University of Alberta

An Exploration of Children's
Serial Motor Test Scores

by

Karin Eileen Eldred

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

In

Rehabilitation Science – Physical Therapy

Edmonton, Alberta
Spring, 2008

Your file Votre référence
ISBN: 978-0-494-45807-5
Our file Notre référence
ISBN: 978-0-494-45807-5

ABSTRACT

Longitudinal research has revealed intra-individual and inter-individual variability of motor scores among typically developing children. Exploration of this variability was conducted using cluster analytic techniques, correlation coefficients, and a comparison of test classifications over time. Data was analyzed from two longitudinal studies using the same cohort of children from 9 to 23 months of age and from 4 to 7 years of age. Cluster analysis was found to be a useful method to distill the variability observed across serial gross motor percentile rank scores into clinically meaningful subgroups. The analyses using correlation coefficients and comparative test classifications confirmed intra-individual variability in both percentile rank scores and test classifications within the sample over time. Clinical implications are discussed. The results will assist in interpreting longitudinal observations of gross motor performance and contribute to the theoretical foundations of screening for physical disabilities in young children.

ACKNOWLEDGEMENTS

I am indebted to Kesheng (Frank) Tao and especially to J. D. Jobson for their help and patience as I worked through the statistical analyses portions of my thesis. I am grateful to the families and therapists who participated in both the infant and preschool studies. I feel privileged to have received an Alberta Heritage Foundation for Medical Research summer studentship award which first exposed me to the world of research in pediatric physical therapy as an undergraduate student. Finally, I would like to thank my supervisor and mentor Johanna Darrah for her belief in my abilities and for her dedication to generating high quality research in the field of pediatric physical therapy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Problem Statement

Longitudinal research has revealed substantial inter- and intra-individual variability of standardized test scores among infants and young children in many developmental domains of early childhood, including motor development. Interpretation and analysis of this variability is influenced by researchers' theoretical perspectives and few tools exist to help clinicians interpret the significance of this data. A greater understanding of intra- and inter-individual change in motor development test scores will help clinicians understand the nature of development and facilitate more effective assessment and intervention practices.

Aim of the Study

This study analyzed data collected in two longitudinal studies using the same cohort of children from infancy to school age. The aim of the study was to determine whether distinct trajectories could be identified within serial assessments of gross motor performance using cluster analysis techniques. The specific objectives of the study were to:

1. Determine whether trajectories of gross motor performance could be grouped into distinct clusters.
2. Determine if the clusters identified have any clinically relevant characteristics.

3. Identify any differences between clusters based on demographic data, reported illness/hospitalization, or scores of cognitive abilities.
4. Determine whether any cluster is associated with suspicious or delayed gross motor development as determined by an exit assessment at 7 years of age.

During the course of the analyses, Objective 4 was modified due to concerns that the exit assessment at 7 years of age, the Movement Battery for Children – Performance Test (MABC-Test, Henderson & Sugden, 1992), was over-identifying children as having suspicious or delayed motor performance. Consequently, ethical approval was obtained to contact the participant's families when the children were 8 years of age in order to identify children who had received medical diagnoses or intervention from a motor therapist. Results of the parent phone calls were then compared with cluster membership.

Additionally, although clusters were generated using gross motor percentile rank scores on the Peabody Developmental Motor Scales (PDMS, Folio & Fewell, 1983) and Peabody Developmental Motor Scales – Second Edition (PDMS-2, Folio & Fewell, 2000), a combination of fine and gross motor subsection scores determined MABC Test classifications. Because the subsections of the MABC Test were not designed to be administered or interpreted in isolation, inclusion of the fine motor (manual dexterity) items in identifying children with suspicious or delayed gross motor development could have potentially confounded the results. Thus, a decision was made to examine

the similarities and differences between test scores and classifications across time instead of relying on the MABC Test as a gold standard outcome measure. In order to produce a concise paper with a larger sample size, only the analyses regarding the percentile rank scores and test classifications between the MABC-Test and the PDMS-2 were used. Fine motor, gross motor, and total motor percentile rank scores on the PDMS-2 were compared with the MABC-Test scores. The classifications using test cut-offs were also compared over time using the PDMS-2 and the MABC Test.

Overview of the Thesis

The thesis follows a paper format and consists of a literature review (Chapter 2) and two papers. The first paper (Chapter 3) presents the results and clinical implications of the cluster analyses using serial gross motor percentile rank scores of children aged 9 months to 5.5 years of age. The second paper (Chapter 4) explores the relationship between percentile rank scores and between test classifications of preschool aged children across serial assessments. Chapter 5 provides a general discussion of the results, clinical implications, and directions for future research.

References

Folio, M. R., & Fewell, R. P. (1983). *Peabody Developmental Motor Scales and activity cards: A manual.* Allen: DLM Teaching Resources.

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

Henderson, S. E., & Sugden, D. A. (1992). *Movement Assessment Battery for Children.* London: The Psychological Corporation.

CHAPTER 2

LITERATURE REVIEW

Clinical Scenario

Jonas has a normal birth and health history. He was referred to physical therapy by his family doctor because he was not walking at 16 months of age. A physical therapist administered a standardized and norm referenced test of gross motor development with excellent psychometric properties. Jonas' percentile rank score was 10, well below the test cut-off for suspicious motor development, indicating he was "at risk" for delay. The therapist gave his mother a few suggestions and asked them to return to the clinic in two months. At 18 months, the therapist repeated the test of gross motor development and Jonas' score had risen to the 50^{th} percentile, well within normal range.

What can explain the variability within Jonas' score trajectory? Was it measurement error by the therapist or problems with the test itself? Is the variable pattern of scores normal or indicative of deviant gross motor development? How should the therapist interpret these scores and what action should she take? Were the intervention suggestions responsible for the improvement in score or would Jonas have scored higher without the therapist's input? These questions merit consideration as they highlight some key issues faced by professionals who screen children for developmental delay.

Synopsis of Literature Review

This literature review discusses issues related to studying the variability of developmental test scores within longitudinal studies. Past research demonstrating the presence of variability in a variety of developmental domains is presented. The origins of variability, including both measurement and theoretical perspectives, are discussed. The implications of variability in relation to screening for developmental delay are outlined, followed by a discussion about the potential for interpreting and exploring variability of serial test scores using different analytic techniques.

Variability

In the clinical scenario, Jonas' therapist is faced with interpreting two types of variability. The therapist determined Jonas' percentile rank on the test of gross motor development using normative tables which represent inter-individual (between-subject) variability. This allowed the therapist to compare Jonas' performance to others his same age. Tests based on inter-individual variability are used extensively in screening infants and young children for delays in motor, social, and emotional development (Siegler, 2002). Secondly, the therapist is faced with the challenge of interpreting the intra-individual (within-subject) variability of Jonas' test scores, as exhibited by his change in score over the two assessment points. Few guidelines exist to help therapists interpret this type of variability in gross motor development.

The intra-individual variability demonstrated within Jonas' gross motor scores is not a unique phenomenon. Longitudinal research into the stability of scores across serial assessments of typically developing infants' gross motor performance has revealed comparable results. Ninety-four of 102 typically developing infants demonstrated variable patterns of gross motor scores over five assessments using the Peabody Developmental Motor Scales (Darrah, Hodge, Magill-Evans, & Kembhavi, 2003), and 31.1% of typically developing infants received a score on the Alberta Infant Motor Scale below the 10^{th} percentile on at least one occasion when tested across 13 assessment times (Darrah, Redfern, Maguire, Beaulne, & Watt, 1998). Similarly, Coryell, Provost, Wilhelm, and Campbell (1989) documented significant variation in test scores across five test ages using the Bayley Motor Scales in 'normal' infants.

Intra-individual variability has been demonstrated in a variety of other developmental domains, including weight gain (Giani, Filosa, & Causa, 1996; Mei, Grummer-Strawn, Thompson, & Dietz, 2004), language (Thal, Bates, Goodman, & Jahn-Samilo, 1997), and cognitive development (McCall, Hogarty & Hurlburt, 1972; Moffit, Caspi, Harkness, & Silva, 1993). In an analysis of longitudinal data from the California Child Health and Development Study, Mei and colleagues (2004) documented the prevalence of shifts across major percentiles in growth rate (height-for-age, weight-for-age, weight-for-height, and body mass index) during infancy and early childhood. They concluded that these shifts are "normal phenomena that affect large numbers of children, particularly

during infancy" (p. e626). Similarly, research into the continuity of cognitive development of individual children has revealed variation within trajectories of intelligence quotient (IQ) scores that was greater than would be expected from classical test theory (Gilmore & Thomas, 2002; Moffit et al., 1993).

Origins of Variability – A Great Debate

Longitudinal research in a variety of developmental domains has helped to establish the existence of intra- and inter-individual variability (Siegler, 2002). Explaining the source of this demonstrated variability provides an ongoing debate in the literature.

Measurement Issues

Measurement error has been the assumed cause of variability in intra-individual test scores over time (van Geert & van Dijk, 2002). Poor test-retest reliability could account for the variability observed within an individual child's percentile ranks over serial assessments, however, considerable effort and research has gone into the creation of developmental tests with excellent psychometric properties. Evaluative measures in particular are designed to have high test-retest reliability to ensure the changes observed on repeated assessments reflect changes in development and not measurement error (Tieman, Palisano, Sutlive, 2005). In a review of the continuity of cognitive development within individuals, Wohlwill (1980) reasons that a lack of stability of a test score over a brief time period indicates a defect in the test while instability over a longer period of time indicates intrinsic change within an individual. In addition

to using tests with high reliability, utilization of confidence intervals and standard error of measurement when interpreting test results enables researchers and clinicians to account for measurement error when reporting findings (McNeely, 2006; Stratford, 2004).

Many researchers strive to reduce intra-individual variability in order to reduce assumed measurement error by using statistical techniques to smooth developmental trajectories and average scores, but perhaps intra-individual variability should be viewed as a pervasive phenomenon, one which should be studied using descriptive and exploratory techniques (van Geert & van Dijk, 2002). This view is supported by Siegler (2002), who argues that analyzing intra-individual variability is central to understanding infant development and should not be discounted as error variance. Increasingly, developmental scientists are acknowledging that measurement error alone cannot account for intra-individual variability.

Theoretical Issues

Theoretical perspectives influence how clinicians and researchers interpret inter- and intra-individual variability. Early investigations into the emergence of motor behaviors were influenced by the work of researchers seeking to explain the seemingly predictable emergence and progression of motor skills (Adolph, 2002). Steeped in the contemporary debate of nature versus nurture, these researchers posited that the emergence of motor skills is driven by the maturation of the central nervous system (Adolph, 2002; Gesell, 1946; McGraw, 1945).

They believed motor development was a genetically driven process, common to all infants, whose outcome was largely unaffected by the environment or individual experience (Adolph, 2002; Kamm, Thelen, & Jensen, 1990; Thelen, 1995). Within this neuromaturational framework, motor behavior was thought to originate from pre-programmed patterns present at birth which unfold in predetermined patterns (Piek, 2002). Variation in the order or timing of the emergence of motor milestones was interpreted as an indication of impaired brain function and abnormal development (Touwen, 1978). Additionally, early identification of deviance or delay was important as remedial practice of missing or deviant skills was necessary to move onto the next stage in the developmental sequence (Case-Smith, 1996; Kamm et al., 1990). Based on a neuromaturational theoretical perspective, percentile rank scores and classification of motor development on standardized tests are expected to be essentially invariant over serial assessments.

In the 1980s, the rediscovery of Russian scientist Leonard Bernstein's work stimulated a paradigm shift in the study of movement science (Adolph, 2002). Bernstein proposed the central nervous system has no privileged status in motor control. Rather, motor behaviors are the result of the interaction of several subsystems within the person, including biomechanical properties, individual motivation, as well as characteristics of the task itself (Bernstein, 1967). Similarly, Newell (1991) articulated the concept that three different types of constraints (biological, environmental and task related) influence motor behavior

at any time. Gibson (1979) also contributed to the evolution of this new perspective by emphasizing the dynamic relationship between perception and action. This shift from conceiving motor development as orderly and sequential to a dynamic systems theory (DST) model, whereby motor development is described as being non-linear and based on the interaction of various subsystems within the child, task and environment, was due in large part to the work of Esther Thelen (Adolph, 2002; Thelen, 1995; Thelen & Fisher, 1982). Thelen (1995) applied DST to the field of motor development, concluding that behaviors can be highly variable, particularly in times of transition when new skills are being explored, but will stabilize as the individual learns the most efficient solution based on subsystem constraints and supports. In their study of infant stepping, Thelen and Fischer (1982) demonstrated that the stepping reflex could be repressed by adding weights to infants' legs or stimulated by submerging their lower bodies in water. They postulated that the integration of the stepping reflex was related to anthropometric changes and not maturation of the CNS.

The basic tenets of DST, as they apply to motor development, are summarized in the following list (Case-Smith, 1996; Piper & Darrah, 1994):

1. The development of motor skills is nonlinear and dependent upon the interaction of many factors including the child, environment and task. The spontaneous interaction of these factors can result in a motor behavior that is more than the sum of its parts. Both intrinsic and extrinsic rate limiting factors can constrain motor performance.

2. Motor performance is influenced by the task itself. Perception and movement continuously interact through feedforward and feedback systems as a task is performed.
3. Transition stages are important as new movement strategies emerge in response to changing task demands.

From a DST perspective, intra-individual variability is seen as a normal and essential part of motor development (Piek, 2002). The trajectory of an individual's motor development is inherently variable and dependent on a myriad of factors (Adolph, 2002; Kamm et al., 1990; Thelen, 1995) and a missing milestone or variable skill performance is not necessarily cause for concern (Darrah et al., 1998; Thelen, 1995). Based on DST, variability across both serial percentile rank scores and classification of motor development is anticipated.

Implications of Variability – Screening for Developmental Delay

Despite the variability demonstrated in longitudinal studies of development, the assumption of stability forms the basis for many screening programs (Darrah et al., 1998). The accumulating evidence about the variable nature of development underscores the need for a new approach to screening children for developmental delay. Isolated assessments must be replaced with practices such as Dworkin's "developmental surveillance." Developmental surveillance describes a process which is broader in scope than screening and which involves longitudinal observations of children's abilities (Dworkin, 1989). At each clinical visit, standardized test scores should be interpreted within the

context of the child's overall well being including a comprehensive developmental history, attention to parent concerns and skilled clinical observation (American Academy of Pediatrics, 2006). Such practices would decrease the risk of falsely identifying typically developing children as delayed, helping to distinguish between children who consistently score below accepted cut-offs and those who are exhibiting normal variability within their trajectory of development. Intervention programs and community services can then be provided to the individuals who most need them.

Secondly, variability also has implications for the physical measures used to assess gross motor development and screen for developmental delay. Although it remains essential to use a test with excellent psychometric properties, it may also be important to establish the theoretical basis from which the test emerged. If development is best expressed by a nonlinear model, then tests based on nonlinear theory should be used to evaluate motor development. The theoretical perspective from which many standardized tests of gross motor development are developed is unreported; however most are based on milestones and norms embedded in neuromaturational theory (Case-Smith, 1996; Wiart & Darrah, 2001).

Most standardized tests of gross motor development are based on cross sectional data, enabling interpretation of performance based on an isolated assessment. Normative data tables have been developed to allow therapists to determine a child's rank order based on their motor performance on a single occasion, but little information is available to help therapists evaluate

performance over time. In order to utilize the data collected in serial assessments, clinicians need to have tests that include guidelines for interpretation of these longitudinal observations.

Lastly, the natural history of the emergence of many mild and moderate disabilities is unknown, making detection difficult for clinicians. Variability within developmental trajectories further compounds these difficulties. How can clinicians predict delays if normal development is characterized by variation and the occasional deviant score? More information is needed to assist in the interpretation of intra-individual variability and identification of developmental delay.

Clinical Scenario Revisited

Based on the available research, Jonas' therapist will be challenged to interpret the variability in Jonas' gross motor performance. As the therapist used a standardized and norm referenced test of gross motor development with excellent psychometric properties, she can distinguish between the effect of measurement error and true change in the child's score by generating confidence intervals around his scores. She can conclude that the variability of Jonas' scores has been observed in other typically developing children, both within gross motor development as well as other domains. However, she cannot know what may have caused his improvement in score or how his scoring pattern relates to his future gross motor abilities. Lastly, the therapist should consider changing her

approach to monitoring Jonas' and other clients' gross motor skill development from one-time screening to developmental surveillance.

Interpreting and Exploring Variability through Research

Sparse research is available to help clinicians and researchers interpret intra and inter-individual variability. Developmental scientists have identified the need for the application of descriptive and exploratory techniques to study patterns of variability (Darrah et al., 2003; Gilmore & Thomas, 2002; Siegler, 2002; van Geert & van Dijk, 2002). Various analysis techniques are suggested to study individual change over time including visual inspection of raw data and using standard deviation and the coefficient of variation (van Geert & van Dijk, 2002). The use of correlation coefficients and cluster analysis techniques may also have potential for interpreting variability within serial test scores.

Within the realm of cognitive development research, correlation coefficients have traditionally been utilized to evaluate stability of test scores over time. Test scores on early childhood or infant measures are compared with later performance using correlation coefficients which represent how well the individual tested maintains their standing relative to the group (Wohlwill, 1980). This approach has also been used to evaluate stability on tests of motor development (Darrah, Magill-Evans, Volden, Hodge & Kembhavi, 2007; Harris, Megens, Backman & Hayes, 2005; Palisano, 1986). However, authors like Siegel (1989) question the use of correlation coefficients, noting that this information may not be as useful to clinicians as information about changes in classification of

an individual child's performance. She suggests using ranges of scores instead of specific scores, classifying children into delayed or normal groupings based on test scores. Changes in classification over time can be then be analyzed, providing clinically important information about the stability of classification for individual children.

Cluster analysis is another exploratory analytic technique which may hold promise in developmental pediatric research (Steele & Aylward, 2007). Cluster analysis describes a group of exploratory multivariate techniques with a strong tradition of organizing objects or individuals into groups based on the similarities and differences between selected variables (Hair, Anderson, Tatham, & Black, 1998). In longitudinal research, cluster analysis enables researchers to evaluate individual patterns of development which may otherwise be masked by group means (Steele & Aylward, 2007). These techniques have been used by developmental researchers to search for subtypes of Developmental Coordination Disorder (DCD) (Macnab, Miller & Polatajko, 2001), written expression (Wakely, Hooper, de Kruif, & Swartz, 2006), and peer social preference (Brendgen, Vitaro, Bukowski, Doyle, & Markiewicz, 2001) in elementary school aged children. The techniques have also been used to identify patterns of cognitive development in typical (Moffit et al., 1993) and very low birth weight children (Koller, Lawson, Rose, Wallace, & McCarton, 1997), as well as to describe growth patterns in children with cerebral palsy (Stevenson et al., 2006).

Analyzing variability with new techniques could yield insights into the nature of human development while providing interesting hypotheses and more powerful testing procedures (Siegler, 2002; van Geert & van Dijk, 2002). Clinically, analyzing inter- and intra-individual variability of serial test scores will help define parameters of typical and atypical development and may improve screening predictions (Darrah et al., 2003).

References

American Academy of Pediatrics. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118*(1) 405-420.

Adolph, K. E. (2002). Babies' steps make giant strides toward a science of development. *Infant Behavior & Development, 25*, 86-90.

Bernstein, N. (1967). *The coordination and regulation of movement.* Oxford: Pergamon Press.

Brendgen, M., Vitaro, F., Bukowski, W. M., Doyle, A. B., & Markiewicz, D. (2001). Developmental profiles of peer social preference over the course of elementary school: Associations with trajectories of externalizing and internalizing behavior. *Developmental psychology, 37*(3), 308-320.

Case-Smith, J. (1996). Analysis of current motor development theory and recently published infant motor assessments. *Infants and Young Children, 9*(1), 29-41.

Coryell, J., Provost, B., Wilhelm, I. J., & Campbell, S. K. (1989). Stability of Bayley Motor Scale scores in the first year of life. *Physical Therapy, 69*(10), 834-841.

Darrah, J., Hodge, M., Magill-Evans, J., & Kembhavi, G. (2003). Stability of serial assessments of motor and communication abilities in typically developing infants-Implications for screening. *Early Human Development, 72*(2), 97-110.

Darrah, J., Magill-Evans, J., Volden, J., Hodge, M., & Kembhavi, G. (2007). Scores of typically developing children on the Peabody Developmental Motor Scales - Infancy to preschool. *Physical & occupational therapy in pediatrics, 27*(3), 5-19.

Darrah, J., Redfern, L., Maguire, T. O., Beaulne, A. P., & Watt, J. (1998). Intra-individual stability of rate of gross motor development in full-term infants. *Early Human Development, 52*(2), 169-179.

Dworkin, P. H. (1989). Developmental screening - expecting the impossible? *Pediatrics, 83*(4), 619-621.

Gesell, A. (1946). The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of child psychology* (pp. 295-331). New York: Wiley.

Giani, U., Filosa, A., & Causa, P. (1996). A non-linear model of growth in the first year of life. *Acta Paediatrica, 85*(1), 7-13.

Gibson, E. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin Company.

Gilmore, R. O., & Thomas, H. T. (2002). Examining individual differences in infants' habituation patterns using objective quantitative techniques. *Infant Behavior and Development, 25*, 399-412.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). New Jersey: Prentice Hall.

Harris, S. R., Megens, A. M., Backman C. L., & Hayes, V.E. (2005). Stability of the Bayley II Scales of Infant Development in a sample of low-risk and high-risk infants. *Developmental Medicine and Child Neurology, 47,* 820-823.

Kamm, K., Thelen, E., & Jensen, J. L. (1990). A dynamical systems approach to motor development. *Physical Therapy, 70,* 763-775.

Koller, H., Lawson, K., Rose, S. A., Wallace, I., & McCarton, C. (1997). Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics, 99*(3), 383-389.

Macnab, J. J., Miller, L. T., & Polatajko, H. J. (2001). The search for subtypes of DCD: Is cluster analysis the answer? *Human Movement Science, 20*(1-2), 49-72.

McCall R.B., Hogarty P.S. & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. *American Psychologist, 72*(8), 728-748.

McGraw, M. B. (1945). *The neuromuscular maturation of the human infant.* New York: Columbia University Press.

McNeely, M.L., & Warren, S. (2006). Value of confidence intervals in determining clinical significance. *Physiotherapy Canada, 58*(3), 205-211.

Mei, Z., Grummer-Strawn, L. M., Thompson, D., & & Dietz, W. H. (2004). Shifts in percentiles of growth during childhood: Analysis of longitudinal data from the California child health and development study. *Pediatrics, 113*(6), e617-e627.

Moffit, T. E., Caspi, A., Harkness, A. R., & Silva, P. A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology and Psychiatry, 34*(4), 455-506.

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology, 42*, 213-237.

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.

Piek, J. P. (2002). The role of variability in early motor development. *Infant Behavior and Development, 25*, 452-465.

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

Siegel, L. S. (1989). A reconceptualization of prediction from infant test scores. In M. H. Bornstein, & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives* (pp. 89-104). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Siegler, R. S. (2002). Variability and infant development. *Infant Behavior and Development, 25*, 550-557.

Steele, R. G. & Aylward, B. S. (2007). The use of cluster analytic techniques in developmental and behavioral pediatric research. *Journal of Developmental and Behavioral Pediatrics, 28*(4), 327-329.

Stevenson, R. D., Conaway, M., Chumlea, W. C., Rosenbaum, P., Fung, E. B., Henderson, R. C., et al. (2006). Growth and health in children with moderate to severe cerebral palsy. *Pediatrics, 118*(3), 1010-1018.

Stratford, P.W. (2004). Getting more from the literature: Estimating the standard error of measurement from reliability studies. *Physiotherapy Canada, 56*(1), 27-30.

Thal, D., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late and early talking toddlers. *Developmental Neuropsychology, 13*(3), 239-273.

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist, 50*(2), 79-95.

Thelen, E., & Fischer, D. M. (1982). Newborn stepping: An explanation for a "disappearing reflex". *Developmental Psychology, 18*, 760-775.

Tieman, B. L., Palisano, R. J., & Sutlive, A. C. (2005). Assessment of motor development and function in preschool children. *Mental Retardation & Developmental Disabilities Research Reviews, 11*(3), 189-196.

Touwen, B. (1978). Variability and stereotypy in normal and deviant development. *Clinical Developmental Medicine, 67*, 99-110.

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*, 340-374.

Wakely, M. B., Hooper, S. R., de Kruif, R. E., & Swartz, C. (2006). Subtypes of written expression in elementary school children: A linguistic-based model. *Developmental Neuropsychology, 29*(1), 125-159.

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

Wohlwill, J. F. (1980). Cognitive development in childhood. In O. G. Brim, & J. Kagan (Eds.), *Constancy and change in human development* (pp. 359-444). Cambridge, Mass.: Harvard University Press.

CHAPTER 3

EXPLORING SERIAL GROSS MOTOR PERCENTILE RANK SCORES OF TYPICALLY DEVELOPING CHILDREN USING CLUSTER ANALYSIS

Introduction and Aim of Study

Longitudinal research into the stability of gross motor (GM) percentile rank scores of typically developing children on standardized tests has documented substantial intra-individual (within child) variability (Darrah, Hodge, Magill-Evans, & Kembhavi, 2003; Darrah, Redfern, Maguire, Beaulne, & Watt, 1998). Darrah et al. (1998) tracked the GM percentile rank scores of 45 typically developing infants for 13 months on the Alberta Infant Motor Scale; the absolute value of changes in these scores within individual infants' score profiles over an average of 12 assessments ranged from 34 to 87 percentile points. In another study, Darrah et al. (2003) assessed 102 typically developing infants five times from 9 to 21 months of age using the Peabody Developmental Motor Scales. The authors generated a graph of each infant's GM percentile rank scores over time with 95 per cent confidence intervals to account for measurement error. Ninety-six (94%) of the infants' graphs were characterized by variability on serial percentile rank scores that exceeded the confidence intervals, demonstrating fluctuating patterns of GM percentile rank scores. The same investigators recently completed a study that followed many of the same children to 5.5 years of age to determine whether the variability of GM percentile rank scores observed in infancy persisted into preschool ages (J. Darrah, personal communication, May 7,

2007). Multi-level models were used to describe the longitudinal changes of predicted curves over time and these analyses revealed continued intra-individual variability of children's GM percentile rank scores into preschool years.

The intra-individual variability demonstrated in these studies of GM percentile rank scores over time makes it challenging to interpret the clinical significance of changes in a child's performance over time using standardized, norm-referenced measures. If fluctuations of percentile rank scores are normal phenomena, does a decrease in percentile rank score between assessment points indicate a need to worry or does it represent that child's natural score profile? Are there common patterns of variability that can be identified across children? Is one pattern more predictive of long term GM challenges? How can clinicians interpret intra-individual variability of serial percentile rank scores?

The present analyses apply cluster analysis techniques to data collected within two longitudinal studies of GM development. The primary objective of the analyses was to determine if the intra-individual variability of percentile rank scores identified in the two studies could be grouped into distinct clusters that are clinically relevant. The influences of children's cognitive abilities and a number of other variables (gender, ethnicity, family income, parents' education, and reported illnesses/hospitalizations) on cluster membership were also examined.

Background

Pediatric physical therapists frequently use standardized measures to assess children's GM development over time. Many of these measures are norm-

referenced and a child's raw score can be converted to either a standard score or percentile rank and compared to the normative sample. Traditionally a change in a child's percentile rank, especially a change that exceeds the standard error of measurement associated with the test, has been viewed as either a deterioration (percentile decrease) or improvement (percentile increase) in the child's performance (Folio and Fewell, 1983). For example, if a child received a percentile rank score of 40 at 9 months of age, it was expected that his or her subsequent percentile scores would be of a similar value. This expectation of a constant rate of development was derived from a neuromaturational approach to motor development, which espouses that the emergence of motor skills is driven primarily by the maturation of the central nervous system at a relatively constant rate (Gesell, 1928). In contrast to this approach, dynamic systems theory (DST), explains the emergence of motor behaviors as highly variable and non-linear, depending on the spontaneous self-organization (Thelen, 1995) of myriad factors within the child, the task and the environment (Newell, 1991). A child may learn many new skills in a short period of time and then not add any new motor skills for some time. From this perspective a child's percentile rank scores on a standardized test may fluctuate dependent on whether the testing occurred during a time of quiescence or skill emergence.

Traditionally, intra-individual variation over serial assessments on the same standardized test has most often been attributed to measurement error or 'noise' (Berenthal & Boker, 1997; Fischer & Pare-Balgoev, 2000; van Geert &

van Dijk, 2002). Researchers assumed variability of a child's developmental test performance over time represented primarily measurement error and could be reduced by the development of measures with strong validity and reliability (McCall, Hogarty, & Hurlburt, 1972; Moffit, Caspi, Harkness & Silva, 1993; van Geert & van Dijk, 2002). In order to minimize the error resulting from fluctuations in test scores of individual performance, developmental scientists used statistical techniques to smooth developmental trajectories and average scores (van Geert & van Dijk, 2002). Discussions have appeared recently in developmental literature about the error interpretation of intra-individual variability. Some authors suggest that variability is a true representation of developmental trajectories (Berenthal & Boker, 1997; Giani, Filosa & Causa, 1996, van Geert & van Dijk, 2002), and Siegler (2002) argues that analyzing intra-individual variability is central to understanding infant development and should not be discounted as error variance.

Evaluation of intra-individual variability is occurring in many areas of developmental research. Mei, Grummer-Strawn, Thompson, and Dietz (2004) documented the prevalence of shifts across major percentiles in growth rate (height-for-age, weight-for-age, weight-for-height, and body mass index) in individual children during infancy and early childhood. Fluctuating patterns of weight gain over time were also documented within the first year of life by Giani et al. (1996), who concluded that shifts in percentiles are a physiological phenomenon in typically developing children. Similarly, research into the

continuity of cognitive development of individual children has revealed variation within trajectories of IQ scores that is greater than would be expected from classical test theory (Gilmore & Thomas, 2002; Moffit et al., 1993). Variability of individual percentile rank scores has also been reported in the development of early communication and fine motor skills (Darrah et al., 2003; Fenson et al., 1994). Methods to identify and analyze intra-individual variability of scores of developmental domains are appearing in the literature; multivariate techniques and methods such as using the moving max-min graph and the critical frequency method have been proposed to aid in exploration of intra-individual variability in developmental data (van Geert & van Dijk, 2002).

The present work utilizes cluster analysis techniques to distill the intra-individual variability within GM percentile rank scores into distinct and clinically manageable subgroups. Cluster analysis describes a group of exploratory multivariate techniques with a strong tradition of organizing objects or individuals into groups based on the similarities and differences between selected variables (Hair, Anderson, Tatham, & Black, 1998). These techniques have been used by developmental researchers to search for subtypes of Developmental Coordination Disorder (DCD) (Macnab, Miller & Polatajko, 2001), written expression (Wakely, Hooper, de Kruif, & Swartz, 2006), and peer social preference (Brendgen, Vitaro, Bukowski, Doyle, & Markiewicz, 2001) in elementary school aged children. The techniques have also been used to identify patterns of cognitive development in typical (Moffit et al., 1993) and very low birth weight

children (Koller, Lawson, Rose, Wallace, & McCarton, 1997), as well as to describe growth patterns in children with cerebral palsy (Stevenson et al., 2006).

Method

Gross motor data derived from two longitudinal studies using the same cohort of children were used for the analyses. The first (infant) study evaluated the stability of GM percentile scores obtained at 9, 11, 13, 16, and 21 months of age (Darrah et al., 2003) The second (preschool) study included many of the children from the infant study and assessed the children's GM skills at 4, 4.5, 5 and 5.5 years of age.

Sample

One hundred and twenty full term infants (37 weeks gestation or greater) and their families participated in the infant study. This volunteer sample was recruited at ages 4 weeks to 8 months from Moms and Babies groups and public health centers. The parents had no concerns about their infant's development at the time of recruitment and screening and agreed to participate. Of the 120 children recruited for the infant study, 83 continued in the preschool study. Thirteen of these families dropped out of the preschool study and four children had some missing GM data. The present analyses used data of 66 children (28 females) with complete GM data from both studies (nine assessment points). Informed consent was obtained from all families for both studies, and the studies were approved by the University of Alberta Health Research Ethics Board (Panel B).

Procedure and Measures

Parents completed a short questionnaire at each visit asking about their child's health, hospitalization, intervention, special programs or other factors that may have affected performance. At the beginning of each study parents also completed a more detailed questionnaire regarding their child's gender, ethnicity, family income and education levels of each parent. When the children were 8 years of age, their families were contacted by phone and asked a series of questions regarding their child's current motor abilities. Parents were asked whether their child had received a medical diagnosis or any intervention by a physical or occupational therapist.

In the infant study GM data were collected by physical and occupational therapists in the infants' homes using the GM subscales of the Peabody Developmental Motor Scales (PDMS) (Folio & Fewell, 1983). The time points for assessment were chosen to represent the midpoint of each age band based on normative tables within the PDMS. As the Peabody Developmental Motor Scales 2nd Edition (PDMS-2) (Folio & Fewell, 2000) was published between the end of the infant study and the beginning of the preschool study, therapists were trained in administration of this new edition for the preschool study. For both studies, assessments were completed within 2 weeks of the predetermined assessment ages. Therapists achieved at least an 80% item by item agreement during initial training and during data collection every 10th assessment was observed and coded independently by a second therapist. The GM inter-rater reliability coefficient for

the infant study therapists was 0.93 and 0.92 for the preschool study. For all assessments, therapists scored individual items and a research assistant tabulated the raw scores and converted them to their associated percentile ranks.

Both the PDMS and PDMS-2 were designed to identify delays in fine motor and GM development of children from birth to 6 years of age (Folio & Fewell, 1983; Folio & Fewell, 2000). The original version was developed using a normative sample of 617 children and consists of 173 GM items scored on a three point ordinal scale. The PDMS-2 has many of the same items as the original PDMS and was developed using a normative sample of 2,003. The PDMS-2 consists of 151 items scored on a three point ordinal scale. As per administration guidelines within the test manuals, refusal by a child to complete an item is coded as a 0. A recent review of measures used to test motor development in preschool children recommended the use of the PDMS-2 due to its excellent reliability and validity (Tieman, Palisano & Sutlive, 2005).

To obtain descriptive information about the children's cognitive abilities, the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990) was administered at 4.5 years of age by assessors trained by a clinical psychologist. This brief, individually administered test of verbal and nonverbal intelligence is designed for use with persons aged 4 -90 years. The K-BIT has two subscales: the Vocabulary subscale tests both word knowledge and verbal concept formation and the Matrices section tests nonverbal skills and problem solving. Standardized on 2,022 subjects, the K-BIT has undergone extensive investigations of reliability

and validity within a variety of populations. For preschool children, construct and discriminant validity was established with the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), and the Adjustment Scales for Children and Adolescents (Canivez, Neitzel, & Martin, 2005). Test-retest reliability coefficients of 0.92 and 0.95 are reported in the manual for the IQ composite score and the internal consistency of the IQ composite score is 0.94 (Kaufman & Kaufman, 1990).

Analyses

The percentile rank scores of each child's nine GM assessments from the combined infant and preschool data were plotted to create individual profiles. Cluster analysis was then used to group children's profiles into distinct groups of high internal (within group) homogeneity and high external (between group) heterogeneity. Using SPSS 14.0 for Windows, two common hierarchical agglomerative clustering methods, Ward's and between groups average linkage, were compared. The algorithm of Ward's method combines profiles with the aim of minimizing the within-cluster sum of squares at each stage in the cluster process, while the algorithm for between groups average linkage method combines profiles using the average distance from all profiles in one cluster to all individuals in another (Hair et al., 1998).

Euclidean distance was chosen as the proximity measure for these analyses because the investigators' primary interests were in both the level of performance for each variable as well as the comparative patterns of variables

between individual profiles (Hair et al., 1998; Jobson, 1992; Macnab et al., 2001). The final number of clusters was determined by examining the agglomeration schedules and dendrograms (hierarchical trees) generated by SPSS. When two very different clusters were combined, the agglomeration coefficient increased substantially compared with the previous stage, indicating that the previous stage was a potential stopping point (Hair et al., 1998; Jobson, 1992). The dendrogram provided a visual representation of the relative distances between clusters at each stage of the clustering process and was used to confirm the final number of clusters. Once the final number of clusters was confirmed, a K-means iterative partitioning method was used to fine tune the clusters. This divisive clustering process uses the average value of each variable within each cluster as cluster seeds to generate a specified number of clusters based on the hierarchical procedure (Hair et al., 1998; Jobson, 1992).

The clustering process uses mathematical formulae to organize individuals into subgroups. Because we were interested in the clinical relevance of the clusters, each potential cluster solution was reviewed by two physical therapists (KE, JD) to evaluate the clusters for clinical relevance and distinctiveness. Factors such as the shape of the profiles and the number of times scores within each profile were at or below the cut-off point for suspicious GM development (16^{th} percentile) on the PDMS and PDMS-2 were discussed. Gender, ethnicity, family income, parental education, reported illness/hospitalization occurrences and scores of cognitive abilities were compared across the clusters using Chi-

square tests or analysis of variance techniques (ANOVA). These variables were chosen because risk factors for poor developmental attainment in Canada include: poor health, male gender, low household income, maternal education below the high school level, and maternal immigrant status (To et al., 2004).

Results

Twenty-eight girls and 38 boys had complete GM data for the two studies described. Most children were White (89%) and the median family income was $60,000 - $69,999. The median education level for both parents was completion of a college diploma or university degree. The children's average composite standard score on the K-BIT was 108.18 (*SD* = 18.11). No child had received a medical diagnosis at the conclusion of the infant study, but by the end of the preschool study one child had received a diagnosis of Benign Rolandic Epilepsy.

Of the 66 families in the sample, 58 were successfully contacted by phone when the children were 8 years of age. The child with Benign Rolandic Epilepsy had received a second diagnosis of moderate cognitive delay and was seeing an OT weekly at school. One child received a diagnosis of Attention Deficit Disorder and had brief OT intervention to work on her pencil grasp. Another child received a diagnosis of 'gross motor planning deficiency,' had a brief PT assessment, and received OT regularly through his school program. One child received OT intervention for two six week sessions to work on his fine motor and GM skills and two other children had a brief assessment with an OT to assess their fine motor skills in the classroom.

After comparing the output of the two clustering methods, the output of Ward's method was chosen because the dendrogram and agglomeration schedule for Ward's method indicated a clear stopping point of 4 distinct clusters while the demarcation points were less clear using the average linkage method. Graphical representations of the percentage change in agglomeration coefficients and dendrograms for each method are displayed in Figures 3-1 through 3-4. Visual inspection of the profiles in each cluster revealed that the clusters generated by Ward's method were also clinically distinct from one another. The individual profiles of percentile rank scores for all participants are graphed together in Figure 3-5 and the four clusters generated using Ward's method are displayed in Figures 3-6 to 3-9. Profiles of all children in each cluster are graphed with a bold line representing the cluster average. The range of individual percentile rank score changes over time within each cluster confirms persistent intra-individual variability in all clusters at both infant and preschool ages.

Cluster 1 consists of 22 children with percentile rank scores ranging from 13 to 91 percentiles over the assessment period. Their percentile rank scores all remain above the 16^{th} percentile cut-off point except for four children each with one instance of a score at or below the 16^{th} percentile (Table 3-3). This cluster is described as having 'robust scores.' Cluster 2 has 14 children and their pattern of percentile rank scores is characterized by lower percentile rank scores in preschool compared to infant assessment ages; this cluster profile is described as 'decreasing scores.' Percentile rank scores range from 4 to 82 percentiles and 13

children in this cluster had a score at or below the 16th percentile on at least one occasion. Cluster 3 is described as 'increasing scores' and includes the scoring profiles of 11 children. They show an opposite scoring pattern to Cluster 2 with lower percentile rank scores for infant compared to preschool assessments. The range of scores is from 1 to 88 and 10 children in this cluster had scores at or below the 16^{th} percentile. Cluster 4 consists of 19 children. Their profile is characterized by primarily low percentile rank scores in both infant and preschool assessments although the absolute range of percentile rank scores is from 1 to 71 percentiles. Eighteen children in this cluster had more than one score at or below the 16^{th} percentile over the nine assessment ages and this cluster is described as 'low scores.'

A significant difference was found between clusters for total illnesses reported $F(3,46) = 3.96$, $p < 0.05$, $\eta^2 = 0.21$. A post hoc Tukey Test was conducted, which revealed a significant difference between Cluster 3 ($M = 3.14$, $SD = 1.07$) and Cluster 4 ($M = 1.31$, $SD = 1.25$). No significant difference was found between Clusters 1, 2, and 3 for total reported illnesses. Cluster 4 had the lowest average number of reported illnesses, followed by Cluster 2 ($M = 2.38$, $SD = 1.12$) and Cluster 1 ($M = 2.41$, $SD = 1.33$). A review of the raw data revealed the majority of reported illnesses in Cluster 3 were documented at the 9, 11, and 13 month assessment points.

No significant differences were found between clusters based on hospitalization, gender, ethnicity, family income, mother or father's years of

education, or cognitive ability scores on the K-BIT (Tables 3-1 and 3-2). The child with a diagnosis of gross motor planning deficiency who received a brief PT assessment and regular OT at school was in Cluster 3 and the other two children with diagnoses and OT intervention were in Cluster 4. The remaining three children who received OT intervention were also in Cluster 4.

Discussion

Cluster analysis was used in this study as a method to distill the intra-individual variability observed within individual children's serial GM percentile rank scores into a manageable number of clinically distinct groups. Because clusters can be created from any data set using these techniques (Jobson, 1992; Macnab et al., 2001), it was very important to examine the clusters for their clinical relevance in addition to their statistical uniqueness. For all research using these techniques, the final cluster solution has to be both statistically significant and clinically meaningful.

Cluster analysis techniques yielded a clinically manageable set of four clusters with clinically distinct patterns. Clusters 1, 3, and 4 exhibit patterns of emergence of motor skills familiar to clinicians. Children in Cluster 1 maintained robust scores across the infant and preschool assessment ages. It is important to note that despite their strong overall scores, no child had a pattern of consistently extremely high percentile ranks. Clinically this finding suggests that if a child scores at the 90th percentile once, it is unlikely that he or she will remain scoring at the 90th percentile on serial assessments. Therapists need to expect some

variation within a child's percentile rank scores, and they need to be able to explain this fluctuation in percentile rank scores to parents. Many parents interpret percentile rank scores as per cent scores and view anything below the 50^{th} percentile as unsatisfactory. It is important that therapists assist parents in understanding that percentile rank scores below the 50^{th} percentile are well within typical development and should not be viewed as worrisome. This approach supports the concept of a large bandwidth of acceptable scores on a standardized measure rather than a hierarchical approach that views scores on the 90^{th} percentile as 'stronger' than scores on the 60^{th} percentile. It is also interesting to note clinically that even children with this 'sustained robust score' pattern can receive a score at or below the cut-off point of the 16^{th} percentile on the measure used.

Cluster 3 represents the children who demonstrate a pattern of percentile rank scores that increased in preschool assessments compared to their infant assessments. The cluster pattern suggests children in this group may demonstrate an increased rate of emergence of motor skills in preschool ages, reflected by increasing percentile rank scores. Based on the initial low infant percentile rank scores clinicians may initiate intervention with children in this group. If children with this pattern of GM skill development receive intervention at young ages, it could be assumed that the change in pattern of scores in preschool ages reflects the effect of intervention. Given the results of the ANOVA for number of reported illnesses, it is possible that illness may have had a negative impact on

early GM performance in this group. However, there was no significant difference among total reported illnesses of Clusters 1, 2, and 3; the significant difference was between Clusters 3 and 4. Thus, the increasing GM percentile rank scores could reflect a distinct pattern in the absence of illness. Our analyses suggest that this pattern of change can occur naturally without intervention.

Cluster 4 represents the group of children of most concern to clinicians because of their low percentile rank scores in infancy and preschool and high frequency of scores below the tests' recommended cut-off point. This cluster has a more linear pattern of low scores. However, intra-individual variability is still evident within many of the children's score profiles. Two of these children had a medical diagnosis and were receiving ongoing OT intervention. Three other children from this group had had some OT involvement by age 8. The remaining 14 children had no history of physical therapy or occupational therapy intervention. It would be interesting to follow the children in this cluster to older ages and see whether any are identified as having a motor disability such as DCD. An understanding of the early GM profile of children with DCD could assist in earlier identification. Conversely, it is also conceivable that most of the children in this cluster are typically developing and demonstrate another pattern of motor skills emergence that is within typically developing limits.

Cluster 2 consists of children with patterns of primarily strong scores in infancy but with a steady decrease of percentile rank scores into preschool. Clinicians may not be as familiar with this pattern of scores because these

children would not have been flagged by low percentile rank scores in infancy. No child from this cluster had a diagnosis or intervention by a motor therapist, yet most clinicians would be concerned by this pattern of scores. Long-term follow-up of this group could provide important information about motor development. Will they self right, as did children from Cluster 3, or will they present with motor problems later in childhood?

Overall, the patterns of scoring in all four clusters suggest that the clinical interpretation of percentile rank scores is not clear cut. Caution should be exercised when using an isolated test score to evaluate GM development. As several scoring patterns are possible, a low score does not necessarily predict future low scores, nor would a high score predict future high scores. Some typically developing children have profiles which exhibit a steady improvement in scores, while others demonstrate an overall decline in scores. Standardized tests should be administered on several occasions and the scores should be interpreted within the context of the child's overall well being including a comprehensive developmental history, recent illness, parent concerns and skilled clinical observation (American Academy of Pediatrics, 2006).

With the exception of reported illnesses, none of the other descriptive data collected could explain cluster membership as there were no significant differences between clusters based on gender, ethnicity, family income, parents' education, reported hospitalizations or cognitive abilities. Outside of these variables, we did not attempt to explain the reasons for membership in a specific

cluster. Further research could be done to explore the interactive effects of fine motor and GM performance, communication ability, cognition, and demographic variables on cluster membership. This was a descriptive study aimed at determining whether the variability observed in longitudinal percentile rank scores of typically developing children could be organized into clinically meaningful clusters. Further research needs to be done to evaluate the shapes and patterns of percentile rank score profiles of at-risk or atypically developing infants and young children.

Measurement error cannot account exclusively for the substantial intra-individual variability of GM percentile rank scores observed in these data. Standardized GM tests with excellent reliability and validity were chosen (Palisano, 1986; Tieman et al., 2005) and inter-rater reliability of the study assessors was monitored regularly in both studies. The variability in patterns of scores also suggests that the fluctuations in percentile rank scores were not due to test item discrepancies as the scores did not increase or decrease in a systematic pattern.

Conclusion

Cluster analyses provided an effective method to manage the intra-individual variability observed in the data. Combined with clinical interpretation, the process facilitated exploration of some of the practical implications of the intra-individual variability observed in serial testing of children's GM skills and provided a manageable number of patterns that may be of use clinically. Future

work is needed to evaluate the long term significance of cluster membership and to replicate and validate these clusters in a different sample of typically developing children. This information would be useful in helping clinicians make decisions about management of children who display various patterns of GM percentile rank scores. Further exploration of intra-individual variability in patterns of GM percentile rank scores has the potential to reveal insights of both theoretical and clinical importance.

Figure 3-1. Graphical Representation of the Percentage Change in Agglomeration Coefficient using Ward's Method

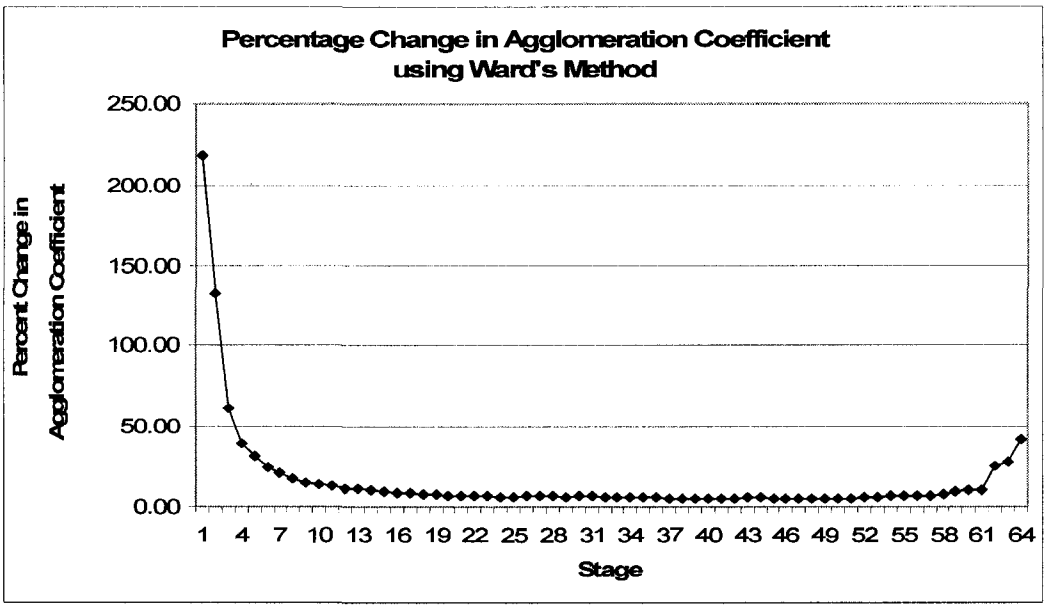


Figure 3-2. Graphical Representation of the Percentage Change in Agglomeration Coefficient using the Average Linkage Method

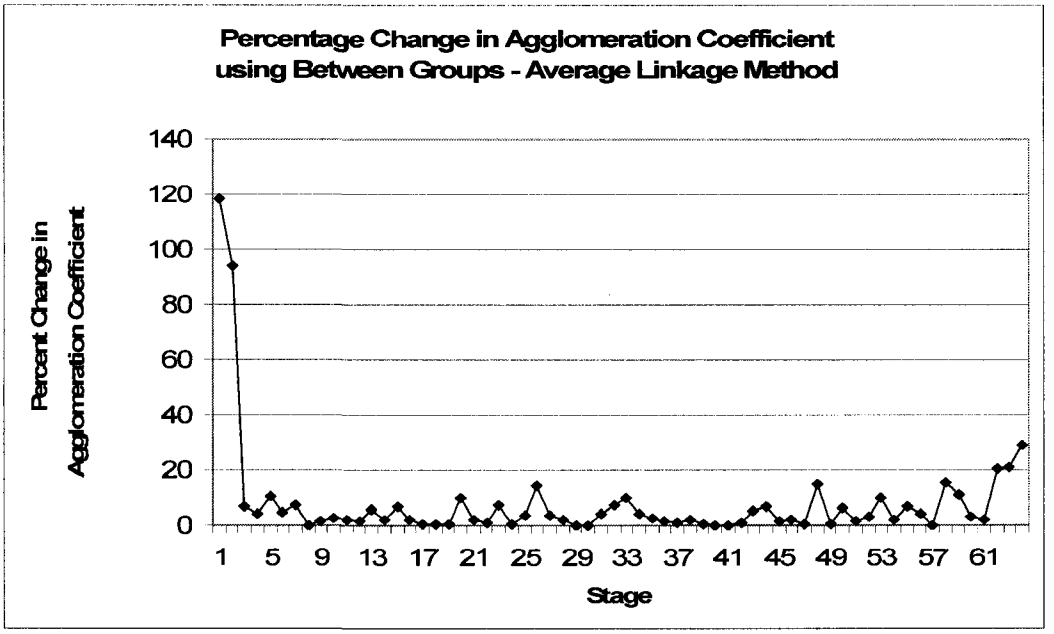
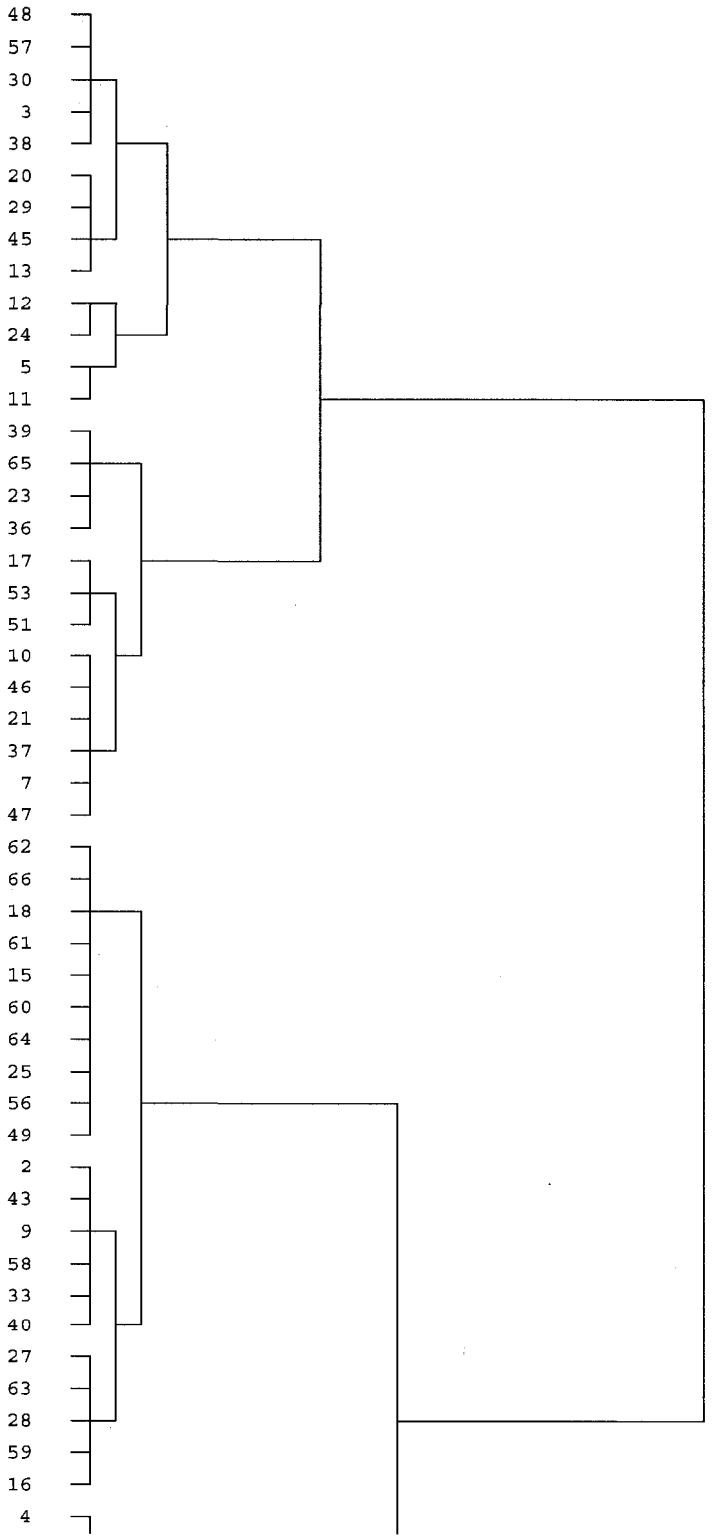


Figure 3-3. Dendrogram using Ward's Method



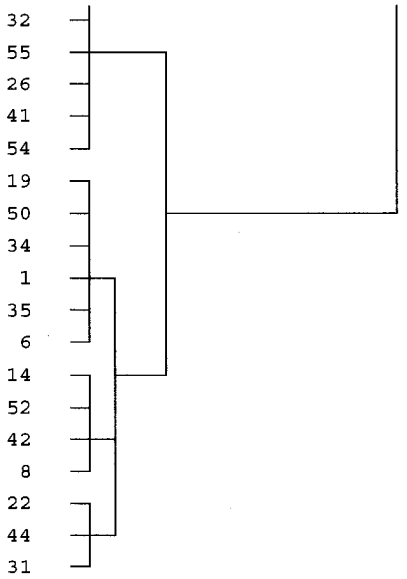
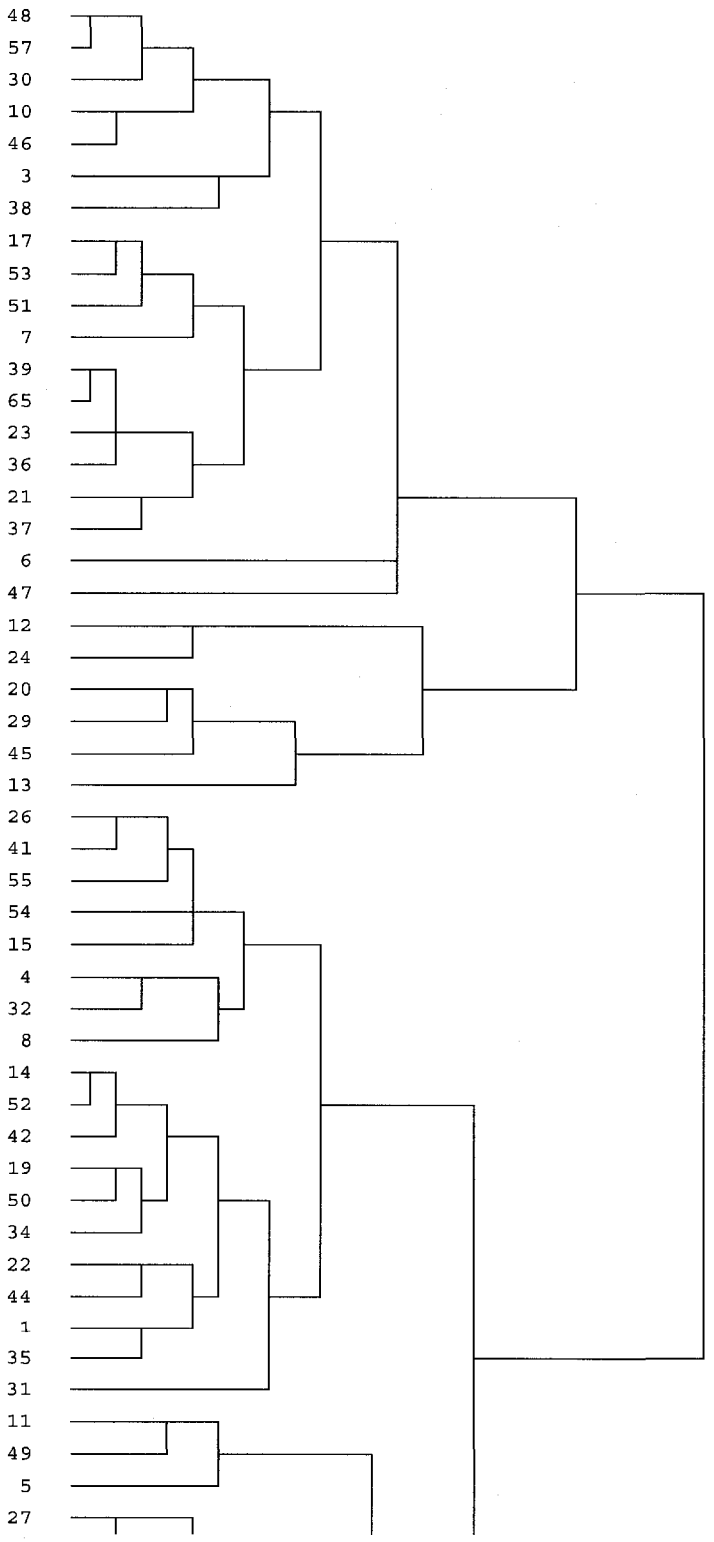

32
55
26
41
54
19
50
34
1
35
6
14
52
42
8
22
44
31

Figure 3-4. Dendrogram using Average Linkage Method



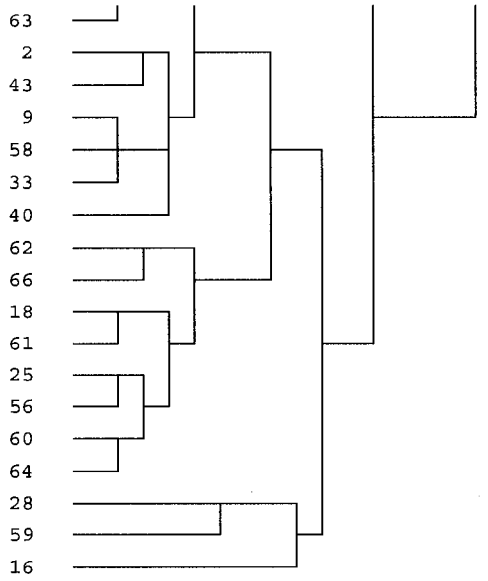

63
2
43
9
58
33
40
62
66
18
61
25
56
60
64
28
59
16

Figure 3-5. Percentile Rank Score Profiles of All Participants

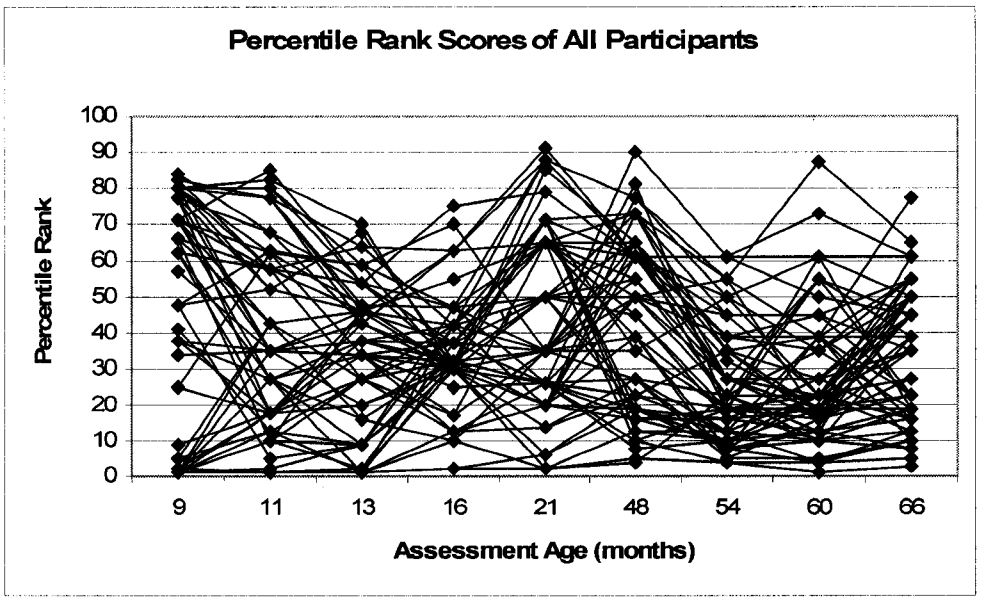


Figure 3-6. Cluster 1 – Robust Scores

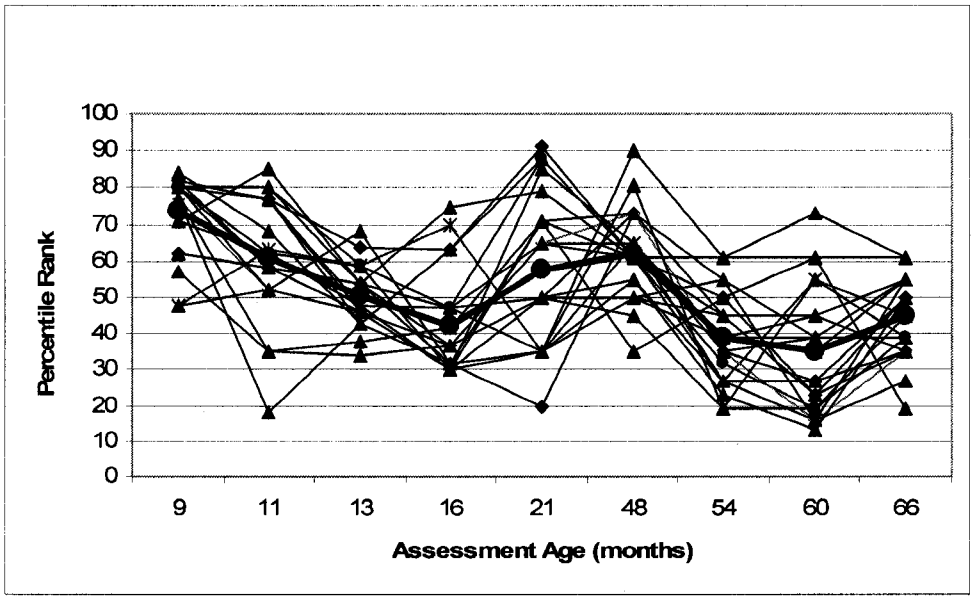


Figure 3-7. Cluster 2 – Decreasing Scores

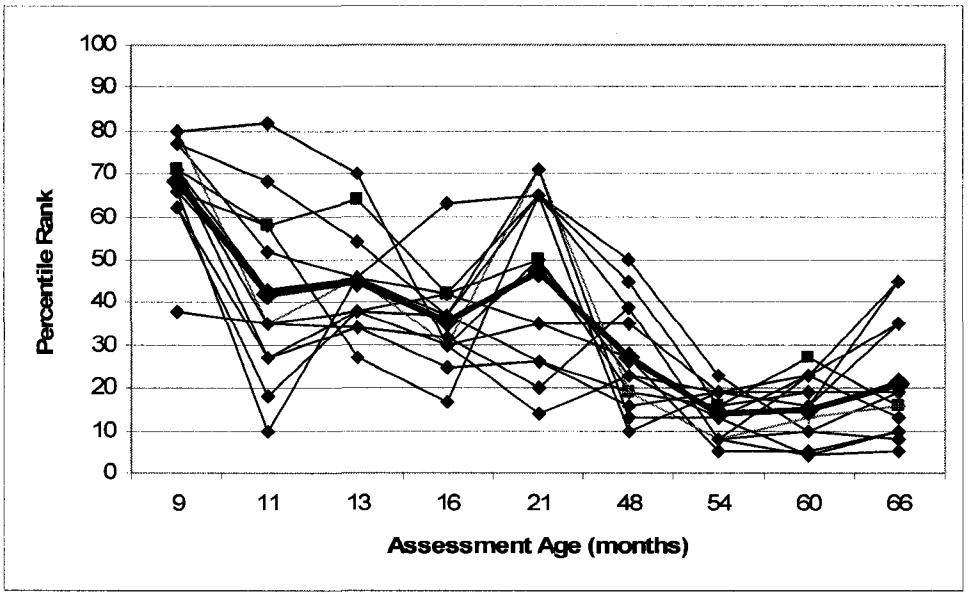


Figure 3-8. Cluster 3 – Increasing Scores

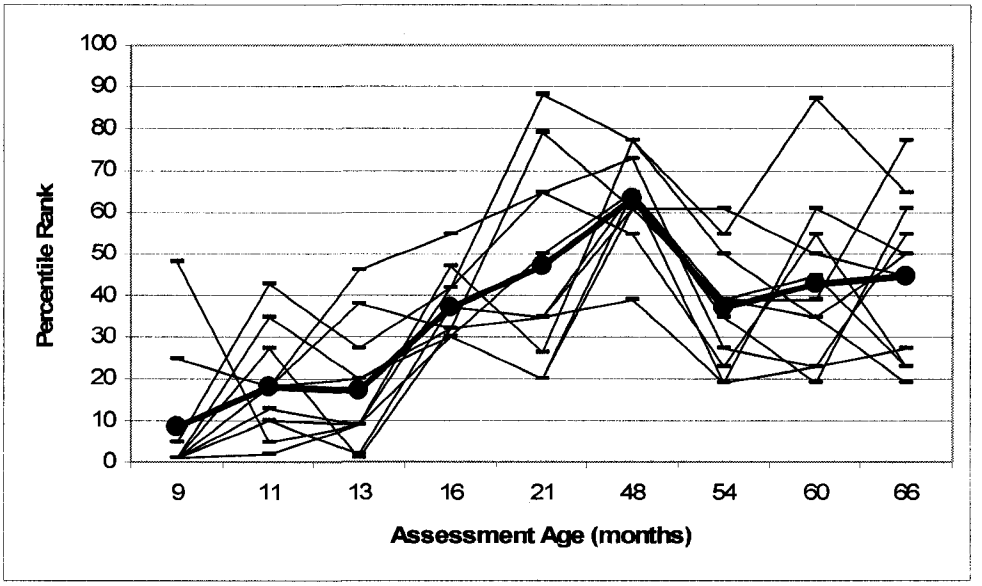


Figure 3-9. Cluster 4 – Low Scores

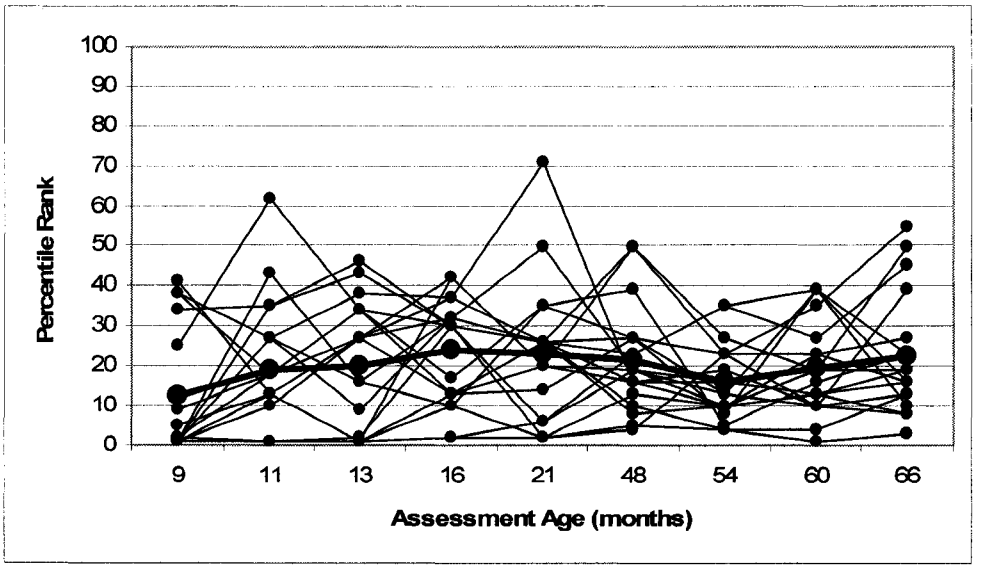


Table 3-1

Analysis of Variance between Clusters

Variable	*df*	*F*	η^2	*p*
Total Illnesses Reported	3	3.96*	0.21	0.01
Total Reported Hospitalizations	3	0.79	0.05	0.51
KBIT Standard Score – age 4.5	3	0.13	0.01	0.94
Years of Education – Father	3	1.55	0.07	0.21
Years of Education - Mother	3	0.52	0.03	0.67

*$p < 0.05$.

Table 3-2

Chi-Square Tests

Variable	Categories	*df*	Pearson X^2
Gender	Male/female	3	1.55
Ethnicity	White/Non-white	3	2.23
Family Income	Above/below $60,000 per year	3	0.15

Table 3-3

The Number of Times a Child Scored at or Below the 16th Percentile Rank Score on the PDMS or PDMS-2 by Cluster Membership

Cluster	*N*	Number of times a child scored ≤ 16th percentile rank score									
		0	1	2	3	4	5	6	7	8	9
1	22	18	4	-	-	-	-	-	-	-	-
2	14	1	4	3	5	1	-	-	-	-	-
3	11	1	4	2	4	-	-	-	-	-	-
4	19	1	0	5	1	4	4	-	-	3	1
Total	66	21	12	10	10	5	4	-	-	3	1

References

American Academy of Pediatrics. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118*(1) 405-420.

Bertenthal, B. I., & Boker, S. M. (1997). New paradigms and new issues: A comment on emerging themes in the study of motor development. *Monographs of the Society for Research in Child Development, 62*(3), 141-151.

Brendgen, M., Vitaro, F., Bukowski, W. M., Doyle, A. B., & Markiewicz, D. (2001). Developmental profiles of peer social preference over the course of elementary school: Associations with trajectories of externalizing and internalizing behavior. *Developmental psychology, 37*(3), 308-320.

Canivez, G. L., Neitzel, R., & Martin, B. E. (2005). Construct validity of the Kaufman Brief Intelligence Test, Wechsler Intelligence Scale for Children – 3rd ed., and Adjustment Scales for Children and Adolescents. *Journal of Psychoeducational Assessment, 23*(1), 15-34.

Darrah, J., Hodge, M., Magill-Evans, J., & Kembhavi, G. (2003). Stability of serial assessments of motor and communication abilities in typically developing infants - Implications for screening. *Early Human Development, 72*(2), 97-110.

Darrah, J., Redfern, L., Maguire, T. O., Beaulne, A. P., & Watt, J. (1998). Intra-individual stability of rate of gross motor development in full-term infants. *Early Human Development, 52*(2), 169-179.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1-173.

Fischer, K. W., & Pare-Blagoev, J. (2000). From individual differences to dynamic pathways of development. Comment. *Child development, 71*(4), 850-853.

Folio, M. R., & Fewell, R. P. (1983). *Peabody Developmental Motor Scales and activity cards: A manual.* Allen: DLM Teaching Resources.

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

Gesell, A. (1928). *Infancy and human growth.* New York: McMillan; 1928.

Giani, U., Filosa, A., & Causa, P. (1996). A non-linear model of growth in the first year of life. *Acta Paediatrica, 85*(1), 7-13.

Gilmore, R. O., & Thomas, H. T. (2002). Examining individual differences in infants' habituation patterns using objective quantitative techniques. *Infant Behavior and Development, 25*, 399-412.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). New Jersey: Prentice Hall.

Jobson, J. D. (1992). *Applied mulitvariate data analysis: Categorical and multivariate methods*. New York: Springer-Verlag.

Kaufman, A., & Kaufman, N. (1990). *Kaufman brief intelligence test (K-BIT).* Circle Pines, MN: AGS Publishing.

Koller, H., Lawson, K., Rose, S. A., Wallace, I., & McCarton, C. (1997). Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics, 99*(3), 383-389.

Macnab, J. J., Miller, L. T., & Polatajko, H. J. (2001). The search for subtypes of DCD: Is cluster analysis the answer? *Human Movement Science, 20*(1-2), 49-72.

McCall R.B., Hogarty P.S. & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. *American Psychologist, 72*(8), 728-748.

Mei, Z., Grummer-Strawn, L. M., Thompson, D., & Dietz, W. H. (2004). Shifts in percentiles of growth during childhood: Analysis of longitudinal data from the California child health and development study. *Pediatrics, 113*(6), e617-e627.

Moffit, T. E., Caspi, A., Harkness, A. R., & Silva, P. A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology and Psychiatry, 34*(4), 455-506.

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology, 42*, 213-237.

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.

Siegler, R. S. (2002). Variability and infant development. *Infant Behavior and Development, 25*, 550-557.

Stevenson, R. D., Conaway, M., Chumlea, W. C., Rosenbaum, P., Fung, E. B., Henderson, R. C., et al. (2006). Growth and health in children with moderate to severe cerebral palsy. *Pediatrics, 118*(3), 1010-1018.

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist, 50*(2), 79-95.

Tieman, B. L., Palisano, R. J., & Sutlive, A. C. (2005). Assessment of motor development and function in preschool children. *Mental Retardation & Developmental Disabilities Research Reviews, 11*(3), 189-196.

To, T., Guttman, A., Dick, P. T., Rosenfield, J. D., Parkin, P. C., & Cao, H., et al. (2004). What factors are associated with poor developmental attainment in young Canadian children? *Canadian Journal of Public Health. 95*(4), 258-263.

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*, 340-374.

Wakely, M. B., Hooper, S. R., de Kruif, R. E., & Swartz, C. (2006). Subtypes of written expression in elementary school children: A linguistic-based model. *Developmental Neuropsychology, 29*(1), 125-159.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children.* (3rd ed.). San Antonio: The Psychological Corporation.

CHAPTER 4

A COMPARISON OF SERIAL PERCENTILE RANK SCORES AND TEST CLASSIFICATIONS USING THE PEABODY DEVELOPMENTAL MOTOR SCALES – SECOND EDITION AND THE MOVEMENT ASSESSMENT BATTERY FOR CHILDREN

Introduction and Purpose

Pediatric physical therapists use standardized tests of motor development to identify motor delays in young children. Two commonly used tests are the Peabody Developmental Motor Scales – Second Edition (PDMS-2, Folio & Fewell, 2000) and the Movement Assessment Battery for Children – Performance Test (MABC-Test, Henderson & Sugden, 1992). Both tests evaluate fine and gross motor development and include normative tables to rank children's performance relative to age-matched peers. The test format, scoring criteria and age ranges of each test differ.

The PDMS-2 was designed to assess the motor skills of children from birth through 6 years of age (Folio & Fewell, 2000). The normative sample consisted of 2,003 children aged 0 to 71 months of age. Two hundred forty-nine items are organized into three global indices of motor performance; the gross motor quotient, the fine motor quotient and the total motor quotient. The gross motor quotient consists of four subtests: reflexes, stationary, locomotion, and object manipulation. The fine motor quotient consists of two subtests: grasping and visual-motor integration. The total motor quotient represents the combination

of the gross and fine motor quotients and is described as the best estimate of a child's overall motor abilities. The gross and fine motor quotients can be administered and scored separately. Each item is scored on an ordinal scale from 0 to 2 using criteria outlined in the test manual. Raw scores can be converted to age equivalents, standard scores, motor quotient scores, and percentile rank scores. Tables are included within the test manual to help therapists interpret subtest standard scores and quotient scores by providing classification categories describing motor performance.

The MABC Test was developed to identify and describe impairments of motor function in children from 4 to 12 years of age (Henderson & Sugden, 1992). The normative sample consists of 1,234 children. The test contains four age bands, each with eight items grouped under three subsections: manual dexterity, ball skills and static/dynamic balance. The subsections of the MABC Test are not designed to be administered or interpreted in isolation. Each item is scored on a scale from 0 to 5 and the item scores are summed to produce a total impairment score, which can then be converted to a percentile rank score. Based on their overall performance on the MABC Test, children are classified as having normal (>15th percentile), suspicious (between 6th and 15th percentile), or delayed motor performance ($\leq$ 5th percentile).

Therapists testing a child with the PDMS-2 at preschool age may need to change measures as the child reaches the maximum age of the test (71 months). For example, a teacher or parent of a child tested at 5 years of age in a preschool

program with the PDMS-2 may request further testing when the child transitions to grade school at 6 years of age. In addition, different measures may be used in different programs to determine funding eligibility. Although the convergent validity of the PDMS-2 and the MABC Test has been established when administered on a single occasion (van Hartingsveldt, Cup & Oostendorp, 2005; van Waelvelde, Peersman, Lenoir & Smits-Engleman, 2007), there is a lack of research comparing children's performance on these tests across serial assessments.

Traditional research into the stability of serial assessments of children's cognitive abilities suggested that a stronger relationship (as indicated by higher correlation coefficients) is present between scores from adjacent testing times, between test scores taken at older ages, and between children with very low test scores (Bayley, 1949; McCall, Hogarty & Hurlburt, 1972; Rose, 1989; Siegel, 1989). Higher correlation coefficients between adjacent test scores have been reported in several longitudinal studies of motor development. When the original version of the Peabody Developmental Motor Scales (PDMS, Folio & Fewell, 1983) and the Bayley Motor Scale (BMS, Bayley, 1969) were administered to a mixed sample of 23 typically developing infants and 21 premature infants at 12, 15, and 18 months of age, the correlation coefficients were highest between adjacent 3 month intervals (Palisano, 1986). Similarly, Darrah, Hodge, Magill-Evans and Kembhavi (2003) reported higher correlation coefficients between

adjacent ages as compared to non-adjacent ages when the PDMS was administered to typically developing infants at 9,11,13,16 and 21 months of age.

The assumption that stronger relationships exist between test scores taken at older ages has had limited support within the motor literature. Research using the BMS in typically developing infants aged one through 15 months of age suggested that later motor scores were more stable than earlier motor scores (Bayley, 1965). In contrast, an analysis of BMS scores taken at 2, 3, 4, 8, 12, and either 24 or 36 months of age using typically developing infants found the infants' scores varied significantly between all testing times (Coryell, Provost, Wilhelm, & Campbell, 1989). When scores on the PDMS taken at 9, 11, 13, 16 and 21 months of age in typically developing infants were compared with their scores at 4 years of age, the scores at 21 months of age were not found to have a stronger relationship than scores from younger assessment ages (Darrah, Magill-Evans, Volden, Hodge & Kembhavi, 2007). Similarly, Palisano (1986) did not find higher correlation coefficients between 15 and 18 month assessments using the PDMS compared to 12 and 15 month assessments.

The final assumption of a stronger relationship between children with very low test scores is supported by longitudinal studies of motor performance. As Netelenbos (2005) reasons, children with atypical motor development may have globally deficient disorders which consistently depress test scores over time resulting in a stronger relationship between scores for these children. Coryell et al. (1989) reported test scores in their 'non-normal' outcome group did not vary

significantly compared with test scores of their 'normal' outcome group using the BMS. Palisano (1986) reported higher correlation coefficients between PDMS fine motor scores for a subgroup of premature infants compared to full term infants. As children with atypical development age, deficits tend to become more evident (Hadders-Algra, 2002). This may result in consistently poor performance by these children on standardized test items compared with their peers which boosts correlation coefficients.

Correlation coefficients provide information about a test's stability using group data, but this information may not be as useful to clinicians as information about changes in classification of an individual child's performance (Siegel, 1989). Most standardized tests of motor development provide guidelines or cut-off scores to interpret test results and identify children as having typical or delayed development. Ideally there would be agreement among motor tests regarding cut-offs for classification of motor ability and identification of delay, but recent research suggests otherwise. Netelenbos (2005) suggests different motor tests regularly classify different children as being at risk for motor delays. For example, when van Waelvelde et al. (2007) compared the MABC Test with the PDMS-2, they found that while the total scores of the two tests correlated well, the agreement between tests in identifying children with motor difficulties was poor. Similarly, Crawford, Wilson, and Dewey (2001) found low levels of agreement between the MABC Test and the Bruininks-Oseretsky Test of Motor Proficiency (Bruninks, 1978) in identifying children with Developmental

Coordination Disorder (DCD). Consistency in test classification over time between the MABC Test and the PDMS-2 has not yet been researched. Clearly, further investigation regarding the stability of test classification between these tests over time is both warranted and needed by therapists to make informed clinical decisions.

The present analyses explore the relationship among serial PDMS-2 scores and the relationship between PDMS-2 scores at different ages and MABC Test scores at 7 years of age. The relationship between test classifications at different ages is also examined. The specific objectives were to:

1. Evaluate the relationship strength of children's percentile rank scores on the PDMS-2 at 4, 4.5, 5, 5.5 years of age and the relationship strength between the PDMS-2 assessments at these ages and the MABC Test at 7 years of age.
2. Examine the similarities and differences in classification across the five assessment points using cut-offs for suspicious motor development on the two tests.

Method

Sample

A total of 104 children participated in the study. Eighty-three participants were recruited as infants for a previous longitudinal study from public health centers and Mom and Baby classes in the Edmonton area and 21 were recruited as preschoolers using advertisements in local newspapers and city wide poster

campaigns. Over the course of the study, 14 families dropped out and 5 children missed at least one motor assessment. Thus, the present analyses include the motor data of 85 children.

Procedure and Measures

Parents completed a short questionnaire at each assessment asking about their child's health, hospitalization, intervention, special programs or other factors that may have affected performance. Parents also completed a more detailed questionnaire at the start of the study regarding their child's gender, ethnicity, annual family income and education levels of each parent. Each child was assessed using the PDMS-2 at 4, 4.5, 5 and 5.5 years of age by a pediatric motor therapist. Children with scores of one standard deviation (16^{th} percentile) or less below the mean were classified as having suspicious motor development. At 7 years of age, each child was assessed using the MABC Test (Henderson & Sugden, 1992). Each child's motor performance was classified as normal (above the 15^{th} percentile) or suspicious/delayed (at or below the 15^{th} percentile). At all assessment ages, therapists administered the test items and recorded the scores for individual items, and a research assistant totaled the raw scores and converted them to percentile rank scores. Therapists were trained in administration of the tests as recommended in the test manuals and attained an 80% item by item agreement before testing began. During data collection, every 10^{th} assessment was observed by a second therapist; inter-rater reliability for the therapists was 0.92 for PDMS-2 gross motor scores, 0.93 for PDMS-2 fine motor scores, and

0.99 for the MABC Test. When the children were 8 years of age, each family was contacted by phone and asked whether their child had received a medical diagnosis or intervention by an occupational therapist or physical therapist.

Analyses

Descriptive statistics were generated for sample characteristics (ethnicity, gender, family income, and parent education level) and gross motor, fine motor and total motor percentile rank scores for the five testing times. Correlation coefficients were calculated for comparisons between total PDMS-2 percentile rank scores and the MABC Test percentile rank scores, PDMS-2 fine motor percentile rank scores and the MABC Test percentile rank scores, and PDMS-2 gross motor percentile rank scores and the MABC Test percentile rank scores. PDMS-2 fine and gross motor percentile rank scores were compared to the MABC Test in addition to the total PDMS-2 percentile rank scores as these subsections of the PDMS-2 can be administered separately. The 95% confidence interval for each correlation coefficient was also calculated to represent the upper and lower ranges of the true correlation values. Overlapping confidence intervals were considered an indication that the differences in magnitude for the correlation coefficients could be due to measurement error. The children who scored at or below a test cut-off at any of the five assessment ages were identified and a table was constructed to review their scores over time. Information gathered from parent phone calls regarding diagnoses and intervention by an occupational or

physical therapist was also added to the table and the results were visually analyzed by two physical therapists (KE, JD).

Results

Data from 35 girls and 50 boys were analyzed. The majority of the children were White (87%) with the remainder consisting of Chinese, South Asian, North American Indian, Other and Mixed Ethnicities. The median annual family income was $70, 000 - $79,000. Both mothers and fathers had a median education level of a university or college degree.

Descriptive statistics for the fine, gross, and total motor percentile ranks on the PDMS-2 at 4, 4.5, 5 and 5.5 years of age and for the MABC Test at 7 years of age are provided in Table 4-1. The correlation coefficients with 95% confidence intervals for PDMS-2 fine motor, gross motor, and total percentile rank scores from 4, 4.5, 5, and 5.5 years of age compared with MABC Test total percentile rank scores are presented in Tables 4-2 to 4-4. The correlation coefficients ranged from 0.30 to 0.71. Correlation coefficients of 0.00 to 0.25 suggest little or no relationship, 0.25 to 0.50 indicate a fair relationship, 0.50 to 0.75 imply a moderate to good relationship, and those over 0.75 suggest a good to excellent relationship between variables (Portney & Watkins, 2000). The confidence intervals for the correlation coefficients of all percentile rank score comparisons overlapped suggesting a stronger relationship was not found between scores from adjacent testing times or between scores taken at older ages.

Thirty-four children (40%) scored below a test cut-off at least once over the duration of the study. Nineteen children (22%) were classified as having suspicious or delayed motor performance using the MABC Test at seven years of age. Seven of these children never received scores below the PDMS-2 suspicious cut-off at any assessment age. The percentile rank scores below the cut-offs across the five assessment ages are summarized in Table 4-5.

Of the 85 families in the sample, 71 were successfully contacted by phone when the children were 8 years of age. Four children received diagnoses and nine had received intervention by a motor therapist. One child had diagnoses of Benign Rolandic Epilepsy and moderate cognitive delay and was seeing an OT regularly at school. Another had diagnoses of Attention Deficit Hyperactive Disorder (ADHD) and Oppositional Defiant Disorder (ODD) and had a brief OT assessment but no intervention. A third child had a diagnosis of Attention Deficit Disorder (ADD) and had brief OT intervention to work on her pencil grasp. The fourth child received a diagnosis of 'gross motor planning deficiency,' had a brief PT assessment, and received OT regularly through his school program. One child received OT intervention for two six week sessions to work on his fine and gross motor skills and two other children had a brief assessment with an OT to assess their fine motor skills in the classroom. Two children worked with an OT on their writing skills; one through the school and the other through private OT services. These findings are summarized in Table 4-5.

Discussion

The fair to moderate correlation coefficients obtained from this study suggest that there is not a strong relationship between test scores on the PDMS-2 over time, nor is there a strong relationship between test scores on the PDMS-2 over time and MABC Test at 7 years of age for typically developing children. Although the correlation coefficients were all significant at the 0.01 level, none demonstrated good to excellent relationship strength. Additionally, because the 95% confidence intervals overlapped for all percentile rank score comparisons, the results do not support the assumptions from the cognitive literature of stronger relationships between scores from adjacent testing times or between test scores taken at older ages.

These results are similar to other correlative studies within the motor literature, which also did not find stronger relationships between test scores taken at older ages (Coryell et al. 1989; Darrah et al., 2007; Palisano, 1986). However, these results are different from past motor studies which found stronger relationships between adjacent test scores compared with non-adjacent test scores (Darrah et al., 2003; Palisano, 1986). It is possible that the use of 95% confidence intervals in this study may explain the difference. Had confidence intervals not been utilized, the correlation coefficients between adjacent testing times on the PDMS-2 would be judged higher than non-adjacent testing times.

Clinically, the fair to moderate relationship strength between percentile rank score comparisons implies that individual children change ranks over time

and that test scores are not stable across assessments. Therapists should expect variability among test scores of motor performance when screening for motor problems over time. Assumptions regarding a child's motor performance cannot be made from previous assessment results and repeated testing is required. Because the strength of the relationships between test scores on the PDMS-2 were not markedly different from the strength of the relationships between test scores of the PDMS-2 at earlier ages and the MABC Test at 7 years of age, switching from the PDMS-2 to the MABC Test as a child reaches the ceiling age on the PDMS-2-does not present a measurement disadvantage.

The correlation coefficients were calculated using scores from a sample of typically developing children and these results cannot be extrapolated to children with atypical motor development. The assumption from the cognitive literature of a stronger relationship between children with very low scores cannot be refuted or supported by these analyses. Netelenbos (2005) suggests that the practice of using mixed samples can artificially increase correlation coefficients for studies with large proportions of atypical children compared to those based on randomly selected samples. He advocates for separate reporting of results from random and mixed samples. Thus, future research comparing longitudinal test scores of the PDMS-2 and MABC Tests in a sample of atypically developing children would be valuable when compared with the present results.

The examination of similarities and differences in test classification over time revealed a lack of agreement both between the PDMS-2 assessments as well

as between the PDMS-2 and MABC Test assessments in identifying individual children as having suspicious/delayed motor performance. A large proportion (40%) of the sample was classified as having suspicious motor performance on at least one occasion, implying that typically developing children demonstrate variability in test classification as well as variability in test scores over time. Low mean and median values for gross motor percentile rank scores at 4 and 4.5 years of age may have contributed to the high proportion of children identified by the PDMS-2 at these ages. The reason for these low aggregate gross motor scores requires further evaluation. Furthermore, a high number of children scored below the cut-off score on the MABC Test, resulting in a prevalence of 22% for suspicious or delayed motor performance among the sample at 7 years of age. Interestingly, Hadders-Algra (2007) reported a similar rate of identification (23%) of children performing below the 15th percentile on the MABC Test at a mainstream primary school. Further evaluation of this high prevalence rate is warranted to confirm its validity as the MABC Test is often used as a gold standard for screening children for mild to moderate motor disabilities such as DCD (Geuze, Jongmans, Schoemaker & Smits-Engelman, 2001).

Of the children who received diagnoses, only the child with Benign Rolandic Epilepsy and moderate cognitive delay was consistently identified at all testing times. This child's percentile rank scores were also consistently very low, supporting the concept that children with the most severe developmental impairments are most easily identifiable (Dworkin, 1989). The child with ADD

was not identified at any assessment and the child with ADHD and ODD was not consistently identified across assessment ages. However, tests of motor development may not be the most appropriate screen for these types of diagnoses. The child with 'gross motor planning deficiency' was only identified at two assessments, highlighting the difficulty of detecting children with milder deficits in motor development (Williams & Holmes, 2004).

Clinically, these results highlight the importance of interpreting test classification in conjunction with clinical judgment, functional ability of the child, and concerns of parents and educators (American Academy of Pediatrics Committee on Children with Disabilities, 2001). Typically developing children can be expected to occasionally score below test cut-offs and be classified as having suspicious motor development, thus repeated testing is necessary to monitor skill development. While children with severe impairments may be consistently classified as having suspicious motor development, children with mild or moderate delays or those who receive intervention by a motor therapist may fluctuate in their classification on motor tests. However, further research with a sample of atypically developing children is needed to confirm this impression.

Conclusion

Based on the results of this study, physical therapists should expect variability in percentile rank scores and test classifications for typically developing children over time. Repeated testing of motor abilities is necessary to

determine a child's current motor abilities as earlier percentile rank scores and motor skill classifications are not indicative of future scores or classifications. Test scores and classifications should be interpreted within the context of the child's functional abilities and parent concerns. Replication of these procedures and analyses using data from a sample of children with mild or moderate motor disabilities is warranted. When comparing tests of motor development, investigations into both correlation and classification provide valuable information to therapists and researchers alike.

Table 4-1

Descriptive Statistics for Percentile Rank Scores (PR) on the PDMS-2 and MABC Test

Variable	Age	Mean	Median	Standard Deviation	Range
MABC PR	7	39.29	40	27.14	0-96
PDMS2 Total PR	5.5	43.01	42	18.19	0-100
	5	37.07	35	17.57	0-86
	4.5	35.27	35	17.62	1-73
	4	42.84	39	24.29	2-96
PDMS2 Gross Motor PR	5.5	33.13	35	17.91	3-77
	5	26.95	23	17.67	1-87
	4.5	26.53	19	16.00	4-61
	4	41.36	39	22.91	4-90
PDMS2 Fine Motor PR	5.5	59.04	58	16.51	0-79
	5	56.56	58	20.63	0-84
	4.5	52.55	58	23.90	0-92
	4	47.21	50	28.93	1-98

Table 4-2

Correlations between Total PDMS-2 Percentile Rank Scores at Different Ages and MABC Test Percentile Rank Scores at Age 7 with 95% Confidence Intervals

Age	4.5	5	5.5	7
4	0.71 (0.59,0.80)	0.50 (0.32,0.65)	0.51 (0.33,0.65)	0.47 (0.29,0.62)
4.5		0.57 (0.41,0.70)	0.53 (0.36,0.67)	0.49 (0.31,0.64)
5			0.59 (0.43,0.71)	0.49 (0.31,0.64)
5.5				0.47 (0.29,0.62)

Note. All correlations are significant at the 0.01 level

Table 4-3

Correlations between Fine Motor PDMS-2 Percentile Rank Scores at Different Ages and MABC Test Percentile Rank Scores at Age 7 with 95% Confidence Intervals

Age	4.5	5	5.5	7
4	0.62 (0.47,0.74)	0.45 (0.29,0.61)	0.45 (0.29,0.61)	0.47 (0.29,0.62)
4.5		0.53 (0.36,0.67)	0.48 (0.30,0.63)	0.30 (0.09,0.48)
5			0.62 (0.47,0.74)	0.41 (0.22,0.57)
5.5				0.39(0.19,0.56)

Note. All correlations are significant at the 0.01 level

Table 4-4

Correlations between Gross Motor PDMS-2 Percentile Rank Scores at Different Ages and MABC Test Percentile Rank Scores at Age 7 with 95% Confidence Intervals

Age	4.5	5	5.5	7
4	0.64 (0.5,0.75)	0.48 (0.30, 0.63)	0.51 (0.33,0.65)	0.35 (0.15,0.52)
4.5		0.52 (0.35,0.66)	0.59 (0.43,0.71)	0.47 (0.29,0.62)
5			0.46 (0.28,0.61)	0.37(0.17,0.54)
5.5				0.49(0.31,0.64)

Note. All correlations are significant at the 0.01 level

Table 4-5

Percentile Rank Scores of Children who scored below Test Cut-offs and/or with a Diagnosis or Motor Therapist Intervention

	Percentile Rank Scores below Cut-off					Diagnosis	Intervention
Age	4	4.5	5	5.5	7	8	8
ID							
2	12	10					
5	12	12					
8	10		10	12	15		
9		16			5	n/a	n/a
10		16					
11			13				
21			13		6		
26		16	10				
33						ADD	Brief OT intervention
42	2	1	0	0	0	Benign Rolandic Epilepsy Cognitive delay	Regular OT at school
45		13			5		
59		16					
62					4		
66		16	16		1		OT for motor skills
68		5					Brief OT assessment
71					4		Brief OT assessment
73					8		
75	16	16				Gross Motor Planning Deficiency	Regular OT at school
76	13	7			7		
77					3		
79	13						
82	10						
101			10		9		
102					5		
104	7		12	8	15		
107		16				n/a	n/a
108					4		
120	7		16			n/a	n/a
203			7		5		School OT for writing
209		8					
212					11	n/a	n/a
216			16		5	ADHD, ODD	Brief OT assessment

217	12						
218							Private OT for writing
219	10					n/a	n/a
222	8	10	8		1		
Total	13	15	12	3	19	4	9

Note. N/a indicates family was unable to be contacted

References

American Academy of Pediatrics Committee on Children with Disabilities. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics, 108*(1), 192-196.

Bayley, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *The Journal of Genetic Psychology, 75,* 165-196.

Bayley, N. (1965). Comparisons of mental and motor test scores for ages 1-15 months by sex, birth order, race, geographical location and education of parents. *Child Development, 36*, 379-411.

Bayley, N. (1969). *Bayley Scales of Infant Development* (1st ed.). New York: Psychological Corporation.

Bruninks, R. H. (1978). *The Bruninks-Oseretsky Test of Motor Proficiency*. Circle Pines, MN: American Guidance Service.

Coryell, J., Provost, B., Wilhelm, I. J., & Campbell, S. K. (1989). Stability of Bayley Motor Scale scores in the first year of life. *Physical Therapy, 69*(10), 834-841.

Crawford, S. G., Wilson, B. N., & Dewey, D. (2001). Identifying Developmental Coordination Disorder: Consistency between tests. *Physical & occupational therapy in pediatrics, 20*(2-3), 29-50.

Darrah, J., Hodge, M., Magill-Evans, J., & Kembhavi, G. (2003). Stability of serial assessments of motor and communication abilities in typically

developing infants-Implications for screening. *Early human development, 72*(2), 97-110.

Darrah, J., Magill-Evans, J., Volden, J., Hodge, M., & Kembhavi, G. (2007). Scores of typically developing children on the Peabody Developmental Motor Scales - Infancy to preschool. *Physical & occupational therapy in pediatrics, 27*(3), 5-19.

Dworkin, P. H. (1989). Developmental screening: Expecting the impossible? *Pediatrics, 83*(4), 619-622.

Folio, M. R., & Fewell, R. P. (1983). *Peabody Developmental Motor Scales and activity cards: A manual.* Allen: DLM Teaching Resources.

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

Geuze, R.H., Jongmans, M.J., Schoemaker, M.M. & Smits-Engelman, B.C. (2001). Clinical and research diagnostic criteria for Developmental Coordination Disorder. *Human Movement Science, 20,* 7-47.

Hadders-Algra, M. (2002). Two distinct forms of minor neurological dysfunction: Perspectives emerging from a review of data of the Groningen Perinatal Project. *Developmental Medicine and Child Neurology, 44*, 561-571.

Hadders-Algra, M. (2007). Atypical performance: How do we deal with that? *Developmental Medicine & Child Neurology, 49*(6), 403.

Henderson, S. E., & Sugden, D. A. (1992). *Movement Assessment Battery for Children.* London: The Psychological Corporation.

McCall, R. B., Hogarty, P. S., & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. *American Psychologist, 72*(8), 728-748.

Netelenbos, J. B. (2005). Teachers' ratings of gross motor skills suffer from low concurrent validity. *Human Movement Science, 24*(1), 116-137.

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.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Application to practice.* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Rose, S. A. (1989). Measuring infant intelligence: New perspectives. In M. H. Bornstein, & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives.* (pp. 171-188). Hillsdale NJ: Lawrence Erlbaum Associates.

Siegel, L. S. (1989). A reconceptualization of prediction from infant test scores. In M. H. Bornstein, & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives* (pp. 89-104). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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.

van Waelvelde, H., Peersman, W., Lenoir, M., & Smits-Engelsman, B. C. (2007). Convergent validity between two motor tests: Movement-ABC and PDMS-2. *Adapted Physical Activity Quarterly, 24*(1), 59-69.

Williams, J. & Holmes, C.A. (2004). Improving the early detection of children with subtle developmental problems. *Journal of Child Health Care, 8(1),* 34-46.

CHAPTER 5

CONCLUSION

Summary of Results

Variability within serial motor percentile rank scores of children aged 9 months to 7 years of age was explored in these analyses using cluster analytic techniques, correlation coefficients, and a comparison of test classifications over time. Motor percentile rank scores were obtained using the Peabody Developmental Motor Scales (Folio & Fewell, 1983) at 9, 11, 13, 16 and 21 months of age, the Peabody Developmental Motor Scales – Second Edition (PDMS-2, Folio & Fewell, 2000) at 4, 4.5, 5 and 5.5 years of age and the Movement Assessment Battery for Children – Performance Test (MABC-Test, Henderson & Sugden, 1992) at 7 years of age. Cluster analysis proved to be a useful method to distill the variability exhibited within profiles of gross motor percentile rank scores of typically developing children aged 9 months to 5.5 years into clinically meaningful subgroups. Investigation into the relationship between percentile rank scores and between test classifications of typically developing children on the PDMS-2 from 4 to 5.5 years of age and the Movement Assessment Battery for Children – Performance Test (MABC-Test, Henderson & Sugden, 1992) at 7 years of age provided further evidence of intra-individual variability on standardized tests of motor development.

The results will assist in interpreting longitudinal observations of gross motor performance and contribute to the theoretical foundations of screening for

physical disabilities in young children. Because the results document intra-individual variability in both percentile rank scores and test classifications across time, they provide further evidence in support of the application of dynamic systems theory to motor development. From a dynamic systems theory perspective, the intra-individual variability revealed by these analyses is seen as a normal and essential part of motor development (Piek, 2002).

Clinical Implications

The results of these analyses provide useful information to clinicians who use standardized tests of motor development as part of their practice. Clinicians should expect variability within serial scores of motor development. Several scoring patterns are possible and caution should be used when using an isolated test score to evaluate motor development. Low scores do not necessarily predict future low scores, nor do high scores predict future high scores on standardized tests. Accurate identification of children with mild or moderate motor disabilities presents a challenge to clinicians because typically developing children can be expected to be occasionally classified as having suspicious motor development, thus contributing to high false positive rates on screening tests. Thus, repeated testing is required to monitor motor skill development and test scores and classifications should be interpreted within the context of the child's functional abilities, family concerns, developmental history and clinical observation.

Dissemination of Results

Chapters 3 and 4 will be submitted to peer-reviewed journals within 8 weeks of the thesis defense. A presentation will be made at Peace Country Health's Pediatric Skills Fair and Conference in Grande Prairie, Alberta, in November 2007.

Implications for Future Research

These analyses explored data collected from a volunteer sample of typically developing children. Further research is needed to replicate and validate these results in a different sample of typically developing children. Similar methods should be applied to a sample of children with suspected motor delays and children at risk for motor delays, such as low birth weight infants, and the results should be compared with the above outcomes. Other factors such as fine motor percentile rank scores could be added to the clustering process to determine their effect on cluster membership. The study methods also could be replicated using several different tests of motor development at each assessment age to confirm that the variability is truly a developmental characteristic rather than a function of a specific standardized measure. The cluster analytic techniques used in this study also could be applied to longitudinal data from other developmental domains and the resulting patterns could be compared to the above results.

References

Folio, M. R., & Fewell, R. P. (1983). *Peabody Developmental Motor Scales and activity cards: A manual.* Allen: DLM Teaching Resources.

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

Henderson, S. E., & Sugden, D. A. (1992). *Movement Assessment Battery for Children.* London: The Psychological Corporation.

Piek, J. P. (2002). The role of variability in early motor development. *Infant Behavior and Development, 25*, 452-465.

APPENDIX A

SCRIPT USED FOR PHONE FOLLOW-UP OF PRESCHOOL STUDY PARTICIPANTS

Hello. My name is Karin Eldred. I am a graduate student at the University of Alberta and a member of the research team of the Preschool Study that you participated in with your child (child's name). For my Master's thesis I am looking at some of the preschool data. I am interested in the patterns of children's motor scores. In order to better understand what the scoring patterns mean, it would be helpful for me to know how the children are doing right now in their motor abilities. Are you willing to talk to me about (child's name) current motor skills? It will only take about 5-10 minutes of your time.

If the answer is 'no', Karin will thank them for their time and hang up. If the answer is 'yes', she will ask the following questions:

Questions:

1. Is (child's name) about ____ years old now? What Grade is he or she in?
2. Can you tell me about (child's name)'s motor skills right now?
   What does he/she enjoy doing?
3. Is (child's name) able to do motor skills appropriate for his/her age? (e.g. riding a bike, soccer, shooting baskets, catching a ball) How is he/she doing in physical education?
4. -Does (child's name) participate in organized team sports?
5. Do you have any concerns about (child's name) motor skills?

If the parent has concerns, Karin will listen to them and then tell them that Dr. Darrah will follow up with their concerns.

6. Has (child's name) received any OT or PT intervention?
7. Does (child's name) have a diagnosis of any kind?
8. Is there anything else you wanted to tell me about (child's name)?

Do you have any questions?

On behalf of the Preschool research team I'd like to thank you for your participation in the study. We are presently looking at the results and hope to publish the results this year. Would you like to be notified about where the results are published?

Thanks again.