7)	34(6)	1(6)	25(8)	10(15)	7(11)	26(14)	26(12)	14(11)	11(6)	3(3)	2(4)	36(14)	87(12)
2	37(3)	36(7)	32 (7)	1 (6)	23 (7)	9 (12)	4 (5)	17 (9)	7(6)	6(9)	8(4)	3(3)	2(4)	26(8)	66(8)
3	33 (6)	35 (8)	30 (8)	0 (12)	34 (12)	7 (9)	7 (11)	20 (10)	38 (13)	33(15)	10 (5)	3 (3)	3 (5)	41 (17)	95 (14)
4	33(6)	39 (7)	33 (7)	4 (1)	21(6)	8 (10)	2 (3)	13 (6)	27(12)	9 (11)	11 (6)	3 (3)	3 (5)	37 (15)	71(8)
5	31(7)	39(7)	29(8)	1 (6)	28(9)	9 (12)	8 (11)	23 (12)	14(9)	13 (11)	10 (5)	4 (4)	3 (5)	34 (11)	85(11)
6	32 (9)	31 (12)	37 (6)	1 (10)	37 (14)	10 (17)	4 (7)	24 (13)	11 (9)	4 (8)	11 (6)	5 (11)	4 (9)	43 (18)	104 (17)
7	32 (9)	38 (9)	29(11)	2 (8)	37 (14)	9 (15)	3(5)	20(10)	11(9)	7 (10)	10(5)	4 (4)	4 (9)	37 (15)	94(14)
8	31 (9)	34(11)	30 (10)	3 (5)	35 (13)	10 (17)	6 (9)	26 (14)	7 (7)	4(8)	15 (12)	4 (4)	4 (9)	40(17)	101(16)
9	36(4)	39(7)	34 (6)	1 (6)	23(7)	10 (15)	8 (11)	26 (14)	10 (7)	8(10)	10 (5)	4 (4)	2 (4)	30(9)	79(10)
10	44 (5)	40(8)	30 (10)	3(5)	28(9)	9 (15)	7 (11)	26 (14)	5 (6)	4 (8)	15 (12)	3 (3)	3 (6)	35 (12)	89(12)
11	38 (6)	47 (7)	34 (8)	5 (2)	23 (7)	10 (17)	3 (5)	22 (12)	6(7)	6(10)	12 (6)	4 (4)	3 (6)	33(11)	78(10)

Rater 2: Day 1

	Pegs pref	Pegs non-pref	Lace	Trail	Manual Dexterity	Catch	Throw	Aim & Catch	OBB best	OBB other	Walking HT	Hop best	Hop other	Balance	Total test score
Subject	Raw score				Component score										
1	32 (7)	37 (7)	33 (7)	1 (6)	27 (9)	10 (15)	8 (11)	26 (14)	26 (12)	15 (12)	10 (5)	3 (3)	3 (5)	37 (15)	90 (13)
2	37 (3)	36(7)	32 (7)	1 (6)	23 (7)	9 (12)	4 (5)	17 (9)	6 (5)	4 (7)	8 (4)	3 (3)	2 (4)	23 (7)	63 (7)
3	33 (6)	35 (8)	30 (8)	0 (12)	34 (12)	7 (9)	7 (11)	20 (10)	38 (13)	32 (15)	10 (5)	3 (3)	3 (5)	41 (17)	95 (14)
4	32 (7)	39 (7)	33 (7)	4 (1)	22 (7)	8 (10)	2 (3)	13 (6)	28 (12)	9 (11)	11 (6)	3 (3)	3 (5)	37 (15)	72 (8)
5	31 (7)	38 (7)	28 (9)	1 (6)	29 (10)	9 (12)	8 (11)	23 (12)	13 (9)	13 (11)	10 (5)	4 (4)	3 (5)	34 (11)	86 (12)
6	32 (9)	31 (12)	37 (6)	1 (10)	37 (14)	10 (17)	4 (7)	24 (13)	11 (9)	4 (8)	11 (6)	5 (11)	4 (9)	43 (18)	104 (17)
7	32 (9)	38 (9)	28 (11)	2 (8)	37 (14)	9 (15)	4 (7)	22 (12)	10 (8)	7 (10)	11 (6)	4 (4)	4 (9)	37 (15)	96 (15)
8	31 (9)	33 (11)	30 (10)	3 (5)	35 (13)	10 (17)	6 (9)	26 (14)	7 (7)	3 (7)	15 (12)	4 (4)	4 (9)	39 (16)	100 (16)
9	36 (4)	40 (6)	34 (6)	1 (6)	22 (7)	10 (15)	8 (11)	26 (14)	10 (7)	6 (9)	10 (5)	4 (4)	2 (4)	29 (9)	77 (9)
10	44 (5)	39 (9)	30 (10)	4 (3)	27 (9)	9 (15)	7 (11)	26 (14)	5 (6)	4 (8)	15 (12)	3 (3)	3 (6)	35 (12)	88 (12)
11	38 (6)	47 (7)	34 (8)	5 (2)	23 (7)	10 (17)	3 (5)	22 (12)	5 (6)	4 (8)	12 (6)	4 (4)	3 (6)	30 (9)	75 (9)

Rater 1: Day 2

	Pegs pref	Pegs non-pref	Lace	Trail	Manual Dexterity	Catch	Throw	Aim & Catch	OBB best	OBB other	Walking HT	Hop best	Hop other	Balance	Total test score
Subject	Raw score				Component score										
1	33(6)	41(6)	31(8)	2(5)	25(8)	10(15)	5(6)	21(11)	15(9)	5(8)	10 (5)	3 (3)	3 (5)	30(9)	76(9)
2	33(6)	40(6)	31(8)	2(5)	25(8)	8(10)	6(8)	18(9)	5(4)	4(7)	7(3)	3(3)	3(5)	22(6)	65(7)
3	33 (6)	32(10)	25(11)	2(5)	32(11)	9(12)	7(11)	23(12)	35(13)	21(13)	10 (5)	3 (3)	3 (5)	41 (17)	96(15)
4	39(3)	40(6)	29(8)	9(1)	18(5)	8 (10)	5(6)	16(8)	23(11)	4(7)	11 (6)	4(4)	3 (5)	33(11)	67(7)
5	29(8)	32(10)	28 (9)	5(1)	28(9)	10(15)	7(11)	26(14)	15(9)	9(11)	11(6)	4 (4)	2(4)	34 (11)	88(12)
6	40(6)	41(8)	33(8)	2(8)	30(10)	9(15)	8(12)	27(15)	26(12)	13(12)	12(6)	5(11)	3(6)	47(19)	104 (17)
7	30(10)	37(9)	29(11)	0(12)	42(18)	7(10)	5(9)	19(10)	7(7)	5(9)	11 (6)	4 (4)	3(6)	32(10)	93(14)
8	32(9)	32(12)	29(11)	2(8)	40(16)	10 (17)	5(9)	26 (14)	17(10)	5(9)	15 (12)	4 (4)	3(6)	41(17)	107(18)
9	31(7)	43(6)	27(9)	1 (6)	28(9)	10(15)	9(14)	29(16)	8(6)	6(9)	10 (5)	3(3)	3(5)	28(9)	85(11)
10	39(6)	37(9)	31(9)	3(5)	29(10)	10(17)	6(9)	26 (14)	5 (6)	4 (8)	15 (12)	3 (3)	3 (6)	35 (12)	90(12)
11	33(8)	46(7)	31(9)	9(1)	25(8)	9(15)	4(7)	22 (12)	7(7)	4 (8)	10(5)	3(3)	3 (6)	29(9)	73(9)

