Chien, Chi-Wen (2007) ***Using the Rasch model to validate the Peabody Developmental Motor Scales-second edition in infants and pre-school children.*** **Masters (Research) Thesis, James Cook University.**

https://doi.org/10.25903/pmf0%2D7n42

USING THE RASCH MODEL TO VALIDATE THE PEABODY DEVELOPMENTAL MOTOR SCALES-SECOND EDITION IN INFANTS AND PRESCHOOL CHILDREN

Thesis submitted by

CHI-WEN CHIEN BSc

in January 2007

for the degree of Master of Education with Honours

in the School of Education

James Cook University

STATEMENT OF ACCESS

I, the undersigned, author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and I do not wish to place any further restriction on access to this work.

______________________	18/Sep./2007
Signature	Date

STATEMENT OF SOURCES

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Signature

18/Sep./2007

Date

STATEMENT OF THE CONTRIBUTION OF OTHERS

This study was financially supported by the student stipend supports and the Internal Research Award from the School of Education, the Graduate Research Scheme fund from the Faculty of Arts, Education and Social Science, and the Graduate Research International Travel Award from the Graduate Research School, at the James Cook University.

The editorial assistance, including grammar checking and modification of sentence structures, was provided by Ms. Vanessa Carson who was a psychology student at the James Cook University. In addition, Professor Michael Linacre provided part of statistical support and technical counseling about the use of the WINSTEPS software.

There are several kindergartens that participated in the study to contribute to collection of normal children, including the Pei-You, Sin-Wang-Ai, Yi-Guo, Bao-Bei-Syong, and Chen-Sing kindergartens in Taiwan. Moreover, the Yu-Min Hospital, Chung Shan Medical University Hospital, Cheng Hsin Rehabilitation Medical Center, and National Taiwan University Hospital were involved in the study to contribute to recruitment of children with motor delays or difficulties.

ACKNOWLEDGEMENTS

I am grateful to a number of people who have provided assistance and support with this study. In particular, I would like to thank my supervisor, Professor Trevor Bond of The Hong Kong Institute of Education, for his support and continued supervision in the study. His insights into Rasch model as well as his critical comments on the data analyses have been the invaluable influence on the study. If Professor Bond had not made the critical meetings available to me, the study would not have developed to produce informative and valuable findings.

Professor Ching-Lin Hsieh in the National Taiwan University was my previous supervisor at work and became a friend. His supports of my data collection in the rehabilitation department of the National Taiwan University Hospital allowed me to recruit successfully the children with disabilities. Additionally Ms. Vanessa Carson is the host of my home stay family, taking care of my daily life when I studied in Australia. Since then, I have deeply appreciated her continuous care as well as guidance in my communication skills, as currently she is my closest friend in Australia.

Finally, my family, Ms. Chia-Yi Liao, and many friends in Taiwan need to be acknowledged for their support and encouragement during the study period. All of the children and their parents involved in the study are also deeply appreciated from my heart.

ABSTRACT

The current research utilized the Rasch model analysis to examine the dimensional structure of the Peabody Developmental Motor Scales-Second Edition (PDMS-2). Furthermore, the three-point rating scales, differential item functioning (DIF), and hierarchical structures of the PDMS-2 items were examined for their utility in discriminating various levels of motor development, for items' function stability across gender and disease entity, and for positioning in a valid hierarchy of difficulty.

The study tested a total of 419 children in Taiwan (including 342 normal children and 77 children with motor delays or difficulties) using the PDMS-2. The three-point rating scales of 180 PDMS-2 items exhibited problems (such as infrequently used categories and disordering step calibration), and thus the rating scales for these items were collapsed to allow only dichotomous responses. Each of the six PDMS-2 subtests formed a unidimensional scale after 21 misfitting items were removed. Additionally, the gross motor, fine motor, and overall motor ability were constructed by combining certain subtests, supporting the theorized dimensionality of the PDMS-2. Fifty-eight items within the subtests demonstrated DIF between children with/out motor problems, while only 35 items demonstrated DIF across gender. The hierarchical order of the PDMS-2 items established using Rasch model showed considerable similarity to the original hierarchy that was ranked by age. The PDMS-2 items had wide coverage but inadequate targeting of the children in the study.

The findings from the research indicated that the reduced PDMS-2 test encompassing dichotomous rating scales in the particular items, is a valid measure of motor development in infants and preschool children. However, further work is needed to improve the inadequate targeting by adding more suitable items.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION AND STATEMENT OF THE PROBLEM

CHAPTER TWO

LITERATURE REVIEW

CHAPTER THREE

METHODOLOGY

CHAPTER FOUR

RESULTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

AND

STATEMENT OF THE PROBLEM

1.1 Context of the present study

Accurately evaluating a child's level of development and functional abilities is a prerequisite for being proficient pediatric professionals (Case-Smith & Allen, 2005), such as pediatricians, occupational and physical therapists, developmental psychologists, and special educators. It is usual for many of these pediatric professionals to gather an overall understanding of their clients with an assortment of standardized assessment devices, e.g., the Peabody Developmental Motor Scales-Second Edition (PDMS-2) (Folio & Fewell, 2000) for assessing motor development. The PDMS-2, being a standardized assessment tool, provides useful information; however many clinicians apparently have reservations about the use of the PDMS-2. The PDMS-2 requires 45-60 minutes to administer (Folio & Fewell, 2000), but this timeframe is not usually achievable, especially when assessing children with motor delays or difficulties. This presents a problem in that additional assessment periods are often required. Unfortunately, those children's parents or caregivers often dislike or even refuse to extend the assessment time, believing that such a time-consuming evaluation is too demanding on their children. This highlights that using a time efficient assessment tool is crucial to complete many evaluations successfully.

Clinicians have also reported a trend towards more time-consuming assessment tools (Lin & Tseng, 2001). Clinicians are beginning to recognize that they are unable to complete lengthy developmental assessment tests within their predetermined time limits. According to a survey conducted in Taiwan (Lin & Tseng, 2001), the frequency of assessment tools being used by pediatric occupational therapists was much less in Taiwan than in the United Sates, Canada, and Australia. This may be due to the problem of the time-consuming evaluations in clinical settings, or perhaps that Taiwanese therapists tended to identify children's problems subjectively and performed treatment without standardized evaluation. This research aims at moving on pediatric assessment tool away from lengthy testing and towards a more concise framework to reduce the administration time of assessment.

A concise assessment tool, in principle, should make administration easier for clinicians, in order to increase the likelihood of completion and require an appropriately minimal effort by the clinician, child, and/or caregiver. However, the development of a concise assessment tool often condenses the original test and may compromise the range and specificity of its content, leading to inaccurate evaluations, e.g., increasing measurement error and contributing to floor and ceiling effects (Ware, 2003). General health and rehabilitation fields have utilized computerized adaptive testing (CAT), which offers a potential solution to the conflict between conciseness and accuracy (Dijkers, 2003; Jette & Haley, 2005; Revicki & Cella, 1997; Ware, 2003). CAT, entails computer selection of questions, directly tailored to the child's ability, which routinely can reduce the required assessment time (Gershon, 2005). Moreover, CAT allows the child's level of ability to be estimated as precisely as stipulated by the assessment requirements.

CAT applications require a well-developed item bank, which contains a set of calibrated items along a hierarchical unidimensional scale providing a comprehensive difficulty range. Consequently, the Rasch measurement model has been used extensively as the mathematical foundation for CAT (Gershon, 2005; Haley, Ni, Fragala-Pinkham, Skrinar, & Corzo, 2005). The Rasch model has an advantage of being able to substantiate whether the items on an assessment tool contribute to a single unidimensional construct, calibrating the difficulty of each item as well as examining the appropriateness of the rating scale's scoring system (Bond & Fox, 2001, 2007; Wright & Masters, 1982; Wright & Stone, 1979).

Since the structure and psychometric properties of the PDMS-2 remain unclear, this study adopted the use of the Rasch model to examine the validity of the PDMS-2. This constituted the main rationale for this thesis. Justification for research in this area is demand driven, highlighting the need for standardized gross/fine motor tests with sound psychometric properties in Taiwan (Lin & Tseng, 2001). The validation of the PDMS-2 using the Rasch model ought to provide more accurate measurement of motor skills for routine clinical use, and, in the future, lead to the development of a CAT version that could offer ever more efficient assessment of motor development.

1.2 The current problems

An assessment tool, to be clinically useful, must be scientifically sound in terms of psychometric properties (Streiner, 1995; Wade, 1992). Although the psychometric properties of the PDMS-2 have been examined using conventional test theory (Connolly, Dalton, Smith, Lamberth, McCay, & Murphy, 2006; Folio & Fewell, 2000; van Hartingsveldt, Cup, & Oostendorp, 2005; Wang, Liao, & Hsieh, 2006), there are at least four issues that still require further examination: (1) the structure of the PDMS-2 has not been thoroughly resolved, (2) the PDMS-2 consists of ordinal, rather than interval, level scales, (3) difficulty estimates of each item have not been genuinely calibrated, and (4) the appropriateness of the currently-used three-point rating scales for each of the PDMS-2 items awaits thorough validation.

In order to address the above-mentioned issues, this research applied the Rasch model in examining the structure as well as the psychometric properties of the PDMS-2. The process was expected to provide a more accurate measurement of motor development for routine clinical use with children with/out motor delays or difficulties in Taiwan.

CHAPTER TWO

LITERATURE REVIEW

2.1 Importance of assessing motor development in children

Motor abilities of infants (birth to 2 years) and preschool children (2 to 6 years) are fundamental to a child's general development (Edwards & Sarwark, 2005; Gallagher & Cech, 1988). From the newborns' reflex movement, to the infants' first attempts at crawling and walking, through to the toddlers' climbing skills, motor ability plays a crucial role in children's exploration of the ever-changing environment. Successful exploration of the new environment permits children to learn the relationships between objects in the environment and their characteristics. Through ongoing interaction with the environment, children develop higher-levels of functional skills enabling them to adapt to environmental changes. Most children usually develop their motor abilities and engage in active exploration of their environment with vigor and enthusiasm; however many children experience motor delays or difficulties at an early age.

Normal development of motor ability contributes significantly to the degree to which a child participates successfully in motor activities (Gallagher & Cech, 1988). However a significant percentage of children suffer congenital motor delays or difficulties, such as Cerebral Palsy, Down's syndrome, Spina Bifida, and other conditions affecting normal motor development. Such abnormal motor development causes deprivation of normal participation in early motor activities. If such deprivation continues unabated, some degree of high-level functional impairment may ensue and in turn, lead to impaired cognition (Fennell & Dikel, 2001), lower levels of academic achievement, as well as limited capacities in dealing with activities of daily living (American Psychiatric Association, 1994; Kadesjo & Gillberg, 1998), and even serious pervasive developmental delays (Wachs & Sheehan, 1988).

To alter the developmental course of children with motor delays or difficulties, early intervention with multidisciplinary services from domains such as pediatrics, occupational therapy, physical therapy, developmental psychology, and education, is imperative to minimize the likelihood of compounding consequences (Blauw-Hosper & Hadders-Algra, 2005). Studies have found that early identification of, and intervention with, children with motor delays or difficulties at an early stage can create opportunities to improve their motor development and prevent possible complications (Blauw-Hosper & Hadders-Algra, 2005; Buccieri, 2003; Shonkoff & Hauser-Cram, 1987). Furthermore, the effect of early intervention can strengthen the relationship between parent and child by raising the parents' confidence in their ability to care for their children outside of ongoing therapy.

However, early identification of children with motor delays or difficulties is dependent on the accurate assessment of children's motor abilities (Case-Smith, 1998; Gallagher & Cech, 1988) to screen for children at risk and to implement timely intervention to minimize further developmental delay (Edwards & Sarwark, 2005; Gallagher & Cech, 1988). Clinicians also acknowledge that accurately assessing motor development is critically important not only for planning intervention programs but also for evaluating and monitoring any individual's ongoing progress (Burton & Miller, 1998; Edwards & Sarwark, 2005; Gallagher & Cech, 1988).

2.2 Considerations for selecting appropriate motor development tests

Currently, a variety of tests are employed to assess motor development in children (Burton & Miller, 1998). Nonetheless, professionals from many disciplines are repeatedly confronted with decisions regarding how best to select the most appropriate and beneficial instruments for a presenting child (Hubert & Wallander, 1988; Tieman, Palisano, & Sutlive, 2005). Three major considerations are suggested best to facilitate instrument selection: 1) *considerations of assessment objectives*, 2) *practical considerations*, and 3) *psychometric considerations* (Hubert & Wallander, 1988). Each consideration is briefly described below.

Firstly, the *assessment objectives* should be considered prior to test selection and administration (Burton & Miller, 1998). Most motor development tests are classified as possessing the ability to assess at least three performance indicators: (a) to identify whether the child has a motor developmental delay, (b) to determine program eligibility and/or to plan an intervention and, (c) to evaluate change over time. Clinicians should familiarize themselves with the purposes of an assessment tool by reading the test manual prior to assessment and conferring with informed colleagues to confirm the test's applicability (Wiart & Darrah, 2001). If a test, e.g., the Denver Developmental Screening Test (Frankenburg & Dodds, 1967, 1969) is designed to identify only those children who might have a certain type of developmental delay, therapists should not use that test to monitor a child's change over time. Otherwise, test information might adversely result in irrelevant test interpretation and inappropriate therapy programs that waste the time of both the child and therapist. Furthermore, reliance on such inaccurate information might render both evaluation and intervention ineffective.

The second consideration in selecting an appropriate motor development test involves *practical considerations*. Numerous practical considerations can determine the successful completion of assessment, such as the suitability of the assessment tool for children with certain handicaps, the ease of administration, and the comprehensiveness of the test manual (Hubert & Wallander, 1988). Additionally, Gallagher and Cech (1988) suggested that a practical motor development test should encompass a broad repertoire of motor abilities, in order to provide comprehensiveness of content for assessing children with motor delays or difficulties. One of the practical considerations in selecting a motor development test for children is the age range targeted by the test. The selected test needs to satisfy goodness of fit with the relevant age group to maintain high validity in evaluation.

Psychometric considerations (i.e., evidence of reliability and validity) are paramount factors that assist in the determination of test selection. An assessment tool that is scientifically sound in terms of reliability and validity contributes significantly to accurate assessment. Then developmental tests of motor ability need to possess sound reliability and validity (Burton & Miller, 1998). Generally, reliability is the extent to which an assessment tool is consistently free from error, in order to have confidence that a score is close to the "true score" (Nunnally & Bernstein, 1994). Validity, in general terms, is the extent to which a test actually measures what it is intended to measure (Burton & Miller, 1998; Wiart & Darrah, 2001). Inferences made from test results rely highly on the establishment of reliability and validity. Motor development tests need to be thoroughly examined for reliability and validity prior to their general clinical application.

2.3 Peabody Developmental Motor Scales-Second Edition (PDMS-2)

The selection of an appropriate motor development test is dependent not only on clinical practicality and stringency of psychometric properties, but also, and perhaps more importantly, the clinicians' intended purposes for assessment. With this in mind a variety of motor development tests have been developed targeting specific applications (Burton & Miller, 1998), such as the Bayley Scales of Infant Development (BSID) (Bayley, 1969), the Movement Assessment Battery for Children (MABC) (Henderson & Sugden, 1992), and the Peabody Developmental Motor Scales (PDMS) (Folio & Fewell, 1983). However, a trend of highlighting early intervention with children with disabilities, resulted in the PDMS being one of the most commonly-used tests used to assess motor development in infants and preschool children (Brown, Rodger, Brown, & Roever, 2005; Burtner, McMain, & Crowe, 2002; Burton & Miller, 1998; Lawlor & Henderson, 1989), and the PDMS is regarded as providing useful and comprehensive information for early intervention (Case-Smith & Allen, 2005; Tieman et al., 2005).

The original version of the PDMS was published in 1983 by Folio and Fewell. The PDMS, unlike the aforementioned MABC which can be used only as a screening test, was developed for a varied purposes, including determining the relative developmental skill level of a child, identifying skills that were not completely developed or not in the child's repertoire, and planning an instructional program to develop those skills (Folio & Fewell, 1983). In addition, the PDMS is a standardized, norm-referenced test that assesses fine motor (FM) and gross motor (GM) attributes of children from birth to 83 months of age. The repertoire of motor skills and the comprehensive age range incorporated in the PDMS are broader than that of the BSID, which does not provide for separate assessment of FM and GM functions, and can be used with children ranging in age from 0-3.5 years only. Furthermore, the psychometric properties of the PDMS have been routinely examined with conventional test theory (Folio & Fewell, 1983; Gebhard, Ottenbacher, & Lane, 1994; Kolobe, Palisano, & Stratford, 1998; Palisano, 1986; Palisano, Kolobe, Haley, Lowes, & Jones, 1995; Palisano & Lydic, 1984; Stokes, Deitz, & Crowe, 1990). All of these properties help to establish the PDMS as a well-credentialled test to assess motor development for early intervention.

The PDMS has had extensive usage and scientific grounding. However, when the PDMS was developed, Folio and Fewell, the test developers, did not adhere to any specific theoretical perspective as its conceptual foundation. Following practical application the developers saw a need to revise the PDMS, and developed the PDMS-Second Edition (PDMS-2) in 2000. The revised PDMS-2 adopted a general developmental framework and incorporated the use of both qualitative and quantitative approaches to assessment (Folio & Fewell, 2000). Qualitative refers to how well the child performs the skill relative to using the correct components, and quantitative refers to how many of those skills the child is able to perform.

The revised PDMS-2 differed from the original version in that the PDMS-2 can be divided into six new subtests: Reflex, Stationary, Locomotion, Object Manipulation, Grasping and Visual-Motor Integration, but the PDMS-2 still incorporates the FM and GM attributes. The Grasping and Visual-Motor Integration subtests when combined, form the FM scale, and the Reflex, Stationary, Locomotion, and Object Manipulation subtests constitute the GM scale. A combination of the results of the FM and GM scales forms a total motor quotient, which can be further used to indicate the overall motor ability of the child. Additionally, each PDMS-2 item employs a three-point rating scale with specific scoring criteria. The specific scoring criteria for such three-point rating scales were not part of the original PDMS and were seen as an integral adjustment required to assist examiners' subjective assessments.

2.3.1 Clinical merits of the PDMS-2

According to the developers Folio and Fewell (2000), the new PDMS-2 has extensive clinical application, which meets most clinicians' needs for assessing motor development. Furthermore, the PDMS-2 includes the broad scope of motor repertoire by measuring the FM and GM functions, and could assist either early detection or longitudinal monitoring of motor delays and difficulties as it can be employed across the age range from birth to 72 months. In addition, the PDMS-2 has the practical value of being suitable to assess various populations of children, who otherwise would be at risk of motor delays or known physical disabilities (Folio & Fewell, 2000). Simplifying administration procedures and providing a comprehensive test manual are also important clinical merits considered to shorten the testing time and to provide technical and psychometric information for clinical use, respectively.

2.3.2 Psychometric properties of the PDMS-2

An assessment tool, to be clinically practical, must be scientifically sound in terms of its psychometric properties (Hobart, Lamping, & Thompson, 1996). Using conventional test theory, Folio and Fewell (2000) have found the PDMS-2 to have satisfactory reliability (i.e., internal consistency [$0.71 \leq$ Cronbach's alpha ≤ 0.98], test-retest reliability [$0.73 \leq$ Pearson r correlation coefficient ≤ 0.96], and inter-rater reliability [$0.73 \leq r \leq 0.96$]) and acceptable validity (i.e., content validity, concurrent validity [$0.84 \leq r \leq 0.91$, by correlation with the original PDMS] and criterion validity [$0.55 \leq r \leq 0.86$, by correlation with the Mullen Scales of Early Learning: AGS version]).

Recently, high test-retest reliability (0.84 ≤ Spearson ρ correlation coefficient ≤ 0.98), high inter-rater reliability (0.94 ≤ ρ ≤ 0.99), acceptable convergent validity (ρ = 0.69 with the MABC test), and discriminant validity of the FM scale in the PDMS-2 were re-confirmed in children with/out mild FM problems (van Hartingsveldt et al., 2005). Wang et al., (2006) also confirmed that the PDMS-2 had high test-retest reliability (0.88 ≤ intraclass correlation coefficient ≤ 1.0) and further confirmed its responsiveness, namely the ability to detect change, in children with cerebral palsy. Moreover, modest relationships ($0.67 \leq r \leq 0.76$ and $0.22 \leq r \leq 0.32$, respectively) between the PDMS-2 and the BSID-Second Edition were found in children with developmental delays (Provost, Heimerl, McLain, Kim, Lopez, & Kodituwakku, 2004) and typically developing infants (Connolly et al., 2006).

In addition, the construct validity (i.e., dimensional structures) of the PDMS-2 has been supported using confirmatory factor analysis (Folio & Fewell, 2000), and the item characteristics such as item difficulty and discrimination power for each PDMS-2 item, were further examined in the use of a two-parameter model in item response theory (IRT) (Folio & Fewell, 2000); however, the details of the IRT analyses were not reported.

2.4 Item response theory: the Rasch model

Item Response Theory (IRT) was originally developed in the field of education, but its application has recently been extended into medical and healthcare assessment (Cella & Chang, 2000; Haley et al., 2005; Hays, Morales, & Reise, 2000; McHorney & Cohen, 2000; McHorney, Haley, & Ware, 1997; Revicki & Cella, 1997). IRT, also generally known as the latent trait theory, models the relationship between subjects' underlying ability or trait and their responses to test items (Zhu & Kurz, 1994). According to the number of parameters accounted for, IRT can be divided into three models: the Rasch model articulated by Rasch (1960), a 1-parameter model, and the others developed by Lord and Novick (1968) and Birnbaum (1968) which are 2- and 3-parameter models. Although these models are considered part of IRT, the Rasch model and the 2- and 3-parameter IRT models differ in their philosophical underpinning (Andrich, 2004; Hays et al., 2000).

The Rasch model is based on a simple logistic model, which assumes that the probability of passing an item is dependent only on the ability of the subject and the difficulty of the item. On the other hand, the other IRT models mentioned previously add additional parameters for item discrimination and subjects' guessing, to find a model that best explains the observed rating scale data. The Rasch model, however, aims to determine the extent to which such observed data satisfy stringent model assumptions; thereby the Rasch model is claimed to conform better with the idea of fundamental measurement (Bond & Fox, 2001, 2007; Wright & Stone, 1979).

The Rasch model has been extensively used in developing new motor ability instruments (Campbell, Wright, & Linacre, 2002) and validating existing measures of motor skills assessment (Avery, Russell, Raina, Walter, & Rosenbaum, 2003; Ludlow & Haley, 1992; Zhu & Cole, 1996; Zhu & Kurz, 1994). In one such example was Zhu and Kurz (1994) used the Rasch model to calibrate and analyze a clinical measure assessing gross motor competence. They demonstrated many advantages of the Rasch model in examining the assessment of motor skills, such as validating the construct of the clinical measure and examining the appropriateness of the rating scale. Campbell et al. (2002) used the Rasch model not only to examine the psychometric qualities of the Test of

Infant Motor Performance, but also to generate a shorter version of the test by eliminating the redundant items. Avery et al. (2003) used the Rasch model to validate a well-known scale called "Gross Motor Function Measure (GMFM)" and created an interval-level measure, by passing the limitation of using ordinal-level scaling in conventional test theory.

In consideration those above studies, the Rasch model has demonstrated its suitable potential for application within the area of motor development and has shown the following advantages in establishing the psychometric qualities of motor development tests.

Primarily the Rasch model has the advantage of being able to examine whether items from a scale measure the unidimensional construct (Rasch, 1960; Wright & Masters, 1982). Unidimensionality means that all test items within the instrument, measure a single latent trait or ability, comprising a clinically meaningful continuum of activities ranging from relatively easy to difficult to accomplish (Bond, 2004; Bond & Fox, 2001, 2007). For example, if all items from a motor development test fit the intended structure of the Rasch model, unidimensionality of motor development in the test is preserved. Therefore, researchers and clinicians can feel confident when interpreting the summary scores of the test items as the unidimensional trait or ability, i.e., motor ability in this case. Additionally, the Rasch model provides fit statistics, which help identify items that may not be contributing to a unidimensional construct. Given the potential reasons for test items being identified as misfitting, the Rasch model enables test developers to either revise or eliminate those items (Campbell, Osten, Kolobe, & Fisher, 1993). Likewise, the Rasch model provides similar fit statistics for detecting misfitting persons and allows researchers to investigate whether the persons responded to the items according to the underlying construct that the test measures or whether additional answering mechanisms apply (Hays et al., 2000; Meijer, 2003).

The Rasch model also possesses the considerable advantage of transforming ordinal scores into the logit scale, an interval-level measurement scale, if the test items fit the Rasch model expectations (Bond & Fox, 2001, 2007; Rasch, 1960; Wright & Masters, 1982). In contrast to ordinal rating scales, the logit scale can accurately reflect the true magnitude of the difference between two estimates, thereby making possible reliable and accurate comparisons of changes among subjects or changes over time for an individual subject. Furthermore, item difficulty and person ability are genuinely calibrated on a common logit scale (Bond & Fox, 2001, 2007; Perline, Wright, & Wainer, 1979). As a result, the Rasch model can provide a helpful item-person map for researchers and clinicians to compare how well the items are matched to the given sample.

Rasch model analysis is particularly useful in examining whether items have construct meanings, or whether differential item functioning (DIF) exists across different populations (Bond & Fox, 2001, 2007; Chen, Bode, Granger, & Heinemann, 2005). DIF, sometimes referred to as a bias, occurs when an item does not provide consistent measurement across different groups of people who otherwise have the same values on the underlying trait (Holland & Wainer, 1993). The bias can result from any number of factors, including important demographic characteristics of a respondent, such as sex and disease entity. DIF analysis using the Rasch model can help identify items with significant bias to ensure that the test provides an accurate assessment of motor development across subgroups included in the sample (e.g., boys vs girls). DIF analysis should also enable identification of which items have different levels of difficulty between children with/out motor delays or difficulties. Those identified items are potentially useful indicators of screening out children at risk of motor delays or difficulties in clinical settings.

Finally, the Rasch model allows researchers and clinicians to examine whether or not the scoring points of a rating scale are appropriate (Bond & Fox, 2001, 2007; Pesudovs & Noble, 2005; Zhu & Kurz, 1994). Many motor development tests employ multiple-point rating scales (Zhu & Kurz, 1994), such as the 10-point rating scales of the Tufts Assessment of Motor Performance (Ludlow & Haley, 1992), 4-point rating scales of the GMFM (Avery et al., 2003), and 3-point rating scales of the PDMS-2 (Folio & Fewell, 2000). However, these multiple-point rating scales assume that equal distances between response options represent equal distances along the dimension of motor development. This assumption, although universally made, is almost certainly erroneous in most cases (Massof & Fletcher, 2001; Pesudovs & Noble, 2005). The Rasch model provides a useful analysis for testing the rating scale assumptions and, in particular, allowing modification of the rating scale structure.

In summary, the application of the Rasch model has demonstrated various advantages of validation in tests of motor development. Furthermore, the Rasch model also holds promise for improving tests by detecting misfit items/persons as well as modifying the rating scales.

2.5 Drawbacks of the PDMS-2

Although the psychometric properties of the PDMS-2 were previously examined using conventional test theory and a 2-parameter IRT model (Connolly et al., 2006; Folio & Fewell, 2000; Provost et al., 2004; van Hartingsveldt et al., 2005; Wang et al., 2006), the PDMS-2 has not been validated from the perspective of the Rasch model. Because the Rasch model approaches the idea of fundamental measurement and has demonstrated superior advantages in validating motor development tests, it was selected to examine the shortfall in research in relation to the following four outstanding issues of validation.

2.5.1 Structure of the PDMS-2

The structure of the PDMS-2 has not been thoroughly examined, even though the overall structure of the six PDMS-2 subtests was supported with confirmatory factor analysis. Traditional factor analysis, however, operates on interval scale scores (Zou, Tuncali, & Silverman, 2003), whereas the PDMS-2 scores are ordinal by nature. Moreover, factor analysis usually assumes that all of the test items should be of approximately the same difficulty, but the difficulties of items from the same dimensional structure are more or less ordered from easy to difficult (Waugh & Chapman, 2005). Thereby, the examination on the basis of the factor analysis suggests that the structuring of the PDMS-2 may be inconclusive and misleading for researchers and clinicians to interpret the results of the PDMS-2. Hence, the dimensional structure of the PDMS-2 six subtests, of FM and GM scales and of the entire PDMS-2 test requires further validation using a more stringent measurement method, such as the Rasch model.

2.5.2 Ordinal scales

Ordinal scales might preclude not only the use of traditional factor analysis, but also adversely affect the application of standard parametric statistical inferences

(Zou et al., 2003). Since raw scores of the PDMS-2 are calculated from ordinal, rather than interval level scales, the same amount of difference in raw scores at two different places is not necessarily equal-interval in a subtest. Furthermore, the raw scores do not have the property of additivity which is a mathematical property required for inferential statistics. For these reasons the validity and reliability results of the PDMS-2, based on the ordinal-rating raw scores in previous studies (Connolly et al., 2006; Folio & Fewell, 2000; Provost et al., 2004; van Hartingsveldt et al., 2005; Wang et al., 2006), could be questionable.

Although raw scores from the PDMS-2 subtests can be further converted into age-specific standardized scores, the standardized scores might have been established on an inappropriate normative population which might not encompass representatives of targeted children (Burton & Miller, 1998; Jacobusse, van Buuren, & Verkerk, 2006). Thus this may adversely affect the comparison of the PDMS-2 results within age. Moreover, such standardized scores fail to have a common metric that allows comparison of developmental changes across age (Jacobusse et al., 2006). In order to deal with these weaknesses, application of the Rasch model will assist transformation from ordinal raw scores to the logit scale providing a basis for comparisons both within and across age categories. This has potential benefits for subsequent parametric statistic analyses.

2.5.3 Item hierarchy

The third issue focuses on the claim that PDMS-2 items are arranged in a developmental sequence of increasing age. Such item hierarchy was established on the robust assumption that children's motor development is absolutely orderly and sequential with age; however this assumption has not always been supported (Burton & Miller, 1998; Shumway-Cook & Woollacott, 1995). Although Folio and Fewell used the 2-parameter IRT model to confirm the present hierarchy of the PDMS-2 items in 2000, the 2-parameter IRT model is not viewed as a definition of measurement (Andrich, 2004; Hays et al., 2000). The 2-parameter model assumes that a person's ability depends on the item's difficulty as well as the item's discrimination power (Andrich, 2004; Rayan, 1983), whereas it remains questionable whether the 2-parameter model can calibrate items by difficulty, when some items have widely varying discrimination powers (Rayan, 1983).

2.5.4 Appropriateness of the three-point rating scales

Finally, the PDMS-2 items employ three-point rating scales with scoring criteria which vary from item to item. The 3-point rating scales allow partial credit for developing motor skills and are particularly advantageous for evaluating the progress of children with developmental delay whose skill acquisition might develop at a very slow rate. However, the appropriateness of three-point rating scales has not been investigated specifically. Whether children can be differentiated by their responses as clearly as the three points allow, it still awaits empirical examination.

2.6 Research questions

To address those aforementioned issues covering the PDMS-2, the study applied Rasch model to examine the structure and psychometric properties of the PDMS-2. The following research questions based on the PDMS-2 are highlighted and await thorough exploration which is intended should increase confidence in the accuracy of such a measure critical to motor development and essential to clinical competence.

2.6.1 Main questions

Question 1: Are the three-point rating scales of each PDMS-2 item appropriate to differentiate children's responses?

Question 2: Is each of the PDMS-2 subtests, FM and GM scales, or the entire PDMS-2 test able to construct a unidimensional measure?

Question 3: For what components of the PDMS-2 can an objective Rasch model interval scale be established?

Question 4: If the PDMS-2 items fit the Rasch model expectations, to what extent does the hierarchy calibrated by the Rasch model differ from the original hierarchy?

2.6.2 Additional questions

Question 5: How well do the PDMS-2 items match to the given sample in the study?

Question 6: What items show DIF between male and female subjects and between children with/out motor delays or difficulties?

Question 7: How reliable are the estimates of the PDMS-2 items and the sample ability in this study?

CHAPTER THREE

METHODOLOGY

3.1 Subjects

Rasch analysis typically requires a sample size larger than 200 to estimate reliably the unidimensionality of a scale (Wright & Stone, 1979), and the sample should characterize a broad range of the given abilities (Bond & Fox, 2001, 2007; Wright & Stone, 1979). In this study, two convenient samples of children were recruited to represent a comprehensive range of motor ability. One, a normative sample, including at least 300 normal children across six different age bands (i.e., birth to 11 month, 12-23 months, 24-35 months, 36-47 months, 48-59 months, and 60-71 months) was recruited from day care centers and kindergartens located throughout central Taiwan. Each age band included at least 50 subjects, and was equally balanced for gender. All subjects of this sample were invited to participate in the study if they met the following criteria: (1) full-term infants (i.e., infants born between 36 weeks and 42 weeks), (2) birth weight of 2500 gm. or more, and (3) absence of any sensory, motor deficits, major diseases (e.g., cancer or heart disease) or body impairments (e.g., amputations or fractures) that would reduce or limit their abilities to perform movement tasks, according to their parents' and caregivers' reports.

The other sample included approximately 60 children with motor delays or difficulties, who were recruited from rehabilitation departments located in central and northern Taiwan. The sample of children contained at least ten subjects at each age band.

All subjects of the sample met the following criteria: (1) diagnosis of a motor-related disorder, such as Cerebral Palsy, Down's syndrome, muscle dystrophy, Spina Bifida, developmental coordination disorder, or developmental delay resulting from Autism, mental retardation or others, and 2) absence of other major physical diseases (e.g., cancer or heart disease) or body impairments (e.g., amputation or fracture) which would reduce or limit their abilities to perform movement tasks. The parents or caregivers of all the subjects in the study were required to give informed consent. The self-reported socioeconomic and educational levels of the children's parents were recorded. In addition, the study was approved by the human ethics committee of James Cook University and the participating hospitals.

3.2 Instruments

The PDMS-2 is a criterion- and norm-referenced scale that contains 249 items. There are eight items in the Reflex subtest, 30 items in the Stationary subtest, 89 items in the Locomotion subtest, 24 items in the Object Manipulation subtest, 26 items in the Grasping subtest, and finally 72 items in the Visual-Motor Integration subtest. Each item employs a three-point rating scale with the following general scoring criteria: 0, if the child cannot or will not attempt the item, or the attempt does not show that skill is emerging; 1, if the child's performance shows a clear resemblance to the item mastery criteria but does not fully meet the criteria; 2, if the child performs the item according to the criteria specified for mastery.

The PDMS-2 developers have reduced the testing time by simplifying the administration procedure (Folio & Fewell, 2000). Such simplified administration procedure involves continuing the PDMS-2 from where the child gets the highest score on three consecutive items (i.e., the basal level) until s/he fails to score at all on three consecutive items (i.e., the ceiling level). However, stopping testing items below the basal level and those above the ceiling level needs to be based on a robust assumption that motor development is orderly, sequential, and reflective in the order of the PDMS-2. However this assumption has not always been supported (Shumway-Cook & Woollacott, 1995). In order to calibrate genuine difficulty estimates for each item, therefore the administration procedure for this research was varied slightly from that proposed by the original developers. In this study, the PDMS-2 items were divided into six age bands, and each child from the normative sample was assessed with the items across three age bands (i.e., the child's current age band plus one age band below and one above the child's current age). For children with motor delays or difficulties, however, due to their motor problems, two age bands below the child's current age but only about half above the child's current age were assessed.

The order of presentation of the PDMS-2 items was reorganized by grouping the items requiring the same test equipment/position (e.g., all the items involving pen and paper) in order to make the administration smooth as well as assisting in maintaining a child's concentration and motivation (Hinderer, Richardson, & Atwater, 1989). In addition, because no Chinese version of the PDMS-2 was currently available in Taiwan, the instructions for each PDMS-2 items were translated by the author.

3.3 Procedure

Each subject was administered the PDMS-2 individually in a quiet room in the absence of the parent or caregiver. For younger children aged from birth to 2 years or children with motor delays or difficulties, their parents/caregivers were allowed to be involved in the PDMS-2 administration. A well-trained, experienced occupational therapist who is the author, administered the PDMS-2. The testing of the PDMS-2 was commenced with the first item corresponding to the child's age. Then, the group of items requiring the same test equipment/position was assessed sequentially. The administration of the PDMS-2 concluded after the child's performance on all of the PDMS-2 items across the adjacent three or three and half age bands had been assessed.

According to the guidelines for administering the PDMS-2 in the test manual, testing children with motor delays or difficulties usually requires a longer period of time than normal children. Therefore to assure complete data collection, the PDMS-2 testing in this study was broken into several shorter sessions if the child had a short attention span or if other conditions made it more convenient to administer separate sessions of the PDMS-2 at different times. Folio and Fewell (2000) recommended that the entire PDMS-2 items be completed within a 5-day interval period to minimize the confounding effects of possible spontaneous growth.

3.4 Data analysis

In this study, the Rasch model was used to examine the structure and psychometric properties of the PDMS-2. With the family of Rasch models, each has different assumptions about the scaling of response options, and so selection of the suitable Rasch model is a prerequisite for conducting the data analysis. Prior to data analysis, a decision process proposed by Avery et al. (2003), was adopted to assist selection of an appropriate model for the current study (Figure 1).

According to the proposed decision making process by Avery et al., the partial credit Rasch model appeared to be the better choice over other Rasch models, since the

PDMS-2 items have more than two response options (i.e., 3-point rating scales) and whether the steps between/within items are evenly spaced were unknown. The partial credit Rasch model makes no assumption about the relative difficulty of response options within items, other than their ordinality (Avery et al., 2003) and enables researchers to yield a person estimate and a difficulty estimate associated with accomplishing each step in an orderly sequence of response options. Moreover, it is the least constrained Rasch model for polytomous data (Bond & Fox, 2001, 2007). Consequently, the partial credit Rasch model was selected as the most suitable Rasch model to perform the following analyses in the current study. All of the Rasch analyses in this study were conducted using WINSTEPS computer software version 3.61.1 (Linacre, 2006).

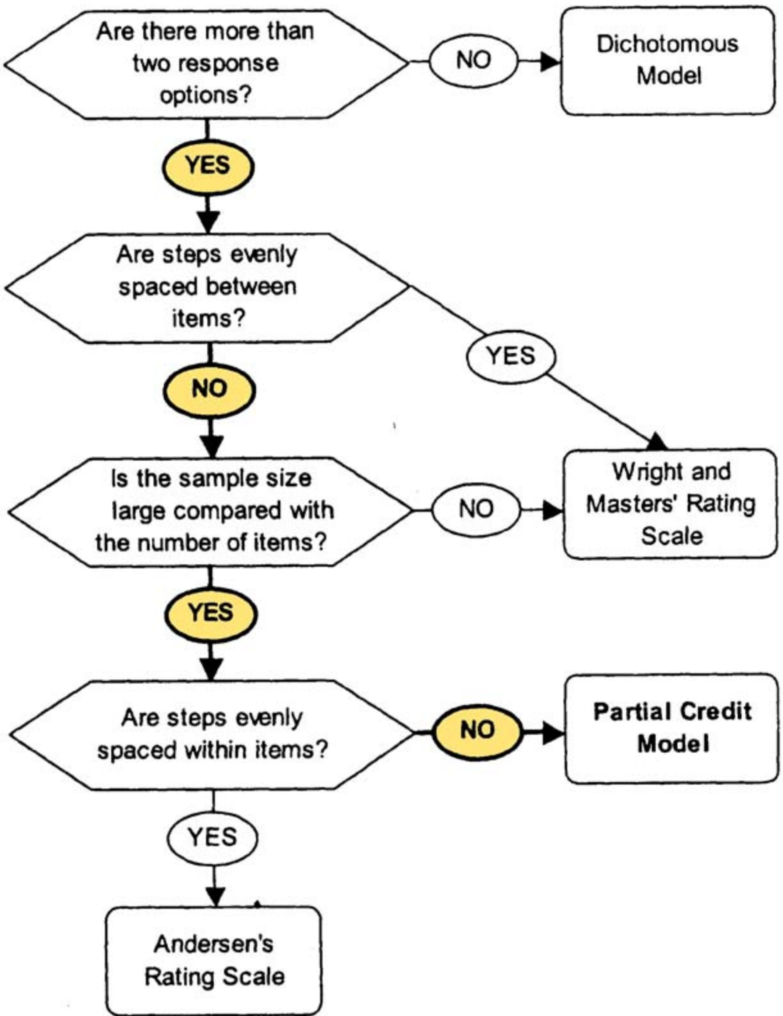

Figure 1: The decision process (represented as the highlighted parts) of selecting the most appropriate Rasch model for the current study. (Avery et al. 2003)

3.4.1 Appropriateness of the three-point rating scales

Examining the appropriateness of the three-point rating scales for the PDMS-2 items was firstly accomplished by determining whether the three rating scale categories in the PDMS-2 met Linacre's three essential criteria for optimising rating scale category effectiveness (Linacre, 2002). The criteria are as follows: 1) at least 10 observations are obtained per rating scale category, 2) average measures for each category increase across the rating scales, and 3) the outfit mean square (MnSq) value for each category is less than 2.0. The outfit MnSq value is the unweighted fit statistic, being affected more by unexpected responses distant to the person, item, or rating scale category measure. Outfit MnSq values greater than 2.0 indicate more unexplained than explained variance, and therefore the categories presenting outfit MnSq values greater than 2.0 were considered to be "category misfit" (Linacre, 2002). The present study followed Linacre's suggestions to compute the frequency for each of the three rating scale categories within all PDMS-2 items. The average measures for the three rating scale categories were also used to determine whether the rating scale categories increased monotonically. Outfit MnSq values for each rating scale category were compared with the threshold value of 2.0.

In addition, the order of step (threshold) calibrations of each PDMS-2 item was examined to see whether or not the step calibrations increase monotonically. The step calibrations represented the increments in difficulty as the scoring points progressed from a rating of 0 to 1 to a rating of 1 to 2. In the case of the PDMS-2, there are two step calibrations (i.e., the step calibrations of 0 to 1 and 1 to 2), among its three-point rating scales. The step calibration of 1 to 2 should be more difficult than that of 0 to 1, implying that the step calibrations increase monotonically. If the step calibrations do not function as expected, i.e., are “disordered step calibrations”, this would indicate that one or more rating scale categories might not be the most probable responses for certain parts of the measured variable (Bond & Fox, 2001).

Where the three-point rating scales within the PDMS-2 items exhibited infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordering step calibrations, the rating scales within the PDMS-2 items were reorganized by collapsing the problematic categories with adjacent ones, in order to optimise the rating scale category effectiveness.

3.4.2 Unidimensionality

Following optimization of the rating scales in the PDMS-2, this study examined whether items from each PDMS-2 subtest, FM and GM scales, or the entire PDMS-2 test could be used to construct unidimensional measures. Primarily the unidimensionality of the six individual subtests was examined. After establishing the unidimensionality of each subtest, the FM and GM scales of the PDMS-2, and the entire PDMS-2 test were subsequently assessed for their potential to construct unidimensional measures. Of particular interest was whether the combination of the six subtests could constitute a single construct such as FM, GM, or overall motor ability, respectively. The Rasch item fit statistics and the principle component analyses (PCA) of residuals were used to test the unidimensionality of the PDMS-2.

Rasch item fit statistics monitor the compatibility of the raw item data with the Rasch model expectations, and, generally, include two types of fit statistics: the outfit statistics mentioned previously and the information weighted infit statistics, also known as “Infit statistics” (Bond & Fox, 2001, 2007). Infit statistics are sensitive to unexpected behaviour that affects responses to items in close proximity to the child’s motor ability (Bond & Fox, 2001, 2007; Wright & Masters, 1982). For example, a 9-month-old infant with normal motor ability unexpectedly has a score that indicates a severe impairment on an easy motor task (e.g., failing to roll over in lying position). On the other hand, outfit statistics are sensitive to unexpected behaviour on items far beyond a children’s motor ability. For instance, a 9-month-old infant unexpectedly has a full score on a difficult motor task (e.g., standing up independently). This study examined the unidimensionality of the PDMS-2 using both infit and outfit statistics.

In general, infit and outfit statistics can be reported as the MnSq as well as the standardized Z values (Zstd) (Bond & Fox, 2001, 2007). The infit and outfit MnSq statistics in the range of 0.6 to 1.4 and their Zstd values ranging from -2.0 to 2.0 are generally deemed acceptable when considering possible misfit or overfit of items (Bond & Fox, 2001, 2007; Chen et al., 2005; Coster, Haley, Andres, Ludlow, Bond, & Ni, 2004a; Coster, Haley, Ludlow, Andres, & Ni, 2004b). When items display infit and outfit MnSq < 0.6 and/or their Zstd < -2.0, this would indicate that these items over fit the model expectations and therefore might be redundant. However, such over fit situations seem reasonable for developmental tests in measuring developmental sequences of certain abilities, such as the motor ability in the PDMS-2. This study predefined the misfit criteria as the infit and outfit MnSq > 1.4 and/or their Zstd > 2.0, for the items that presented potential departures from unidimensionality. These misfitting items were eliminated from the test in a stepwise manner by inspecting a series of the infit and outfit MnSq values and their Zstd values. The most-misfitting item, according to the predefined misfit criteria, was removed at each step. Successive Rasch analyses were re-run until all remaining items showed acceptable fit statistics.

When items fit the Rasch model's expectations, the residuals (observed scores minus expected scores) should be randomly distributed (Linacre, 1998; Smith, 2002). Several principal component analyses (PCA) on the residuals were conducted to detect whether any dominant component remained among the Rasch residuals in each PDMS-2 subtest, FM and GM scales, and the overall PDMS-2 test, respectively. The unidimensionality requirement was confirmed in the absence of a dominant component explaining > 10 % of the residual variance (Smith, 2002).

3.4.3 Person fit

Evaluation of person fit − the extent to which the responses of any participant conformed to the Rasch model expectations − is also important in test validation, regardless of item fit assessment (Hays et al., 2000). Two person fit statistics (i.e., the outfit and infit statistics) have been developed for this purpose (Linacre, 2006; Meijer, 2003), and these person fit statistics are explained in the same way as the item fit statistics. Person fit statistics, however, focus on detecting misfitting persons, instead of misfitting items. The identified misfitting persons might suggest aberrant respondents, response carelessness, cheating, or other situations such as a lack of motivation or concentration (Meijer, 2003).

In order to obtain more accurate results of the item calibrations, this study identified the misfitting persons prior to the item fit examination. The criteria for detecting misfitting persons were predefined as the same item misfit criteria, namely the infit and outfit MnSq > 1.4 and/or their Zstd > +2.0. The identified misfitting persons were subsequently set aside from the calibration sample.

3.4.4 *Differential item functioning*

Differential item functioning (DIF) analysis subsequently was performed to evaluate whether item calibrations were stable across gender (boys and girls) and disease entity (children with/out motor delays or difficulties). The Rasch model allows each item to be calibrated in log-odd units, also called logits, which is an indication of the estimated difficulty values for that item. The item difficulty estimates between boys and girls and those between children with/out motor delays or difficulties were illustrated in scatter plots. The paired 95% control lines were also computed and plotted in a scatter plot, using the method demonstrated by Bond and Fox (2001, 2007). Therefore, the items that fall outside of the control lines suggest gender-specific or disease-specific items, whereas the items that fall within the control lines suggest invariant items.

3.4.5 *Item and person reliability*

Rasch analysis provides item and person reliability indicators for further examination of the psychometric properties of the PDMS-2. The item reliability indicated the replicability of item placement along the pathway if these same items were given to another suitable sample. Item reliability can be interpreted on a 0 to 1 scale, much the same way as Cronbach's alpha is interpreted. Likewise, the person reliability can also be interpreted as Cronbach's alpha, indicating the replicability of the person ordering if the same sample was given another set of items, measuring the same construct.

In addition to the reliability indicators, the person separation index (G_p) indicating the number of distinct ability strata of persons discerned within each subtest, was used to describe the reliability of the PDMS-2 for the sample. The larger the G_p index, the more distinct levels of motor ability can be discerned in the PDMS-2 test. The number of distinct ability strata was computed using the formula: number of distinct strata = $(4G_p + 1) / 3$ (Wright & Masters, 1982). The number of distinct ability strata was used to indicate that the persons could be separated into statistically distinct ability strata. For example, a person separation index of 1.5 could discern two ability strata (high, low); whereas an index of 2.0 could discern three ability strata (high, middle, low); and so on. Additionally, the separation index for items is estimated in the same manner so that the number of distinct difficulty strata are calculated and interpreted in the same way.

3.4.6 Item hierarchy and targeting

For the PDMS-2 items which fit the Rasch model expectations, the difficulty estimates for those good-fit items were calibrated on objective interval scales, reflecting the genuine hierarchical order of the PDMS-2 items. With the partial credit Rasch model items can be arranged from easy to difficult, according to their corresponding average item difficulty estimates or logits. With the PDMS-2 items that have three rating scale categories, the average item difficulty estimates are difficult to interpret because the step calibration estimates used to compute the average are not the same across items (Coster et al., 2004a, b). As a consequence, the highest, average, and the lowest category estimates on each separate item, respectively, were adopted and three different item hierarchies were produced. Subsequently, three scatter plots with the paired 95% control lines (such as highest vs average, average vs lowest, and lowest vs highest) for each subtest were plotted using a method demonstrated by Bond and Fox (2001, 2007). These scatter plots provide visual illustrations to investigate which items show measurably different difficulty estimates among the highest, average, and the lowest category estimates. Items that fall outside of the control lines suggest items with measurably different difficulty estimates among these the category estimates.

In addition, Pearson *r* correlation coefficients were used to examine the level of agreement among the three different item hierarchies. The three item hierarchies were also compared concurrently with the item hierarchy that was ranked by age in the PDMS-2, using Pearson *r* correlation method, in order to examine the extent to which the item hierarchies differ in difficulty from that by age.

Finally a comparison of the item difficulty together with person ability, can be illustrated by utilising item-person maps. Item-person maps can determine whether the PDMS-2 items encompass (target) the range of person ability in the sample of this study.

CHAPTER FOUR

RESULTS

4.1 Subjects

This study recruited two convenience samples (i.e., one normative sample and one clinical sample) in Taiwan to examine the rating scales, dimensionality, DIF, and item hierarchy of the PDMS-2. The normative sample was used to produce high-quality measures for item calibration for examination of the PDMS-2 dimensionality and item hierarchy. The clinical sample was recruited to attain item calibrations for children with disabilities for the DIF analysis. Furthermore, the combination of normative and clinical samples was proposed to examine improved application of the PDMS-2 rating scales.

A total of 342 normal children were recruited to constitute a normative sample for the study. The normative sample comprised 183 boys and 159 girls ranging in age from 2 months to 77 months (mean=37.6 months, standard deviation [SD]=21.2). Each of the six age bands included nearly 50 children and approximated a gender-balanced sample. Their demographic characteristics, parents' socioeconomic and educational levels were tabulated in Table 1.

The clinical sample included 77 children with motor delays or difficulties, consisting of 55 boys and 22 girls. The age of the clinical sample ranged from 3 to 79 months (mean=44.2 months, SD=16.5). Although the number of the clinical sample was larger than the intended size (i.e., 60 children), gender composition was biased (55 boys and 22 girls). Fewer than ten children were included within the two lowest age bands of birth to 1 year and 1-2 years, while each of the other age bands had more than 10 children in each. Nearly half (49.4%) of the children in the clinical sample had diagnoses of developmental delay and, secondly, 20.1% of them suffered from sensorimotor disorders. The diagnoses of the remaining children were Cerebral Palsy (13.0%), Autism (6.5%), Down's syndrome (5.2%), mental retardation (2.5%), and Spina Bifida (1.3%). Table 1 summarizes additional characteristics of the clinical sample and of their parents' socioeconomic and educational levels.

Table 1: Demographical characteristics of the normative and clinical samples

	Normative sample n=342		Clinical sample n=77	
Demographic variable	Boys n=183	Girls n=159	Boys n=55	Girls n=22
Average age in months (mean ± SD)	38.2 ± 20.8	37.0 ± 21.6	46.7 ± 15.1	38.0 ± 18.4
Age in months (n)				
0-11	22	26	1	2
12-23	36	25	4	3
24-35	25	23	9	6
36-47	34	27	10	2
48-59	28	26	23	7
≥ 60	38	32	8	2
Education of parent/spouse (n)				
Elementary	0/1	0/0	1/0	1/0
Junior high	1/3	0/3	5/8	1/2
Senior high	48/46	42/42	11/17	6/8
College	117/103	110/950	34/24	14/80
Postgraduate	17/29	06/17	3/4	0/4
Unreported	0/1	1/2	1/2	0/0
Family monthly income, USD* (n)				
0~700	9	6	4	2
700~1,500	26	40	17	8
1,500~2,000	54	35	15	5
2,000~2,700	41	36	3	2
2,700~4,600	37	26	11	5
4,600 or above	10	7	3	0
Unreported	6	9	2	0

* 1 USD (US dollar) = 1.33 AUD (Australian dollar) or 32.95 NTD (new Taiwan dollar).

4.2 Appropriateness of the three-point rating scales

An observation of the frequency of the rating scale categories in the PDMS-2 items illustrated that a large percentage of the items (61%) were found to have fewer than 10 observations in particular rating scale categories (Table 2). This indicated that some rating scale categories, especially in the middle category, were not frequently used across the items. The less frequently used categories influenced the disordering of step calibrations in 148 items among the six subtests. The step calibration between category 1 and category 2 was found to be less difficult than that between category 0 and category 1 in the items with disordering step calibrations. Moreover, one Reflex item, nine Stationary items, nine Locomotion items, two Object Manipulation items, five Grasping items, and five Visual-Motor Integration items showed only one step calibration due to the limited observations in the middle or the lowest category.

Although only four items had average measures that did not increase with the rating scale categories, there were a total of 55 items in the subtest, which showed category misfit. These observations further indicated that the three-point rating scales might not be appropriate for parts of the PDMS-2 items in the study. In particular, the middle category response for the majority of items was the response least likely to be made by the sample according to the results of the usage frequency and disordering step calibrations. Consequently, it appeared that the middle category was redundant for the three-point rating scale efficiency in some parts of the PDMS-2 items. Accordingly, the original three-point rating scale (0-1-2) was reorganized as a new dichotomous scale (0-0-1) by collapsing the middle category down and combining it with the lowest category. Following the rating scale reorganization, two alternative versions using the new dichotomous scale were created. The first version reorganized 180 PDMS-2 items out of 249 items that used the problematic three-point rating scales (These specific items can be found in Appendices 1 to 6). The second version adopted the new dichotomous scale into the entire PDMS-2 items.

Table 2 shows a comparison of the appropriateness between the original and the two new versions of rating scales for each of the subtests. The two new versions were found to have a substantial reduction in infrequently used categories, less evidence of lack of monotonicity in the average measures, category misfit, and disordering step calibrations for the items across the six PDMS-2 subtests. Moreover, the two versions yielded comparable person and item reliabilities in the Locomotion, Object Manipulation, Grasping, and Visual-Motor Integration subtests. In the Reflex and Stationary subtests, however, the version where only the items with problematic three-point rating scales were reorganized showed better person reliability than the version in which the entire PDMS-2 items were reorganized. Therefore, the version in which only the items with problematic three-point rating scales were reorganized was preferred and thus adopted in the following Rasch analyses of the study.

Table 2: Items with category problems across three different rating scale versions

Subtest/ scaling scales	Count <10 (n*)	Average Measures† (n)	Category Fit >2.0 (n)	Step Calibration‡ (n)	Person Reliability	Item Reliability
Reflex						
0-1-2	7	1	1	4	0.86	0.99
0-0-1§	4	0	1	0	0.77	0.98
0-0-1¶	4	0	0	0	0.68	0.98
Stationary						
0-1-2	18	1	3	7	0.96	1.0
0-0-1§	10	0	4	1	0.94	1.0
0-0-1¶	10	0	4	0	0.78	1.0
Locomotion						
0-1-2	55	0	21	58	0.99	1.0
0-0-1§	9	0	14	1	1.00	1.0
0-0-1¶	7	0	13	0	0.99	1.0
Object Manipulation						
0-1-2	11	1	10	17	0.97	1.0
0-0-1§	3	0	4	1	0.97	1.0
0-0-1¶	3	0	6	0	0.96	1.0
Grasping						
0-1-2	17	1	5	14	0.91	0.99
0-0-1§	4	0	7	0	0.91	0.99
0-0-1¶	4	0	6	0	0.87	0.99
Visual-Motor Integration						
0-1-2	44	0	15	48	0.99	1.0
0-0-1§	11	0	16	3	0.99	1.0
0-0-1¶	11	0	14	0	0.99	1.0

* indicates the number of items that exhibited infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordering step calibrations.

† indicates items whose average measures in each rating scale category did not increase monotonically.

‡ indicates items whose step calibrations in each rating scale category did not increase monotonically.

§ indicates only the items with problematic rating scales that were reorganized to the rating scale of 0-0-1.

¶ indicates the entire PDMS-2 items that were reorganized to the rating scale of 0-0-1.

4.3 Person fit

Once the mode of collapsing categories had been adopted, Rasch analysis was then performed to evaluate whether children's response strings on the PDMS-2 items conformed to the Rasch model's expectations. Person fit evaluation focused on identifying the misfitting children in the normative sample of the study only. The clinical sample was excluded from this analysis, because it was expected that the disease entity of motor delays or difficulties will not conform to normal motor development indicators and unlikely fit the Rasch model's expectations. In this case, person misfit is likely to be indicative of the motor delay or difficulties.

The person fit analyses for the normative sample revealed one response (3.0%) in the Reflex subtest, 25 responses (7.7%) in the Stationary subtest, 30 responses (9.1%) in the Locomotion subtest, 34 responses (11.5%) in the Object Manipulation subtest, 27 responses (12.1%) in the Grasping subtest, and 30 responses (9.7%) in the Visual-Motor Integration subtest with person misfit. This indicated that these children's responses did not fit the underlying Rasch model. The children's responses, even though they were part of a normative sample, were removed subsequently from the corresponding subtests. The remaining children's responses in the reduced sample have shown acceptable model fit, and thus the reduced sample was used to calibrate the PDMS-2 items in the next step. Given that parameter with Rasch model allows for the estimation of item difficulties independent of the distribution of person abilities, then using the reduced sample does not impair the conclusion that may be made from results.

4.4 Dimensionality

Following removal of the misfitting children's responses, the unidimensionality of six individual PDMS-2 subtests was examined with the reduced normative sample. The subscales were combined into FM, GM and then the entire PDMS-2 scales. Furthermore, the FM and GM scales of the PDMS-2 and the entire PDMS-2 scale were assessed for their potential to construct unidimensional measures. The WINSTEPS control files for each dimensionality examination are included in Appendices 7 to 8.

4.4.1 Unidimensionality of the six individual subtests

Firstly, the unidimensionality of each of the six individual PDMS-2 subtests was examined using the partial credit Rasch model. Successive Rasch analyses identified one of the eight Reflex items, four of the 30 Stationary items, 13 of the 89 Locomotion items, seven of the 24 Object Manipulation items, four of the 26 Grasping items, and 14 of the 72 Visual-Motor Integration items that exceeded the infit and outfit MnSq > 1.4 with Zstd > 2.0. Tables 3 to 8 demonstrate the identified misfitting items in each subtest. However, the number of children in the Reflex subtest was relatively small (n=21), because 71 children achieved full scores on the Reflex subtest. Because these extreme responses were automatically eliminated from the sample in the Rasch analysis, the results for the Reflex subtest were less convincing given such a small sample. For this reason, although the Reflex Item 3 was identified as misfitting (Table 3), the item was retained in the Reflex subtest due to the limited sample size.

Table 3: Misfitting items in the Rasch analysis of the 8-item Reflex subtest (n=21)

Misfitting items*	Infit statistics		Outfit statistics	
(Item no.)	MnSq	Zstd	MnSq	Zstd
Landau reaction (3)	1.75	1.7	0.95	0

* The item was considered as misfitting in the Reflex subtest; however the item was retained due to the limited sample size.
Note
[1] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[2] Two items (Items 1 and 2) were dropped from the Rasch item analysis, because most children passed the items.

In addition, four items were identified as misfitting in the Stationary subtest (Table 4). Of the misfitting items, the rotating head item (Item 1), had only the infit MnSq value of greater than 1.4 with the remaining fit statistics for the item not exceeding the misfitting criteria. Even though the rotating head item had a high misfitting value, the clinical importance of retaining this item was warranted because it is the easiest item to perform in the Stationary subtest (Foilo & Fewell, 2000). Only three items, standing on tiptoes I (Item 22), standing on tiptoes II (Item 24), and imitating movement (Item 26) items were thus eliminated from the Stationary subtest in the study.

Table 4: Misfitting items in the Rasch analyses of the 30-item Stationary subtest (n=302)

Misfitting items*	Infit statistics		Outfit statistics	
(Item no.)	MnSq	Zstd	MnSq	Zstd
Imitating movement (26)	2.19	4.7	9.90	1.3
Standing on tiptoes II (24)	1.95	3.9	2.20	0.3
Standing on tiptoes I (22)	1.57	2.4	1.86	0.2
Rotating head (1)	3.01	1.6	0.13	-0.3

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.
Note
[1] The misfitting items above the dashed line were eliminated from the Stationary subtest, but the item below the dashed line was retained.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Four items (Items 2, 3, 6, and 7) were dropped from the Rasch item analysis, because most children passed the items.

Likewise, the Locomotion subtest had five of 13 misfitting items (Table 5), which were identified as misfitting in one fit statistic. The author's consideration of the clinical importance for the five items, i.e., the stepping I (Item 28), running speed and agility (Item 85), standing up (Item 33), extending arm II (Item 15), and walking line II (Item 56) items justified their retention in the Locomotion subtest. Moreover, while the jumping up II item (Item 55) demonstrated misfitting statistics in both infit MnSq and Zstd values, the author concluded that this item might be more sensitive for children with developmental delay; therefore it was retained in the Locomotion subtest.

Table 5: Misfitting items in the Rasch analysis of the 89-item Locomotion subtest (n=299)

Misfitting items* (Item no.)	Infit statistics		Outfit statistics	
	MnSq	Zstd	MnSq	Zstd
Bouncing (25)	3.12	4.2	2.51	0.1
Scooting (21)	2.21	2.9	3.23	0.4
Creeping down stairs (39)	2.62	3.8	2.32	0.1
Pivoting II (29)	2.01	2.4	9.90	2.3
Rolling forward (79)	1.81	4.7	3.45	3.9
Jumping hurdles II (84)	1.48	2.3	1.83	0.6
Running balance/ coordination (74)	1.35	2.9	2.87	3.1
Jumping up II (55)	1.61	2.7	0.94	0
Standing up (33)	2.39	1.8	0.23	-0.1
Extending arm II (15)	1.43	1.0	0.67	-0.1
Walking line II (56)	1.41	1.7	0.50	-0.1
Stepping I (28)	1.25	0.5	2.37	0.1
Running speed and agility (85)	1.22	1.5	2.19	1.6

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.

Note

1 The misfitting items above the dashed line were eliminated from the Locomotion subtest, but the items below the dashed line were retained.

2 Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.

3 Two items (Items 1 and 2) were dropped from the Rasch item analysis, because most children passed the items.

Although the Rasch analysis showed that the Object Manipulation subtest had seven misfitting items (see Table 6), none of these items had more than two fit statistics greater than the misfit criteria. The author considered the potential clinical importance of these items, and no items were eliminated from the Object Manipulation subtest.

Table 6: Misfitting items in the Rasch analysis of the 24-item Object Manipulation subtest

Misfitting items (n=253) (Item no.)	Infit statistics		Outfit statistics	
	MnSq	Zstd	MnSq	Zstd
Kicking ball I (4)	1.48	1.6	0.23	-0.2
Hitting target-overhand I (18)	1.48	2.0	0.57	-0.1
Catching ball II (10)	1.37	2.6	0.63	-0.1
Rolling ball (2)	1.30	0.4	4.09	0.3
Throwing ball-overhand III (15)	0.88	-0.8	1.88	0.3
Hitting target-overhand II (20)	0.94	-0.4	1.84	0.2
Catching ball V (22)	1.02	0.1	1.61	0.2

Note
[1] Although these items were considered as misfitting in Object Manipulation subtest, all were retained due to clinical importance.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Item 1 was dropped from the Rasch item analysis, because most children passed the item.

In the Grasping subtest, there were four items showing misfit in the Rasch analysis (see Table 7). The manipulating paper (Item 16), pulling string (Item 8), and grasping marker II (Item 22) items each had two fit statistics that were greater than the misfitting criteria. However, the grasping marker II item holds significant clinical importance for the Grasping subtest, as it fills the item gap between 16 months of age and 41 months of age. The additional misfitting item, namely the grasping cloth item (Item 2) is clinically important because it is one of the easiest items to perform in the Grasping subtest (Foilo & Fewell, 2000). Therefore, the only items removed from the Grasping subtest were the manipulating paper item, and the pulling string item.

Table 7: Misfitting items in the Rasch analysis of the 26-item Grasping subtest (n=142)

Misfitting items*	Infit statistics		Outfit statistics	
(Item no.)	MnSq	Zstd	MnSq	Zstd
Manipulating paper (16)	1.62	1.0	1.50	0
Pulling string (8)	2.11	2.1	0.67	0
Grasping marker II (22)	1.62	3.0	0.93	-0.1
Grasping cloth (2)	1.64	0.9	0.11	-0.4

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.

Note
[1] The misfitting items above the dashed line were eliminated from the Grasping subtest, but the items below the dashed line were retained.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Three items (Items 1, 3, and 4) were dropped from the Rasch item analysis, because most children passed the items.

Finally, the Visual-Motor Integration subtest had 14 items showing misfit in the Rasch analysis (Table 8), however six items of these had only one misfitting statistic. These six items were the copying circle (Item 55), removing top (Item 45), clapping hands (Item 21), manipulating string (Item 23), extending arm (Item 13), and copying cross (Item 61) items. Because they showed little misfit and might have clinical importance based on the author's own clinical experience, these six items were retained in the Visual-Motor Integration subtest. On the other hand, the connecting dots (Item 67), lacing string (Item 58), transferring cube (Item 15), tapping spoon (Item 34), scribbling (Item 37), turning pages II (Item 41), removing socks (Item 26), and imitating vertical strokes (Item 44) items were removed from the Visual-Motor Integration subtest.

Table 8: Misfitting items in the Rasch analysis of the 72-item Visual-Motor Integration subtest (n=280)

Misfitting items*	Infit statistics		Outfit statistics	
(Item no.)	MnSq	Zstd	MnSq	Zstd
Connecting dots (67)	1.69	4.1	1.64	1.5
Lacing string (58)	1.53	2.1	3.13	0.8
Transferring cube (15)	1.79	2.4	0.63	0
Tapping spoon (34)	1.52	2.4	0.86	0
Scribbling (37)	2.58	2.2	0.71	0
Turning pages II (41)	1.63	2.5	2.19	0.3
Removing socks (26)	1.70	1.3	1.79	0.1
Imitating vertical strokes (44)	1.44	1.6	4.46	0.3
Manipulating string (23)	1.60	1.9	0.91	0
Extending arm (13)	1.45	0.6	0.38	-0.2
Copying cross (61)	1.27	2.1	1.28	0.2
Copying circle (55)	1.03	0.1	9.90	1.3
Removing top (45)	1.33	1.6	9.90	0.5
Clapping hands (21)	1.15	0.5	6.40	0.2

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.

Note

[1] The misfitting items above the dashed line were eliminated from the Visual-Motor Integration subtest, but the items below the dashed line were retained.

[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.

[3] Four items (Items 1, 2, 4, and 8) were dropped from the Rasch item analysis, because most children passed the items.

In summary, three of 30 Stationary items, seven of 89 Locomotion items, two of 26 Grasping items, and eight of 72 Visual-Motor Integration items but no items in Reflex and Object Manipulation subtests were removed from the corresponding subtests, according to the Rasch analysis fit results in light of clinical considerations.

Further all the reduced PDMS-2 subtests showed very little common variance remaining (i.e., nearly zero percent) in their corresponding PCA of the residuals after the Rasch dimension had been removed. There was no obvious clustering of items in the PCA of residuals. The PCA results were interpreted to indicate that each subtest had acceptable model fit, although some misfitting items were not removed from subtests due to clinical importance. The requirement of unidimensionality for each PDMS-2 subtest was supported by combining the fit statistics with the PCA of residuals in the Rasch model.

4.4.2 Unidimensionality of the combined gross motor scale

Following the examination of unidimensionality for the six individual subtests, the unidimensional structure of the GM scale was examined, by combining all the items in the Reflex, Stationary, Locomotion, and Object Manipulation subtests. A series of Rasch analyses identified 22 of the most misfitting items, including three Reflex items, two Stationary items, seven Locomotion items, and 10 Object Manipulation items (Table 9). After removing these items, 16 items showed misfit on just one fit statistic. However, the PCA of residuals provided for the evidence that there were no dominant residual factors indicating the unidimensional nature of the combined GM scale. From the author’s perspective, the 16 slight misfitting items held potential clinical importance and, thus, were not removed from the combined GM scale.

Table 9: Misfitting items in the Rasch analysis of the 151-item combined GM scale (n=306)

Misfitting items* (Item no.)[†]	Infit statistics MnSq	Infit statistics Zstd	Outfit statistics MnSq	Outfit statistics Zstd
Bouncing (L25)	3.12	3.8	2.67	0.1
Landau reaction (R3)	2.17	2.9	9.90	1.1
Pivoting II (L29)	1.97	2.1	9.90	0.7
Scooting (L21)	1.71	2.2	2.38	0.2
Rolling forward (L79)	1.70	4.3	2.37	2.6
Hitting target-overhand II (O20)	1.73	4.6	1.84	1.1
Catching ball II (O10)	1.65	3.4	1.87	0.7
Kicking ball (O23)	1.48	2.1	2.81	2.3
Catching ball (O1)	1.90	2.0	9.90	0.5
Creeping down stairs (L39)	1.66	2.0	1.68	0.1
Rolling ball (O2)	1.77	1.9	9.90	0.6
Kneeling (S19)	2.83	5.1	1.46	0.1
Throwing ball-overhand III (O15)	1.33	2.4	2.42	0.8
Throwing ball-underhand III (O19)	1.36	2.6	2.15	0.8
Protecting reaction-backward (R8)	2.84	2.0	0.35	-0.1
Righting reaction-forward (R7)	3.09	2.0	0.68	0
Hitting target-underhand (O16)	1.79	4.1	1.26	0.3
Walking line II (L56)	1.67	2.7	0.78	-0.1
Catching ball III (O14)	1.38	2.1	1.55	1.2
Catching ball IV (O17)	1.41	3.0	1.11	0.2
Running balance/coordination (L74)	1.25	2.1	1.76	1.7
Standing on tiptoes I (S22)	1.42	2.6	1.10	0.1
Positioning flex (R2)	2.35	1.6	0.27	-0.2
Extending arm II (L15)	1.91	1.7	0.91	0
Thrusting arm (L3)	1.79	1.1	0.13	-0.3
Kicking ball I (O4)	1.74	1.8	0.48	-0.1
Standing up (L33)	1.71	1.1	0.16	-0.2
Rolling II (L16)	1.67	1.4	0.21	-0.2
Rotating head (S1)	1.62	0.9	0.08	-0.3
Pulling to sit (S13)	1.50	0.8	0.36	-0.2
Throwing ball-overhand II (O11)	1.43	1.6	0.70	-0.1
Flinging ball (O3)	1.42	1.0	1.24	0
Push-ups (S30)	1.36	2.1	1.35	0.7
Walking line backward I (L75)	1.31	2.1	1.02	0

Continued

Table 9: Misfitting items in the Rasch analysis of the 151-item combined GM scale (n=306) [Continued]

Misfitting items* (Item no.)[†]	Infit statistics		Outfit statistics	
	MnSq	Zstd	MnSq	Zstd
Standing on tiptoes II (S24)	1.27	2.1	1.26	0.6
Walking down stairs II (L53)	1.16	0.6	9.52	0.7
Stepping I (L28)	1.18	0.3	6.09	0.3
Running speed and agility (L85)	0.98	-0.1	2.15	1.7

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.
[†] indicates the item no. in the corresponding subtest, e.g., R is Reflex; S is Stationary; L is Locomotion; and O is Object Manipulation.
Note
[1] The misfitting items above the dashed line were eliminated from the combined GM scale, but the items below the dashed line were retained.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Six items (S2, S3, S6, S7, L1, and L2) were dropped from the Rasch item analysis, because most children passed the items.

4.4.3 Unidimensionality of the combined fine motor scale

The FM scale, which is a collection of the items from the Grasping and Visual-Motor Integration subtests, was also examined for unidimensionality. Four Grasping items and nine Visual-Motor Integration items were considered to be the most misfitting on the FM attribute and for this reason were eliminated from the combined FM scale (Table 10). The PCA on the residuals of the combined FM scale (without the 13 most misfitting items) revealed little remaining common variance, supporting the unidimensionality of the FM attribute. Although the additional three Grasping items and seven Visual-Motor Integration items also exhibited misfit in just one fit statistic each, they were considered to hold clinical importance from the author’s own clinical perspective. Given that the ten misfitting items seem not to adversely affect the unidimensionality of the combined FM scale, they were retained in the combined FM scale.

Table 10: Misfitting items in the Rasch analysis of the 98-item FM subscale (n=299)

Misfitting items* (Item no.)†	Infit statistics MnSq	Infit statistics Zstd	Outfit statistics MnSq	Outfit statistics Zstd
Manipulating paper (G16)	2.79	4.8	9.90	0.6
Connecting dots (V67)	1.65	4.1	1.56	1.7
Turning pages II (V41)	1.50	2.1	1.89	0.4
Grasping marker III (G25)	1.50	4.1	1.58	3.3
Tapping spoon (V34)	1.47	2.5	0.84	0
Imitating vertical strokes (V44)	1.43	1.5	5.33	0.5
Grasping marker II (G22)	1.74	4.0	1.09	0.1
Lacing string (V58)	1.50	2.2	2.86	0.9
Transferring cube (V15)	1.65	2.3	0.59	0
Removing socks (V26)	1.46	1.0	1.66	0.1
Manipulating string (V23)	1.69	2.2	0.96	0
Grasping marker I (G21)	1.70	2.5	0.66	0
Scribbling (V37)	2.74	2.5	0.79	0
Grasping cloth (G2)	1.53	1.0	0.05	-0.6
Regarding hands (V5)	1.53	0.9	0.06	-0.4
Copying cross II (V61)	1.28	2.2	1.29	0.3
Placing pegs (V33)	0.90	-0.3	9.90	0.9
Copying circle (V55)	0.90	-0.4	9.90	1.0
Clapping hands (V21)	1.23	0.9	7.73	0.3
Removing top (V45)	1.35	1.7	6.56	0.3
Shaking rattle II (G13)	1.23	0.5	2.90	0.1
Building bridge (V54)	1.19	0.7	2.19	0.1
Touching fingers (G26)	1.20	1.3	1.55	0.9

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.
† indicates the item no. in the corresponding subtest, e.g., G is Grasping and V is Visual-Motor Integration.
Note
[1] The misfitting items above the dashed line were eliminated from the combined FM scale, but the items below the dashed line were retained.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Seven items (G1, G3, G4, V1, V2, V4, and V8) were dropped from the Rasch item analysis, because most children passed the items.

4.4.4 Unidimensionality of the entire PDMS-2 scale

Finally, whether all the PDMS-2 items could constitute a single construct of overall motor ability, was examined. A series of Rasch analyses identified 74 misfitting items (i.e., 29.7%) in a total of 249 items (Table 11). Of the items identified as misfitting, 50 items (20%) were identified as the most misfitting, because they showed at least two fit statistics exceeding the fit criteria. The 50 items consisted of two Reflex items, two Stationary items, eight Locomotion items, 10 Object Manipulation items, eight Grasping items, and 20 Visual-Motor Integration items. Once the 50 most misfitting items were excluded, the Rasch analysis of the remaining items showed acceptable model fit and, more importantly, nearly zero percent of the residual variance was accounted for by the first factor in the PCA. As a consequence, the PCA results together with the fit statistic results suggested that, although the 24 misfit items were retained because of the author's clinical considerations, the unidimensionality of the overall motor ability was upheld in the reduced PDMS-2 scale,

Table 11: Misfitting items in the Rasch analysis of the 249-item PDMS-2 scale (n=310)

Misfitting items*	Infit statistics		Outfit statistics	
(Item no.)[†]	MnSq	Zstd	MnSq	Zstd
Bouncing (L25)	3.43	3.7	1.79	0.1
Manipulating paper (G16)	2.67	4.4	9.90	1.5
Placing hand (V3)	2.72	2.2	9.90	0.9
Holding rattle (G5)	2.61	2.5	9.90	1.4
Kneeling (S19)	2.27	2.1	3.27	0.3
Landau reaction (R3)	2.07	2.9	6.15	0.2
Turning pages II (V41)	1.69	2.6	2.93	0.4
Grasping marker I (G21)	1.98	3.5	2.38	0.1
Rolling forward (L79)	1.46	3.1	2.09	2.6
Grasping marker II (G22)	1.97	4.8	1.62	0.6
Hitting target-overhand II (O20)	1.63	4.2	1.81	1.4
Catching ball III (O14)	1.45	2.5	1.83	1.8
Catching ball II (O10)	1.61	3.7	1.67	1.0
Grasping marker III (G25)	1.48	3.9	1.59	3.3
Removing top (V45)	1.42	2.1	2.50	0.2
Pivoting II (L29)	1.48	1.1	9.90	0.6
Rolling ball (O2)	1.58	1.8	7.28	0.3
Tapping spoon (V34)	1.95	3.9	1.48	0.1
Scribbling (V37)	2.67	2.6	1.56	0.1
Scooting (L21)	1.71	2.1	1.71	0.1
Inserting shapes I (V35)	2.03	2.5	1.69	0.1
Removing socks (V26)	2.58	2.1	6.84	0.3
Placing pegs (V33)	1.58	1.4	9.90	1.1
Clapping hands (V21)	1.58	1.7	2.66	0.1
Positioning flex (R2)	3.78	2.2	0.66	0
Kicking ball V (O23)	1.40	1.9	3.19	3.3
Connecting dots (V67)	1.36	2.4	2.17	2.2
Hitting target-underhand (O16)	1.66	3.8	1.41	0.6
Transferring cube (V15)	2.57	3.3	0.90	0
Retaining cubes II (V22)	2.34	3.1	1.12	0
Rolling II (L16)	1.72	1.5	2.47	0.2
Rolling III (L17)	1.41	1.0	2.25	0.2
Pulling string (G8)	1.42	1.0	1.60	0.1
Catching ball I (O1)	1.45	1.2	1.74	0.1

Continued

Table 11: Misfitting items in the Rasch analysis of the 249-item PDMS-2 scale (n=310) [Continued]

Misfitting items* (Item no.)†	Infit statistics MnSq	Infit statistics Zstd	Outfit statistics MnSq	Outfit statistics Zstd
Combining cubes (V20)	2.07	2.5	0.71	0
Stirring spoon (V30)	1.70	2.4	1.29	0
Creeping down stairs (L39)	1.83	2.5	0.72	0
Throwing ball-overhand II (O11)	1.64	2.1	0.81	0
Lacing string (V58)	1.61	2.9	1.28	0.2
Throwing ball-overhand III (O15)	1.33	2.5	2.33	0.9
Throwing ball-underhand III (O19)	1.32	2.3	2.10	0.9
Folding paper I (V50)	1.47	2.4	0.68	-0.2
Copying cross II (V61)	1.18	1.5	2.40	2.8
Push-ups (S30)	1.42	2.4	1.62	1.3
Jumping up II (L55)	1.51	2.4	1.01	0
Folding paper III (V72)	1.31	2.1	1.49	2.7
Folding paper I (V70)	1.16	1.3	2.65	3.5
Coloring between lines (V71)	1.40	2.5	1.75	1.2
Shaking rattle II (G13)	1.29	0.6	9.90	2.4
Grasping pellets II (G17)	1.90	2.1	0.41	-0.1
Protecting reaction-backward (R8)	2.79	2.0	0.79	0
Tracking rattle I (V1)	2.70	1.2	0.07	-0.6
Perceiving rattle (V4)	2.70	1.2	0.07	-0.6
Righting reaction-forward (R7)	2.17	1.5	0.68	0
Grasping rattle II (G7)	1.98	1.6	0.45	-0.1
Placing pellets (V36)	1.95	0.9	0.17	-0.1
Manipulating rattle (G6)	1.79	1.4	0.09	-0.3
Extending arm II (L15)	1.78	1.6	0.44	-0.2
Thrusting arm (L3)	1.75	0.9	0.11	-0.3
Extending arms and legs I (L9)	1.73	1.5	0.28	-0.2
Rotating head (S1)	1.65	0.8	0.08	-0.3
Grasping cloth (G2)	1.65	0.8	0.08	-0.3
Walking line II (L56)	1.47	1.9	0.61	-0.1
Grasping pellets I (G15)	1.45	1.1	0.72	-0.1
Flexing legs (L10)	1.47	0.8	0.12	-0.2
Lowering (L27)	1.41	0.8	0.20	-0.1
Catching ball IV (O17)	1.37	2.7	1.14	0.3
Standing on tiptoes I (S22)	1.32	2.2	1.03	0

Continued

Table 11: Misfitting items in the Rasch analysis of the 249-item PDMS-2 scale (n=310) [Continued]

Misfitting items* (Item no.)†	Infit statistics		Outfit statistics	
	MnSq	Zstd	MnSq	Zstd
Imitating vertical strokes (V44)	1.00	0	7.05	0.7
Walking down stairs II (L53)	1.27	0.9	5.52	0.5
Walking line backward I (L75)	1.27	1.9	1.96	0.6
Catching bounced ball (O24)	1.37	1.3	1.89	0.8
Running speed and agility (L85)	1.00	0	1.61	1.0
Running balance/ coordination (L74)	1.16	1.3	1.59	1.3

* indicates items arranged in order of the extent to which they misfit in the Rasch analysis.
† indicates the item no. in the corresponding subtest, e.g., R is Reflex; S is Stationary; L is Locomotion; O is Object Manipulation; G is Grasping; and V is Visual-Motor Integration.
Note
[1] The misfitting items above the dashed line were eliminated from the entire PDMS-2 scale, but the items below the dashed line were retained.
[2] Underlying values indicate that the MnSq or Zstd values were beyond the misfitting criteria, i.e., MnSq > 1.4 or Zstd >2.0.
[3] Six items (S2, S3, L1, L2, G1, and G3) were dropped from the Rasch item analysis, because most children passed the items.

4.5 Differential item functioning

DIF analysis was subsequently conducted to determine if the item calibrations were stable across disease entity (children with/out motor delays or difficulties) and gender (boys and girls) in the PDMS-2. Since parts of the PDMS-2 items still use the three-point rating scales, the partial credit Rasch model produced not only the average difficulty estimates for these items but also the highest and lowest difficulty estimates for the steps of 0-1 and 1-2 in the items. For this reason, the highest, average, and the lowest category estimates on each separate item, respectively, were adopted to perform three separated DIF analyses. This analysis was performed in order to identify any substantial DIF items exist across disease entity or gender on the basis of different difficulty estimates.

4.5.1 Disease entity

Figures 2 and 3 display the item calibrations of normal children against those of children with motor delays or difficulties for the six PDMS-2 subtests, in terms of the average item difficulty estimates. The dashed identity line shows perfect agreement on the difficulty of items, whereas the items that fall outside of the paired 95% control lines suggest disease-specific items. Only a few items in the Reflex (one item), Stationary (two items), Object Manipulation (four items), Grasping (four items), and Visual-Motor Integration (three items) subtests were found to fall outside of the control lines, indicating significant DIF. As might be expected, a considerable number of 32 Locomotion items, however, had the significant DIF between children with/out motor delays or difficulties.

In addition to using the average difficulty estimates, DIF analyses based on the highest and lowest category estimates were performed in each subtest. The DIF scatter plots were similar to Figures 2 and 3, and therefore were not illustrated separately in the study. Despite the items having DIF based on the average item difficulty estimates, twelve additional items in the highest/lowest category estimates were also found to show significant DIF across the disease entity. All of the items showing DIF between normal children and children with motor delays or difficulties in each subtest were detailed in Table 12.

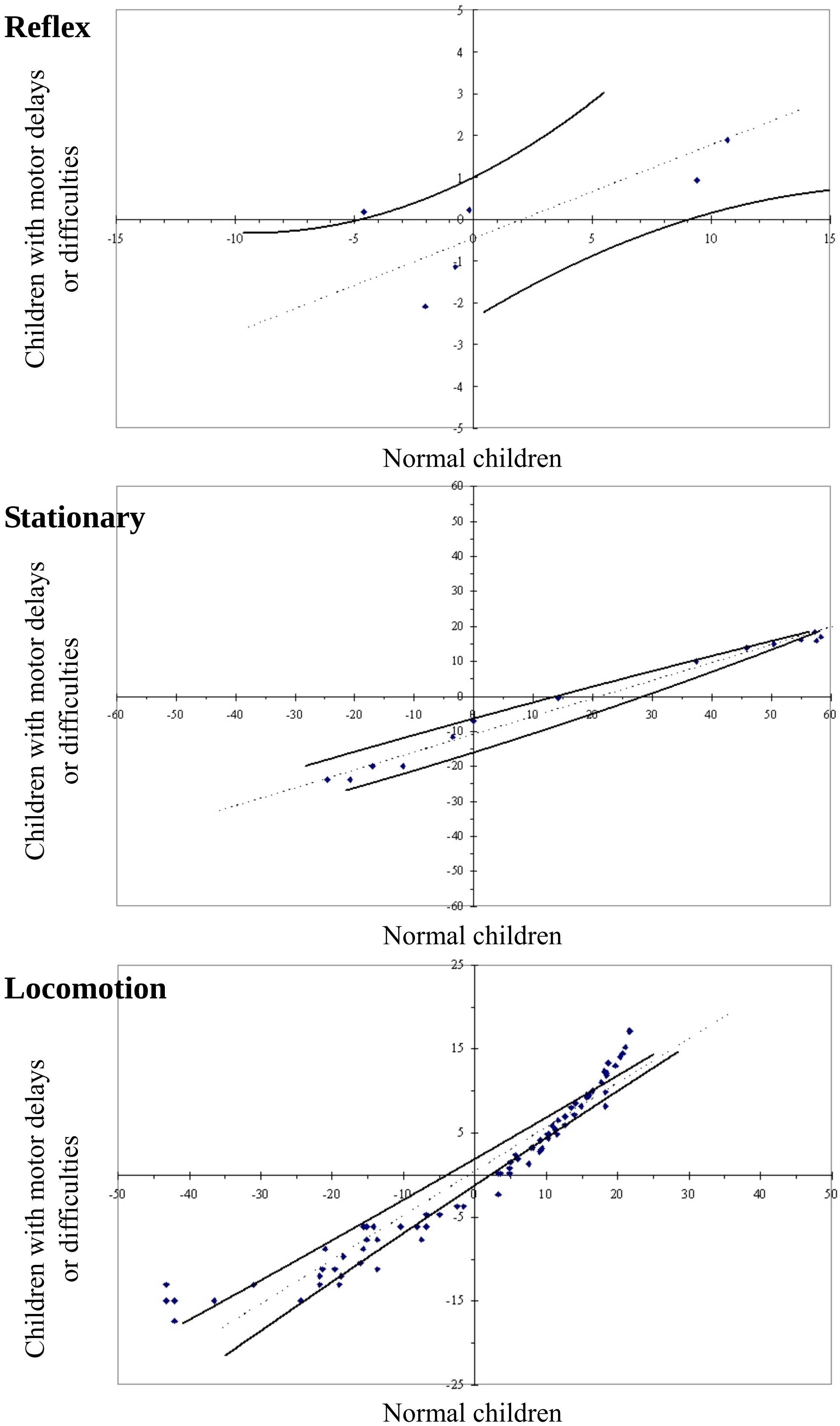

Figure 2: DIF between normal children and children with motor delays or difficulties in the Reflex, Stationary, and Locomotion subtests. The area between the two control lines indicates 95% confidence interval.

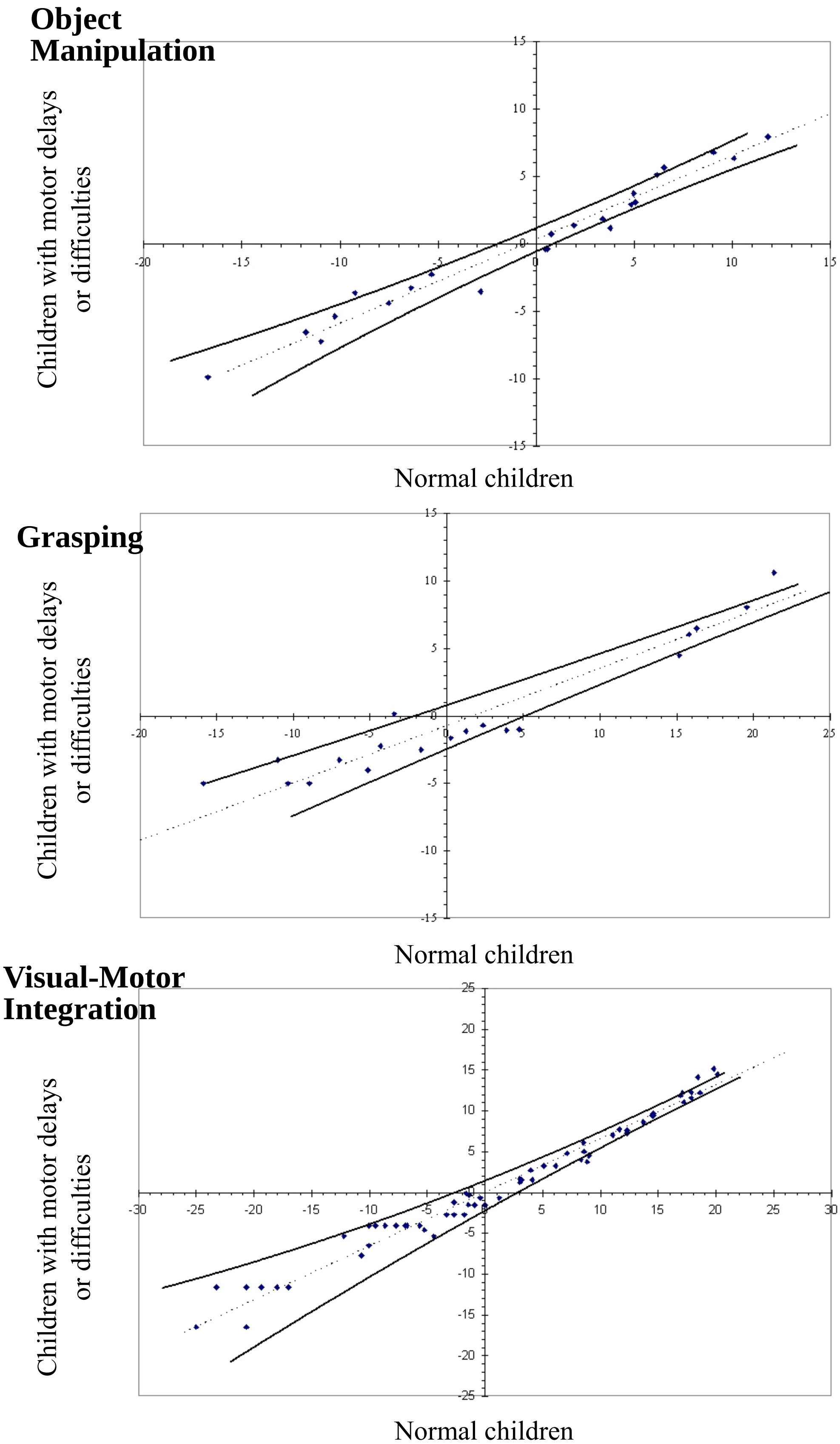

Figure 3: DIF between normal children and children with motor delays or difficulties in the Object Manipulation, Grasping, and Visual-Motor Integration subtests. The area between the two control lines indicates 95% confidence interval.

Table 12: The items showing DIF across disease entity in six PDMS-2 subtests

Subtest/ Item no.	Item content	Subtest/ Item no.	Item content
Reflex		Locomotion	
3	Landau reaction	3	Thrusting arm
		4	Bearing weight
Stationary		5	Extending trunk
25	Standing on 1 foot IV*	6	Symmetrical posture
28	Sit-ups I	7	Propping on forearms
29	Sit-ups II	16	Rolling II
		28	Stepping I
Object Manipulation		30	Standing II
4	Kicking ball I	32	Stepping II
10	Catching ball II	35	Walking II
14	Catching ball III	36	Standing and moving balance
15	Throwing ball-overhand III	38	Walking III
17	Catching ball IV*	40	Walking up stairs I
20	Hitting target-overhand II*	41	Walking fast
		42	Walking backward I
Grasping		43	Walking down stairs I
13	Shaking rattle II	44	Walking backward II
19	Grasping cube IV*	49	Jumping forward I
20	Grasping cubes	50	Jumping up I
21	Grasping marker I	51	Jumping down I*
22	Grasping marker II*	52	Walking up stairs II
25	Grasping marker III	61	Jumping forward II*
		67	Jumping forward III*
Visual-Motor Integration		72	Jumping forward on 1 foot
30	Stirring spoon*	75	Walking line backward I
49	Stringing beads I	76	Jumping forward IV
61	Copying cross II*	77	Hopping
64	Copying square	78	Walking line backward II
66	Building steps*	81	Jumping forward V
72	Folding paper III	82	Turning jump*
		83	Hopping forward
		85	Running speed and agility
		86	Skipping I
		87	Jumping sideways
		88	Skipping II
		89	Hopping speed

* indicates items showing DIF in the highest/lowest category estimates but not in the average item estimates.

4.5.2 Gender

The same DIF analyses were performed to determine if the item calibrations were stable across gender (i.e., boys vs girls) in the normative sample. In terms of the average item difficulty estimates, seven Stationary items, 10 Locomotion items, four Object Manipulation items, seven Grasping items, and three Visual-Motor Integration items were found to show significant DIF. Moreover, the highest/lowest category estimates of another four items, each of which was in four different subtests (Table 13), respectively, showed significant DIF. All of the items showing DIF across gender in each subtest were listed in Table 13.

Table 13: The items showing DIF across gender in six PDMS-2 subtests

Subtest/ Item no.	Item content	Subtest/ Item no.	Item content
Stationary		Object Manipulation	
20	Standing on 1 foot I	7	Throwing ball-overhand I
21	Standing on 1 foot II	9	Kicking ball III
23	Standing on 1 foot III	10	Catching ball II
25	Standing on 1 foot IV*	21	Bouncing ball
27	Standing on 1 foot V	23	Kicking ball V*
28	Sit-ups I		
29	Sit-ups II	Grasping	
30	Push-ups	5	Holding rattle
		18	Grasping pellets III
Locomotion		20	Grasping cubes
36	Standing and moving balance	22	Grasping marker I
38	Walking III	23	Unbuttoning button
43	Walking down stairs I	24	Buttoning button
53	Walking down stairs II	25	Grasping marker III
69	Running form		
76	Jumping forward IV*	Visual-Motor Integration	
81	Jumping forward V	3	Placing hand
85	Running speed and agility	13	Extending arm
86	Skipping I	18	Poking finger
87	Jumping sideways	54	Building bridge*
88	Skipping II		

* indicates items showing DIF in the highest/lowest category estimates but not in the average item estimates.

4.6 Item and person reliability

The item and person reliability indices for each subtest are reported in Table 14. The results demonstrated the excellent reliability for person measures in all but the Reflex subtests. The person reliability for the Reflex subtest was in the acceptable range only ($0.7 \leq$ Reliability ≤ 0.8) (Duncan, Bode, Min Lai, & Perera, 2003), and the Reflex subtest differentiated only two (high and low) strata of children's reflex motor ability. However, all but the Reflex subtests differentiated more than five strata of children's motor development. From the item perspective, the items within each subtest had excellent reliability and had more than 10 distinct strata. Although the person reliability and strata in the Reflex subtest was not satisfactory, the person reliability in the other five subtests and the item reliability in all subtests yielded very precise estimates. Furthermore, the sizable results of distinct strata shown in items of the subtests indicated that each subtest covered a wide range of item difficulty that might be suitable for measuring children with a broad range of motor abilities.

Additionally, the item and person reliability indices for the combined GM, FM, and overall motor scales (Table 14) revealed high reliability and sizable strata. The results indicated that the estimates of items within the combined scales as well as the children in the sample had good reliability and wide coverage.

Table 14: Item and person reliability statistics

Subtest	Item		Person	
	Reliability	Strata	Reliability	Strata
Reflex (n=33)	0.98	10	0.71	2
Stationary (n=326)	1.00	38	0.93	5
Locomotion (n=327)	1.00	35	1.00	21
Object Manipulation (n=296)	1.00	27	0.97	8
Grasping (n=222)	1.00	21	0.95	6
Visual-Motor Integration (n=309)	1.00	26	0.99	15
GM scale (n=336)	1.00	34	1.00	26
FM scale (n=310)	1.00	26	0.99	17
Overall motor scale (n=338)	1.00	29	1.00	30

4.7 Targeting

The item difficulty estimates and calibration for person for each subtest are plotted in the item-person maps of Figures 4 to 7. Each figure shows the distribution of children's ability measures in terms of motor ability and the item difficulty for each subtest on the same linear continuum, based on the partial credit Rasch model. The item-person maps show the wide-ranging person distributions of the Stationary subtest as -51.6 to 60.1 logits; the Locomotion subtest as -42.7 to 23.6 logits; the Object Manipulation subtest as -14.6 to 12.8 logits; the Grasping subtest as -17.0 to 22.2 logits; and the Visual-Motor Integration subtest as -25.5 to 22.1 logits. However, the person and item mean measures, indicated in the figures by letter M, are separated by 34.1 logits (i.e., 26% for the total range measurement) in the Stationary subtest, 9.7 logits (12%) in the Locomotion subtest, 3.3 logits (11%) in the Object Manipulation subtest, 12.6 logits (30%) in the Grasping subtest, and 10.1 logits (21%) in the Visual-Motor Integration subtest. The high proportion (> 20%) of difference between the person and item mean measures indicated inadequate item to person targeting in the Stationary, Grasping, and Visual-Motor Integration subtests. In particular, the results indicated that the person distributions were skewed to the upper end of the item distributions, suggesting a need for more difficult items. The person distribution in the Reflex subtest was also wide (from -10.70 to 9.98), but its mean measure of -1.0 logits was slightly negatively skewed to the item mean measure. In addition, significant ceiling effects were found in the Reflex (65.1%), Grasping (34.8%), and Visual-Motor Integration (9.6%) subtests, and a floor effect was identified in the Object Manipulation subtest (12.1%).

The item-person maps also show possible gaps between Items 4 and 7 in the Reflex subtest, between Items 19 and 20 in the Stationary subtest, between Items 8 and 9 in the Object Manipulation subtest, between Items 21 and 22 in the Grasping subtest, and between Items 10 and 16 in the Visual-Motor Integration subtest, indicating that new items are needed to fill the gaps between these pairs of items. However, the gaps are less obvious in the construction of the GM, FM and overall motor scales. The GM scale does not show significant gaps (Figure 8), thus compensating for the gaps that occurred in the Reflex, Stationary and Object Manipulation subtests. The combining of the Grasping and Visual-Motor Integration subtests of the FM scale (shown in Figure 9), also filled the gaps in the individual subtests. Owing to similarity, the item-person map of the overall motor scale was not included.

Each combined scale produced wide-ranging person distributions in the GM scale -30.5 to 22.9, the FM scale -22.0 to 23.6, and the overall motor scale -33.2 to 23.8. However, the means of the person and item measures were separated by 9.8 logits (14%) in the GM scale, 11.9 logits (25%) in the FM scale, and 11.2 logits (16%) in the overall PDMS-2 scale, indicative of inadequate item to person targeting in the FM scale.

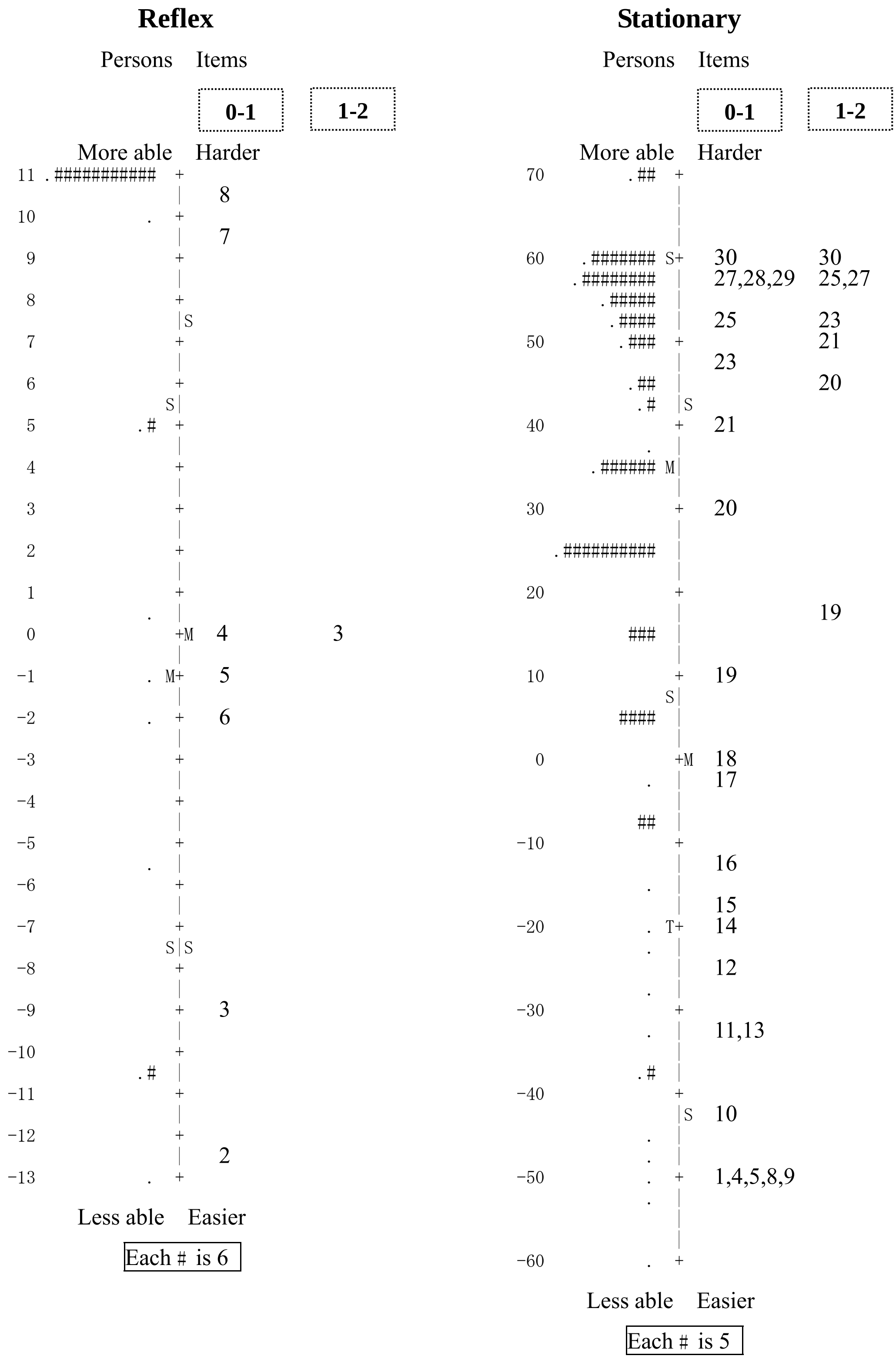

Figure 4: Item-person maps of the Reflex and Stationary subtests: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

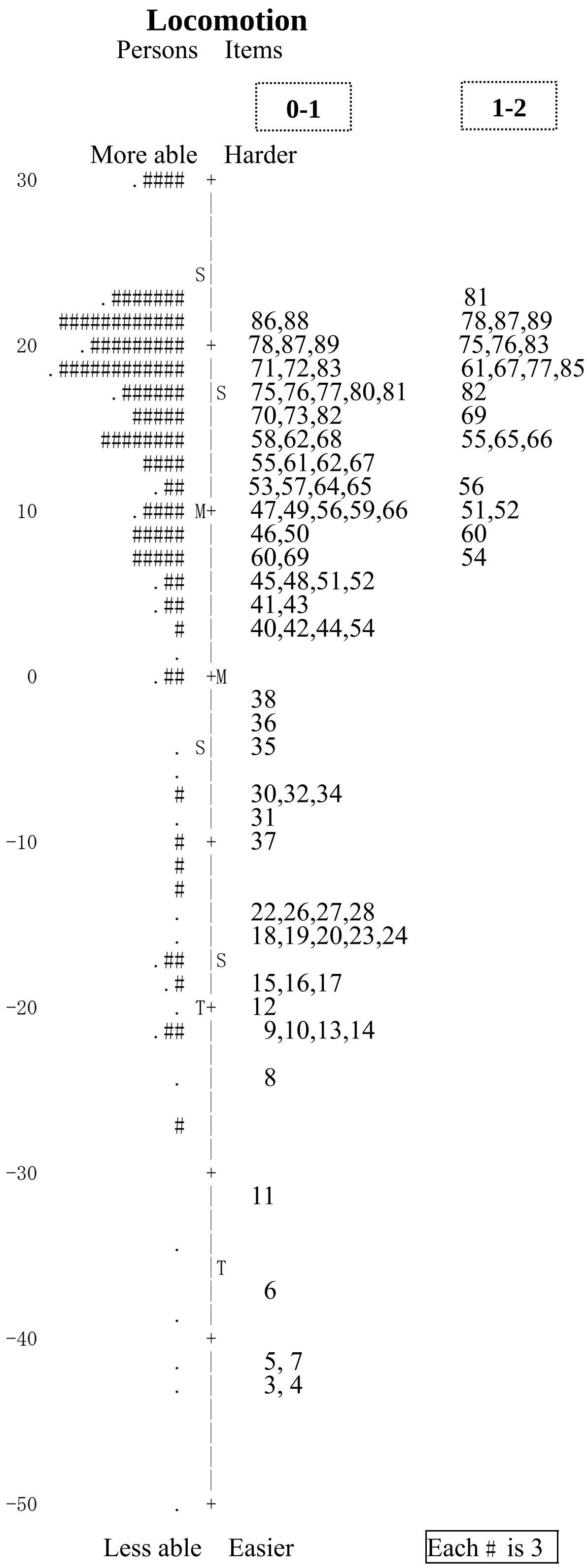

Figure 5: Item-person map of the Locomotion subtest: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

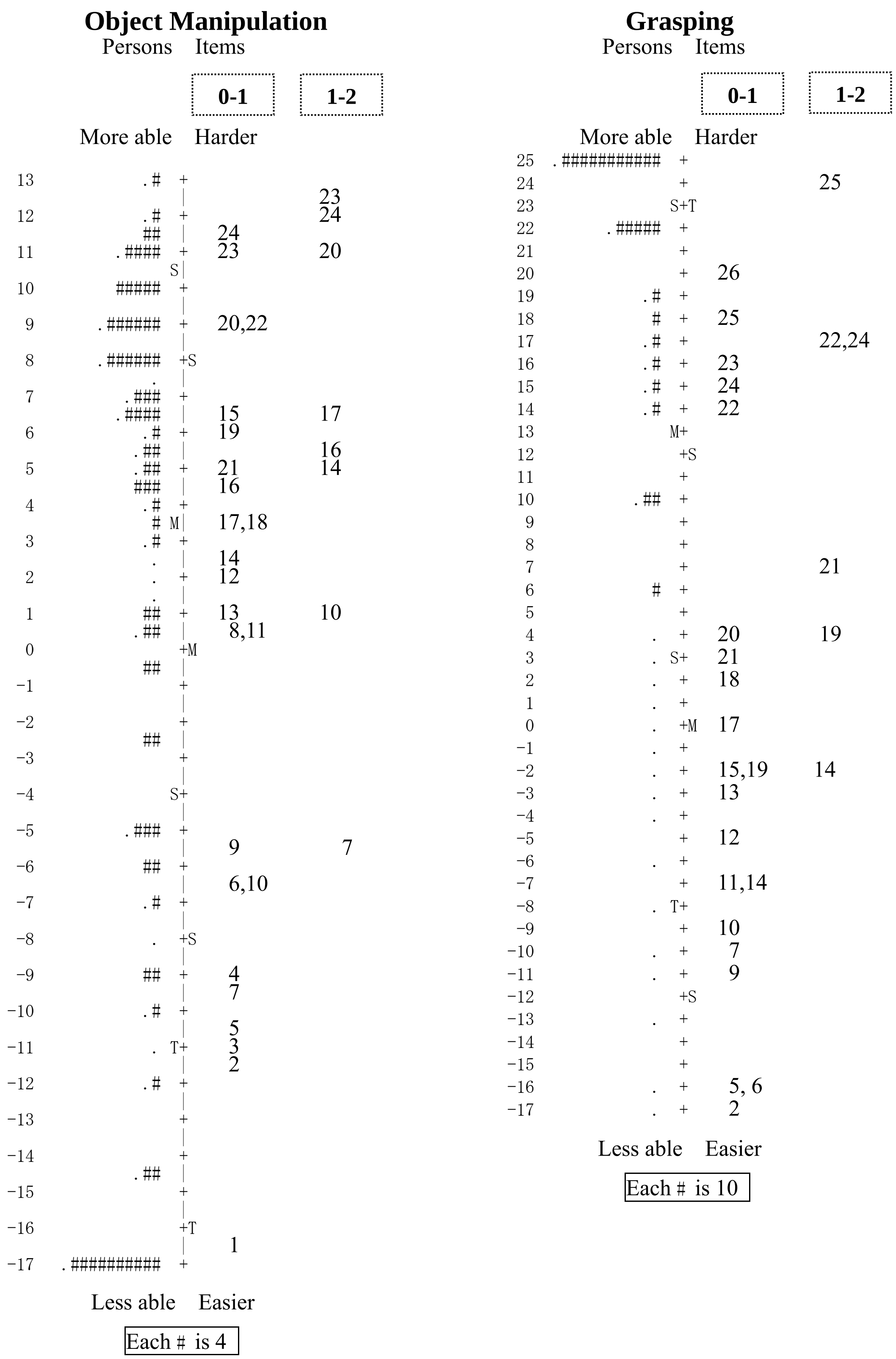

Figure 6: Item-person maps of the Object Manipulation and Grasping subtests: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

Visual-Motor Integration

	Persons		Items 0-1	Items 1-2
	More able		Harder	
23	###########	+		
22	#########	+		72
21	.######	S+		
20	######	+	71	68,71
19	.########	+		61,65,70
18	.#####	+	64	66
17	######	+	65,66,68,69,72	
16	.####	+	70	
15	.##	+	60,63	
14	###	+	57,59,61	
13	.#	+S		
12	.##	+	55,56	54
11	.###	+	51,54	
10	.##	M+		
9	##	+	47,49,52	53,62
8	.##	+	48,53	50
7	.##	+	46	
6	.###	+		
5	.###	+		45
4	.##	+	42,43,50	
3	##	+	39,40	30
2	##	+		
1	.##	+	38,45,62	31
0	##	+M	33,35	
−1	.#	S+	28,32	22
−2	.	+	36	
−3	.	+	25,27	23
−4	.#	+	29,31	
−5	.#	+	22,30	
−6	.	+	21	20
−7	####	+	23,24	
−8	.	+	19,20	
−9	.	+	17	
−10	.	+	12,14,18	
−11	.	+	13	
−12		T+	16	
−13		+S		
−14		+		
−15	###	+		
−16		+		
−17	.	+	10	
−18	.	+	9	
−19		+	11	
−20		+		
−21	#	+	3,5,6,7	
−22		+		
−23		+	2,8	
−24	.	+		
−25	.	+	1,4	
−26		+T		
	Less able		Easier	Each # is 3

Figure 7: Item-person map of the Visual-Motor Integration subtest: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

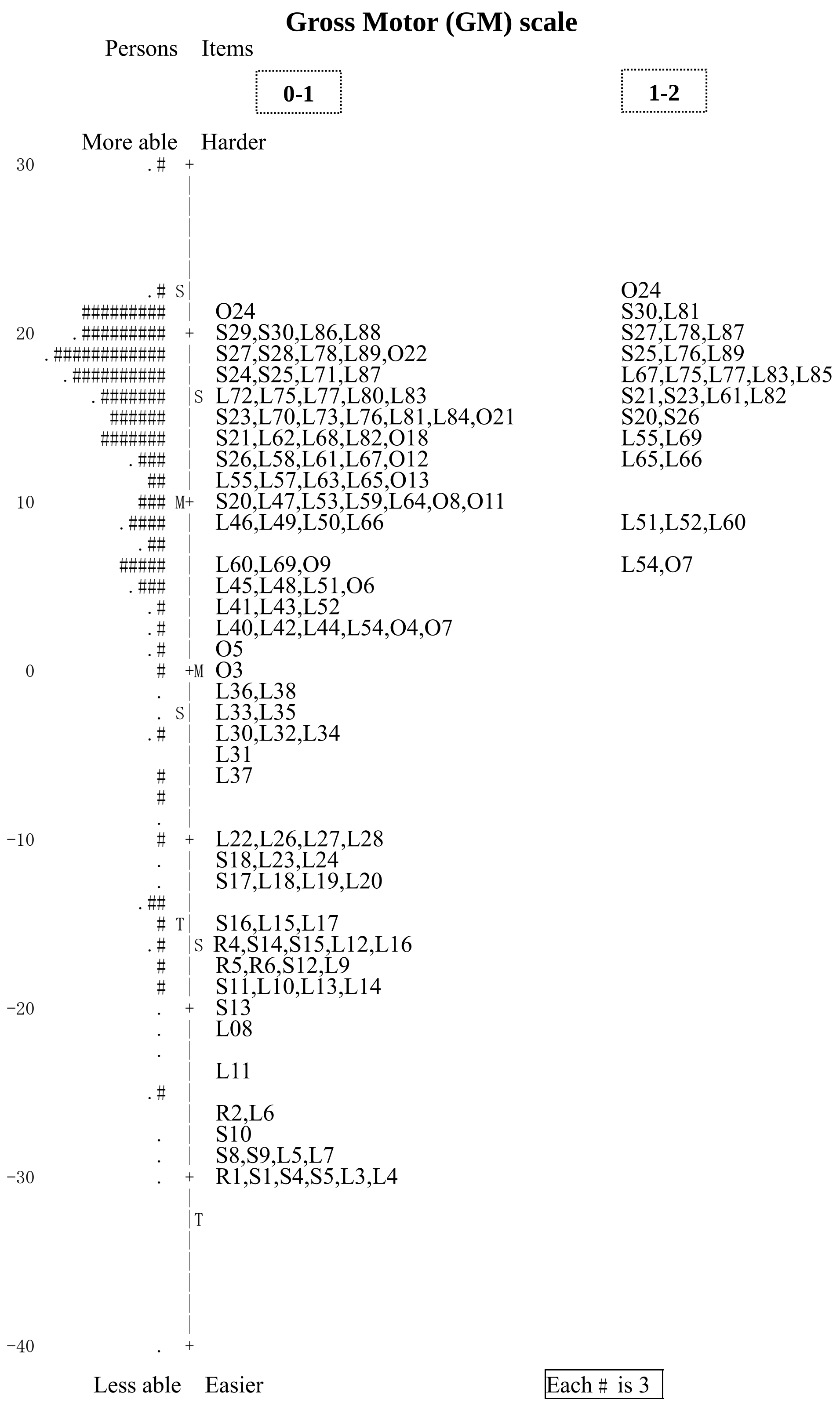

Figure 8: Item-person map of the GM scale: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

Note: R denotes Reflex; S denotes Stationary; L denotes Locomotion; and O denotes Object Manipulation.

Fine Motor (FM) scale

	Persons		Items 0-1	1-2
	More able		Harder	
24	.###########	+		V72
23	####	S+		
22	.#	+		V71
21	.#####	+	V71	V61,V68,V70
20	.####	+	V64	V65
19	####	+	G26,V65,V68,V69,V72	V66
18	.##	+	V66,V70	
17	.#	+		G24
16	.##	+	G23,V57,V59,V60,V61,V63	
15	##	+	G24	
14	.	+	V55,V56	V54
13	.##	+S	V51,V54	
12	.#	M+		
11	.	+	V47,V48,V49,V52	V53,V62
10	.#	+	V53	V50
9	.#	+	V46	
8	###	+		
7	.#	+		V45
6	.#	+	V42,V43,V50	
5	.#	+	V39,V40	V30
4	.#	+	V38	
3	.#	+	V45,V62	V31
2	.	+	V33,V35	V22
1		S+	V28,V32	
0	.	+M	V36	G19
-1	.	+	G20,V27	
-2	.	+	V25,V31	
-3	#	+	G18,V22,V29	
-4	#	+	V21,V30	V20
-5	#	+	G17,V24	
-6	.	+	G15,G19,V19,V20	G14
-7	.	+	G08,V17	
-8	.	+	G13,V14	
-9	.	+	G12,V12,V13,V18	
-10	.	T+		
-11	.	+	G11,G14,V16	
-12	.	+	G10	
-13	.	+S	G7,G9	
-14		+		
-15	.	+		
-16	.	+	V9,V10	
-17		+	V11	
-18		+	V3,V5,V6,V7	
-19	.	+	G5,G6	
-20	.	+	G2,V2,V8	
-21		+		
-22	.	+	G4,V1,V4	
-23		+		
	Less able		Easier	

Each # is 5

Figure 9: Item-person map of the FM scale: person ability measures in relation to item difficulty calibrations including thresholds for response categories on a rating scale (0-1 and 1-2). Higher measures indicate higher person ability and higher item difficulty.

Note: G denotes Grasping and V denotes Visual-Motor Integration.

4.8 Item hierarchy

Rasch analysis calibrates difficulty estimates for each item and produces the item hierarchical structures for each subtest. The average estimates are then generated for items in the partial credit Rasch model. However, 69 PDMS-2 items retain three-point rating scales in the study, and thus, in addition to the average estimates, the highest and lowest category estimates were produced on the 69 three-point rating scale items. Investigation of the scatter plots in the corresponding subtests found 33 items that have significantly different difficulty estimates between the highest and lowest category estimates (refer to Tables 15-17). Moreover, parts of the 33 items also demonstrated differences in difficulty estimates between the average and the highest category estimates or/and between the average and the lowest category estimates.

Despite of the individual items showing different difficulty estimates between the categories, the high level of agreement among three item hierarchies, arranged according to the highest, average, and the lowest estimates, respectively, was found within the six subtests ($0.98 \leq r \leq 1.0$). In addition, the three item hierarchies were found to be strongly correlated to the original item hierarchy that was ranked by age in each subtest ($0.87 \leq r \leq 0.97$). As well as being similar to the original item hierarchy, the results indicated that the three item hierarchies were nearly identical to each other.

Table 15: Item and step measures for the items in the Reflex, Stationary, and Object Manipulation subtests

Subtest/ Items	Item measure Average	Item measure SE	Step measure Lowest (0-1)	Step measure Highest (1-2)	Subtest/ Items	Item measure Average	Item measure SE	Step measure Lowest (0-1)	Step measure Highest (1-2)
Reflex									
1	Dropped		-	-	5	-0.76	0.76	-	-
2	-12.46	1.06	-	-	6	-2.02	0.84	-	-
3	**-4.61**	**0.59**	**-9.03**	**-0.20**	7	9.38	1.12	-	-
4	-0.19	0.75	-	-	8	10.66	1.20	-	-
Stationary									
1	-50.75	1.47	-	-	16	-11.91	1.87	-	-
2	Dropped				17	-3.53	1.90	-	-
3	Dropped				18	-0.02	1.87	-	-
4	-50.75	1.47	-	-	19	14.15	1.24	9.66	18.64
5	-50.75	1.47	-	-	**20**	**37.38**	**0.58**	**30.19**	**44.56**
6	Dropped				**21**	**45.83**	**0.37**	**40.60**	**51.06**
7	Dropped				22	Removed due to misfit			
8	-48.82	1.36			23	50.42	0.28	47.68	53.17
9	-48.82	1.36	-	-	24	Removed due to misfit			
10	-42.78	2.54	-	-	25	54.99	0.21	53.78	56.20
11	-31.53	1.05	-	-	26	Removed due to misfit			
12	-24.58	2.03	-	-	27	57.26	0.17	56.67	57.86
13	-33.52	1.07	-	-	28	57.58	0.24	-	-
14	-20.77	2.07	-	-	29	58.32	0.25	-	-
15	-17.02	1.99	-	-	30	59.58	0.19	59.53	59.63
Object Manipulation									
1	-16.73	1.05	-	-	13	0.77	0.33	-	-
2	-11.75	0.52	-	-	**14**	**3.77**	**0.20**	**2.82**	**4.71**
3	-10.97	0.50	-	-	15	6.50	0.22	-	-
4	-9.24	0.44	-	-	16	4.96	0.17	4.54	5.38
5	-10.27	0.47	-	-	**17**	**4.82**	**0.17**	**3.33**	**6.32**
6	-6.39	0.39	-	-	18	3.38	0.28	-	-
7	**-7.52**	**0.29**	**-9.35**	**-5.70**	19	6.16	0.22	-	-
8	0.44	0.33	-	-	20	10.06	0.16	9.44	10.68
9	-5.34	0.38	-	-	21	5.04	0.26	-	-
10	**-2.83**	**0.25**	**-6.54**	**0.87**	22	9.03	0.21	-	-
11	0.55	0.33	-	-	23	11.79	0.21	11.38	12.21
12	1.91	0.32	-	-	24	11.85	0.21	12.11	11.59

[1] One Reflex item and four Stationary items were dropped from the analysis because most children passed the items.
[2] Items with no step measures indicated dichotomous items (i.e., the rating scale of 0-1).
[3] The bolded items have significantly different category estimates either between the highest and lowest categories or between the highest, average, and the lowest categories.

Table 16: Item and step measures for the items in the Locomotion subtest

Items	Item measure		Step measure		Items	Item measure		Step measure	
	Average	SE	Lowest (0-1)	Highest (1-2)		Average	SE	Lowest (0-1)	Highest (1-2)
1	Dropped				46	9.09	0.34	-	-
2	Dropped				47	10.28	0.36	-	-
3	-43.28	1.18	-	-	48	5.64	0.39	-	-
4	-43.28	1.18	-	-	49	9.41	0.34	-	-
5	-42.11	1.02	-	-	50	9.08	0.33	-	-
6	-36.53	2.17	-	-	**51**	**7.92**	**0.25**	**6.05**	**9.80**
7	-42.11	1.02	-	-	**52**	**7.50**	**0.27**	**5.04**	**9.96**
8	-24.39	1.05	-	-	53	11.36	0.36	-	-
9	-20.98	0.60	-	-	**54**	**4.97**	**0.30**	**3.10**	**6.83**
10	-21.76	0.65	-	-	55	13.46	0.22	12.75	14.18
11	-30.97	2.09	-	-	56	10.25	0.24	9.45	11.06
12	-19.62	0.56	-	-	57	11.49	0.36	-	-
13	-21.36	0.62	-	-	58	13.89	0.30	-	-
14	-21.76	0.65	-	-	59	10.11	0.35	-	-
15	-18.43	0.53	-	-	60	8.08	0.24	7.37	8.78
16	-19.01	0.55	-	-	**61**	**15.70**	**0.22**	**13.46**	**17.94**
17	-18.72	0.54	-	-	62	14.85	0.29	-	-
18	-16.04	0.61	-	-	63	12.59	0.33	-	-
19	-15.64	0.65	-	-	64	10.85	0.36	-	-
20	-15.64	0.65	-	-	65	12.63	0.23	11.65	13.61
21	Removed due to misfit				**66**	**11.64**	**0.25**	**9.37**	**13.92**
22	-14.22	0.72	-	-	**67**	**16.01**	**0.18**	**13.42**	**18.60**
23	-15.21	0.68	-	-	68	14.03	0.30	-	-
24	-15.21	0.68	-	-	**69**	**11.08**	**0.26**	**7.07**	**15.09**
25	Removed due to misfit				70	15.62	0.27	-	-
26	-13.69	0.73	-	-	71	18.28	0.22	-	-
27	-14.22	0.72	-	-	72	18.18	0.22	-	-
28	-13.69	0.73	-	-	73	15.70	0.27	-	-
29	Removed due to misfit				74	Removed due to misfit			
30	-7.48	0.81	-	-	75	18.48	0.16	17.82	19.13
31	-8.11	0.79	-	-	**76**	**18.44**	**0.17**	**16.63**	**20.25**
32	-6.80	0.86	-	-	77	18.14	0.16	17.59	18.69
33	Removed due to misfit				78	21.07	0.16	-	-
34	-6.80	0.86	-	-	79	Removed due to misfit			
35	-4.93	1.09	-	-	80	17.70	0.24	-	-
36	-2.53	1.05	-	-	**81**	**19.69**	**0.19**	**16.63**	**22.75**
37	-10.45	0.75	-	-	82	16.45	0.19	15.12	17.78
38	-1.59	0.88	-	-	83	18.65	0.15	18.27	19.03
39	Removed due to misfit				84	Removed due to misfit			
40	3.26	0.53	-	-	85	18.29	0.23	-	-
41	4.82	0.43	-	-	86	21.58	0.24	-	-
42	3.26	0.53	-	-	87	20.69	0.16	19.35	22.04
43	4.82	0.43	-	-	88	21.71	0.25	-	-
44	3.53	0.51	-	-	89	20.32	0.16	20.38	20.26
45	5.96	0.38	-	-					

[1] Two Locomotion items were dropped from the analysis because most children passed the items.
[2] Items with no step measures indicated dichotomous items (i.e., the rating scale of 0-1).
[3] The bolded items have significantly different category estimates either between the highest and lowest categories or between the highest, average, and the lowest categories.

Table 17: Item and step measures for the items in the Grasping and Visual-Motor Integration subtests

Subtest/ Items	Item measure		Step measure		Subtest/ Items	Item measure		Step measure	
	Average	SE	Lowest (0-1)	Highest (1-2)		Average	SE	Lowest (0-1)	Highest (1-2)
Grasping									
1	Dropped				**14**	**-4.32**	**0.54**	**-6.94**	**-1.70**
2	-17.00	1.20	-	-	15	-1.68	0.63	-	-
3	Dropped				16	Removed due to misfit			
4	Dropped				17	0.26	0.52	-	-
5	-15.86	0.96	-	-	18	2.37	0.48	-	-
6	-15.86	0.96	-	-	**19**	**1.25**	**0.40**	**-1.65**	**4.16**
7	-10.37	0.81	-	-	20	3.89	0.52	-	-
8	Removed due to misfit				**21**	**4.73**	**0.37**	**2.91**	**6.56**
9	-11.01	0.80	-	-	22	15.16	0.23	13.66	16.66
10	-8.97	0.89	-	-	23	15.79	0.29	-	-
11	-7.01	1.07	-	-	24	16.29	0.20	15.68	16.89
12	-5.14	0.84	-	-	**25**	**21.35**	**0.25**	**18.39**	**24.32**
13	-3.44	0.70	-	-	26	19.57	0.30	-	-
Visual-Motor Integration									
1	-25.00	1.37	-	-	37	Removed due to misfit			
2	-23.26	1.27	-	-	38	1.22	0.40	-	-
3	-20.67	1.09	-	-	39	3.15	0.37	-	-
4	-25.00	1.37	-	-	40	3.01	0.37	-	-
5	-20.67	1.09	-	-	41	Removed due to misfit			
6	-20.67	1.09	-	-	42	4.08	0.36	-	-
7	-20.67	1.09	-	-	43	3.95	0.36	-	-
8	-23.26	1.27	-	-	44	Removed due to misfit			
9	-18.04	1.09	-	-	**45**	**3.00**	**0.27**	**1.11**	**4.90**
10	-17.02	0.93	-	-	46	7.07	0.36	-	-
11	-19.37	1.18	-	-	47	8.57	0.37	-	-
12	-10.07	0.78	-	-	48	8.29	0.37	-	-
13	-10.72	0.83	-	-	49	8.84	0.37	-	-
14	-9.51	0.71	-	-	**50**	**6.11**	**0.26**	**4.14**	**8.09**
15	Removed due to misfit				51	11.05	0.35	-	-
16	-12.22	0.87	-	-	52	8.98	0.37	-	-
17	-8.65	0.60	-	-	53	8.50	0.26	7.96	9.04
18	-10.07	0.78	-	-	54	11.60	0.25	11.46	11.75
19	-7.73	0.51	-	-	55	12.28	0.35	-	-
20	-6.89	0.34	-7.61	-6.17	56	12.28	0.35	-	-
21	-5.66	0.48	-	-	57	13.70	0.34	-	-
22	**-2.69**	**0.35**	**-4.66**	**-0.72**	58	Removed due to misfit			
23	**-5.26**	**0.37**	**-7.43**	**-3.09**	59	14.45	0.32	-	-
24	-6.77	0.47	-	-	60	14.55	0.32	-	-
25	-3.34	0.55	-	-	**61**	**16.89**	**0.19**	**14.51**	**19.27**
26	Removed due to misfit				**62**	**5.08**	**0.30**	**0.93**	**9.23**
27	-2.71	0.57	-	-	63	14.55	0.32	-	-
28	-1.45	0.54	-	-	64	18.41	0.24	-	-
29	-4.46	0.51	-	-	65	17.85	0.17	17.31	18.40
30	**-1.40**	**0.30**	**-5.42**	**2.63**	66	17.10	0.18	16.89	17.31
31	**-1.68**	**0.33**	**-3.87**	**0.52**	67	Removed due to misfit			
32	-0.91	0.50	-	-	68	18.61	0.17	17.23	19.99
33	-0.44	0.47	-	-	69	17.19	0.26	-	-
34	Removed due to misfit				**70**	**17.85**	**0.18**	**16.26**	**19.44**
35	-0.02	0.45	-	-	**71**	**20.11**	**0.16**	**20.41**	**19.81**
36	-1.80	0.57	-	-	**72**	**19.78**	**0.24**	**17.08**	**22.47**

[1] Three Grasping items were dropped from the analysis because most children passed the items.

[2] Items with no step measures indicated dichotomous items (i.e., the rating scale of 0-1).

[3] The bolded items have significantly different category estimates either between the highest and lowest categories or between the highest, average, and the lowest categories.

4.9 Additional validity examination

Test results on a developmental assessment should increase progressively with age, reflecting the sensitivity of the test to the development of motor ability (Campbell, Kolobe, Osten, Lenke, & Girolami, 1995). The Pearson *r* correlation coefficient between children's ages and their ability measures on each subtest ranged from 0.77 to 0.92. The moderate to high correlations indicated that children's ability measures in the subtests increased systematically with increasing age. In addition, the children's ability measures in the six subtests were highly inter-correlated ($0.85 \leq r \leq 0.97$), despite a moderate correlation between Reflex and Object Manipulation (r=0.63). Children's ability measures in each subtest were also highly correlated with the combined GM, FM, and overall motor scales ($0.86 \leq r \leq 1.0$). Subsequently, the GM ability measures correlated highly with the FM ability measures (r=0.98). These observations suggested acceptable convergent validity of each subtest and combined scale.

The children were evaluated using only the items that were appropriate to their age and abilities; therefore not all items in the subtests were used. WINSTEPS software for implementing the Rasch model analysis has the advantage of estimating children's performance or item difficulty, by linking the unused items within the subtests. The children's ability and item difficulty estimates were compared against a standard scoring rule for developmental tests. The standard scoring rule involved replacing the unused items with full scores below their abilities or zero scores for items above their abilities. A high level of agreement ($r \geq 0.99$) between the children's ability estimates (or item difficulty estimates) using the two procedures was found, despite replacing the unused items through use of the standard scoring rule or linking these items by the Rasch model analysis.

CHAPTER FIVE

DISCUSSION

A motor development test with sound psychometric properties can offer clinicians and researchers more accurate assessment to identify children at risk of motor delays or difficulties and thus enhance early intervention with at risk children. Although the PDMS-2 is currently used in Taiwan clinics, it was developed and validated using a group of children in the United States. The psychometric properties of the PDMS-2 rarely have been validated in Taiwan where cultural context might influence its relevance and psychometric properties. Furthermore, the psychometric properties (e.g., the appropriateness of the rating scales, dimensionality, and item hierarchy) of the PDMS-2 have not been thoroughly examined from the perspective of the Rasch model.

To address the aforementioned issues, this study applied the Rasch model to validate the psychometric properties of the PDMS-2 (including the rating scale function, dimensionality, and item hierarchy) in Taiwanese children. The DIF, item to person targeting, as well as item and person reliabilities were also examined. The Rasch analysis of the PDMS-2 provides psychometric evidence to validate its clinical usefulness in assessing children's motor development. Moreover, the results from the study justify the effectiveness of the reduced PDMS-2 in Taiwanese children. In the following sections, seven research questions concerning the PDMS-2 are addressed, on the basis of the results of the study.

5.1 Question 1

Are the three-point rating scales of each PDMS-2 item appropriate to allow differentiated children's responses?

According to the results of the study, the three-point rating scales of only 69 out of 249 PDMS-2 items can be used to differentiate children's motor development appropriately. The infrequently used middle categories of the remaining 180 items provided little additional information about children's motor development and are therefore redundant. For this reason the response categories of the 180 items were recoded as dichotomous in the study. The reliability indices of the PDMS-2 test with reorganized dichotomous scales were found to be comparable to those of the original PDMS-2 that employed the three-point rating scales. Moreover, some well-known children's motor assessments, such as the BSID (Bayley, 1969), Comprehensive Development Inventory for Infants and Toddlers (CDIIT) (Liao, Wang, Yao, & Lee, 2005), and Alberta Infant Motor Scale (Piper & Darrah, 1994) also adopt dichotomous scales, suggesting that dichotomous rating scales sufficiently differentiate children's motor development.

Additionally, the scoring criteria of the original three-point rating scales for 249 PDMS-2 items vary from item to item. Such complicated scoring criteria make it difficult as well as inconvenient for raters to make judgments about children's test performances. This study simplified the three-point rating scales of most items into the dichotomous form. The simplified dichotomous rating scales are expected to enable easier and more convenient administration for clinicians and researchers and are thus recommended for clinical use.

5.2 Question 2

Are each of the PDMS-2 subtest, FM and GM scales, or the entire PDMS-2 test able to construct a unidimensional measure?

Six individual subtests

After the rating scale performance was optimized for each item and the misfitting children's responses were eliminated, the unidimensionality of the six individual subtests was established by removing those misfitting items that did not contribute to the corresponding unidimensional constructs. The summation scores of the remaining items in each subtest may be interpreted as representing a unidimensional trait to quantify the reflex, stationary, locomotion, object manipulation, grasping, and visual-motor integration abilities, respectively. Therefore, this study confirmed that each subtest, after removing some misfitting items, constitutes a single construct, comprising a clinically meaningful continuum of motor skills, ranging from relatively easy through to difficult to accomplish.

There were three possible reasons why the misfitting items, which were removed from the subtests, demonstrated misfit in the corresponding subtests. Firstly, the requirement when performing the misfitting items might be more complex or might have influences confounded with aspects (e.g., movement experience) other than the targeted ability. For example, two "Standing on tiptoes" items in the Stationary subtest require children to demonstrate standing on tiptoes with the hands held overhead for 3 seconds. The children might have never before performed these movements, leading to awkward (or misfitting) performance on the items. In addition to the "Standing on tiptoes" items, the children's fewer opportunities to experience the following activities (i.e., “Imitating movement”, “Rolling forward”, “Jumping hurdles II”, and “Running balance/coordination”) might influence item misfit. Particularly the “Jumping hurdles II” item that requires children to jump over a string tied 10 inches off the floor, is more challenging than the “Jumping hurdles I” item (6 inches off the floor). Children with less experience in jumping might find the height level challenging and be afraid of falling. Thus the degree of difficulty and the fear of falling might affect their performance adversely on the “Jumping hurdles II” but not “Jumping hurdles I” item. Children’s experiences also appear to have a significant impact on their performance in the three misfitting Visual-Motor Integration items (i.e., “Connecting dots”, “Lacing string”, and “Imitating vertical strokes”). The children who have not yet had previous experience with these activities might not perform these items in a manner consistent with their overall ability on the test. However, it is noted that the above-mentioned reasons remain speculative and future studies are warranted to confirm the real impact of children’s motor mastery on these misfitting items.

Secondly, the questionable scoring criteria might present an additional reason for items, such as the “Scribbling” item in the Visual-Motor Integration subtest, not to show good fit. In the “Scribbling” item, if a child made at least 1 scribble more than 1 inch long, s/he would obtain the full score of 2, while the child that obtained a 1 or 0 score made a scribble less than 1 inch or no scribble at all. Based on the author’s administration experiences, however, it is frequently observed that 9 to 12 month-old infants were able to make a 1-inch-long scribble by accidence, e.g., postural reflex movements. Because examiners are not required to identify whether such performance is unintentional or not, the unintended performance would result in this item being categorized as misfitting in a Rasch analysis. As a consequence, it is suggested that the scoring criteria for the “Scribbling” item be reviewed by increasing the 1-inch-long scribble criteria to 2- or perhaps 3-inch scribble categories to minimize false, unintentional performance.

Finally the third reason why the items might have showed misfit is that the items were difficult to administer, especially to younger children below 2 years of age. Often, younger children do not comply to adults’ verbal commands and tend to behave in their own preferred ways. The administration of the PDMS-2 items, however, includes several verbal commands, which might be impractical for younger children. For example, one of the misfitting items in the Locomotion subtest was the “Scooting” item, which is designed for 9-month-old children. When administering the item, examiners are required to demonstrate scooting by using both hands to propel the body forward on the buttocks, then place a toy in front of children, and say “Scoot like I did and get the toy.” Unfortunately, children who have developed the scooting motor skills might not be compliant with the verbal instruction, so they may not attempt the scooting movement, instead preferring to get the toy by crawling or walking. For this reason the children’s performances that did not comply with instruction on the scooting item were scored as zero, according to the test manual; however the substitute scoring would result in item misfit. Similar administration difficulties for younger children were also observed in other items (including the “Bouncing”, “Creeping down stairs” and “Pivoting II” items in the Locomotion subtest, the “Manipulating paper” and “Pulling string” items in the Grasping subtest, as well as the “Transferring cubes”, “Tapping spoon”, “Turning pages II”, and “Removing socks” items in the Visual-Motor Integration subtest). It is thus suggested that future researchers should revise the verbal instructions of these items to match the appropriate level of development and ease understanding for younger children.

Gross motor scale

The PDMS-2 test developers, Folio and Fewell (2000), proposed combining the Reflex, Stationary, Locomotion, and Object Manipulation subtests to constitute a discrete GM scale. The unidimensionality of the GM scale was supported by the study, after removing twenty-two misfitting items. However, it is noted that nearly half (ten items) of the removed misfitting items were from the Object Manipulation subtest, suggesting that the Object Manipulation subtest might not, in effect, be crucial in the combined GM scale. The items in the Object Manipulation subtest all involve catching, throwing, or kicking balls. To perform the ball-related activities, children are required not only to use their large muscle systems physically (i.e., the gross motor ability) but also mentally to integrate the relative position between the balls and their bodies, such as the involvement of visual-motor integration or motor planning abilities. Therefore, the Object Manipulation subtest has the potential to form its own unidimensional scale for measuring children’s ability to manipulate balls, but it might not be so informative when combined with the other three subtests in constructing an independent GM scale, according to the results of the current study.

In addition to the ten Object Manipulation items removed from the GM scale, Figure 10 demonstrates that five additional items showed acceptable-fit in the Reflex, Stationary, and Locomotion subtests, respectively, but they were removed from the combined GM scale. On the contrary, three items that were excluded from the Stationary and Locomotion subtests showed acceptable fit in the combined GM construct. The findings illustrated that the four subtests, which are able to form their own unidimensionality, were not automatically included within the GM construct, as the Figure 10 shows parts of the individual constructs were located outside the overall frame of the GM construct.

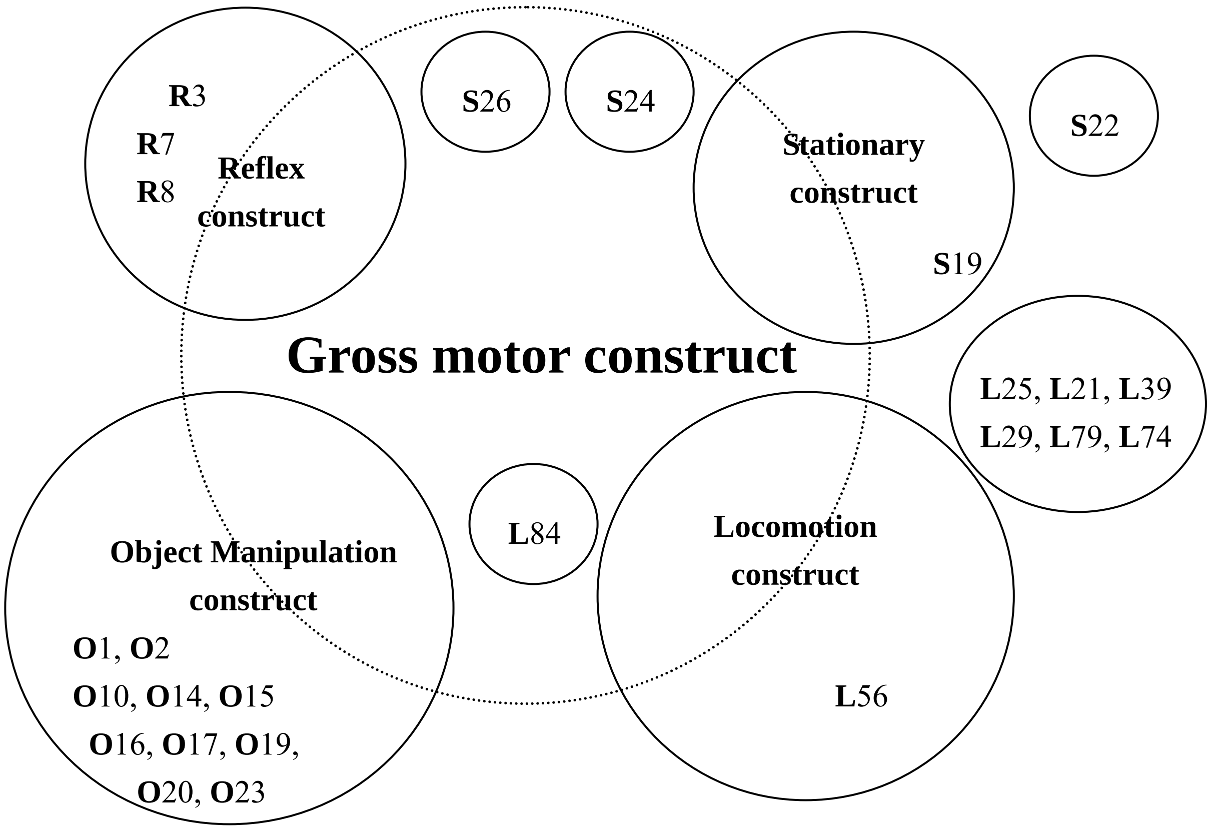

Figure 10: Relationship between the GM construct and the four individual constructs

Note: R denotes Reflex; S denotes Stationary; L denotes Locomotion; and O denotes Object Manipulation. The number indicates item number in the corresponding subtest

Fine Motor scale

Previous studies have used the Rasch model to investigate the unidimensionality of some well-known GM scales routinely (Avery et al., 2003; Campbell et al., 2002; Hands & Larkin, 2001; Zhu & Cole, 1996; Zhu & Kurz, 1994), but unidimensional examinations related to the FM construct using the Rasch model, are scarce. The unidimensionality of the FM scale in the PDMS-2 combining both the Grasping and Visual-Motor Integration subtests was upheld in this study. It is noted that four items showing acceptable fit in the corresponding subtests were not included within the FM construct. Three of these were the series of “Grasping marker” items in the Grasping subtest and one was the “Manipulating string” item in the Visual-Motor Integration subtest. Similar to the findings for the GM construct, Figure 11 illustrated that the FM construct was not accurately covered by the two individual constructs. Nine misfitting items, however, consistently fell outside the two subtests and the FM scale as well, indicating that these items do not contribute to the individual constructs or to the FM scale.

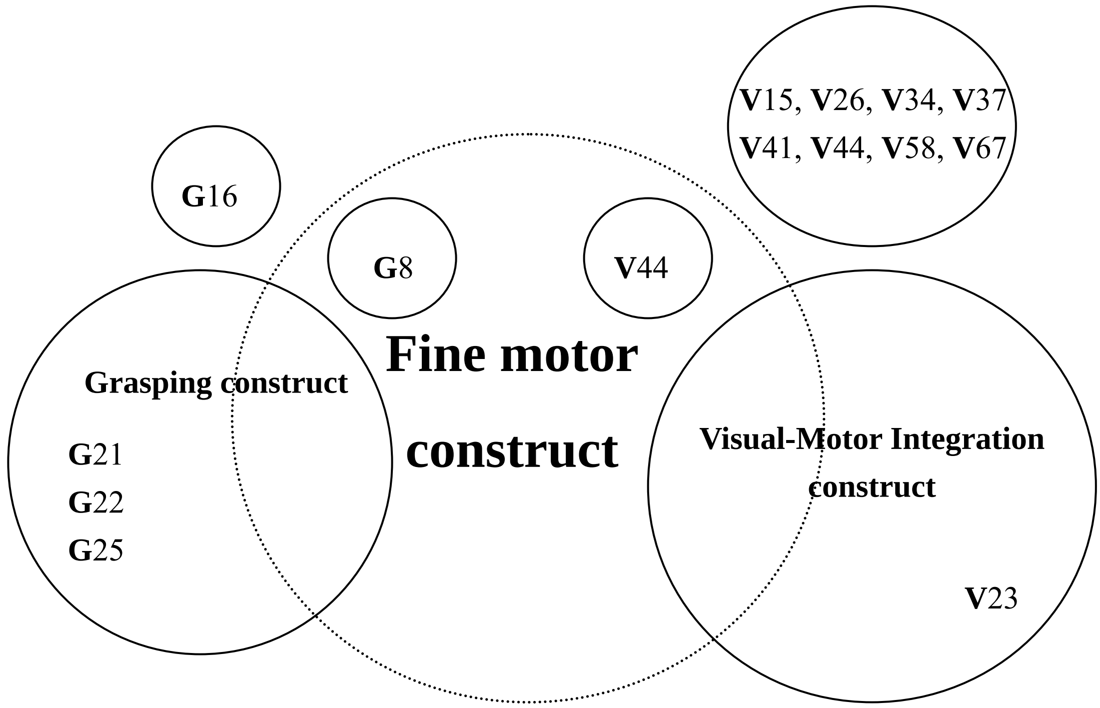

Figure 11: Relationship between the FM construct and the two individual constructs

Note: G denotes Grasping and V denotes Visual-Motor Integration. The number indicates item number in the corresponding subtest

The entire PDMS-2 indicating overall motor ability

Lastly, the entire PDMS-2 scale was examined to confirm whether it could constitute a single construct representing the overall motor ability. Rasch analysis supported the unidimensionality of the entire PDMS-2 scale representing the overall motor ability, after removing part of the misfitting items. However, the number of items (i.e., 50 items) that were removed from the PDMS-2 scale was somewhat large. Moreover, the majority of the removed items were from the Object Manipulation (10 items) and Visual-Motor Integration (20 items) subtests, thus implying that the two constructs were less informative in their contributions to measure the overall motor ability. The results offer clinicians and researchers only preliminary evidence of summing the scores from the reduced PDMS-2 scale into a score that indicates the overall motor ability. However, future studies are warranted to confirm the extent to which the overall motor ability can be upheld when the six PDMS-2 subtests are combined.

5.3 Question 3

For what components of the PDMS-2 can an objective Rasch model interval scale be established?

To the extent that the test items fit the Rasch model requirements, the ordinal raw scores can be converted into the Rasch-transformed logit scale and thus to an objective, interval scale (Bond & Fox, 2001, 2007; Rasch, 1960; Wright & Masters, 1982). This study removed most of the misfitting items in each subtest and then confirmed that the remaining items, in each subtest, met the Rasch model expectations. As a consequence, the ordinal raw scores of each PDMS-2 subtest were transformed to logit scales. The Rasch-transformed logit scores in each PDMS-2 subtest subsequently were found to reflect the true magnitude of differences between two raw scores more accurately. For example, Figure 12 illustrates the non-linear relationship between the raw score and the Rasch-transformed logit score within the Locomotion subtest. The same amount of difference in raw scores at two different places (e.g., the same score differences of 20 between 10 and 30 as well as between 60 and 80), were not equal-interval, in terms of their Rasch-transformed logit scores (i.e., in this case, 15.7 and 5.6 logits, respectively).

Other subtests also show similar relationships between the raw scores and Rasch-transformed logit scores; thus their relationship figures, except for the Stationary subtest, were not illustrated separately in the current study. The Stationary subtest showed discontinuity in the Rasch-transformed scores (refer Figure 13), and, from a theoretical perspective this would present a reasonable result because the children tested experienced an enormous improvement in stationary ability from sitting up, kneeling, and then to standing on one foot. The raw scores, however, did not reflect any developmental discontinuity. For these reasons the Rasch-transformed logit scores are recommended for clinical use and may reflect real differences when used to compare within an individual's progress or children with different motor abilities.

In addition to reflecting of real differences, the Rasch-transformed logit scores can be viewed as an interval-level of measurement. Since most statistical techniques assume that the data are on an interval scale, the Rasch-transformed logit scores have potential benefits for parametric statistical analyses. The study provided conversion tables (Appendices 9 to 12) showing the raw scores with their Rasch-transformed logit scores in each PDMS-2 subtest, to allow prospective users to derive Rasch-transformed logit scores from raw scores easily. However, it is noted that the standard error (SE) of raw scores in the Stationary and Locomotion subtests were extremely large (refer Appendices 9 and 11), indicating extremely imprecise Rasch estimations for these raw scores on the basis of the data in this study. Most of these items were for extreme (easy) items, where few of the children in this sample could be expected to fall, i.e., the likely explanation for such imprecision in estimations (very low estimates + very large SEs) was the limited number of the children with these raw scores. The limited sample size may also have influenced to the imprecise estimations contributing to the unusually wide range of item difficulty in the Stationary subtest (from -51.6 to 60.1 logits). Future studies should recruit more very young children to address this problem.

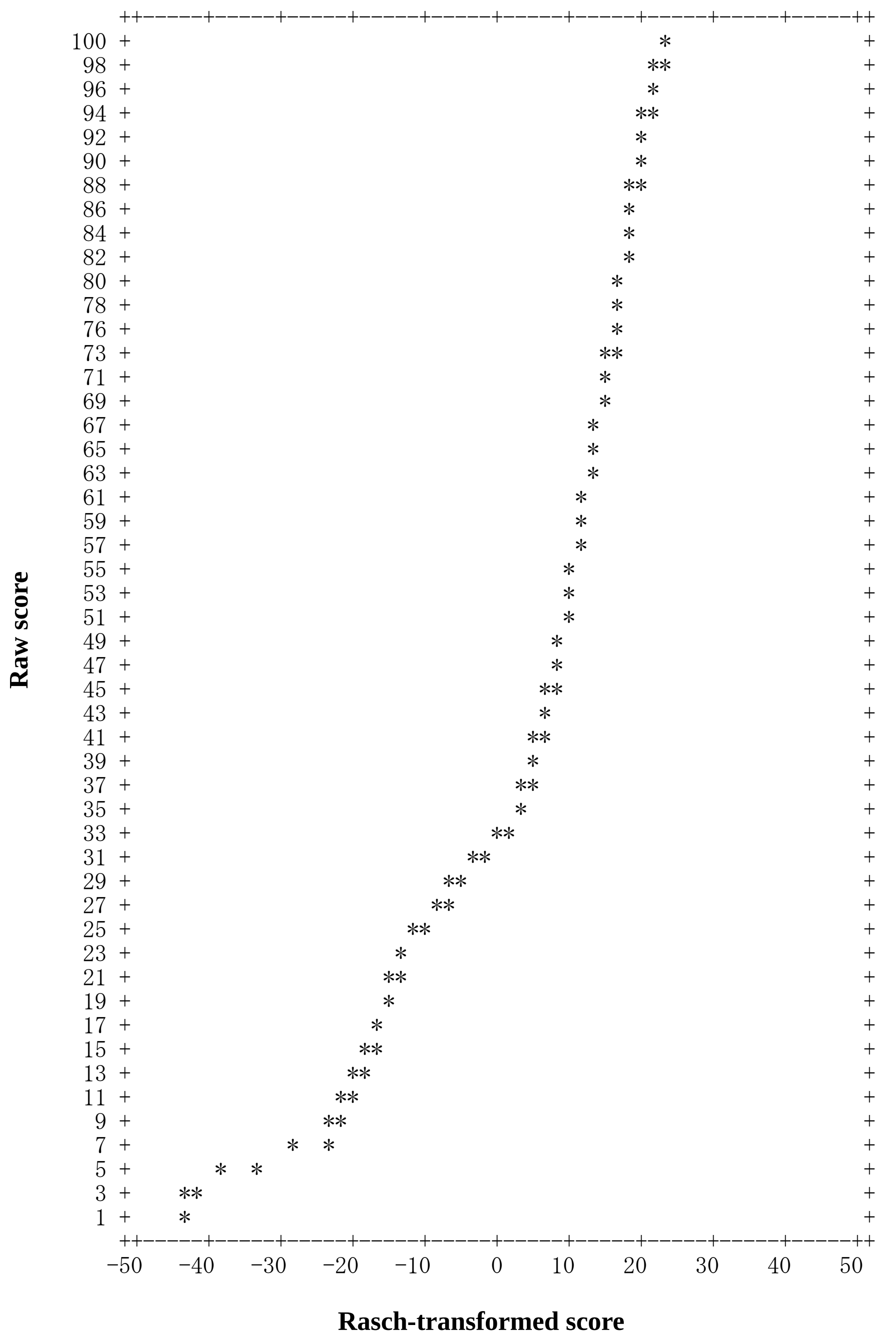

Figure 12: Relationship between raw scores and Rasch-transformed scores for the Locomotion subtest

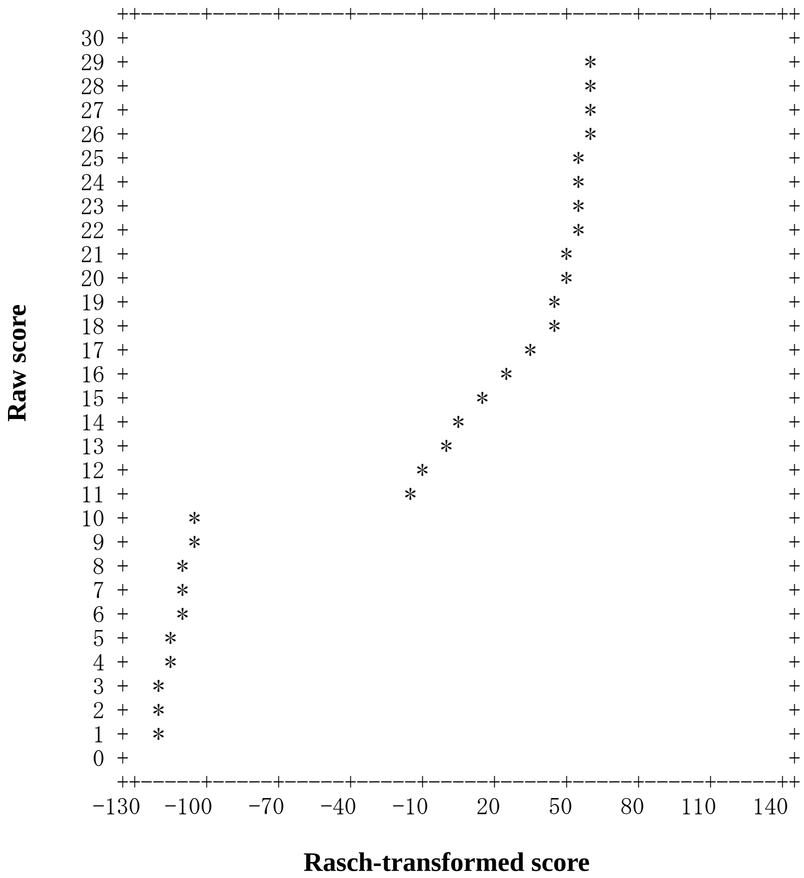

Figure 13: Relationship between raw scores and Rasch-transformed scores for the Stationary subtest

5.4 Question 4

If the PDMS-2 items fit the Rasch model expectations, to what extent does the hierarchy calibrated by the Rasch model differ from the original hierarchy?

Items that are ordered by age in developmental tests need to be justified by the item's placement by rank ordering from easy to difficult (Burton & Miller, 1998). The PDMS-2 test developers used the 2-parameter IRT model to validate the present hierarchy of the PDMS-2 items that are ordered by age, however, the 2-parameter IRT model does not approximate the definition of measurement (Andrich, 2004; Hays et al., 2000). This study utilized the Rasch model to authenticate the order of difficulty of the PDMS-2 items. On the whole, the item hierarchies within the subtest level were found to be extremely similar to those ranked by age in each subtest, justifying the effectiveness of the hierarchical order of the PDMS-2 items.

However, Folio and Fewell (2000) did not present the detailed reports and statistics of the 2-parameter IRT analyses. Therefore, it is difficult to conduct an in-depth comparison between the items' difficulty estimates obtained by the Rasch analysis in the current study and those obtained by the 2-parameter IRT analyses. However, adapting an eyeball approach to checking identified that some individual items appeared to have clearly different hierarchy between the ordering by item difficulty and ordering by age. Item 62, for example, in the Visual-Motor Integration subtest was ranked in a higher order (i.e., 62nd) according to age. On the basis of item difficulty, however, Table 17 indicated that item 62 was less difficult (i.e., 5.08 logit) than item 61 (16.89 logit) or even item 46 (7.07 logit). One consideration of the differences was that culture might account for the variations because the original item hierarchy ordered by age was developed with American children, whereas the current study was developed with Taiwanese children.

5.5 Question 5

How well do the PDMS-2 items match to the given sample in the study?

The item-person maps place children's ability measures, in terms of motor ability and item difficulty in each subtest, on the same linear continuum and allow the researcher to examine the extent to which the PDMS-2 items match the range of children's motor abilities. The study found that the items in the Stationary, Grasping, and Visual-Motor Integration subtests as well as the FM scale were not well matched to the children with higher motor abilities in this sample, indicating inadequate item to person targeting. In other words, these items were too easy and would be suitable only when assessing children with lower motor abilities. The significant ceiling effects, which were identified in Grasping and Visual-Motor Integration subtests, further confirmed the limitations of specific items in identifying children with higher motor abilities. Interestingly, the Grasping and Visual-Motor Integration subtests are both related to fine motor attributes but showed significant ceiling effects in the study. The suggestion that Taiwanese children might have more advanced manual dexterity than do American children (Chow, Henderson, & Barnett, 2001; Hsu, Cherng, Yu, & Chen-Sea, 2004), could present one explanation for the significant ceiling effects in the American-developed Grasping and Visual-Motor Integration subtests.

Additionally, the Reflex subtest appeared to exhibit a ceiling effect, however this would be considered clinically reasonable for that subtest, because the Reflex subtest was designed specifically for children aged from birth through to 11 months. In order to calibrate accurate difficulty estimates for the PDMS-2, the study stipulated that each child be assessed using the items within her/his current age band and one age band either side. Thus the children aged between 12 and 23 months, for example, would be assessed using the items located within three age bands of birth to 11 months, 12 to 23 months and 24 to 35 months. However, the 12 to 23 months old children were expected to achieve full scores in the Reflex subtest, because that subtest was designed for the age band below their level of development in the Reflex subtest, thus resulting in a pseudo ceiling effect.

Likewise, a pseudo floor effect in the Object Manipulation subtest was also expected, because the subtest was designed to assess only those children aged 12 months and older. Most children aged below 12 months have not yet developed the abilities required to perform the Object Manipulation subtest, and therefore the absence of scores produced a false representation. Thus, when put into a clinical and theoretical perspective, the Object Manipulation and Reflex subtests did not demonstrate substantive floor or ceiling effects.

5.6 Question 6

What items show DIF between male and female subjects and between children with/out motor delays or difficulties?

DIF occurs when items have different levels of difficulty based on sample characteristics rather than ability, and these DIF items are usually eliminated from the test because theoretically, they may cause items misfit overall (Andrich, 1988; Linacre & Wright, 1998). This study identified a total of 35 such items showing DIF across gender. However these items were not eliminated from the PDMS-2 in this study because the number of the DIF items in each subtest was relatively small. It was also reasonable to retain these items in the subtests, given that gender differences are expected in children's motor development (Schickedanz, Schickedanz, Forsyth, & Forsyth, 2001). Additionally, this study showed that the items in each subtest demonstrated acceptable model fit, indicating no significant DIF impact on the item fit. For these reasons, retaining the items showing DIF across gender was recommended in this study.

DIF analysis enables identification of not only the gender-specific items but also disease-specific items. This study found many items with DIF across disease entity, and, in particular, more than half of them were present in the Locomotion subtest. These DIF items highlighted the different levels of difficulty between children with/out motor delays or difficulties. In other words, children with motor delays or difficulties might complete these DIF items in a manner quite dissimilar to that of normal children. For example, children with motor disabilities might perform poorly in specific activities within the DIF items, such as the series of sit-ups, catching balls, grasping marker and locomotion-related activities. As a result, these DIF items could be useful specific indicators for screening children at risk of motor delays or difficulties when used in clinical settings, and for this reason they were not removed from the PDMS-2.

5.7 Question 7

How reliable are the estimates of the PDMS-2 items and the sample ability in this study?

The results showed that the item reliability in each subtest demonstrated high estimates. Moreover, the measures of children's motor abilities, in all but the Reflex subtests, demonstrated excellent estimation reliability and can differentiate sufficient strata. The person reliability within the Reflex subtest did not distinguish as well as other subtests because many children in this sample were able to master this subtest completely and achieved full scores. Furthermore, the rating scales of all but one item within the Reflex subtest were reorganized in the dichotomous form, thus decreasing the subtest's ability to produce accurate estimates of children's reflex motor abilities. Consequently, the Reflex subtest had only modest person reliability in the study.

5.8 Limitations

The present study encountered three issues that called for careful interpretation. First, the size of the sample was small in the Reflex subtest, which made the item fit analysis (i.e., the unidimensionality examination) less convincing, and for this reason one misfitting item was retained. Future investigations that confirm the unidimensionality of the Reflex subtest would benefit by using a larger sample. Second, sixteen items in the subtests were dropped following item fit analyses, because most children in this sample managed to pass these items. All of the items that were discarded were originally designed for children within the age range of birth to 12 months old specifically. Although this study has included a reasonable number of children (i.e., 48 subjects) in that age bracket, future studies should recruit more children aged below 12 months old to enable the item fit analysis for discarded items to be substantiated. Third, although most of the misfitting items that were removed from the individual subtests as well as from the combined scales in the study, some items exhibited misfit on just one particular fit statistic, were considered to hold potential clinical benefits by the author and were therefore retained. The retention of these misfitting items did not affect the unidimensionality of each subtest and combined scale significantly, according the PCA of residuals. More replication studies, however, are warranted to ensure the unidimensionality of each subtest retaining the misfitting items in the future. Until then, the misfitting items that are retained need to be interpreted with caution when calculating a summation score meant to be indicative of a single construct of children's motor ability.

In addition, this study reported only the children's chronological ages and did not identify their developmental stages using alternative assessments, such as the Children Developmental Inventory (Ireton, 1992) or the Pediatric Evaluation of Disability Inventory (Haley, Coster, Ludlow, Haltiwange, & Andrellos, 1992). It is thus suggested that future studies might gather more characteristics (e.g., cognitive function and adaptive behaviours) of the children to provide a more holistic understanding of their developmental stages. Furthermore, this study only used the Rasch model to examine the dimensionality of the PDMS-2. Factor analysis with the polychoric correlation matrix as input (Jöreskog & Sörbom, 1993; Wallenhammar, Nyfjäll, Lindberg, & Meding, 2004), however, is currently available to perform dimensionality examination for an ordinal scale, such as the PDMS-2. Hence, some might suggest that future studies using the polychoric correlation matrix model of factor analysis are warranted to support the results of the current study.

5.9 Implications of the study

Based on Rasch analysis of the PDMS-2, the study simplified most items' three-point rating scales, into the dichotomous form to optimize their clinical utility in discriminating various levels of children's motor development. After removing most of the misfitting items, the summation scores of each reduced PDMS-2 subtest and combined scale provided certain measures for clinicians to indicate a unidimensional trait, which quantified the children's motor abilities. By converting ordinal raw scores into Rasch-transformed logit scores, the interval-level logit scores were helpful in reflecting real differences when monitoring an individual's progress and comparing different children's motor abilities. To facilitate further the clinical value of the reduced PDMS-2 scale in developmental diagnosis however, a national normative study is warranted to establish norms for Taiwanese children of different ages and then to define the specificity and sensitivity of the reduced PDMS-2 scale to facilitate the identification of children at risk of motor delays or difficulties. Following this, the identified items with DIF across children with/out motor problems in this study might enable an alternative short-form test for rapid screening use in a clinical setting.

Ceiling effects existed in the Grasping and Visual-Motor Integration subtests as well as in the combined FM scale. Some possible item gaps were also observed in each subtest. Furthermore, the PDMS-2 did not include certain motor skills related to the traditional Chinese culture, such as the use of chopsticks. Although the reduced PDMS-2 has weaknesses, the items have been calibrated along a hierarchical unidimensional scale in the current study, constituting the foundation of the CAT application. Thus the reduced PDMS-2 scale has great potential in developing a CAT version that offers clinicians more efficient assessment of children's motor development in the future.

An item bank that spans a comprehensive range of motor abilities and includes culture relevant items might provide an improved test that can deal with the above-mentioned weaknesses, prior to the CAT application. After the items in the motor item bank are scaled along a single dimension from easy to difficult motor ability, a more powerful CAT application will emerge. The new CAT would provide an efficient evaluation that will reduce the administration time required of clinicians and researchers and the demand on children, whilst providing a much needed in-depth, culturally sensitive assessment for Taiwanese children.

BIBLIOGRAPHY

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Wasgington DC: American Psychiatric Association.

Andrich, D. (1988). *Rasch models of measurement*. Newbury Park, CA: Sage Publications.

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care, 42*(1 Suppl), 7-16.

Avery, L. M., Russell, D. J., Raina, P. S., Walter, S. D., & Rosenbaum, P. L. (2003). Rasch analysis of the Gross Motor Function Measure: Validating the assumptions of the Rasch model to create an interval-level measure. *Archives of Physical Medicine and Rehabilitation, 84*(5), 697-705.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-545). Mass: Addison-Wesley.

Blauw-Hosper, C. H., & Hadders-Algra, M. (2005). A systematic review of the effects of early intervention on motor development. *Developmental Medicine and Child Neurology, 47*(6), 421-432.

Bond, T. G. (2004). Validity and assessment: A Rasch measurement perspective. *Metodologia de las Ciencias del Comportamiento, 5*(2), 179-194.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Brown, G. T., Rodger, S., Brown, A., & Roever, C. (2005). A comparison of Canadian and Australian paediatric occupational therapists. *Occupational Therapy International, 12*(3), 137-161.

Buccieri, K. M. (2003). Use of orthoses and early intervention physical therapy to minimize hyperpronation and promote functional skills in a child with gross motor delays: A case report. *Physical and Occupational Therapy in Pediatrics, 23*(1), 5-20.

Burtner, P. A., McMain, M. P., & Crowe, T. K. (2002). Survey of occupational therapy practitioners in southwestern schools: Assessments used and preparation of students for school-based practice. *Physical and Occupational Therapy in Pediatrics, 22*(1), 25-39.

Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Champaign, IL: Human Kinetics.

Campbell, S. K., Kolobe, T. H., Osten, E. T., Lenke, M., & Girolami, G. L. (1995). Construct validity of the Test of Infant Motor Performance. *Physical Therapy, 75*(7), 585-596.

Campbell, S. K., Osten, E. T., Kolobe, T. A., & Fisher, A. G. (1993). Development of the Test of Infant Motor Performance. In G. C.V. & G. E. Gresham (Eds.), *New developments in functional assessment; Physical Medicine and Rehabilitation Clinics of North America* (pp. 541-550). Philadelphia: WB Saunders.

Campbell, S. K., Wright, B. D., & Linacre, J. M. (2002). Development of a functional movement scale for infants. *Journal of Applied Measurement, 3*(2), 190-204.

Case-Smith, J. (1998). *Pediatric occupational therapy and early intervention*. Boston:

Butterworth-Heinemann.

Case-Smith, J., & Allen, A. S. (2005). *Occupational therapy for children* (5th ed.). St. Louis: Elsevier Mosby.

Cella, D., & Chang, C. H. (2000). A discussion of item response theory and its applications in health status assessment. *Medical Care, 38*(9 Suppl), 66-72.

Chen, C. C., Bode, R. K., Granger, C. V., & Heinemann, A. W. (2005). Psychometric properties and developmental differences in children's ADL item hierarchy: A study of the WeeFIM instrument. *American Journal of Physical Medicine and Rehabilitation, 84*(9), 671-679.

Chow, S. M., Henderson, S. E., & Barnett, A. L. (2001). The Movement Assessment Battery for Children: A comparison of 4-year-old to 6-year-old children from Hong Kong and the United States. *American Journal of Occupational Therapy, 55*(1), 55-61.

Connolly, B. H., Dalton, L., Smith, J. B., Lamberth, N. G., McCay, B., & Murphy, W. (2006). Concurrent validity of the Bayley Scales of Infant Development II (BSID-II) motor scale and the Peabody Developmental Motor Scale II (PDMS-2) in 12-month-old infants. *Pediatric Physical Therapy, 18*(3), 190-196.

Coster, W. J., Haley, S. M., Andres, P. L., Ludlow, L. H., Bond, T. L., & Ni, P. S. (2004a). Refining the conceptual basis for rehabilitation outcome measurement: Personal care and instrumental activities domain. *Medical Care, 42*(1 Suppl), 62-72.

Coster, W. J., Haley, S. M., Ludlow, L. H., Andres, P. L., & Ni, P. S. (2004b). Development of an applied cognition scale to measure rehabilitation outcomes. *Archives of Physical Medicine and Rehabilitation, 85*(12), 2030-2035.

Dijkers, M. P. (2003). A computer adaptive testing simulation applied to the FIM instrument motor component. *Archives of Physical Medicine and Rehabilitation,* *84*(3), 384-393.

Duncan, P. W., Bode, R. K., Min Lai, S., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation, 84*(7), 950-963.

Edwards, S. L., & Sarwark, J. F. (2005). Infant and child motor development. *Clinical Orthopaedics and Related Research, 434*, 33-39.

Fennell, E. B., & Dikel, T. N. (2001). Cognitive and neuropsychological functioning in children with cerebral palsy. *Journal of Child Neurology, 16*(1), 58-63.

Folio, M. R., & Fewell, R. R. (1983). *Peabody Developmental Motor Scales and activity cards*. Austin, TX: PRO-ED.

Folio, M. R., & Fewell, R. R. (2000). *Peabody Developmental Motor Scales: Examiner's manual* (2nd ed.). Austin, TX: PRO-ED.

Frankenburg, W. K., & Dodds, J. B. (1967). The Denver developmental screening test. *Journal of Pediatrics, 71*(2), 181-191.

Frankenburg, W. K., & Dodds, J. B. (1969). *The Denver developmental screening test*. Denver: University of Colorado Medical Center.

Gallagher, R. J., & Cech, D. (1988). Motor assessment. In T. D. Wachs & R. Sheehan (Eds.), *Assessment of young developmentally disabled children* (pp. 241-254). New York: Plenum Press.

Gebhard, A. R., Ottenbacher, K. J., & Lane, S. J. (1994). Interrater reliability of the Peabody Developmental Motor Scales: Fine motor scale. *American Journal of Occupational Therapy, 48*(11), 976-981.

Gershon, R. C. (2005). Computer adaptive testing. *Journal of Applied Measurement, 6*(1), 109-127.

Haley, S. M., Coster, W. J., Ludlow, L. H., Haltiwanger, J. T., & Andrellos, P. J. (1992). *Pediatric Evaluation of Disability Inventory (PEDI)*. New England Medical

Center Hospitals, Inc. and PEDI Research Group.

Haley, S. M., Ni, P., Fragala-Pinkham, M. A., Skrinar, A. M., & Corzo, D. (2005). A computer adaptive testing approach for assessing physical functioning in children and adolescents. *Developmental Medicine and Child Neurology, 47*(2), 113-120.

Hands, B., & Larkin, D. (2001). Using the Rasch measurement model to investigate the construct of motor ability in young children. *Journal of Applied Measurement, 2*(2), 101-120.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9 Suppl), 28-42.

Henderson, S. E., & Sugden, D. A. (1992). *Movement Assessment Battery for Children*. Sidcup, UK: Therapy Skill Builders.

Hinderer, K. A., Richardson, P. K., & Atwater, S. W. (1989). Clinical implication of the Peabody Developmental Motor Scales: A constructive review. *Physical and Occupational Therapy in Pediatrics, 9*(2), 81-106.

Hobart, J. C., Lamping, D. L., & Thompson, A. J. (1996). Evaluating neurological outcome measures: The bare essentials. *Journal of Neurology, Neurosurgery and Psychiatry, 60*(2), 127-130.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale HJ: LEA.

Hsu, Y. W., Cherng, R. J., Yu, T. Y., & Chen-Sea, M. J. (2004). Comparison of the performance of preschoolers from Taiwan and USA in the Movement Assessment Battery for Children. *Formosan Journal of Physical Therapy, 29*(5), 307-316.

Hubert, N. C., & Wallander, J. L. (1988). Instrument selection. In D. W. Theodore & S. Robert (Eds.), *Assessment of young developmentally disabled children* (pp. 43-60). New York: Plenum press.

Jacobusse, G., van Buuren, S., & Verkerk, P. H. (2006). An interval scale for development of children aged 0-2 years. *Statistics in Medicine, 25*(13), 2272-2283.

Jöreskog. K., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International, Inc., Lawrence Erlbaum Associates Publishers, Inc.

Jette, A. M., & Haley, S. M. (2005). Contemporary measurement techniques for rehabilitation outcomes assessment. *Journal of Rehabilitation Medicine, 37*(6), 339-345.

Kadesjo, B., & Gillberg, C. (1998). Attention deficits and clumsiness in Swedish 7-year-old children. *Developmental Medicine and Child Neurology, 40*(12), 796-804.

Kolobe, T. H., Palisano, R. J., & Stratford, P. W. (1998). Comparison of two outcome measures for infants with cerebral palsy and infants with motor delays. *Physical Therapy, 78*(10), 1062-1072.

Lawlor, M. C., & Henderson, A. (1989). A descriptive study of the clinical practice patterns of occupational therapists working with infants and young children. *American Journal of Occupational Therapy, 43*(11), 755-764.

Liao, H. F., Wang, T. M., Yao, G., & Lee, W. T. (2005). Concurrent validity of the Comprehensive Developmental Inventory for Infants and Toddlers with the Bayley Scales of Infant Development-II in preterm infants. *Journal of Formosa Medicine Association, 104*(10), 731-737.

Lin, C. K., & Tseng, M. H. (2001). A survey of assessment tools for pediatric occupational therapy. *Formosan Journal of Medicine, 5*(6), 636-644.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2*(3), 266-283.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2006). WINSTEPS [computer program]. Chicago: IL.

Linacre, J. M., & Wright, B. D. (1998). *A user's guide to Bigsteps, Winsteps*. Chicago: MESA Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Mass: Addison-Wesley.

Ludlow, L. H., & Haley, S. M. (1992). Polytomous Rasch models for behavioral assessment: The Tufts Assessment of Motor Performance. In M. Wilson (Ed.), *Objective measurement: Theory into practice*. Monterey: Ablex Publishing Corporation.

Ireton, H. (1992). *Child Development Inventory, Manual.* Behavior Science Systems, Minneapolis.

Massof, R. W., & Fletcher, D. C. (2001). Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Research, 41*(3), 397-413.

McHorney, C. A., & Cohen, A. S. (2000). Equating health status measures with item response theory: Illustrations with functional status items. *Medical Care, 38*(9 Suppl), 43-59.

McHorney, C. A., Haley, S. M., & Ware, J. E., Jr. (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology, 50*(4), 451-461.

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods, 8*(1), 72-87.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York:

McGraw-Hill.

Palisano, R. J. (1986). Concurrent and predictive validities of the Bayley motor scale and the Peabody Developmental Motor Scales. *Physical Therapy, 66*(11), 1714-1719.

Palisano, R. J., Kolobe, T. H., Haley, S. M., Lowes, L. P., & Jones, S. L. (1995). Validity of the Peabody Developmental Gross Motor Scale as an evaluative measure of infants receiving physical therapy. *Physical Therapy, 75*(11), 939-948.

Palisano, R. J., & Lydic, J. S. (1984). The Peabody Developmental Motor Scales: An analysis. *Physical and Occupational Therapy in Pediatrics, 4*, 69-75.

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3*, 237-255.

Pesudovs, K., & Noble, B. A. (2005). Improving subjective scaling of pain using Rasch analysis. *Journal of Pain, 6*(9), 630-636.

Piper, M. C., & Darrah, J. (1994). *Motor assessment of the developing infant*. Philadelphia: WB Saunders.

Provost, B., Heimerl, S., McLain, C., Kim, N. H., Lopez, N. R., & Kodituwakku, P. (2004). Concurrent validity of the Bayley Scales of Infant Development II motor scale and the Peabody Developmental Motor Scales-2 in children with developmental delays. *Pediatric Physical Therapy, 16*(3), 149-156.

Rasch, G. (1960). *Probabilistic models for some intelligent and attainment tests*. Copenhagen, Dermark: Dermarks Paedagogiske Institue.

Rayan, J. P. (1983). Introduction to latent trait analysis and item response theory. In W. E. Hathaway (Ed.), *Testing in the schools. New directions for testing and measurement*. San Francisco: Jossey-Bass.

Revicki, D. A., & Cella, D. F. (1997). Health status assessment for the twenty-first century: Item response theory, item banking and computer adaptive testing.

Quality of Life Research, 6(6), 595-600.

Schickedanz, J. A., Schickedanz, D. I., Forsyth, P. D., & Forsyth, G. A. (2001). *Understanding children and adolescents* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Shonkoff, J. P., & Hauser-Cram, P. (1987). Early intervention for disabled infants and their families: A quantitative analysis. *Pediatrics, 80*(5), 650-658.

Shumway-Cook, A., & Woollacott, M. H. (1995). *Motor control: Theory and practical application*. Baltimore: Williams & Wilkins.

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205-231.

Stokes, N. A., Deitz, J. L., & Crowe, T. K. (1990). The Peabody Developmental Fine Motor Scale: An interrater reliability study. *American Journal of Occupational Therapy, 44*(4), 334-340.

Streiner, D., Norman, GR. (1995). *Health measurement scales*. Oxford, UK: Oxford, University Press.

Tieman, B. L., Palisano, R. J., & 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 Hartingsveldt, M. J., Cup, E. H., & Oostendorp, R. A. (2005). Reliability and validity of the fine motor scale of the Peabody Developmental Motor Scales-2. *Occupational Therapy International, 12*(1), 1-13.

Wachs, T. D., & Sheehan, R. (1988). Developmental patterns in disabled infants and preschoolers. In T. D. Wachs & R. Sheehan (Eds.), *Assessment of young developmentally disabled children* (pp. 3-23). New York: Plenum press.

Wade, D. (1992). *Measurement in neurological rehabilitation.* Oxford, UK: Oxford

University Press.

Wallenhammar, L., Nyfjäll, M., Lindberg, M., & Meding, B. (2004). Health-related quality of life and hand eczema-A comparison of two instruments, including factor analysis. *Journal of Investigative Dermatology, 122*, 1381-1389.

Wang, H. H., Liao, H. F., & Hsieh, C. L. (2006). Reliability, sensitivity to change, and responsiveness of the Peabody Developmental Motor scales-Second Edition for children with cerebral palsy. *Physical Therapy, 86*(10), 1351-1359.

Ware, J. E., Jr. (2003). Conceptualization and measurement of health-related quality of life: Comments on an evolving field. *Archives of Physical Medicine and Rehabilitation, 84*(4 Suppl 2), 43-51.

Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement, 6*(1), 80-99.

Wiart, L., & Darrah, J. (2001). Review of four tests of gross motor development. *Developmental Medicine and Child Neurology, 43*(4), 279-285.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Zhu, W., & Cole, E. L. (1996). Many-faceted Rasch calibration of a gross motor instrument. *Research Quarterly for Exercise and Sport, 67*(1), 24-34.

Zhu, W., & Kurz, K. A. (1994). Rasch partial credit analysis of gross motor competence. *Perceptual and Motor Skills, 79*(2), 947-961.

Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology, 227*(3), 617-622.

APPENDICES

Appendix 1: Rating scale analysis for the 8-item Reflex subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Walking reflex*				
0	1	-5.57	0.90	—
1	2	-5.64	0.05	0
2	44	0.61	0.89	0
2. Positioning reflex*				
0	4	-5.81	0.18	—
1	1	-3.22	0.18	1.24
2	42	0.86	0.96	-1.24
3. Landau reaction				
0	16	-3.16	5.19	—
1	13	1.01	1.49	-1.41
2	18	2.65	1.71	1.41
4. Protecting reaction-forward*				
0	21	-2.86	0.49	—
1	7	1.36	0.96	-0.30
2	19	3.17	0.68	0.30
5. Protecting reaction-side*				
0	22	-2.74	0.48	—
1	3	0.89	0.02	0.74
2	22	3.07	0.51	-0.74
6. Protecting reaction-forward*				
0	21	-2.94	0.19	—
1	1	0.60	0	1.89
2	25	2.84	0.26	-1.89
7. Righting reaction-forward*				
0	40	-0.49	0.77	—
1	3	3.81	0.29	0.62
2	4	4.52	0.65	-0.62
8. Protecting reaction-backward*				
0	40	-0.49	0.76	—
1	5	4.11	0.07	-0.26
2	2	4.49	0.50	0.26

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

Appendix 2: Rating scale analysis for the 30-item Stationary subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Rotating head*				
0	1	-27.88	0.48	—
1	1	-23.67	2.58	-0.06
2	132	4.63	1.87	0.06
2. Aligning trunk*				
0	1	-27.65	0.02	—
1	0	NA	NA	NA
2	133	4.41	1.20	0
3. Aligning head-front*				
0	1	-27.65	0.20	—
1	0	NA	NA	NA
2	133	4.41	1.20	0
4. Aligning head-back*				
0	0	NA	NA	—
1	2	-27.12	0.12	NA
2	132	4.65	1.27	NA
5. Aligning head I*				
0	3	-27.30	0.02	—
1	0	NA	NA	NA
2	131	4.89	0.05	0
6. Extending head*				
0	1	-27.88	0.13	—
1	0	NA	NA	NA
2	133	4.41	0.79	0
7. Aligning head II*				
0	2	-27.76	0.03	—
1	0	NA	NA	NA
2	132	4.66	0.20	0
8. Aligning head III*				
0	3	-27.30	0.14	—
1	1	-24.33	0.01	0.08
2	130	5.12	0.37	-0.08
9. Stabilizing trunk*				
0	3	-27.30	0.22	—
1	2	-23.25	0.18	-0.72
2	129	5.33	0.85	0.72
10. Aligning head IV*				
0	6	-25.47	0.06	—
1	0	NA	NA	NA
2	128	5.56	0.10	0
11. Sitting I*				
0	16	-21.54	0.47	—
1	3	-14.58	0.02	-0.52
2	115	8.24	0.51	0.52
12. Sitting/reaching*				
0	24	-18.68	0.68	—
1	1	-10.52	0	0.88
2	109	9.34	0.33	-0.88
13. Pulling to sit*				
0	13	-22.83	0.20	—
1	3	-15.52	0.12	-1.04
2	118	7.65	0.86	1.04

Continued

Appendix 2: Rating scale analysis for the 30-item Stationary subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
14. Sitting II*				
0	26	-18.02	0.09	—
1	0	NA	NA	NA
2	108	9.52	0.18	0
15. Sitting with toy*				
0	28	-17.40	0.02	—
1	0	NA	NA	NA
2	106	9.87	0.03	0
16. Sitting III*				
0	29	-17.05	0.10	—
1	1	-6.36	0	0.67
2	104	10.19	0.20	-0.67
17. Raising to sit*				
0	41	-13.24	0.45	—
1	3	-1.03	0.01	-0.18
2	90	12.28	0.49	0.18
18. Sitting up*				
0	42	-12.92	0.93	—
1	4	-1.20	0.04	-0.79
2	88	12.58	0.19	0.79
19. Kneeling				
0	72	-6.10	0.11	—
1	15	10.22	0	-2.71
2	115	20.72	0.11	2.71
20. Standing on 1 foot I				
0	99	13.31	0.76	—
1	67	24.93	0.18	-2.74
2	69	28.92	0.57	2.74
21. Standing on 1 foot II				
0	81	22.34	0.69	—
1	51	27.60	0.28	-0.99
2	145	30.97	0.60	0.99
22. Standing on tiptoes I				
0	48	20.53	1.75	—
1	66	26.31	0.45	-1.86
2	163	30.59	0.93	1.86
23. Standing on 1 foot III				
0	110	23.56	0.44	—
1	38	28.71	0.20	-0.31
2	129	31.21	0.53	0.31
24. Standing on tiptoes II*				
0	105	27.16	1.59	—
1	24	30.32	0.80	0.48
2	93	31.51	0.89	-0.48
25. Standing on 1 foot IV				
0	92	26.78	0.94	—
1	46	29.83	0.48	-0.35
2	84	31.83	0.54	0.35
26. Imitating movement				
0	11	22.02	1.71	—
1	49	27.16	1.48	-2.17
2	162	30.47	1.25	2.17

Continued

Appendix 2: Rating scale analysis for the 30-item Stationary subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
27. Standing on 1 foot V				
0	138	27.80	0.76	—
1	33	31.41	0.38	0.01
2	51	32.10	0.65	-0.01
28. Sit-ups I*				
0	139	28.01	0.83	—
1	12	30.13	1.84	1.28
2	71	31.75	9.90	-1.28
29. Sit-ups II*				
0	86	29.25	0.60	—
1	11	32.00	0.62	1.18
2	52	31.92	9.90	-1.18
30. Push-ups				
0	112	29.96	0.84	—
1	18	32.33	0.65	0.12
2	13	32.61	1.26	-0.12

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

NA indicates not available.

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Thrusting legs*				
0	2	-13.43	0.09	—
1	0	NA	NA	NA
2	132	-3.00	0.63	0
2. Turning from side to back*				
0	0	NA	NA	—
1	1	-12.43	0.14	NA
2	133	-3.09	1.27	NA
3. Thrusting arm*				
0	2	-13.43	0.37	—
1	2	-12.21	0.02	0.29
2	130	-2.86	0.64	-0.29
4. Bearing weight*				
0	4	-12.08	9.90	—
1	1	-11.48	0	1.41
2	129	-2.82	0.85	-1.41
5. Extending trunk*				
0	3	-13.19	0.20	—
1	1	-11.96	0	1.26
2	130	-2.86	0.62	-1.26
6. Symmetrical posture*				
0	8	-12.27	0.03	—
1	0	NA	NA	NA
2	126	-2.58	0.16	0
7. Propping of forearms*				
0	5	-12.79	0.04	—
1	0	NA	NA	NA
2	129	-2.79	0.30	0
8. Rolling I*				
0	10	-11.14	9.90	—
1	4	-10.00	0	0.47
2	120	-2.27	5.89	-0.47
9. Extending arms and legs I*				
0	20	-10.41	0.55	—
1	6	-7.82	0.08	0.46
2	108	-1.56	0.68	-0.46
10. Flexing legs*				
0	16	-10.67	2.26	—
1	4	-8.99	0.04	0.79
2	114	-1.90	0.61	-0.79
11. Extending arms and legs II*				
0	11	-11.64	0.15	—
1	0	NA	NA	NA
2	123	-2.40	0.27	0
12. Extending arm I*				
0	27	-9.90	0.10	—
1	0	NA	NA	NA
2	107	-1.46	0.17	0
13. Flexing body*				
0	19	-10.47	1.21	—
1	3	-7.16	0.57	1.19
2	112	-1.81	0.83	-1.19

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
14. Pushing up*				
0	19	-10.61	0.12	—
1	0	NA	NA	NA
2	115	-1.93	0.35	0
15. Extending arm II*				
0	27	-9.72	1.16	—
1	6	-7.64	0.02	0.57
2	101	-1.14	0.52	-0.57
16. Rolling II*				
0	19	-10.22	1.82	—
1	8	-8.38	0.98	0.10
2	107	-1.51	0.67	-0.10
17. Rolling III*				
0	24	-9.88	1.12	—
1	5	-7.94	2.43	0.74
2	105	-1.39	0.67	-0.74
18. Moving forward*				
0	39	-8.93	0.65	—
1	1	-6.54	0	2.45
2	94	-0.73	0.69	-2.45
19. Raising shoulders and buttocks*				
0	38	-9.07	0.29	—
1	5	-6.54	0.01	0.70
2	91	-0.51	0.21	-0.70
20. Creeping I*				
0	45	-8.61	0.08	—
1	0	NA	NA	NA
2	89	-0.40	0.21	0
21. Scooting*				
0	77	-6.03	9.90	—
1	1	-0.17	0.03	2.74
2	53	0.81	9.90	-2.74
22. Pivoting I*				
0	47	-8.40	4.24	—
1	1	-5.38	0	2.35
2	85	-0.30	0.50	-2.35
23. Standing I*				
0	42	-8.83	0.15	—
1	3	-5.68	0.01	1.22
2	89	-0.40	0.25	-1.22
24. Creeping II*				
0	45	-8.61	0.16	—
1	1	-6.38	0.02	2.35
2	88	-0.33	0.11	-2.35
25. Bouncing*				
0	57	-7.46	3.88	—
1	1	-2.46	0.02	2.45
2	74	0.08	3.41	-2.45
26. Cruising III*				
0	46	-8.53	0.24	—
1	2	-4.92	0	1.63
2	86	-0.25	0.52	-1.63

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
27. Lowering*				
0	40	-8.93	0.31	—
1	8	-5.55	0.07	0.09
2	86	-0.25	0.47	-0.09
28. Stepping I*				
0	43	-8.50	2.87	—
1	1	-6.38	0.01	2.37
2	90	-0.57	9.90	-2.37
29. Pivoting II*				
0	49	-8.16	9.90	—
1	7	-3.53	1.31	0.29
2	77	-0.01	1.95	-0.29
30. Standing II*				
0	55	-7.88	0.20	—
1	4	-3.35	0.01	1.01
2	75	0.31	0.74	-1.01
31. Standing III*				
0	54	-7.95	0.23	—
1	5	-3.87	0.02	0.75
2	75	0.34	0.24	-0.75
32. Stepping II*				
0	60	-7.52	0.13	—
1	1	-3.80	0.01	2.54
2	73	0.44	0.14	-2.54
33. Standing up*				
0	65	-7.14	0.33	—
1	3	-1.55	0.03	1.52
2	132	2.69	0.58	-1.52
34. Walking I*				
0	61	-7.46	0.10	—
1	1	-1.97	0.01	2.56
2	139	2.49	0.17	-2.56
35. Walking II				
0	60	-7.52	0.19	—
1	4	-3.18	0	1.11
2	137	2.57	0.20	-1.11
36. Standing and moving balance*				
0	67	-7.02	0.11	—
1	0	NA	NA	NA
2	134	2.67	0.10	0
37. Creeping up the stairs II*				
0	54	-7.96	0.14	—
1	1	-4.68	0	2.43
2	145	2.21	0.14	-2.43
38. Walking III*				
0	67	-7.03	0.13	—
1	1	-1.97	0	2.69
2	132	2.72	0.18	-2.69
39. Creeping down stairs I*				
0	72	-6.42	4.00	—
1	8	-1.81	0.05	0.57
2	120	3.02	3.43	-0.57

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
40. Walking up stairs I*				
0	76	-6.38	0.40	—
1	6	-1.03	0.04	0.94
2	118	3.20	0.52	-0.94
41. Walking fast*				
0	81	-6.07	0.36	—
1	11	-0.06	0.15	0.30
2	107	3.56	0.54	-0.30
42. Walking backward I*				
0	80	-6.16	0.18	—
1	6	-0.91	0.02	0.98
2	115	3.36	0.26	-0.98
43. Walking down stairs I*				
0	88	-5.62	0.31	—
1	6	-0.91	0.11	1.05
2	106	3.64	0.28	-1.05
44. Walking backward II*				
0	85	-5.85	0.18	—
1	2	-0.06	0.02	2.18
2	114	3.38	0.23	-2.18
45. Running*				
0	95	-5.18	2.02	—
1	8	0.73	0.11	0.81
2	96	3.83	0.80	-0.81
46. Standing VI*				
0	121	-3.79	1.11	—
1	13	1.93	0.44	0.43
2	66	4.84	0.53	-0.43
47. Walking sideways*				
0	140	-2.95	0.97	—
1	6	2.98	0.23	1.27
2	54	5.21	0.88	-1.27
48. Walking line I*				
0	88	-5.54	2.51	—
1	14	0.25	0.13	0.11
2	100	3.78	1.37	-0.11
49. Jumping forward I*				
0	124	-3.70	0.44	—
1	11	2.18	0.03	0.69
2	150	6.60	1.53	-0.69
50. Jumping up I*				
0	113	-4.22	0.63	—
1	18	2.01	0.07	0.05
2	153	6.47	1.23	-0.05
51. Jumping down I				
0	104	-4.63	1.29	—
1	38	2.01	0.32	-0.99
2	143	6.72	0.71	0.99
52. Walking up stairs II				
0	91	-5.39	0.72	—
1	52	1.68	0.23	-1.70
2	140	6.76	0.70	1.70

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
53. Walking down stairs II*				
0	101	0.55	9.90	—
1	3	2.85	0.08	2.15
2	127	7.08	0.84	-2.15
54. Walking backward III				
0	33	-2.63	9.90	—
1	27	0.85	0.13	-0.78
2	172	5.95	0.96	0.78
55. Jumping up II				
0	120	1.16	0.76	—
1	28	5.73	0.27	-0.15
2	82	8.03	0.78	0.15
56. Walking line II				
0	80	-0.17	1.61	—
1	25	3.19	0.60	-0.27
2	128	7.11	0.83	0.27
57. Walking up stairs III*				
0	97	0.36	0.52	—
1	6	2.43	1.00	1.41
2	128	7.14	0.48	-1.41
58. Jumping down II*				
0	128	1.50	2.19	—
1	7	5.52	0.03	1.49
2	96	7.64	2.71	-1.49
59. Walking on the tiptoes I*				
0	77	-0.32	1.69	—
1	15	2.90	1.47	0.34
2	140	6.82	1.11	-0.34
60. Running speed I				
0	57	-1.25	0.65	—
1	25	2.19	0.22	-0.39
2	147	6.58	0.65	0.39
61. Jumping forward II				
0	127	1.36	0.65	—
1	79	7.19	0.36	-1.83
2	35	9.11	1.04	1.83
62. Jumping down III*				
0	132	1.61	1.21	—
1	19	5.66	0.31	0.40
2	83	8.12	1.28	-0.40
63. Jumping hurdles I*				
0	111	0.84	1.83	—
1	10	5.77	0.47	0.98
2	113	7.40	1.60	-0.98
64. Walking on the tiptoes II*				
0	96	0.39	9.90	—
1	6	4.43	1.29	1.41
2	131	7.02	1.27	-1.41
65. Walking up stairs VI				
0	103	0.55	1.25	—
1	28	5.14	0.24	-0.24
2	221	9.39	0.72	0.24

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
66. Running speed II				
0	20	2.67	0.94	—
1	50	4.88	0.37	-1.39
2	211	9.57	0.69	1.39
67. Jumping forward III				
0	54	3.85	0.65	—
1	104	7.64	0.48	-1.96
2	121	10.82	0.85	1.96
68. Walking line III*				
0	57	4.04	1.14	—
1	14	5.76	0.15	0.77
2	208	9.62	0.74	-0.77
69. Running form				
0	8	1.41	0.95	—
1	75	4.91	0.87	-2.56
2	198	9.78	0.61	2.56
70. Walking line forward*				
0	82	4.62	0.59	—
1	13	7.61	0.57	0.99
2	184	9.97	1.05	-0.99
71. Walking down stairs III*				
0	136	6.00	5.41	—
1	6	7.37	0.59	2.03
2	138	10.51	4.51	-2.03
72. Jumping forward on 1 foot*				
0	123	5.56	0.52	—
1	25	8.61	0.35	0.45
2	131	10.79	0.71	-0.45
73. Jumping up III*				
0	71	4.36	0.58	—
1	24	6.69	0.37	0.25
2	184	10.01	0.67	-0.25
74. Running balance/ coordination				
0	72	4.58	2.34	—
1	92	8.12	1.21	-1.61
2	115	10.74	0.94	1.61
75. Walking line backward I				
0	130	5.80	0.87	—
1	47	9.46	0.72	-0.31
2	102	10.91	2.12	0.31
76. Jumping forward VI				
0	113	5.44	0.96	—
1	83	9.09	0.48	-1.21
2	83	11.36	0.75	1.21
77. Hopping				
0	126	5.57	0.42	—
1	43	8.98	0.16	-0.20
2	110	11.13	0.49	0.20
78. Walking line backward II				
0	142	8.05	0.90	—
1	42	10.98	0.71	-0.31
2	40	11.70	1.26	0.31

Continued

Appendix 3: Rating scale analysis for the 89-item Locomotion subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
79. Rolling forward				
0	46	6.26	1.05	—
1	56	9.40	3.30	-0.85
2	122	10.31	2.56	0.85
80. Galloping*				
0	77	6.70	0.65	—
1	5	9.19	0.46	2.14
2	143	10.61	0.81	-2.14
81. Jumping forward V				
0	63	6.49	0.83	—
1	131	9.96	0.72	-2.39
2	31	11.76	1.09	2.39
82. Turning jump				
0	33	5.68	1.34	—
1	55	7.97	0.55	-1.02
2	136	10.64	0.89	1.02
83. Hopping forward				
0	88	6.94	0.54	—
1	36	9.20	0.12	0.02
2	100	11.31	0.44	-0.02
84. Jumping hurdles II*				
0	54	6.19	1.61	—
1	9	9.01	1.60	1.34
2	161	10.29	1.63	-1.34
85. Running speed and agility				
0	7	3.27	0.67	—
1	89	7.53	1.29	-2.82
2	131	10.65	0.72	2.82
86. Skipping I*				
0	152	8.22	0.88	—
1	25	10.49	1.00	0.37
2	47	11.91	1.72	-0.37
87. Jumping sideways				
0	116	7.58	0.78	—
1	71	10.49	0.49	-1.07
2	37	12.10	0.78	1.07
88. Skipping II*				
0	91	9.48	0.77	—
1	13	10.81	4.34	0.98
2	43	12.06	1.18	-0.98
89. Hopping speed				
0	59	8.84	0.67	—
1	24	10.44	0.46	0.30
2	63	11.73	0.89	-0.30

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

NA indicates not available.

Appendix 4: Rating scale analysis for the 24-item Object Manipulation subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Catching ball I*				
0	3	-7.32	2.22	—
1	0	NA	NA	NA
2	150	-1.92	0.99	0
2. Rolling ball*				
0	17	-6.68	9.90	—
1	3	-7.21	0.16	1.44
2	133	-1.32	0.83	-1.44
3. Flinging ball*				
0	15	-7.62	0.22	—
1	7	-6.15	0.02	0.48
2	131	-1.17	0.59	-0.48
4. Kicking ball I*				
0	29	-6.54	5.20	—
1	8	-4.94	0.04	0.62
2	116	-0.70	0.77	-0.62
5. Throwing ball*				
0	24	-6.90	1.49	—
1	4	-6.08	0.05	1.29
2	126	-0.93	0.39	-1.29
6. Kicking ball II*				
0	50	-5.70	0.62	—
1	4	-3.79	0.07	1.58
2	99	-0.10	0.26	-1.58
7. Throwing ball-overhand I				
0	32	-6.56	0.52	—
1	26	-3.68	0.17	-0.79
2	97	-0.02	0.89	0.79
8. Throwing ball-underhand I*				
0	101	-3.64	0.65	—
1	6	-0.51	0.11	1.60
2	125	2.30	0.77	-1.60
9. Kicking ball III*				
0	54	-5.57	0.37	—
1	10	-2.33	0.04	0.68
2	172	1.50	0.48	-0.68
10. Catching ball II				
0	40	-4.93	1.00	—
1	56	-1.48	0.70	-1.62
2	132	2.16	1.25	1.62
11. Throwing ball-overhand II*				
0	84	-3.40	0.93	—
1	15	-0.79	0.32	0.56
2	124	2.29	6.17	-0.56
12. Throwing ball-underhand II*				
0	105	-2.87	0.35	—
1	15	0.74	0.12	0.72
2	103	2.68	1.44	-0.72
13. Kicking ball IV*				
0	104	-2.86	1.45	—
1	2	-0.99	0.20	2.82
2	122	2.50	0.35	-2.82

Continued

Appendix 4: Rating scale analysis for the 24-item Object Manipulation subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
14. Catching ball III				
0	48	-0.18	1.00	—
1	50	1.83	0.66	-0.71
2	64	3.24	0.90	0.71
15. Throwing ball-overhand III*				
0	132	1.52	2.47	—
1	19	2.79	1.33	1.03
2	136	5.39	1.60	-1.03
16. Hitting target-underhand				
0	87	0.67	0.85	—
1	43	3.03	0.52	0
2	157	5.08	0.91	0
17. Catching ball IV				
0	63	0.23	0.96	—
1	85	2.56	0.36	-1.10
2	140	5.39	0.65	1.10
18. Hitting target-overhand I*				
0	52	-0.19	0.48	—
1	19	1.68	0.26	0.65
2	216	4.47	0.80	-0.65
19. Throwing ball-underhand III*				
0	122	1.34	2.50	—
1	19	2.74	0.50	1.03
2	146	5.28	0.80	-1.03
20. Hitting target-overhand II				
0	162	3.32	0.76	—
1	40	5.72	0.96	-0.26
2	32	6.58	2.45	0.26
21. Bouncing ball*				
0	52	2.12	1.74	—
1	7	2.33	0.20	1.88
2	175	4.86	7.14	-1.88
22. Catching ball V*				
0	131	2.95	0.84	—
1	27	4.82	0.47	0.50
2	74	6.18	2.22	-0.50
23. Kicking ball V				
0	124	4.68	0.90	—
1	22	6.69	0.66	-0.46
2	7	7.09	2.44	0.46
24. Catching bounced ball				
0	133	4.77	0.90	—
1	11	6.41	0.58	0.41
2	9	8.04	1.01	-0.41

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

NA indicates not available.

Appendix 5: Rating scale analysis for the 26-item Grasping subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Grasping reflex				
0	0	NA	NA	—
1	0	NA	NA	NA
2	136	NA	NA	NA
2. Grasping cloth*				
0	0	NA	NA	—
1	2	-7.49	0.08	NA
2	80	0.55	0.79	NA
3. Disappearing reflex				
0	0	NA	NA	—
1	0	NA	NA	NA
2	136	NA	NA	NA
4. Grasping rattle I*				
0	0	NA	NA	—
1	1	-8.74	0.02	NA
2	81	0.47	0.63	NA
5. Holding rattle*				
0	3	-5.05	9.90	—
1	1	-8.74	0.18	1.13
2	78	0.68	0.99	-1.13
6. Manipulating rattle*				
0	3	-7.68	0.04	—
1	0	NA	NA	NA
2	79	0.66	0.28	0
7. Grasping rattle II*				
0	8	-6.15	1.17	—
1	5	-3.65	0.22	-0.02
2	69	1.40	0.77	0.02
8. Pulling string*				
0	21	-4.25	1.90	—
1	6	-1.86	0.11	0.20
2	55	2.36	1.28	-0.20
9. Securing paper*				
0	8	-6.04	1.18	—
1	6	-3.60	1.70	-0.24
2	70	1.59	0.70	0.24
10. Grasping cube I*				
0	14	-5.34	0.18	—
1	1	-4.41	0.02	1.98
2	67	1.62	0.22	-1.98
11. Grasping cube II*				
0	16	-5.14	0.20	—
1	3	-2.86	0	0.86
2	63	1.91	0.33	-0.86
12. Shaking rattle I*				
0	16	-4.88	0.91	—
1	4	-3.20	0.08	0.54
2	62	1.94	0.67	-0.54
13. Shaking rattle II*				
0	22	-3.97	9.90	—
1	6	0.42	9.90	0.23
2	54	2.11	2.09	-0.23
14. Grasping cube III				
0	20	-4.63	0.36	—
1	10	-1.32	0.22	-0.44
2	52	2.60	0.44	0.44

Continued

Appendix 5: Rating scale analysis for the 26-item Grasping subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
15. Grasping pellets I*				
0	23	-4.32	0.28	—
1	6	-0.74	0.06	0.27
2	54	2.60	0.68	-0.27
16. Manipulating paper				
0	17	-4.29	6.92	—
1	18	-0.44	1.99	-1.33
2	49	2.49	1.82	1.33
17. Grasping pellets II*				
0	34	-3.21	0.14	—
1	2	0.59	0.02	1.78
2	47	3.07	0.24	-1.78
18. Grasping pellets III*				
0	39	-2.73	0.37	—
1	7	0.90	0.01	0.50
2	37	3.68	0.39	-0.50
19. Grasping cube IV				
0	30	-3.53	0.59	—
1	23	0.72	0.09	-1.26
2	29	4.09	0.57	1.26
20. Grasping cubes*				
0	50	-1.77	0.25	-
1	2	-0.87	5.19	2.00
2	95	5.84	0.53	-2.00
21. Grasping marker I				
0	47	-1.97	0.44	—
1	16	2.13	1.17	-0.33
2	88	6.22	0.83	0.33
22. Grasping marker II				
0	48	5.50	1.96	—
1	42	8.73	0.42	-1.00
2	133	11.99	2.02	1.00
23. Unbuttoning button*				
0	67	6.13	0.23	—
1	10	8.87	0.05	1.00
2	140	12.11	0.31	-1.00
24. Buttoning button				
0	70	6.23	0.36	—
1	31	10.22	1.74	-0.29
2	116	12.42	0.73	0.29
25. Grasping marker III				
0	65	9.15	0.81	—
1	88	12.95	0.90	-2.53
2	7	12.91	1.33	2.53
26. Touching fingers*				
0	74	9.21	0.27	—
1	5	11.77	0.79	1.72
2	82	13.31	8.62	-1.72

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

NA indicates not available.

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
1. Tracking rattle I*				
0	1	-14.27	0.19	—
1	1	-13.41	0	0.29
2	133	-1.08	0.44	-0.29
2. Tracking rattle-side*				
0	1	-24.27	0.36	—
1	3	-12.40	0.03	-1.00
2	131	-0.92	0.60	1.00
3. Placing hand*				
0	4	-10.97	9.90	—
1	1	-10.72	0.01	0.95
2	130	-0.90	9.90	-0.95
4. Perceiving rattle*				
0	1	-14.27	0.19	—
1	1	-11.09	0.23	0.29
2	133	-1.10	1.01	-0.29
5. Regarding hands*				
0	6	-12.28	0.06	—
1	0	NA	NA	NA
2	129	-0.76	0.74	0
6. Tracking ball-left to right*				
0	4	-12.96	0.25	—
1	1	-11.46	0	0.95
2	130	-0.83	0.35	-0.95
7. Tracking ball-right to left*				
0	5	-12.66	0.05	—
1	0	NA	NA	NA
2	130	-0.83	0.26	0
8. Tracking rattle II*				
0	2	-13.84	0.23	—
1	2	-11.90	0.03	-0.18
2	131	-0.92	0.53	0.18
9. Extending arms*				
0	6	-12.24	0.42	—
1	2	-10.91	0.01	0.32
2	127	-0.60	0.24	-0.32
10. Approaching midline*				
0	6	-12.18	0.47	—
1	3	-10.42	0.03	-0.14
2	126	-0.53	0.33	0.14
11. Fingering hands*				
0	7	-12.16	0.02	—
1	0	NA	NA	NA
2	128	-0.67	0.14	0
12. Bringing hands together*				
0	23	-8.84	0.19	—
1	2	-5.66	0.07	1.45
2	110	0.39	0.49	-1.45
13. Extending arm*				
0	21	-9.11	0.87	—
1	0	NA	NA	NA
2	114	0.17	0.22	0

Continued

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
14. Retaining cubes I*				
0	23	-8.79	0.66	—
1	3	-5.78	0.08	1.03
2	109	0.44	0.30	-1.03
15. Transferring cube*				
0	44	-6.32	7.39	—
1	3	-2.06	2.78	1.39
2	88	1.28	2.35	-1.39
16. Touching pellet*				
0	16	-9.90	0.60	—
1	5	-6.95	0.01	0.16
2	115	0.21	0.38	-0.16
17. Banging cup*				
0	25	-8.45	1.58	—
1	3	-5.20	0.01	1.12
2	107	0.52	0.77	-1.12
18. Poking finger*				
0	20	-9.30	0.25	—
1	3	-5.92	0.05	0.89
2	112	0.29	0.41	-0.89
19. Removing pegs I*				
0	24	-8.72	0.53	—
1	7	-4.71	0.04	0.17
2	104	0.68	0.47	-0.17
20. Combining cubes				
0	29	-7.96	0.64	—
1	11	-2.93	9.90	-0.19
2	95	0.96	2.06	0.19
21. Clapping hands*				
0	39	-6.98	0.27	—
1	1	-4.21	0.01	2.57
2	95	1.10	5.41	-2.57
22. Retaining cubes II				
0	45	-6.40	1.35	—
1	17	-1.36	2.55	-0.71
2	73	1.91	1.15	0.71
23. Manipulating string				
0	29	-7.70	2.36	—
1	23	-3.49	0.93	-1.29
2	82	1.59	0.97	1.29
24. Removing pegs II*				
0	35	-7.42	0.11	—
1	0	NA	NA	NA
2	100	0.88	0.25	0
25. Releasing cube*				
0	46	-6.42	0.24	—
1	4	-3.79	0.68	1.05
2	85	1.63	0.34	-1.05
26. Removing socks*				
0	56	-5.36	9.90	—
1	2	-2.44	0.08	1.84
2	76	1.74	9.90	-1.84

Continued

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
27. Placing pellet*				
0	47	-6.39	0.25	—
1	5	-3.10	0.10	0.79
2	84	1.75	0.18	-0.79
28. Placing cubes I*				
0	52	-6.03	0.21	—
1	5	-2.15	0.07	0.81
2	78	1.96	0.17	-0.81
29. Turning pages I*				
0	35	-6.92	1.21	—
1	9	-5.38	9.90	0.14
2	158	3.36	0.67	-0.14
30. Stirring spoon				
0	43	-6.66	0.62	—
1	39	-0.58	0.76	-1.94
2	119	4.60	0.79	1.94
31. Removing pellets				
0	49	-6.16	8.35	—
1	21	-1.02	0.11	-0.90
2	132	4.27	0.58	0.90
32. Placing cubes II*				
0	58	-5.59	0.29	—
1	1	-0.50	0	2.61
2	142	3.94	0.14	-2.61
33. Placing pegs*				
0	58	-5.49	0.24	—
1	4	-1.19	0.33	1.20
2	142	4.08	6.75	-1.20
34. Tapping spoon				
0	39	-6.90	0.97	—
1	43	-0.60	0.85	-2.21
2	119	4.48	1.42	2.21
35. Inserting shapes I*				
0	56	-5.72	0.95	—
1	7	-0.24	0.30	0.55
2	141	4.10	0.38	-0.55
36. Placing pellets*				
0	53	-5.99	0.06	—
1	1	-0.27	0.04	2.47
2	147	3.81	0.14	-2.47
37. Scribbling*				
0	51	-6.03	1.12	—
1	2	-1.35	0.04	1.78
2	149	3.69	0.90	-1.78
38. Building tower I*				
0	63	-5.18	0.34	—
1	8	-0.01	0.03	0.60
2	131	4.32	0.48	-0.60
39. Inserting shapes II*				
0	77	-4.14	0.94	—
1	12	0.70	0.06	0.40
2	115	4.92	0.40	-0.40

Continued

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
40. Building tower II*				
0	78	-4.16	0.27	—
1	9	1.54	0.10	0.73
2	115	4.79	0.52	-0.73
41. Turning pages II				
0	53	-5.72	0.69	—
1	107	2.53	0.60	-3.86
2	42	6.48	1.59	3.86
42. Inserting shapes III*				
0	91	-3.37	0.33	—
1	5	2.42	0.07	1.51
2	108	5.10	0.46	-1.51
43. Building tower III*				
0	91	-3.34	0.85	—
1	8	2.22	0.10	1.01
2	103	5.11	0.52	-1.01
44. Imitating vertical strokes*				
0	117	-1.94	2.03	—
1	16	3.05	0.14	0.35
2	151	7.87	4.31	-0.35
45. Removing top				
0	20	-0.17	1.90	—
1	35	1.36	0.63	-1.07
2	178	7.18	0.93	1.07
46. Building tower IV*				
0	63	1.13	1.99	—
1	16	3.33	0.30	0.34
2	152	7.84	0.91	-0.34
47. Snipping with scissors*				
0	81	1.64	1.15	—
1	9	4.29	0.04	1.01
2	142	8.18	1.31	-1.01
48. Imitating horizontal strokes*				
0	75	1.40	1.01	—
1	10	4.41	0.10	0.89
2	146	8.00	1.06	-0.89
49. Stringing beads I*				
0	83	1.64	0.36	—
1	5	3.56	0.09	1.65
2	143	8.13	0.65	-1.65
50. Folding paper				
0	42	0.35	0.71	—
1	47	3.34	0.45	-1.22
2	141	8.06	0.84	1.22
51. Building train*				
0	113	2.66	2.52	—
1	5	5.26	0.10	1.71
2	112	8.82	0.94	-1.71
52. Stringing beads II*				
0	90	1.78	0.16	—
1	2	4.08	0.04	2.61
2	139	8.26	0.27	-2.61

Continued

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
53. Building tower V				
0	86	1.82	3.15	—
1	18	4.77	0.17	0.23
2	127	8.46	0.74	-0.23
54. Building tower VI				
0	118	2.72	0.69	—
1	16	6.79	0.53	0.42
2	96	9.22	0.58	-0.42
55. Copying circle*				
0	125	2.88	0.43	—
1	5	6.99	0.64	1.71
2	102	9.11	9.90	-1.71
56. Building wall*				
0	120	2.83	1.52	—
1	11	6.37	0.07	0.88
2	203	10.80	0.75	-0.88
57. Cutting paper*				
0	67	5.49	0.35	—
1	7	7.43	0.04	1.43
2	186	11.14	0.67	-1.43
58. Lacing string				
0	61	5.70	9.90	—
1	26	8.19	1.36	-0.07
2	172	11.20	1.10	0.07
59. Copying cross*				
0	82	5.93	0.33	—
1	1	6.08	0.27	3.52
2	177	11.34	0.20	-3.52
60. Cutting line*				
0	77	5.81	0.72	—
1	8	8.98	0.44	1.36
2	174	11.31	0.52	-1.36
61. Copying cross				
0	83	5.93	0.47	—
1	93	10.45	0.76	-1.69
2	84	12.32	1.17	1.69
62. Dropping pellets				
0	14	-0.03	1.15	—
1	66	3.13	0.42	-2.55
2	234	10.13	0.61	2.55
63. Tracing line*				
0	72	5.77	1.51	—
1	12	7.92	0.06	0.90
2	175	11.30	0.62	-0.90
64. Copying square*				
0	85	8.52	0.80	—
1	20	10.59	0.35	0.64
2	101	12.45	0.65	-0.64
65. Cutting circle				
0	65	8.03	0.59	—
1	33	10.53	0.33	-0.05
2	106	12.30	1.72	0.05

Continued

Appendix 6: Rating scale analysis for the 72-item Visual-Motor Integration subtest (Continued)

Item/ Category label	Observed Count	Average Measure	Outfit MnSq	Step Calibration
66. Building steps				
0	62	8.06	1.09	—
1	22	9.96	0.74	0.41
2	122	12.07	0.92	-0.41
68. Cutting square				
0	68	8.12	0.58	—
1	61	10.92	0.65	-0.90
2	74	12.80	0.68	0.90
69. Building pyramid*				
0	68	8.10	0.52	—
1	2	9.31	0.08	2.98
2	135	11.96	1.10	-2.98
70. Folding paper I				
0	53	7.80	0.67	—
1	68	10.62	0.72	-1.14
2	82	12.59	0.93	1.14
71. Coloring between lines				
0	125	9.43	0.76	—
1	22	11.76	0.46	0.47
2	56	13.07	1.15	-0.47
72. Folding paper II				
0	27	9.42	0.71	—
1	78	12.36	0.86	-1.99
2	22	13.05	1.21	1.99

* indicates items whose rating scales were reorganized due to infrequently used categories, lack of monotonicity in the average measures, category misfit, or disordered step calibrations.

NA indicates not available.

Appendix 7: The WINSTEPS control file for dimensionality examination of the Reflex, Stationary, and Locomotion subtests

Reflex

```
; this file is PDMS-2 analysis
TITLE="Reflex"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-2 B
3 A
4-8 B
9-38 A
39-127 A
128-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
9-249
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Stationary

```
; this file is PDMS-2 analysis
TITLE="Stationary"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-8 A
9-26 B
27-31 A
32 B
33-35 A
36-37 B
38 A
39-127 A
128-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
1-8
34
32
30
39-249
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Locomotion

```
; this file is PDMS-2 analysis
TITLE="Locomotion"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-38 A
39-88 B
89-90 A
91 B
92-94 A
95-97 B
98-99 A
100-102 B
103-105 A
106 B
107 A
108-111 B
112-117 A
118 B
119-121 A
122 B
123 A
124 B
125 A
126 B
127 A
128-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
1-38
63
59
77
67
117
122
112
71
128-249
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Appendix 8: The WINSTEPS control file for dimensionality examination of the Object Manipulation, Grasping, and Visual-Motor Integration subtests

Object Manipulation

```
; this file is PDMS-2 analysis
TITLE="Object Manipulation"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-38 A
39-127 A
128-133 B
134 A
135-136 B
137 A
138-140 B
141 A
142 B
143-144 A
145-146 B
147 A
148-149 B
150-151 A
152-177 A
178-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
1-127
152-249
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Grasping

```
; this file is PDMS-2 analysis
TITLE="Grasping"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-151 A
152 A
153 B
154 A
155-164 B
165 A
166 B
167 A
168-169 B
170 A
171 B
172-173 A
174 B
175-176 A
177 B
178-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
1-151
167
159
178-249
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Visual-Motor Integration

```
; this file is PDMS-2 analysis
TITLE="Visual-Motor Integration"
NAME1=1
ITEM1=8
NI=249
CODES=012
IREFER=*
1-177 A
178-196 B
197 A
198 B
199-200 A
201-206 B
207-208 A
209-210 B
211 A
212-217 B
218 A
219-221 B
222 A
223-226 B
227 A
228-229 B
230-231 A
232-234 B
235 A
236-237 B
238-239 A
240-241 B
242-245 A
246 B
247-249 A
*
IVALUEA=012
IVALUEB=001
IDFILE=*
1-177
244
235
192
211
214
218
203
221
*
WHEXACT=NO
GROUPS=0
DATA=PDMS-419.txt
;IDELQU=Y
&END
```

Appendix 9: Raw score, Rasch-transformed score, and standard error (SE) of the Reflex and Stationary subtests

Raw scores	Rasch-transformed logit scores	SE	Raw scores	Rasch-transformed logit scores	SE
Reflex subtest			Stationary subtest		
0	-13.4	2.1	0	-124.2	316.2
1	-10.7	2.0	1	-122.2	316.2
2	-5.8	3.7	2	-120.2	316.2
3	-2.0	1.2	3	-118.2	316.2
4	-0.8	1.1	4	-116.2	316.2
5	0.4	1.2	5	-114.2	316.2
6	5.0	5.7	6	-112.2	316.2
7	10.0	1.5	7	-110.2	316.2
8	11.9	2.0	8	-108.2	316.2
			9	-106.2	316.2
			10	-104.2	316.2
			11	-14.5	2.7
			12	-7.7	5.8
			13	-1.8	2.0
			14	4.8	8.0
			15	14.2	6.7
			16	24.41	12.7
			17	35.39	9.6
			18	42.56	2.1
			19	46.11	1.8
			20	49.29	1.8
			21	51.8	1.4
			22	53.44	1.2
			23	54.82	1.1
			24	55.98	1.0
			25	56.87	0.9
			26	57.64	0.9
			27	58.4	0.9
			28	59.19	0.9
			29	60.12	1.1
			30	122.24	316.2

Appendix 10: Raw score, Rasch-transformed score, and standard error (SE) of the Object Manipulation and Grasping subtests

Raw scores	Rasch-transformed logit scores	SE	Raw scores	Rasch-transformed logit scores	SE
Object Manipulation subtest			Grasping subtest		
0	-17.6	2.2	0	-18.6	2.0
1	-14.6	2.3	1	-17.0	1.3
2	-12.0	1.2	2	-15.5	1.2
3	-10.8	1.0	3	-13.3	1.8
4	-9.9	1.0	4	-11.0	1.3
5	-9.0	1.0	5	-9.6	1.2
6	-7.9	1.0	6	-8.2	1.1
7	-6.9	1.0	7	-7.0	1.1
8	-5.9	1.0	8	-5.7	1.1
9	-4.8	1.2	9	-4.4	1.2
10	-2.7	1.8	10	-3.2	1.1
11	-0.6	1.1	11	-2.1	1.0
12	0.3	0.9	12	-1.2	1.0
13	1.0	0.8	13	-0.1	1.1
14	1.6	0.8	14	1.1	1.1
15	2.2	0.8	15	2.3	1.1
16	2.8	0.8	16	3.4	1.0
17	3.4	0.7	17	4.4	1.1
18	3.9	0.7	18	5.9	1.4
19	4.4	0.7	19	10.1	3.9
20	4.9	0.7	20	14.1	1.3
21	5.4	0.7	21	15.3	1.0
22	5.9	0.7	22	16.2	0.9
23	6.5	0.8	23	17.0	0.9
24	7.2	0.9	24	18.0	1.1
25	8.1	1.0	25	19.4	1.3
26	9.1	1.0	26	22.2	2.3
27	10.0	0.9	27	25.2	2.2
28	10.8	0.8			
29	11.4	0.8			
30	12.1	0.8			
31	12.8	1.0			
32	14.0	1.8			

Appendix 11: Raw score, Rasch-transformed score, and standard error (SE) of the Locomotion subtest

Raw scores	Rasch-transformed logit scores	SE	Raw scores	Rasch-transformed logit scores	SE
1	-45.4	1.9	51	9.4	0.5
2	-43.9	1.2	52	9.6	0.5
3	-42.7	1.0	53	9.9	0.5
4	-41.5	1.2	54	10.2	0.5
5	-38.8	2.4	55	10.5	0.5
6	-33.7	3.0	56	10.7	0.5
7	-27.8	3.6	57	11.0	0.5
8	-23.9	1.3	58	11.3	0.5
9	-22.6	1.0	59	11.6	0.5
10	-21.7	0.9	60	11.9	0.5
11	-21.0	0.8	61	12.2	0.5
12	-20.3	0.8	62	12.5	0.5
13	-19.6	0.8	63	12.8	0.5
14	-18.9	0.8	64	13.0	0.5
15	-18.2	0.9	65	13.3	0.5
16	-17.4	0.9	66	13.6	0.5
17	-16.7	0.8	67	13.9	0.5
18	-16.1	0.8	68	14.1	0.5
19	-15.6	0.7	69	14.4	0.5
20	-15.1	0.7	70	14.7	0.5
21	-14.6	0.7	71	15.0	0.5
22	-14.0	0.8	72	15.3	0.5
23	-13.4	0.8	73	15.6	0.5
24	-12.6	1.0	74	15.9	0.5
25	-11.4	1.3	75	16.1	0.5
26	-9.7	1.3	76	16.4	0.5
27	-8.3	1.1	77	16.7	0.5
28	-7.3	1.0	78	17.0	0.5
29	-6.4	1.0	79	17.2	0.5
30	-5.3	1.2	80	17.5	0.5
31	-3.7	1.3	81	17.7	0.5
32	-2.0	1.3	82	17.9	0.5
33	0.2	1.6	83	18.1	0.5
34	1.9	1.1	84	18.4	0.5
35	2.8	0.8	85	18.6	0.5
36	3.4	0.8	86	18.8	0.5
37	3.9	0.7	87	19.0	0.5
38	4.4	0.7	88	19.3	0.5
39	4.9	0.7	89	19.5	0.5
40	5.3	0.7	90	19.8	0.5
41	5.7	0.7	91	20.1	0.5
42	6.2	0.6	92	20.3	0.5
43	6.6	0.6	93	20.6	0.6
44	7.0	0.6	94	20.9	0.6
45	7.4	0.6	95	21.3	0.6
46	7.7	0.6	96	21.7	0.6
47	8.1	0.6	97	22.1	0.7
48	8.4	0.6	98	22.7	0.8
49	8.8	0.6	99	23.6	1.1
50	9.1	0.6	100	43.4	316.2

Appendix 12: Raw score, Rasch-transformed score, and standard error (SE) of the Visual-Motor Integration subtest

Raw scores	Rasch-transformed logit scores	SE	Raw scores	Rasch-transformed logit scores	SE
0	-27.0	1.9	41	2.6	0.7
1	-25.5	1.2	42	3.1	0.7
2	-24.3	1.0	43	3.6	0.7
3	-23.3	1.0	44	4.1	0.7
4	-22.3	0.9	45	4.6	0.8
5	-21.5	0.9	46	5.3	0.8
6	-20.8	0.8	47	6.0	0.9
7	-20.1	0.9	48	6.7	0.8
8	-19.3	0.9	49	7.3	0.7
9	-18.4	1.0	50	7.8	0.7
10	-17.1	1.2	51	8.2	0.7
11	-14.6	2.0	52	8.6	0.7
12	-12.2	1.2	53	9.1	0.7
13	-11.1	0.9	54	9.5	0.7
14	-10.4	0.8	55	10.0	0.7
15	-9.7	0.8	56	10.5	0.7
16	-9.1	0.8	57	11.0	0.7
17	-8.5	0.7	58	11.5	0.7
18	-8.0	0.7	59	12.0	0.7
19	-7.5	0.7	60	12.5	0.7
20	-7.0	0.7	61	13.0	0.7
21	-6.5	0.7	62	13.5	0.7
22	-6.0	0.7	63	14.1	0.7
23	-5.6	0.7	64	14.6	0.7
24	-5.1	0.7	65	15.1	0.7
25	-4.6	0.7	66	15.5	0.7
26	-4.2	0.7	67	16.0	0.7
27	-3.7	0.7	68	16.4	0.6
28	-3.2	0.7	69	16.8	0.6
29	-2.8	0.7	70	17.1	0.6
30	-2.3	0.7	71	17.5	0.6
31	-1.9	0.7	72	17.9	0.6
32	-1.4	0.7	73	18.2	0.6
33	-1.0	0.7	74	18.6	0.6
34	-0.6	0.7	75	19.0	0.7
35	-0.1	0.7	76	19.5	0.7
36	0.3	0.7	77	19.9	0.7
37	0.7	0.7	78	20.4	0.7
38	1.2	0.7	79	21.1	0.9
39	1.6	0.7	80	22.1	1.2
40	2.1	0.7	81	23.8	2.0