I, Mohammad Iqbal Maherally, hereby submit this original work as part of the requirements for the degree of Doctor of Education in Curriculum & Instruction.

It is entitled:
The Development and Validation of the Algebra Curriculum Based Measure: A Measure of Preschool Children’s Sorting and Classifying Skills

Student’s name:  Mohammad Iqbal Maherally

This work and its defense approved by:

Committee chair:  Sally Moomaw, Ed.D.
Committee member:  Mary Barbara Boat, Ph.D.
Committee member:  Stephan Pelikan, Ph.D.
The Development and Validation of the Algebra Curriculum Based Measure: A Measure of Preschool Children’s Sorting and Classifying Skills

A dissertation submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Doctor of Education in the Department of Curriculum and Instruction: Teaching and Learning of School Subjects of the College of Education, Criminal Justice, and Human Services by

Mohammad Iqbal Maherally

M.Ed. Mathematics Education, University of Southern Queensland, Australia
B.A. (Honors) Mathematics, University of Delhi, India
February 2014

Committee Chair: Sally Moomaw, Ed.D.
Abstract
The purpose of this study was to develop and validate an assessment tool entitled the Algebra Curriculum Based Measure (ACBM) with the intent of measuring preschool children’s sorting and classifying skills based on one attribute (color, shape, and size) and two attributes (color and shape) simultaneously; and their ability to explain their sorting and classifying strategies. The hypothesis was that five sub-constructs, namely, “Color,” “Shape,” “Size,” “Color and Shape,” and “Explanation,” would significantly contribute to the latent construct sorting and classifying. The ACBM was administered to 120 preschool children enrolled in accredited child care centers in Cincinnati, Ohio, during the 2012-2013 school year. A Confirmatory Factor Analysis (CFA) was used to measure the hypothesized model with raw and square root transformed data. Results indicated a good model fit between the hypothesized and measured models in both cases: $\chi^2 (2, N = 120) = 0.882, p = .643; \text{NFI} = .992; \text{CFI} = 1.000; \text{and RMSEA} = .000$ with raw data, and $\chi^2 (2, N = 120) = 0.749, p = .688; \text{NFI} = .992; \text{CFI} = 1.000; \text{and RMSEA} = .000$ with transformed data. While factor loadings obtained with raw data ranged from .43 to .69, factor loadings obtained with transformed data ranged from .44 to .63. All standardized path coefficients demonstrated both statistical significance at the .001 level (2-tailed) and practical significance ($\beta > .3$) in both cases. The conclusion was that the ACBM is a valid and reliable assessment tool of sorting and classifying for this sample of preschool children.
Dedication

To my wife, Uzma Nooreen, for her unconditional love and constant support. Thank you for always being by my side and for providing me with the strength to never give up and to always persevere during difficult times.
Acknowledgements

First and foremost, I am thankful to GOD for having blessed me with the privilege, courage, and perseverance to pursue and successfully complete this doctoral program in spite of several challenges faced during this remarkable journey.

I would like to extend my thanks to several people, starting with my family. I thank my wife for her constant support and encouragement throughout my studies, and also for helping with data collection while simultaneously collecting data for her own study. I thank my father Aboo Twalib and my mother Hafez for always stressing the importance of education, for having always taken care of my education since I was a little boy, and also for encouraging me to pursue higher studies. I thank my little brother, Mohammad Nabiil, and my sister, Dr. Zaynah Maherally, for providing their support and for their kind words during tough times.

I would also like to thank my dissertation committee. Dr. Sally Moomaw has been acting as my doctoral advisor. Her experience working with young children, has contributed much to my knowledge and understanding about young children. As a committee member, director of Graduate Studies, and my graduate assistant supervisor for over two years, Dr. Mary Boat has always provided her support and advice. She was always available to talk to in spite of her busy schedule. As a committee member, Dr. Stephan Pelikan has contributed much to the initial development of the scoring system of my assessment tool.

I would also like to thank some individuals who have directly or indirectly contributed to my dissertation. As a former committee member and professor for several of the statistics classes that I took, Dr. Wei Pan has made statistics understandable and exciting. I was inspired to undertake a research study involving multivariate statistics after following his lectures. I thank Dr. Charles Moomaw for his valuable assistance in improving the scoring system of my assessment tool. I thank Dr. Teebaye Moradhun, a mathematics content expert with extensive
classroom experience, who took out time from his busy schedule, to verify the scores attributed to the categories of responses for the set of questions requiring the children to explain their sorting strategies. I am grateful to Mohammad Irshaad Abdool, an Information Systems Engineering student, for his precious help with graphics and formatting issues. I am also thankful to Dr. Holly Johnson who was always ready to provide her support and advice at any time. Finally, I greatly appreciate the assistance of the administrators, staff, teachers, parents, and children of the four child care centers where data collection took place.
# Table of Contents

Abstract ........................................................................................................................................... ii
Dedication ....................................................................................................................................... iv
Acknowledgements ............................................................................................................................. v

Chapter 1: Introduction ................................................................................................................. 1
  Background ................................................................................................................................. 3
    Mathematics Learning in the Early Years ..................................................................................... 3
      Hands-on Learning and Mathematics ...................................................................................... 3
      Sorting and Classifying as a Strand of Algebra ...................................................................... 4
    Sorting and Classifying in Mathematics ..................................................................................... 6
    Factors to Consider in Sorting and Classifying Tasks ............................................................... 6

Curriculum Based Measurement as a Way of Measuring Educational Performance .......... 8
  Curriculum Based Measurement in the Content Areas ................................................................. 10

Theoretical Background ............................................................................................................... 12
  Piagetian-Based Theory ............................................................................................................. 12
    Characteristics of Classification ............................................................................................... 13
    Maturation and Physical Experience ...................................................................................... 14
  Constructivist Theory ................................................................................................................ 15
    Piagetian-Based and Vygotskian-based Constructivist Theory ............................................ 16
    Other Perspectives of Constructivism ...................................................................................... 18
  Creative and Playful Learning Theory ....................................................................................... 19

Critical Review of the Literature ................................................................................................. 20
  Foundational Research ............................................................................................................. 20
  Newer Research on Sorting and Classifying .............................................................................. 26
    Research Employing Sorting and Classification Measurement Tools . . . . . . . . . . . . . . 27
    Investigating Young Children’s Behavior Toward Sorting and Classification Dimensions .... 31

Problem Statement ..................................................................................................................... 36
Significance ................................................................................................................................. 39
Chapter 2: Methods

Purpose

Objectives

Variables

Research Questions

Hypothesis

Research Design

Quasi-experimental Design

Sample

Instrumentation

Instrument design

Pilot Study

Overall Pilot Results

Changes Made to the Initial Instrument

Description of the Final Instrument

Scoring of the Final Instrument

Determination of Weightings

Data Management

Data File Structure

Data Quality

Data Coding

Data Entry

Data Checking

Data Cleaning

Confidentiality of Data

Data and Assumption Checks

Missing data

Normality checks

Skewness and Kurtosis

Kolmogorov-Smirnov (K-S) Test

Shapiro-Wilk Test
Histograms……………………………………………………………………………….. 60
Data Transformation……………………………………………………………………. 63
  Square Root Transformation…………………………………………………………. 63
Linearity……………………………………………………………………………………. 65
Correlations Among the Scales on the ACBM………………………………………. 65
Correlation Between Age in Months and Scale scores on ACBM…………………. 65
Outliers……………………………………………………………………………………… 67
  Univariate Outlier Check……………………………………………………………….. 68
  Multivariate Outlier Check……………………………………………………………… 68
Data Analysis Plan………………………………………………………………………… 68
  Analysis of First Research Question…………………………………………………. 69
  Analysis of Second Research Question…………………………………………… 71
Reliability Assessments………………………………………………………………… 71
  Internal Consistency Reliability………………………………………………………. 71
  Construct Reliability…………………………………………………………………… 72
  Inter-rater Reliability…………………………………………………………………… 72
  Test-retest Reliability…………………………………………………………………… 73
Validity Assessments…………………………………………………………………… 74
  Content Validity………………………………………………………………………… 74
  Construct Validity………………………………………………………………………… 75
  Internal Validity………………………………………………………………………….. 75
  Potential Threats to Internal Validity………………………………………………… 75
  Minimizing Potential Threats to Internal Validity…………………………………. 76
  Criterion Validity………………………………………………………………………… 77
Chapter 3: Results…………………………………………………………………………… 78
  First Research Question: What is the Contribution of Each Sub-construct to the Latent
  Construct on the ACBM? ………………………………………………………………. 78
  Evaluation of Model Before Square Root Transformation………………………… 78
  Evaluation of Model After Square Root Transformation………………………….. 79
  Comparison of Models…………………………………………………………………. 79
  Second Research Question: What is the Technical Adequacy of the ACBM in
List of Tables

Table 1. Challenges to Assessing Young Children, along with a Comparison of Traditional versus Curriculum-Based Assessments ................................................................. 11
Table 2. Demographic Information of Participants for Main Study ........................................ 47
Table 3. Scoring Description of the Question “How did you decide which objects to put in your bowls?” ................................................................................................................. 55
Table 4. Comparison of Models Before and After Transformation ........................................ 80
Table 5. Comparison of Goodness of Fit Indices Before and After Transformation .................. 80
Table A1. Table of Specifications for the Algebra Curriculum Based Measure ......................... 117
Table B1. Algebra Curriculum Based Measure ........................................................................ 118
Table B2. Assessment and Scoring Guide ................................................................................ 125
Table B3. Scoring Guide for the Question “How did you decide which objects to put in your bowls?” ................................................................................................................. 130
Table D1. Descriptive Statistics for the Scales of the ACBM and the Total Measure Before Data Transformation ............................................................... 133
Table D2. Kolmogorov-Smirnov Values for the Scales of the ACBM and the Total Measure Before Data Transformation ................................................................. 134
Table D3. Shapiro-Wilk Values for the Scales of the ACBM and the Total Measure Before Data Transformation ................................................................. 135
Table D4. Descriptive Statistics for the Scales of the ACBM and the Total Measure After Data Transformation ................................................................. 136
Table D5. Kolmogorov-Smirnov Values for the Scales of the ACBM and the Total Measure After Data Transformation ................................................................. 137
Table D6. Shapiro-Wilk Values for the Scales of the ACBM and the Total Measure After Data Transformation ................................................................. 138
Table D7. Correlations Among the Scales on the ACBM ......................................................... 139
Table D8. Correlations Between Age in Months and Scores on the ACBM ......................... 140
Table D9. Test for Univariate Outliers on the Scales of the ACBM ........................................ 141
List of Figures

Figure C1. Adjusted Scores (3221) against Raw Scores (1111) at the One-attribute Level…… 131
Figure C2. Adjusted Scores (3211) against Raw Scores (1111) at the Two-attribute Level…… 132
Figure E1. Histograms Representing Total Scores on the ACBM………………………… 143
Figure E2. Scatterplots Comparing Age in Months to Scores on the ACBM……………… 144
Figure E3. Hypothesized CFA Model for the ACBM Before Transformation……………… 145
Figure E4: Hypothesized CFA Model for the ACBM After Transformation……………… 146
Figure E5. Measured Model with Standardized Factor Loadings of the Scales of the ACBM
Before Transformation…………………………………………………………………… 147
Figure E6. Measured Model with Standardized Factor Loadings of the Scales of the ACBM
After Transformation…………………………………………………………………… 148
Chapter 1: Introduction

According to the Second Handbook of Research in Mathematics Teaching and Learning (Lester, 2007), early mathematics education has received considerable attention. Clements and Sarama (2007) point out seven reasons that have increased the attention in mathematics education for young children. These are: (a) increasing attendance of children in early child care and education programs; (b) increased acknowledgement of the importance of mathematics (Kilpatrick, Swafford, & Findell, 2001); (c) unfavorable performance of American students in mathematics, including early mathematics, when compared to other nations; (d) low mathematics achievement of American students living in economically impoverished urban societies (Griffin, Case, & Siegler, 1994; Saxe, Guberman, & Gearhart, 1987; Siegler, 1993; Ysseldyke et al., 2003; Williams, 2005); (e) the shifting of the research focus from young children’s learning ability and knowledge of mathematics (Piaget, Inhelder, & Szeminska, 1960; Piaget & Szeminska, 1952; Thorndike, 1922) to theories related to young children’s intrinsic or developing aptitudes during the early years of life (Baroody, Lai, & Mix, 2006; Clements, Sarama, & DiBiase, 2004; Gelman & Gallistel, 1978; Perry & Dockett, 2002); (f) evidence showing that quantitative and numerical proficiency during the years prior to first grade is a better predictor of future mathematics achievement than intelligence and memory tests (Krajewski, 2005); and (g) prior research demonstrating that knowledge gaps occurred mostly because of the disconnection between children’s casual and inherent knowledge (Ginsburg & Russell, 1981; Hiebert, 1986) and school mathematics.

According to Clements and Sarama (2007), prior findings (National Center for Education Statistics, 2000) indicate that young children’s basic and fundamental mathematical knowledge begins from the early developmental years of life until the age of 5. For instance, preschool
children explore shapes, spatial relationships, patterns, and counting by getting involved in free play. In fact, this process reflects children’s fundamental everyday experiences that build up the inherent, tacit, and abstract basis for more advanced mathematics. Afterwards, they symbolize these ideas by creating mathematical activities with objects.

The National Association for the Education of Young Children (NAEYC) and the National Council of Teachers of Mathematics (NCTM), the governing professional organizations in their respective fields, have joined forces to amalgamate strong and positive characteristics of pre-primary and primary program models to bring to light crucial factors needed for an outstanding mathematics education. As the NAEYC (2010) indicate, the partnership of these two organizations renders the pre-primary mathematics curricula more specific and emphasizes more hands-on experiences by including more manipulative materials in the primary programs. The intent of incorporating pre-primary and primary programs was to strengthen mathematics education in early childhood.

The position statement of NAEYC emphasizes the process of mathematics by stating “The goal of the mathematics program is to enable children to use mathematics through exploration, discovery, and solving meaningful problems” (Bredekamp 1987, p. 71), and NCTM focuses on the content of mathematics by stating “The curriculum is mathematically rich…with concepts and procedures…It provides a common foundation of mathematics to be learned by all students” (NCTM 2000, pp. 3-5). In regards to these two statements, the joint position statement of these two professional organizations “affirm that high-quality, challenging, and accessible mathematics education for 3-to 6-year-old children is a vital foundation for future mathematics learning. In every early childhood setting, children should experience effective, research-based curriculum and teaching practices. Such high-quality classroom practice requires policies,
organizational supports, and adequate resources that enable teachers to do this challenging and important work” (National Association for the Education of Young Children, 2010, pp. 1).

**Background**

**Mathematics learning in the early years.**

Young children learn mathematical concepts long before they enter formal schooling (Moomaw, 2011), where they are taught mathematics through an established curriculum. Along the same lines, as Bullard (2010) states, in their initial learning stage, children informally learn the mathematical content and acquire procedural skills in learning mathematics. In this regard, Moomaw (2011) puts forward that the mathematical knowledge acquired informally enables children to link mathematics to their everyday lives and to develop their thinking abilities in organizing the world around them. She further indicates that since young children are involved in playing with mathematics, the informal mathematical knowledge acquired helps them in developing their play.

**Hands-on learning and mathematics.**

Even though young children are interested in playing with mathematics and show an appreciation toward this discipline, many elementary school children indicate an attitude of dislike toward mathematics (Moomaw, 2011). In this regard, Moomaw (2011) states that due to children’s lack of enjoyment and confidence about learning mathematics in early elementary school, it is of utmost importance to pay particular attention to children aged 3 until they start formal schooling. Seen from this perspective, Moomaw indicates that the learning of mathematics at the preschool and kindergarten level starts through play-based activities. She notes that natural mathematical situations are created when young children interact with each other. She further points out that since young children are very sensitive to instruction, and they
do not generally respond to questions that are of no interest to them, it is important to integrate mathematics into play-based activities. According to her, questions and models that arise in a natural way will tend to get young children’s attention and will prompt them to reflect on new mathematical concepts and relationships.

As Copley (2010) put forward, young children learn about quantity, relationships, and symbols intuitively. She notes that the basis of young children’s construction of mathematical ideas emerges from their environmental exposure, by interacting with other children and adults, and from their everyday observations. The author further indicates that by listening to young children’s ideas, early childhood teachers can provide children with the necessary materials and create an environment that will promote the development of mathematical concepts. An argument put forward by Kamii, Lewis, and Kirkland (2001) is that the use of manipulatives significantly impact children’s thinking in problem solving situations. In this regard, they note that the use of counters is a type of manipulative that generally helps young children in learning mathematics efficiently. For instance, they indicate that young children are confused when introduced to addition and subtraction problems, and thus, as a result, teachers can provide children with counters as an aid and have them pretend, for example, adding four cookies to four other cookies.

**Sorting and classifying as a strand of algebra.**

An important component in the Algebra Standard for Grades PreK-2 (NCTM, 2000) is that PreK-2 students should develop an understanding of patterns, relations, and functions. NCTM stresses that in order to understand relationships, all PreK-2 students should learn to “sort, classify, and order objects by size, number, and other properties” (p. 90). In this regard, Moomaw (2011) notes that by sorting and classifying objects, children develop their understanding about mathematical relationships of same, different, and sharing a particular
attribute. Therefore, the development of sorting and classifying skills in young children becomes critical in promoting young children’s algebraic reasoning skills.

Moomaw (2011) further notes that, before starting to arrange objects into patterns, children must have the ability to focus on particular characteristics of items. Based on these characteristics, children will be able to sort and classify items into groups, and organize them into an ordered system, for example, from smallest to largest. Seen from this perspective, Moomaw puts forward that sorting and classifying, as well as patterning, are significant strands of the Algebra standard at the preschool and kindergarten level. She also indicates that, as children transition to elementary and middle school, they apply the knowledge obtained during their preschool years (about creating relationships with concrete objects) to create relationships with numbers. For example, they use preschool skills and knowledge to sort whole numbers and fractions into different groups, or to identify numerical patterns.

Patterns can be regarded as relationships comprising repetition of an element. For example, in the case of the number system, every odd number is followed by an even number, every whole number is one more than the previous number, and so on. Some common patterns involving numbers include the following: arithmetic sequences (for example, add 4 to the last number of a sequence resulting in a pattern such as 4, 8, 12, 16, 20, …; or subtract 3 each time to result in a pattern such as 21, 18, 15, 12, 9, …); geometric sequences (for example, multiply by 3 each time resulting in a pattern such as 3, 9, 27, 81, 243, …); patterns involving triangular numbers (that is, an arrangement of dots which form a triangle, resulting in the pattern of 1, 3, 6, 10, 15, …); square numbers (squaring whole numbers resulting in the pattern of 0, 1, 4, 9, 16, …); and cube numbers (cubes of the counting numbers resulting in the pattern of 1, 8, 27, 64, 125, …).
Sorting and classifying in mathematics.

Based on a review of how scholars define sorting and classifying, it appears that different meanings may be assigned based on how children group items. In this regard, different authors hold different meanings for sorting and classifying. In a broad sense, sorting may be described as a way of creating groups or sets, whereas classifying may be referred to as the type of language being used to illustrate these groups or sets (Lawson, 2006). This is the definition used in this study. The meanings of these terms may be based on the contexts also. Based on various contradictory definitions, the terms sorting and classifying are used interchangeably.

Sorting and classifying represent foundational skills that an individual needs to develop and acquire because s/he uses these skills in his/her everyday life (Lawson, 2006). For instance, everyday experiences of sorting and classifying are used when one is doing dishes or going to the grocery store. Furthermore, the author indicates that the ability to classify items into groups and to recognize relationships within and among distinct groups enhances logical and accurate thinking. As a matter of fact, sorting and classifying skills are essential aptitudes that allow young children to organize the world around them (Sousa, 2008). Similarly, Lawson (2006) puts forward that sorting and classifying abilities help in promoting students’ development in the sense that these skills enable them to learn about organizational skills, to learn about how to think analytically, and learn how to articulate their way of thinking accurately. These skills generally start to develop around the age of three, and represent components that are essential in a child’s development of his/her understanding of the real world around him/her (Sousa, 2008). Along the same lines, Lawson notes that the ability to classify into groups and to recognize relationships within and among distinct groups enhances logical and accurate thinking.

Factors to consider in sorting and classifying tasks.
Platz (2004) indicates that tangible objects should be used when engaging young children in sorting and classifying tasks because real objects make more sense and add more meaning to young children, thus enhancing young children’s logical thinking skills (Charlesworth, 2000). Essentially, the sequence of administrating sorting and classifying tasks is fundamental for developing logical thinking in young children (Platz, 2004).

When engaging young children in sorting and classifying tasks, it is important to consider characteristics of both children and the activities in which they are involved. Factors such as age of the children, their perceptions about things around them, how they use tactile or kinesthetic tasks with real objects, how they build up information, the quantity of objects used during the activities, the children’s communication skills or mathematical talking, and the fun side of the activities employed are considered to be essential in developing efficient and successful sorting and classifying activities for young children (Platz, 2004).

As Wadsworth (1971) puts forward, age is a critical factor to consider when implementing sorting and classifying tasks with young children. For instance, a 3-year-old might find a sorting task challenging, but it might not be the case for a 4½-year old (Platz, 2004). Therefore, the level of challenge of sorting tasks varies depending on children’s age. Young children’s perceptions are another fundamental factor to take into account when involving them in sorting and classifying tasks (Russell, 1956). The basis of the understanding of young children about sorting and classifying tasks lies in the way they perceive things (Platz, 2004). As Clements (2001) indicates, the way young children build up information differs from how adults do. When children are involved in manipulation, messages are processed in a more effective and promising way to children’s brains (Zaslavsky, 2001; Richardson, 1999), thus enabling them to learn mathematical concepts, including sorting and classifying, more efficiently. Furthermore,
the use of real and tangible objects is important when involving children in sorting and classifying tasks (Platz, 2004).

The quantity or number of objects exposed to children is another critical factor in evaluating their sorting and classifying ability and in raising and maintaining their interest in the activities. For instance, if a 3-year-old is exposed to too many objects to be sorted and classified, s/he will be more likely to demonstrate lack of interest toward sorting and classifying tasks (Platz, 2004). As such, exposure to a large number of objects to be sorted and classified by too many attributes might have an adverse effect on the children’s performance and the activities as a whole. The next essential factor in assessing children’s sorting and classifying is their communication or mathematical talking. Children need not only sort and classify, but they also need to be able to explain the basis of their sorting and classifying strategies. In other words, they should be able to communicate their thinking about their sorting and classifying strategies. As Corwin (1966) notes, children can easily be confused about some phrases when being engaged in sorting and classifying activities, in which case, they should be provided with the chance to communicate their actions in order to illuminate mathematical terms and phrases. The last underlying factor affecting assessment is the enjoyment aspect of the activities employed with the children. Getting children involved in fun activities and play relating to sorting and classifying tasks will enhance effective learning of sorting and classifying abilities (Platz, 2004).

**Curriculum-Based Measurement as a Way of Measuring Educational Performance**

The measurement of change is the most crucial and inherent constituent of education (Espin, Shin, & Busch, 2005). In their review, the authors note that teachers can more accurately assess student learning and the impact of instructional strategies employed on that learning by
measuring change in performance. Despite this fact, measurement of change has not been the main emphasis of educational assessment.

Espin et al. (2005) point out that measurement of change has been a minor focus in education due to the complications encountered in measuring change in performance, including a dearth of statistical methods for manipulating several data points (Willett, 1989) and a significant shortage of the availability of assessment tools that could be used to generate repeated measures within short periods of time (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1994). As such, curriculum-based measures have been created with the purpose of measuring specifically the change in students’ progress. It should be noted that curriculum-based measures are strongly supported by a body of research in terms of validity and reliability.

Review of research by Espin et al. (2005) shows that, over 25 years of research, curriculum-based measurement (CBM) has been proven to be a valid and reliable indicator of students’ performance at the elementary level in the areas of reading, mathematics, spelling, and written expression. The authors further point out that curriculum-based measurements are data collection tools that are specifically designed to inform teachers about students’ level of achievement as well as the degree to which instructional strategies are employed to impact students’ progress. In this regard, Deno (1985) indicates that the measures developed within a CBM are effective, clear, understandable, and inexpensive, and enable repeated measurements of student performance progressively. As Marston (1989) notes, when CBM indicators are correlated with other criterion measures, values generally range from .60 to .90, and test-retest and other forms of reliabilities generally yield values above .80. Treatment validity of CBM measures has also been established (Espin et al., 2005). A large body of research, as reviewed by Espin et al. (2005), reveals that when CBM measures are used by teachers to assess and modify
their instructional strategies, the outcomes show students’ achievement (Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, & Hamlett, 1989a, 1989b, 1989c; Fuchs, Fuchs, Hamlett, & Allinder, 1991; Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Fuchs, Fuchs, Hamlett, & Stecker, 1990; Stecker & Fuchs, 2000; Wesson et al., 1988). Research by Compton (2000) and Shin, Deno, and Espin (2000) indicate that the use and application of statistical techniques, such as hierarchical linear modeling (HLM), to the analysis of CBM allows the creation of student growth curves, which are used to interpret and answer questions regarding the relationships that exist between student progress and instructional variables.

Curriculum based measurement in the content areas.

The review of research conducted by Espin et al. (2005) indicates that early research on the development of curriculum based measurements in the content areas has been carried out by Tindal and Nolet (Nolet & Tindal, 1993, 1994, 1995; Tindal & Nolet, 1995, 1996). They analyzed critical thinking skills such as concept explanation and representation of facts that are important components required for the understanding and application of content area data, from which measures are generated to illustrate these thinking skills. As Espin et al. (2005) point out in their review, the measures show some degree of appropriateness when student learning of a particular unit of study was concerned, but revealed that these measures were not very useful as they did not illustrate performance across study units (Tindal & Nolet, 1995).

Assessing young children’s performance reliably is a challenging task due to the social and emotional components related to their developmental stage (Moomaw, 2008). In this regard, she further indicates interest in curriculum based measures on the part of teachers and researchers since CBMs can be easily administered and used with children at this age. Table 1 below, taken from Moomaw, shows the difficulties associated with young children’s assessment.
and how CBMs can be used to address these difficulties by incorporating a comparison between traditional assessment and curriculum based assessments:

Table 1.  

*Challenges to Assessing Young Children, along with a Comparison of Traditional versus Curriculum-Based Assessments*

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Traditional Assessment</th>
<th>Curriculum-Based Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>Because preschool children are developing and asserting autonomy (Erikson, 1950), issues of compliance often arise during assessment procedures.</td>
<td>The measures can be developed to align with young children’s interests.</td>
</tr>
<tr>
<td>Attention</td>
<td>Preschool children typically have much shorter attention spans than school-age children. This affects the length of time that can be devoted to assessment.</td>
<td>Curriculum-based measures are often much shorter than other types of assessment measures.</td>
</tr>
<tr>
<td>Trust</td>
<td>Since preschool children are also establishing trusting relations (Erikson, 1950) outside of their homes, fear about working with an assessor is a concern.</td>
<td>The measures can be scripted and administered by classroom teachers who have training on scoring procedures.</td>
</tr>
<tr>
<td>Development &amp; Maturation</td>
<td>Maturation and learning occur quickly during the preschool years; therefore, ongoing assessment is critical for educational planning.</td>
<td>Curriculum-based measures are designed for ongoing assessment.</td>
</tr>
<tr>
<td>Alignment to Curriculum</td>
<td>Assessment information is often not related to standard preschool experiences and curricula.</td>
<td>Assessment is directly aligned to curriculum.</td>
</tr>
</tbody>
</table>

Theoretical Background

The present study is based on three theories: Piagetian theory, constructivist theory, and the creative and playful learning theory (Kangas, 2009). These three approaches are further discussed in the following sections in terms of how the study reflects their features and is guided by their theoretical components.

Piagetian-based theory.

Piaget conducted seminal research on children’s ability to sort and classify. As classic research by Piaget has shown, children do not grasp the idea that an item can belong both to a class and a subclass at the same time (Inhelder & Piaget, 1964; Piaget, 1952). Importantly, these scholars relate the formation of classification to language development. In fact, Piaget (1951) demonstrates that when children are involved in learning how to talk, they do not necessarily grasp the concept of collective classifications in their native language; as such, language helps children develop the formation of classes. As far as classes are concerned, they may be defined by both their “intension” and by their “extension” (Inhelder & Piaget, 1964). As the authors define, the inclusion of a class represents a set of the common properties that members of that class share and that differentiate them from other classes. The extension of a class represents the group or set of members comprising that class. One important criterion that coordinates the intension and extension of a class is the understanding of class inclusion (Overton & Lerner, 2010). From a Piagetian theory perspective, the concept of class inclusion results from the grouping of a family of classes, $A$, $A'$, and $B$ such that $B = A \cup A'$ and $A \cap A' = 0$. This follows that since $A$ and $A'$ are disjoint, $A'$ is therefore the complement of $A$ with respect to $B$. For example, the formula can be represented by putting roses (characterized by $A$) and tulips
(characterized by \( A' \)) together into two subclasses, and the operation of putting these two subclasses together to form a superordinate class, \( B \) (which represent flowers).

The operational existence of classes, as put forward by Inhelder and Piaget (1964), are based on the following criteria: (a) the subject can intensively define a class as a more general class and as one or more specific differences; (b) the subject can command their extension in relation to the framework of inclusion, as demonstrated by his/her mastery of the quantifiers “all,” “some,” “one,” and “none.” In regard to the formation of classes, the concepts of the relations of resemblance, complementarity, and class membership play an important role in children’s classificatory behavior. Relations of resemblance can be represented by the following example: based on the common property that two subclasses share, for instance, roses and tulips, they both belong to a superordinate class of flowers, although they differ in terms of being two different types of flowers. Complementarity can be represented by the following example: putting roses and tulips together into two subclasses respectively gives rise to a superordinate class, that is, flowers, where a tulip is the complement of a rose or vice-versa. Finally, class membership can be represented by the following example: a subclass, roses, belongs to a superordinate class, flowers.

**Characteristics of classification.**

As an important note, Inhelder and Piaget (1964) come up with ten important characteristics of classification. These are as follows:

- There are no isolated elements, that is, all elements must be classified. This is to say that if an element \( x \) is unique of its kind; it must give rise to its own specific class, i.e. \( (x) \in (A_x) \).
• There are no isolated classes, that is, each particular class $A$ constituted or satisfied by the property $a$, denotes its complement $A'$ within the set $B$ such that $A \cup A' = B$.

• A class $A$ constitutes all of the elements satisfying the property $a$.

• A class $A$ constitutes only elements satisfying the property $a$.

• All classes with the same rank are considered to be disjoint, i.e. $A \cap A' = \emptyset$ or $An \cap Am = \emptyset$.

• A complementary class $A'$ possess its own characteristics $a_x$, which its complement $A$ does not possess.

• A class $A$ (or $A'$) is constituted in each higher ranking class containing all of its elements, starting with the closest set, $B$ such that $A = B - A'$ (or $A' = B - A$) and $A \cap B = A$. in other words, this implies that “all” of $A$ represent “some” of $B$.

• Extensional simplicity: inclusions (in characteristic 7) are decreased to a minimum and that are consistent with intensional properties.

• Intensional simplicity: criteria that are similar (e.g. colors) differentiate classes of the same rank.

• Symmetrical sub-division: if a class $B_1$ is split into $A_1$ and $A'_1$ and the same criterion is applied to $B_2$, then, in a similar fashion, $B_2$ should be split into $A_2$ and $A'_2$.

Maturation and physical experience.

Based on Piagetian theory, children are exposed to two factors of development: maturation and physical experience. Initially, traditional learning theorists thought that maturation and learning were two separate processes (Gallagher & Reid, 1982). In regards to this, Piaget argues that a child is unable to learn from particular experiences unless his/her cognitive structures have reached a certain level of maturity that will allow him/her to understand
the experience. He puts forward that a factor called equilibration acts as a balance between the effects of maturation and learning. As far as physical experience is concerned, Piaget indicates that there are three types of physical experience: exercise, physical experience proper, and logico-mathematical experience. In the case of exercise, children are exposed to repeated activities that mostly improve their practice instead of enabling them to acquire new knowledge. On the other hand, Piaget uses the term physical experience proper to refer to experiences in which children perform activities to acquire new knowledge by manipulation of objects that lead the child to discover the inherent characteristics of the objects. Finally, the third type of experience is what Piaget called the logico-mathematical experience, in which the child is involved in activities that enable them to gain knowledge indirectly from these experiences by manipulating objects and reflecting on these activities. In addition to object manipulation as an important source of developing logical cognition, constructivist theory suggests the importance of language or perception in enhancing children’s classifying skills since it raises human infants’ perceptions and awareness about the objects they constitute and classify (Piaget, 1969). As such, language significantly contributes to children’s classifying skills (Piaget, 1951).

**Constructivist theory.**

Constructivism is a theory relating to knowledge and learning that characterizes both the meaning of “knowing” and how people obtain knowledge (Fosnot, 2005). According to Fosnot (2005), learning from the constructivist viewpoint is considered as a self-regulatory process that occurs due to disagreement between existing personal models of the world and new understanding. This results in the formation of new representations and models of reality through meaning-making attempts with cultural tools and symbols, and by additionally discussing such
meaning through collaborative social activities, conversations, and discussions in communities of practice.

As a psychological theory, constructivism arises from the expanding field of cognitive science (Fosnot, 2005). For a long time, constructivism has been regarded as a valuable theory of learning in which learners construct mental representations by interacting in suitable types of active cognitive processing while learning which in turn results in the development of cognitive representations (Mayer, 2001, 2008, 2009). In agreement with the constructivist learning theory, teaching should encourage proper cognitive processing during learning (Mayer, 2001, 2008).

As Chaillé (2008) points out, constructivism is a theory of learning that assumes that children build knowledge based on the connection between their own ideas and that of their encounters in the social and physical world. According to the author, children have their own broad background and ideas before each encounter. In the process of getting involved in the encounters and testing their ideas, they construct new theories and ideas, which, in turn, give rise to motivation. In this regard, Chaillé indicates that motivation for learning is considered as natural from the constructivist viewpoint. This is because children are always attempting to make sense of the world and develop their ideas. Thus, from a constructivist perspective, Chaillé indicates that children do not learn when knowledge and information is passed on to them. Moreover, they are not enthusiastic to learn when provided with extrinsic resources like reinforcements and rewards. In fact, learning that occurs as a result of providing reinforcements and rewards is an example based on the behaviorist theory rather than the constructivist theory.

**Piagetian-based and Vygotskian-based constructivist theory.**

notes, these two famous scholars were involved in the development of constructivist theories. Piaget explains the learning process as follows:

- By schemes (organization of information on the way things work)
- By assimilation (allocating new information into schemes)
- By accommodation (altering existing schemes or developing new ones)

From Piaget’s point of view, the motivation for learning arises from the likelihood of the learner to adjust to his environment, thus, establishing a state of equilibrium between schemes and the environment. New learning occurs due to ongoing experiences among existing schemes, assimilation, accommodation, and equilibrium.

In addition, Ozer (2004) attests that discovery is the basis of Piaget’s developmental theory of learning and constructivism; therefore, the best learning environment can be created by allowing children to build knowledge that is significant to them. A constructivist classroom is one that caters to various activities that challenge students to accept individual differences, boost their eagerness to learn, discover new ideas, and build their own knowledge. In line with Piaget’s belief about constructivism, Ozer believes that constructivism in education asserts that human beings can better understand the information they have built up themselves.

Ozer (2004) further notes that, based on constructivist theories, learning is considered as a social progress that comprises language, real world circumstances, and encounters and cooperation among learners. In this regard, the author indicates that Vygotsky believes that constructivism is centered on a social aspect due to the emphasis on culture and social context. Based on Vygotskian theory, learning occurs in collaboration with development, and children undergo cognitive development in the context of socialization and education. Children’s perceptual, attention, and memory abilities are altered by fundamental cognitive tools related to
culture, such as history, social context, traditions, language, and religion. In order for learning to take place, children initially get involved with the social environment on an interpersonal level, after which they internalize this involvement. Their previous ideas, as well as new encounters, have an impact on the children, who then build new ideas. Therefore, for Vygotsky, the zone of proximal development (the difference between what a learner can do without help and what s/he can do with help) proposes that cognitive development is restricted to a certain range at a specific age. However, social experiences, such as help from a mentor, allow students to understand concepts and schemes that they do not know on their own.

**Other perspectives of constructivism.**

Seen from other lenses, different authors have different perspectives about constructivism. For instance, constructivism is usually seen as different from behaviorism, which takes into account external reinforcements in order to control learning (Schwartz, Lindgren, & Lewis, 2009). These authors believe that constructivism is a wide vision of learning rather than a simply way to teach. It provides the ability to take into account students’ capacities to build new knowledge at a time when they are not exposed to teaching, more precisely, at a time when there is no control over specific instructional changes. In addition, Schwartz et al. (2009) indicate that it is possible for direct instruction to be effective only if people have an appropriate amount of prior knowledge that leads them to build new knowledge simply from what they are being told or shown.

Instructional methods that are based on constructivist theory, that is, methods that lead students to discover information by themselves instead of being presented with the information, have provided education researchers with a favorite teaching technique for many decades (Sweller, 2009). According to Sweller (2009), there is no theorist who disagrees to this view of
learning. The author further notes that there are several aspects of constructivism that cannot be disagreed upon. An example is that people must build mental representations of the outside world that they can make use of in order to function in that world. Seen in this light, all learning is considered to be fundamentally constructivist.

**Creative and playful learning theory.**

The creative and playful learning theory literally relates to creative and playful learning occurring in a playful learning environment (PLE) setting, which is referred to as (a) learning that allows, encourages, and enhances creativity of the learner and knowledge co-creation, (b) learning that occurs through the content design of the PLE by using modern technology, and (c) learning that takes place through a range of playful activities involving concrete and hands-on artifacts within the PLE (Kangas, 2009). Other researchers (Säljö, 2004; Vygotsky, 1978; Wells & Claxton, 2002) note that learning is related not only to academic achievement but also to the person in terms of his or her body, mind, spirit, and culture. In fact, playfulness is believed to impact learning positively at different school levels as well as learning at work (Sawyer, 2006). While some researchers (Paavola, Lipponen, & Hakkarainen, 2004) describe creativity as the construction of creative knowledge, where learning occurs through the use of technology and the design of artifacts, games, or media products, other scholars (Cremin, Burnard, & Craft, 2006; Egan, 2005) describe creativity as the use of imagination, which is the ability to think of many possible things. In regards to the creative construction of knowledge, Paavola et al. (2004) stress that knowledge does not occur solely due to traditional learning but has also a tendency to develop from the composition of artifacts. Essentially, children in the present study are mostly involved in playful activities where they construct creative knowledge by sorting and classifying foam objects according to color, shape, size, and color and shape simultaneously.
Critical review of the literature

Preschool age is a developmental period, during which children acquire basic and vital aptitudes that are later developed and established during the course of the elementary school years (Schneider & Sodian, 1991). The following sections describe both historical and more recent research related to children’s concepts of sorting and classification.

Foundational research.

Baker-Ward, Ornstein, & Holden (1984) found that when asked to sort items, 4-year-olds depend on, for the most part, naming and visualizing items that should be remembered. For this reason, 4-year-olds need clear sorting instructions (Sodian, Schneider, & Perlmutter, 1986). Prior research in the early years of young children’s sorting ability revealed their grouping inability based on class membership (Sigel, 1953; Thompson, 1941). Damrauer (1976) put forward that younger children sort and group objects together conforming to concrete settings, create disconnected groups, or create groups on the basis of common stimulus attributes like color and shape. In line with the correlation between children’s age and their sorting abilities, Piaget's theory of intellectual development (Ginsburg & Opper 1969; Inhelder & Piaget, 1964; Piaget, 1967) suggested that children who are in the preoperational stage (under 7 years old) cannot sort objects by more than one attribute at a time (Watson, Hayes, & Vietze, 1979).

Early studies (Gopnik & Meltzoff, 1987; Langer, 1982; Nelson, 1973; Ricciuti, 1965; Starkey, 1981; Sugarman, 1983) investigated how children’s sorting and classifying strategies of objects changed to a large extent from 15 months of age to 21 months of age when they were exposed to a mixed collection of objects to be sorted. At about the age of 1, children sort and group objects by one dimension, for instance, they might group all balls in a single collection (Ricciuti, 1965; Starkey, 1981; Sugarman, 1983). At about 15 months, they might indicate
exhaustive serial touching (Mandler & Bauer, 1988) where they touch all objects in one group followed by touching all objects in another group (Gopnik & Meltzoff, 1987; Nelson, 1973; Ricciuti, 1965; Sugarman, 1983). It is by 18 months that children start to sort and classify objects by creating distinct groups. For example, they will group all boxes in one pile and all balls in another (Gopnik & Meltzoff, 1987; Nelson, 1973; Ricciuti, 1965; Sugarman, 1983). In this regard, it is important to note that in exhaustive grouping, children move objects to separately create defined groups, with each group consisting of objects based on a different dimension (Gopnik & Meltzoff, 1992). Along the same lines of research, Vygotsky (1962) measured children’s logical understanding about the relationship between objects and object characteristics by investigating their sorting abilities. This classical research, known as “Vygotsky blocks,” exposed children to an array of blocks based on the different dimensions of color, shape, and size. Children were found to group the blocks randomly instead of sorting according to the above mentioned dimensions.

Other research (Langer, 1980) showed that, when exposed to an array of objects, 6-month-olds tend to group both the identical and different objects together, rather than grouping the similar ones together. At the age of 12 months, they tend to sort similar objects together based on one attribute and continue to do so as they grow up (Langer, 1980, 1986; Nelson, 1973; Ricciuti, 1965; Starkey, 1981; Sugarman, 1983). Similar studies on children’s learning strategies of classification tasks emphasized (a) how they create a class (Bruner & Olver, 1963; Inhelder & Piaget, 1964; Lovell, Mitchell, & Everett, 1962; Vygotsky, 1962), (b) how they change dimensions when categorizing (Heald & Marzolf, 1953; Lovell et al., 1962), and (c) how they compare the size and contents of distinct dimensions (Dodwell, 1962; Hyde, 1959; Inhelder & Piaget, 1964; Piaget, 1952). In this regard, Inhelder and Piaget (1964) provided an explicit step-
by-step explanation of how children learn about inclusion. They argued that classification starts when the child sorts and classifies two similar objects based on one attribute, which is known as resemblance sorting. As the child grows, his/her sorting and classifying abilities develop progressively when s/he sorts and classifies more than two similar objects based on one attribute, which is known as consistent sorting. Eventually, s/he sorts and classifies all similar objects based on one attribute, which is also known as exhaustive sorting. This hierarchy is embedded in the scoring part of the assessment tool employed in the present study.

Other early findings related to the developmental trends in classification reveal that older children have the ability to classify objects based on properties of the selected dimensions (Bruner, Olver, & Greenfield 1966; Inhelder & Piaget 1964; Vygotsky 1962). Younger children, who do not classify by dimensions, proceed with classification tasks by overall similarity (Kemler, 1982). To illustrate, prior research by Smith and Kemler (1977) exposed small geometric shapes of different colors and sizes to kindergarteners, second graders, and fifth graders. The results revealed that the fifth grade children sorted and grouped the objects based on dimensional relationships, whereas, kindergarten children did so by overall similarity relations. Interestingly, the second grade children did both. Seen from this light, Ward and Vela (1986) concluded that, since young children are more inclined to classify even disconnected stimuli based on the overall similarity principle, their intuitive world is built on overall similarity relationships and the concept of development includes the betterment of an explicit study of stimuli related to their dimensional factors. However, the author put forward that children might not necessarily be more nor less holistically- or similarity- inclined than adults. As Odom and colleagues (Aschkenasy & Odom, 1982: Odom & Cook, 1984) advocated, young children might identify dimensional relationships but be less responsive to those relationships than older
children and adults might be. These researchers showed that when young children regarded a dimension as highly important, and when they were involved in stimulus manipulations to reflect this importance, even 4-year-old children classified based on several dimensions (Ward & Vela, 1986). Therefore, it can be deduced that even young children have the ability to identify and use some dimensional relationships (Ward, 1980), and young children do not always tend to give more holistic responses than older children and adults (Shepp, Burns, & McDonough, 1980; Smith, 1981).

Getting children involved in activities requiring them to sort and classify by one attribute at a time prepares them to engage in more complex experiences such as sorting and classifying by more than one attribute at a time, thus enhancing their understanding about class inclusion (Kofsky, 1966). As the author noted, the child gradually starts to identify objects that do not necessarily belong to different classes; s/he then attempts to create different groups of objects based on one particular attribute, and on another attribute and so on, which is also known as horizontal classification. As the child’s logical aptitudes develop, his/her strategies of selecting dimensions becomes more complicated, where s/he will select single attributes to sort and will try more than one attribute to create subsequent groupings of objects, which is known as hierarchical classification.

Some results have shown that young children aged between 2 and 5 years old do not sort and classify geometric stimuli based on likelihood or similarity (Inhelder & Piaget, 1964). On the other hand, different results have revealed that 2- to 5-year-olds base sorting decisions on the similarity between objects (Denney, 1972a). The author put forward that the conflicts found between her findings and that of Inhelder and Piaget might have occurred due to different attention given to early childhood education and programs in the two different time frames in
which the studies were conducted (Denney, 1972b). She further stressed that preschool children in the United States tended to learn about sorting and classifying based on both similarities and differences at a much younger age than the children studied by Inhelder and Piaget. She also indicated that without training on sorting and classifying objects according to similarities and differences, children would be more likely to group objects based on differences in classification tasks instead of grouping similar objects together. Most probably, if the child were not given any kind of training or specific directions of what and how to sort and classify, s/he would be more apt to do so based on other present stimuli (Denney & Acito, 1974). Interestingly, Frith and Frith (1978) added that people of all age groups have the ability to identify different objects based on similarities. Along the same line of research as Denney, other researchers (Mandler & Stephens, 1967; Vygotsky, 1934, 1962; Inhelder & Piaget, 1964) found that when 2- to 5-year-olds are not given any specific direction for classification tasks, they might sort objects by shifting from one sorting dimension to another sorting dimension. Children aged between 5 years old and 8 years old sort by one attribute at a time, if not provided with any specific instructions. Eventually, after the age of 8, children have already acquired the skills and abilities to sort and classify by an array of attributes at the same time.

Around the beginning of the 1980s, it was a common belief among researchers that young children often did not perform well on experimental activities that assessed their level of understanding of logico-mathematical concepts, such as the science of reasoning related to classification (Thornton, 1982). Inhelder and Piaget (1964) had maintained that young children usually failed in these types of tests due to the limited level of their intelligence structures that are vital for achievement. Several studies challenged this idea by showing that young children demonstrated an acceptable level of performance in some forms of tasks related to Piaget. For
instance, 5-year-olds, who, typically, failed in Piagetian class inclusion questions, performed better when there was a change in the procedures required to complete the tasks (Donaldson, 1978; McGarrigle & Donaldson, 1974) or when part-whole relationships were broken down and made easy (Markman & Siebert, 1976). Other results (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) showed that when young children are required to sort conventional empirical objects, they perform as well as older children in classification tasks where there was a change in the procedures required to group objects. Therefore, it can be deduced that simplifications made to either the tasks’ procedure, the objects, or the relationships linking the tasks enable young children to perform more effectively than expected (Donaldson, 1978; Fodor, 1972; Gelman, 1978; Johnson-Laird & Wason, 1977).

Research by Thompson (1941) attempted to investigate children’s level of achievement and the way they respond to sorting tests, and to analyze the possibility of differences in test effectiveness across different age groups. Three types of tests were administered to 60 children in the first grade through sixth grade. These were: the Weigl Color-Form test (Weigl, 1941), the BRL test (Bolles, Rosen, & Landis, 1938) and the Vygotsky Test of Concept Formation (Hanfmann & Kasanin, 1937). During the Weigl test, children sorted cardboard figures based on shape and color, and were required to explain their sorting strategies afterwards. In the BRL test, they classified objects of varying size, color, shape, and other attributes by creating groups of similar objects, and were required to provide an explanation of their sorting and classifying strategies. Finally, in the Vygotsky test, children sorted wooden objects into four groups. The implementation of these tests to children within this age range revealed a difference in test performance and thinking organization between younger and older students. The author found that younger children had a tendency to create groups of objects with respect to a specific
concrete situation, and recognized more differences among objects than similarities. On the other hand, older students were more able to determine similarities or non-specificities within groups of objects, and were more apt to see a broader domain of potential relationships. Since younger children were more inclined to focus on differences among objects, they could not disregard inconsistencies to center their attention on basic similarities.

**Newer research on sorting and classifying.**

More recent research on sorting and classifying dates from the mid-90s onwards. These studies are mostly focused on measuring sorting and classifying skills and abilities from a different perspective in terms of the materials used, tasks administered to participants, and the procedures employed to conduct the studies and/or experiments.

Measures that have been, and are presently being used to assess sorting and classifying skills of preschool children, include the Dimensional Change Card Sort (Zelazo, Frye, & Rapus, 1996), also known as the DCCS Test, Wisconsin Card Sort Test (Grant & Berg, 1948), also known as the WCST, and the Flexible Item Selection Task (Jacques & Zelazo, 2001), also known as the FIST. Within this section of this review, only the DCCS and the FIST have been used in the studies presented. It is important to note that the DCCS and the WCST are widely and primarily used in the field of psychology. Although they have been mostly used for psychological purposes, these two measures have contributed largely in the field of education as well, especially in assessing children’s sorting and classifying skills. As highlighted by Nyhus and Barceló (2009), the WCST was initially developed by Grant and Berg (1948) to measure abstract reasoning and shifting abilities in cognitive strategies of individuals between 5 years old and 90 years old due to changing situations (Eling, Derckx, & Maes, 2008). Participants are involved in sorting cards according to different dimensions such as color, shape, size, and
number of objects. Prior studies (Brown & Campione, 1971; Kagan & Lemkin, 1961) showed that both younger and older students had different stimuli inclinations and different types of groupings based on two attributes when involved in sorting geometric stimuli differing in terms of color, shape, and size. Presently, the WCST is mostly used as a neuropsychological tool to measure the robustness of the frontal lobe functions (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The DCCS was developed by Zelazo et al. (1996) with the intent of studying children’s redirection of attention (Haas, De Mers, Fulton, Terrana, & Horton, 2010). In fact, the DCCS is a simplified version of the WCST (Berg, 1948). As indicated by Fisher (2011), the DCCS was developed appropriately for children based on the WCST, which was previously commonly used in studying the deterioration of executive control in school children and adults (Milner, 1963). As illustrated further in this literature review, a few studies have used more modern and computerized versions of these measures.

All studies in this review investigated young children’s sorting and classifying skills. Nevertheless, they analyzed these aptitudes from two different perspectives, which make them fall under two important themes. Almost half of these studies investigated sorting and classifying skills and dimensions by using a particular measure. The other half are studies that specifically investigated young children’s behavior and tendency toward different sorting and classifying dimensions or attributes, including activities with and without sorting rules or directions.

**Research employing sorting and classification measurement tools.**

Kloo, Perner, Kerschhuber, Dabernig, and Aichhorn (2008) conducted a study, comprised of two experiments with 150 children aged between 2 and 4 years with whom the Dimensional Change Card Sorting (DCCS) Task was administered. The authors explored the extra-dimensional shifting aptitudes of children, that is, their ability of switching between different
dimensions, and the factors making the DCCS extra-dimensional shifting task challenging for children at this age. Parallel to the extra-dimensional shifting tasks, uni-dimensional (based on same color but different shapes) and bi-dimensional (based on same shape but different colors) reversal shift tasks were employed. In both of these two types of tasks, target cards and test cards were utilized, that comprised of pre- and post-switch phases. In the standard DCCS task, children sorted cards according to one dimension (e.g. shape or color) without assistance and justified their decisions. The reversal shift task followed the same procedure as in the DCCS task, with the exception that children were asked to reverse the sorting procedures (e.g. all bananas go to the apples and vice-versa instead of correct sorting) in the post-switch phase. The two main findings of this study revealed that the extra-dimensional shift sorting tasks are more difficult than the reversal shifting sorting tasks, and the sorting instructions are important in addressing the difficulty of extra-dimensional shift sorting tasks. Essentially, the authors found that since 3-year-olds have a significant level of difficulty in the extra-dimensional task when exposed to two descriptions of one thing simultaneously, clear sorting instructions becomes critical for children at this age.

In regards to investigating preschoolers’ sorting ability on the DCCS task, three experiments were conducted with 143 children aged between 3 and 4 years to demonstrate how labeling can improve preschool children’s sorting ability on the DCCS (Müller, Zelazo, Lurye, & Liebermann, 2008). All three experiments included card sorting demonstrations with pre-switch and post-switch phases. During pre-switch phase, children sorted cards by one dimension, and during post-switch phase they sorted cards by a different dimension. Results of the study revealed no favorable impact of labeling on preschoolers’ task-sorting performance on the DCCS. The authors found that even when children labeled correctly on the post-switch phase,
the majority of them sorted the cards incorrectly. Essentially, a major finding indicated that the correct designation of the appropriate dimension of test cards did not necessarily enhance children’s sorting skills. This finding was consistent with previous results (Towse, Redbond, Houston-Price, & Cook, 2000) where the authors showed that labeling of test cards in the DCCS did not improve 3-year-olds’ card sorting performance.

Along the same lines of research, two experiments were conducted to investigate the possibility that a 3-year-old can succeed on the DCCS task in the presence of a certain amount of environmental exposure with respect to post-switch sorting rules, and to test whether 3-year-olds swap sorting rules during the post-switch condition of the DCCS task when they are exposed to a substantial amount of environmental means for using the new rules (Jordan & Morton, 2012). The first experiment consisted of sixty-five 3-year-olds and the second experiment consisted of fifty 3-year-olds, where children were subject to pre- and post- switch phases. Prior to the experiments, children completed a computerized DCCS task and in both experiments, flankers were used. Flankers are incidental stimuli that are displayed next to a target stimulus and that environmentally support the target’s responses (Jordan & Morton, 2012). The two experiments differed in the following way: In the first experiment, during the pre-switch condition, children were taught to sort cards based on one attribute (e.g. color), and during the post-switch condition, they were instructed to sort cards based on another attribute (e.g. shape); in the second experiment, children sorted cards by either color or shape during the pre-switch phase, and they were instructed to sort the objects based on another attribute during the post-switch phase. Results of experiment one showed that, as the distance between the flankers and the objects to be sorted is decreased, 3-year-olds’ sorting ability improved consistently. In the second experiment, the authors found that children’s overall performance was enhanced by the flankers. In sum,
findings revealed that more challenging criteria had a more adverse effect on children’s knowledge of new sorting rules rather than when they were exposed to less challenging criteria.

By extending previous research (Gentner, 1989; Gentner & Rattermann, 1991) on object matching based on the same dimensions and investigating if these tasks can anticipate the seclusion and abstraction of relations among objects, Bennett and Müller (2010) conducted a study with 97 preschool children aged between 3 and 5 years to investigate how flexibility and abstraction in children within this age group develop. One of the tasks administered to participants was the Flexible Item Selection Task (FIST), which was related to investigating children’s sorting abilities. In this task, children matched and sorted cards with respect to two dimensions, color and shape. Results of the FIST revealed two critical developmental aspects that prevail during preschool years. First, children learned to build association skills with respect to dimensions, after which they learned to develop more complicated similarities between objects and their characteristics (Gentner, 1989; Gentner & Rattermann, 1991). The second finding, which is aligned with previous results (Jacques & Zelazo, 2001; Smidts, Jacobs, & Anderson, 2004) indicated that the developmental aspects in cognitive flexibility contributed in the improvement of the flexible shifting between similarity relations.

Along the same lines of research, Jacques and Zelazo (2001) used the FIST with 197 preschool children aged 2- to 5-years old to investigate the viability of the FIST in regards to preschoolers’ abstraction and cognitive flexibility, and to determine whether there are significant differences in children’s performance related to age based on this particular task. Initially, children were exposed to demonstration trials, where the researcher demonstrated tasks by sorting and grouping similar cards based on particular dimensions, after which children were presented with four cards and were asked to show two similar cards based on one dimension. In
the same activity, they were asked again to show two similar cards based on two different dimensions. Finally, they were presented with three cards and were asked to show two similar cards based on one dimension and were asked again to show two similar cards based on two different dimensions. Results revealed that 2-year-olds experienced difficulty in understanding the basic task instructions, and therefore performed poorly on the final sorting activities where they were asked to sort three cards and two cards based on one and two dimensions respectively. Similarly, 3-year-olds did not perform well in the following cases: first, when they were asked to sort a set of four cards based on one dimension, and second, when they were asked to sort another set of three cards based on one dimension. This finding indicated that they experienced difficulty in selecting a particular attribute to sort, which further indicated that preschool children are more likely to respond to rather visual or tangible characteristics of a dimension (Inhelder & Piaget, 1964; Kendler, 1972; Smith, 1989; Vygotsky, 1934, 1978, 1986). Most of the 4-year-olds performed well in sorting four cards based on one attribute as well as three cards based on one attribute. In other words, they did not experience difficulty in selecting a particular attribute to sort the cards. However, other findings indicated that they experienced difficulty in shifting from one dimension to another on this task, which demonstrates how cognitive flexibility is being developed during preschool years (Gerstadt, Hong, & Diamond, 1994; Inhelder & Piaget, 1964; Jacques, Zelazo, Kirkham, & Semcossen, 1999; Zelazo et al., 1996). Finally, 5-year-olds’ performance was better than 4-year-olds on tasks requiring them to sort one and two cards based on one and two dimensions respectively.

**Investigating young children’s behavior toward sorting and classifying dimensions.**

Langer, Schlesinger, Spinozzi, and Natale (1998) conducted a study with six children aged between 6- and 24-months old to investigate how children at this age interact with objects
they are exposed to. During the experiment, the participants were presented with four objects, one object at a time. Results of this study revealed that 6-month old infants interacted with all types of objects, and at 12 months they started to interact with identical objects simultaneously. They continued to manipulate identical objects at the same time until the age of 2 years. Seen in this light, the authors deduced that the concept of classification starts as early as 1 year old. This finding is consistent with previous findings (Langer, 1980, 1986). In addition, it was also shown that by the age of 2 years, children start to categorize similar objects together (Langer, 1980; Nelson, 1973; Ricciuti, 1965; Starkey, 1981; Sugarman, 1983).

Kloo and Perner (2005) conducted a study consisting of three experiments with ninety-one 3- to 4- year-olds to investigate if separating two attributes with respect to one object and distinguishing between two different objects can improve children’s sorting performance; that is, enhancing their abilities from sorting by one attribute (e.g. color) to sorting by another attribute (e.g. shape). The first experiment was comprised of pre-switch and post-switch phases where the assessor modeled one trial by removing one test card, after which the child sorted five cards by the dimension pointed out by the assessor on his card. In addition to the rules and procedures of the first experiment, two test cards and a pair of objects were used in the second experiment, where children were asked to sort both objects and cards simultaneously based on a particular dimension. The third experiment confirmed the results of the first two experiments by using two sets of cards and two sets of objects, combined and separated. Results of this study revealed that the separation of two dimensions and description of two different objects enhanced 3-year-olds’ performance significantly. These three experiments showed that the separation of sorting dimensions allowed children at this age to shift their attention from one sorting condition to another.
Building on previous research, Bernstein, Zimmerman, Werner-Wilson, and Vosburg (2000) employed classification tasks in a study with 19 children, of which 12 children were assigned to an experimental group and 7 children were assigned to a control group. The mean age of the participants in the experimental group was 4.56 years and the mean age of the participants in the control group was 4.57 years. Two measures were administered to the participants: the race/ethnicity photo task, consisting of photographs of people differing in terms of various attributes including race/ethnicity; and the shape classification task involving cards of different shapes, colors, and sizes. Children sorted photo cards by grouping similar people together and were also asked to find other ways of sorting the cards and to justify their actions. Activities carried out in this study were found to improve children’s classification skills. In particular, by the end of the interventions, the authors found that children were able to sort and classify individuals by different attributes (age, gender, and race/ethnicity), and further pointed out that participants could recognize the similarities and differences among people of different racial/ethnic backgrounds.

Deák, Ray, and Pick (2002) conducted three experiments to investigate how 3- to 4-year-olds use rules to classify and label objects based on their shape or their function. In all the three experiments, materials used were 10 object trios, where each trio consisted of a hybrid object (an object serving more than one purpose), a same-shape object with a different function, and a same-function object with a different shape. The first experiment included forty-eight 3-year-olds and forty-eight 4-year-olds. Children were divided into three groups: a non-specific instruction group (sorting and classifying instructions were not specific), a shape-instruction group (sorting and classifying instructions were focused on shape of objects), and a function-instruction group (sorting and classifying instructions were focused on function of objects). In
the second experiment, forty-two 4-year-olds were tested, but there were no instruction trials. In the third experiment, sixty-four 3-year-olds were involved, with two new instruction sets added. Results of the first experiment showed that most of the 4-year-olds consistently emphasized either shape or function, depending on instructions, without any cues or feedback. However, 3-year-olds mainly classified objects by shape, with a few matching by function. The authors further noted that 75% of 3-year-olds who received function instructions classified objects correctly by function and 97% who received shape instructions correctly sorted by shape. The major finding of the second experiment revealed that 4-year-olds who did not receive specific instructions were more inclined to match objects by shape. Finally, results of the third experiment revealed that most of the 3-year-olds did not immediately understand the instruction that was provided regarding the grouping of objects by function. However, it was found that 3-year-olds had the ability to quickly recall the objects’ functions, apply some matching rules, and understand questions that used redundant wordings that were the same as function instructions. On the whole, the authors note that while 3- and 4-year-olds had the tendency to label objects of the same function correctly, they were less inclined to do so in the case of shape.

Brace, Morton, and Munakata (2006) conducted a study to investigate whether children switch more promptly to a new rule based on events that are less challenging on working memory, such as guided practice with the new rule. In this regard, the authors tested this prediction with forty-eight 3-year-olds. They were assigned to three conditions: an instruction condition, where children followed direct instruction with the new rule; scaffolding condition, which represented guided practice with the new rule; or in a scaffolding-plus-instruction condition, which was a combination of both direct instruction and guided practice. Children who were in the instruction condition sorted cards that were unrelated to each other between the pre-
and post-switch phases. In the scaffolding condition and in the scaffolding-plus-instruction condition, children were initially presented with cards displaying information that were consistent to the dimension that would be used for sorting in the next post-switch phase without conflicting information of the previous rule. Results revealed that, in terms of the post-switch phase, children were found to be more successful in the scaffolding and scaffolding-plus-instruction conditions than in the instruction condition. Twenty-five percent of the participants succeeded in the post-switch phase of the instruction condition, 81% in the scaffolding condition, and 94% in the scaffolding-plus-instruction condition. On the whole, the authors reported that children receiving guided practice by using a new rule were more effective than those receiving direct instruction in using the new rule.

Previous research has shown that card sorting tasks involving pre- and post-switching trials with preschool children produce a cognitive conflict between what children are initially asked to do and what they need to do post switch. Zelazo, Müller, Frye, and Marcovitch (2003) clearly indicate that when rules are used in pre-switch trials, children have the tendency to stick to these rules and neglect rules in post-switch trials, which therefore create a conflict between pre- and post-switch trials in sorting tasks. In this regard, Jordan and Morton (2008) conducted a study involving one hundred and nine 3-year-olds in two experiments, comprised of pre- and post-switch phases, to measure the possibility of improving children’s performance on sorting tasks. The focus was on decreasing the level of conflict on post-switch trials, while preserving the task’s level of difficulty; the necessity for re-describing the stimuli; and the importance of not disturbing children’s attention. Children were administered DCCS sorting tasks through a touch screen computer where neutral and congruent flankers were used. As defined by Sullivan (2012), flankers represent information that is presented at the same time as the target, but which are not
relevant to the task, and, therefore, should be disregarded. Results revealed that 3-year-olds were more inclined to shift sorting conditions when test cards were framed by images that were compatible with post-switch rules than when they were surrounded by images that were incompatible with post-switch rules. Essentially, children did not just match the flankers with the target cards. These findings showed that the use of compatible flankers significantly improved 3-year-olds’ adaptability in the tasks. These findings were consistent with prior evidence that suggested that under various conditions, 3-year-olds can use two pairs of different rules in the DCCS. For example, they can sort test cards (e.g. red rabbits and yellow trucks) by matching them to a pair of target cards (e.g. red truck and yellow rabbit) by using two rules: color and shape (red rabbits test cards go to the red truck target card and yellow trucks test cards go to the yellow rabbit target card for the color rule; red rabbits test cards go to the yellow rabbit target card and yellow trucks test cards go to the red truck target card for the shape rule). These include conditions involving the use of labeling (Towse et al., 2000), guided practice (Brace et al., 2006), and separation of stimulus dimensions (Kloo & Perner, 2005). Jordan and Morton (2008) demonstrated that 3-year-olds had the ability to use contrasting rules in circumstances that required re-description of stimulus and switching of attention to a new stimulus dimension.

**Problem Statement**

Educators and policymakers are concerned about the need for improvement in mathematics learning (Brenneman, Stevenson-Boyd, & Frede, 2009). Some of the policy recommendations outlined by the authors are as follows:

- Policymakers need to make sure curricula, learning standards, and teaching expectations for early mathematics are research-based. They should delineate expectations that can be reached by preschoolers in an appropriate way.
• Policies related to early childhood education need to characterize mathematics as a discipline that goes beyond counting and number.

• Adequate preparation in understanding mathematics content should be given to pre-service and in-service educators; they should also be provided with experience in integrating this content into their teaching practice.

• In order to support a high quality mathematics education, proper accountability systems emphasizing the classroom, the teacher, and the child need to be established.

Due in part to the fact that the number of children enrolled in preschool programs is increasing, valid and reliable means are needed to measure programs’ effectiveness for promoting student learning (Snow & Van Hemel, 2008). In order to plan and tailor learning experiences according to learners’ interests, strengths, and needs, teachers must make use of more informal assessment techniques every day (Brenneman et al., 2009). As these scholars indicate, the appropriate way of doing this in mathematics is to gather the maximum amount of information related to specific mathematical skills or learning indicators. The authors further point out that well-designed and comprehensive assessment tools can help in supporting and expanding learning activities. In addition, Brenneman et al. (2009) put forward that assessments have the power of identifying preschoolers’ skills and how they learn mathematics, which will help teachers to plan instruction based on children’s knowledge-base, level of mathematics understanding, and skills.

Researchers have employed various measures to assess young children’s sorting and classifying skills. Examples include the Dimensional Card Sort Test (DCCS), the Flexible Item Selection Task (FIST), and the Wisconsin Card Sort Test (WCST). All of these measures differ from the ACBM, as explained in the following sections.
The DCCS is widely used to measure preschool children’s executive function (Müller et al., 2008). In the standard version of the DCCS, children are presented with two target cards and are required to first sort by one attribute (e.g. color) and then sort by another attribute (e.g. shape). During the activity, children are told how to sort, such as by putting all blue cards in a blue box and all red cards in a red box (Kloo et al., 2008). As far as the ACBM is concerned, initially, children are not given any specific sorting directions; instead, they are given the opportunity to choose their first/second/third attribute as a criterion for sorting.

The FIST (Jacques & Zelazo, 2001) follows almost the same procedures as the DCCS. At the beginning of the sorting tasks, children are presented with demonstration trials with picture cards to be sorted. Afterwards, they are asked to sort these cards in the same way as they were shown. If the children are successful in this initial stage, then they are subject to two criterial trials. In trial one, children are asked to identify two identical cards (e.g. one pair of little orange socks on each card); and in trial two, they are asked to identify two other identical cards (e.g. two medium purple fish on each card). If children are unsuccessful in one of these selection tasks, they do not receive the subsequent test trials. In contrast to the FIST, the ACBM gives the children the opportunity to first select an attribute of their choice and sort independently, after which, depending on the children’s actions, the assessor either models sorting based on one attribute or provides a prompt to help children to move on with the activity. The initial step of the implementation of the ACBM takes into account, previous important findings suggesting that young children’s basic and fundamental mathematical knowledge begins from the early developmental years of life until the age of 5 (National Center for Education Statistics, 2000) and that young children possess basic mathematics competence prior to formal schooling (Brenneman et al., 2009). Based on these findings, it is considered important to give children the
opportunity to put into action their own sorting abilities and strategies, and this is precisely what the ACBM attempts to measure.

The WCST has been designed mainly for the age range of 6.5 to 89 years (Heaton et al., 1993). This tool differs from the ACBM, which involves 3- to 5-year-old children. Moreover, during the administration of the WCST, the assessor provides verbal feedback to participants, whereas in the administration of the ACBM, participants are provided with a standardized modeling activity or prompt(s) if they are unable to sort independently, instead of verbal feedback.

To the researcher’s knowledge, there is an absence of reliable and valid assessment tools that can be used to measure, gather, and examine preschoolers’ algebraic competence related to sorting and classifying, and that can also provide data specific to the test rather than pertaining to more general mathematics learning. The measurement tool developed for this study has been designed in a way that can be used by teachers for ongoing assessment.

**Significance**

As outlined in the National Council of Teachers of Mathematics (NCTM) standards, within the algebra strand, there is the stated expectation that all students PreK-2 should, "…Sort, classify and order objects by size, number and other properties" (NCTM, 2000, p.90). NCTM (2000) also indicates that algebraic thinking is a crucial and underlying ability that serves all topics of mathematics. These statements clearly show that sorting and classifying forms an integral part in children’s development of algebraic skills.

According to the Second Handbook of Research on Teaching and Learning of Mathematics (Lester, 2007), there is a strong developmental relationship between how young children approach data analysis and classification, counting, and data representation. Yet, little
research has been done on preschoolers’ knowledge of data analysis. As indicated by Clements and Sarama (2007), the foundational part of young children’s development and understanding of data analysis lies in the development of their sorting and classification skills in their preprimary years. For instance, as Russell (1991) indicates, children’s interest in data lies in the elements of sorting and classifying activities. They later get involved in learning how to classify data by creating distinct groups or sets. Clements and Sarama (2007) further point out that in order for young children to develop their problem solving and number sense skills and knowledge, teachers and curricula need to emphasize classification, organization, representation abilities, and how to use information to ask questions and justify answers. As a matter of fact, sorting and classification abilities in young children remain an important component in the development of other mathematical areas such as data analysis, problem solving, number sense, and ultimately the field of algebra.

Classification and seriation skills represent fundamental elements of young children’s development of mathematical reasoning and learning (Kamii, Rummelsburg, & Kari, 2005; Piaget, 1971, 1974). For instance, previous work (Ciancio, Rojas, McMahon, & Pasnak, 2001; Lebron-Rodriguez & Pasnak, 1977) indicates that preschoolers’ classification, seriation, and conservation skills have a significant impact on more advanced mathematics at school level. For this reason, the development of classification skills in young children is critical to mathematics achievement in later years of schooling.

As Clements and Sarama (2007) indicate, children’s initial and informal classification abilities emerge from their intuitive recognition of analogous objects or situations. Children distinguish between objects they suck and objects they do not at 2 weeks of age. They are later involved in sorting and classifying objects based on functional relationships, which represent the
foundations of sorting (Piaget, 1964; Vygotsky, 1934, 1986). As such, when children reach 6 months of age, they sort objects that are different, and when they reach 12 months of age, they start to sort objects that are similar by putting them together based on particular attributes (Langer, Rivera, Schlesinger, & Wakeley, 2003). After reaching 18 months of age, children are able to create sets of identical objects and sets of different objects; and by the age of 2 years they have the ability to create sets of objects that are similar based on some properties, but these objects may not necessarily be identical. An important point to note is that children do not adhere to verbal sorting rules not until they reach the age of 3 years. As a matter of fact, most preschool children sort objects based on a particular attribute by forming categories even when switching of attributes during sorting occurs (Kofsky, 1966; Vygotsky, 1934, 1986). Children usually sort objects consistently based on one attribute and re-classify them based on different attributes when they reach the age of 5 or 6 years.

Clements and Sarama (2007) point out that all young children need to be involved in activities that will provide them with opportunities that will help them to acquire an acceptable level of proficiency in classification and seriation prior to entering primary grades. The complex relationship between the areas of classification and seriation, and number concepts development (Piaget & Szeminska, 1952), governed by logic and reasoning (Piaget, 1964), have been reviewed thoroughly in the literature (Clements, 1984).

NCTM (2006) considers algebra and data analysis as critically connected to the content areas of number, geometry, and measurement. In fact, NCTM (2007) describes algebra as a way of thinking and reasoning in regard to relationships. Algebraic reasoning in young children starts as early as 3 or 4 years of age, when young children start to manipulate pattern blocks to create their own patterns, classify objects based on a particular rule, or when they start to pay attention
to patterns they see and observe in the environment (Epstein, 2003). Later, Epstein (2006)
indicates that children’s inclination to collect and sort objects based on their attributes represents
a major contribution to their ability to represent, analyze, and interpret mathematical data.

As indicated by the review of literature, earlier studies were more centered on
investigating children’s sorting and classifying skills than newer research in this field. However,
these studies did not use assessment tools that were designed to measure only these skills in
children. Earlier studies were more focused on engaging children in sorting and classifying tasks,
where the researcher observed their sorting and classifying strategies and reported the findings
based on these observations. On the other hand, most of the recent studies (1998-2011) have
been found to be more inclined to partially investigate children’s sorting and classifying skills;
they mostly sought information about how children’s psychological aspects impact their sorting
and classifying, how children’s switching of attributes/sorting rules in a particular sorting task
affect their sorting and classifying performance, or how/whether environmental aspects can
affect children’s sorting and classifying skills. These lines of research emphasized the following:
exploring children’s ability of switching between different dimensions; investigating how
labeling can improve preschool children’s sorting ability; investigating how children can succeed
on sorting tasks when exposed to certain environment factors; investigating whether children
swap sorting rules during sorting tasks; exploring children’s flexibility and abstraction in regards
to particular sorting tasks; investigating how children interact with objects they are exposed to;
investigating how the separation of two attributes with respect to one object and distinguishing
between two different objects can improve children’s sorting performance; examining
preschoolers’ use of rules as a way of classifying and labeling objects based on their shape or
their function; inquiring into whether children switch more promptly to a new rule based on
events that are less challenging on working memory; and exploring whether children’s performance on sorting tasks can be improved if the level of conflict on post-switch trials is reduced while the task’s level of difficulty is preserved, the necessity for re-describing the stimuli, and the need not to disturb children’s attention. In contrast to these studies, the main focus of the present study is to measure exclusively children’s sorting and classifying skills and strategies based on one attribute and two attributes simultaneously, by employing the assessment tool, the ACBM.

The intent of the present study is to develop and validate an instrument entitled the Algebra Curriculum Based Measure (ACBM) that will contribute to the understanding of algebra at the preschool level. This assessment tool is designed to measure young children’s development of sorting and classifying skills, including the nature of responses in terms of their sorting and classifying strategies. As such, it will help teachers in determining children’s difficulties, thus helping them to plan their instruction accordingly. Also, the ACBM is a stepping stone for further research in the field of algebra, extending to other fields such as data analysis, problem solving, and number sense, and more advanced mathematics in later school years.
Chapter 2: Methods

Purpose

The purpose of this study was to develop and validate the Algebra Curriculum Based Measure (ACBM) with the intent of assessing preschool children’s sorting and classifying ability. The instrument was administered to 3- to 5-year-olds, who were asked to sort and classify objects based on color, shape, and size and to provide a rationale for how they sorted the objects. The instrument can be used by teachers as a way to monitor student progress or by schools as a way to monitor their mathematics program.

Objectives

The ACBM has been developed based on two inherent objectives: to measure preschool children’s sorting and classifying ability based on one attribute and two attributes simultaneously; and to assess their algebraic reasoning and understanding with respect to sorting and classifying skills. The items measuring these objectives are organized in a Table of Specifications (Appendix A).

Variables

The study has a latent construct and five sub-constructs (or scales). The latent construct is represented by sorting and classifying and the five scales are represented by the following observed variables: “Color,” “Shape,” “Size,” “Color and Shape,” and “Explanation.” In addition, the participants’ ages were investigated across these five scales to determine the degree to which they are correlated.

Research Questions

Two research questions pertain to this study. These are as follows:
Research Question 1: What is the contribution of each sub-construct to the latent construct of sorting and classifying?

Research Question 2: What is the technical adequacy of the ACBM in measuring preschool children’s sorting and classifying skills?

**Hypothesis**

The hypothesis guiding this research study is as follows:

The five sub-constructs will contribute significantly to the latent construct of sorting and classifying, thus making the ACBM a significant indicator of preschool children’s sorting and classifying skills.

**Research Design**

Quasi-experimental design.

This study follows a quasi-experimental research design. No control groups were involved. The instrument was administered to children aged 3 to 5 years on an individual basis, where they were presented with manipulative materials and were required to arrange them into groups. These materials were foam objects of different colors (red, blue, yellow), different shapes (circle, square, heart), and different sizes (small, medium, big). Participants were involved in activities requiring them to sort and classify these objects by one attribute and two attributes simultaneously. Directions that the assessor gave to the child were scripted to ensure procedural reliability. Children were first given the opportunity to spontaneously sort the items based on an attribute of their choice. If they were unable to do this, the assessor modeled sorting by a specific attribute (color). After each activity, the children were asked to explain their sorting and classifying strategies, which enabled the assessor to measure their algebraic understanding and reasoning. After collecting data from the sample, a scoring technique based upon a
hierarchical development reported in the literature was established and responses were scored accordingly. The ACBM and the administration procedure of the study are explained in later sections of the document.

Sample

The target population for this study included all preschool children aged 3 to 5 years in Hamilton County in the state of Ohio. The accessible population was identified from the Greater Cincinnati metropolitan area, where many 3- to 5-year olds are enrolled in licensed preschool or child care programs. The convenience sample selected for this study was children enrolled in four of the preschool or child care programs from the Greater Cincinnati metropolitan area during the 2012-2013 academic year. Prior to conducting this study, approval of the University of Cincinnati Institutional Review Board was sought, including permission letters that were distributed to the parents or legal guardians of children at the center. No children who had obtained parental permission were excluded from the study.

Prior to conducting the main study, a pilot study was carried out. Twenty permission letters were distributed, out of which six signed letters were returned. The data collection for the main study spanned 6 weeks. The final sample size was 120 children. Out of a total of 250 parent permission slips distributed, 145 signed permission slips were returned to the classroom teachers. Of these 145 children, only 120 completed the assessment. The other 25 children did not take part in the study either due to absenteeism or refusal to participate. Table 2 shows demographic information of the participants for the main study.
Table 2

*Demographic Information of Participants for Main Study*

<table>
<thead>
<tr>
<th>Demographic information</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Males: 68</td>
</tr>
<tr>
<td></td>
<td>Females: 52</td>
</tr>
<tr>
<td>Age</td>
<td>Maximum: 64 months</td>
</tr>
<tr>
<td></td>
<td>Minimum: 34 months</td>
</tr>
<tr>
<td></td>
<td>Average: 49 months</td>
</tr>
<tr>
<td>Ethnicity/Race</td>
<td>White: 79</td>
</tr>
<tr>
<td></td>
<td>African-American: 27</td>
</tr>
<tr>
<td></td>
<td>Asian: 10</td>
</tr>
<tr>
<td></td>
<td>Hispanic: 4</td>
</tr>
<tr>
<td>Socio-economic level</td>
<td>Lower-income families: 37</td>
</tr>
<tr>
<td></td>
<td>Middle to upper-income families: 83</td>
</tr>
<tr>
<td>English as second language</td>
<td>None</td>
</tr>
<tr>
<td>Physical and/or cognitive impairment</td>
<td>None</td>
</tr>
</tbody>
</table>

**Instrumentation**

The measurement tool designed for this research study is the Algebra Curriculum Based Measure. This assessment tool is based on the curriculum framework by Moomaw and Hieronymus (2011). As Shinn and Bamonto (1998) put forward, curriculum-based measurement provide teachers with the opportunity of continuously investigating the developmental progress of children’s thinking, which therefore allows them to plan their instruction accordingly. The
primary purpose of the present research study was to develop and validate a curriculum-based measurement tool that would assess young children’s sorting and classifying skills. To that end, it was important to highlight the algebraic understanding and reasoning of preschool children in regards to sorting and classifying, including how they sort and classify objects based on one attribute or two attributes at the same time. At pre-K, algebraic thinking in young children starts to develop through activities involving sorting and classifying concrete objects. As NCTM’s (2000) Principles and Standards for School Mathematics points out, “Classifying and ordering are natural and interesting to children” (p. 37).

Materials that were used for the purpose of this study were concrete foam objects with cross-over attributes of color, shape, and size. Red, blue, and yellow colors were used because at this young age, preschool children are attracted to and are familiar with these colors. In terms of shape, circles, squares, and hearts were represented. Again, these shapes were chosen because children are familiar with them. As far as size is concerned, there were small, medium, and large foam objects involved in the study.

**Instrument design.**

The ACBM consists of 27 foam objects for children to sort:

- 9 circles (small, medium and large; red, blue, and yellow)
- 9 squares (small, medium and large; red, blue, and yellow)
- 9 hearts (small, medium and large; red, blue, and yellow)

Children are asked to sort these items based on one attribute or two attributes simultaneously. Children are also asked how they determined which objects to group together. It is important to note that 27 items was the minimum needed to have crossover attributes of three colors, three shapes, and three sizes. Also, children are asked to first sort by one attribute at a
time because this is generally considered easier for young children. Afterwards, they are asked to sort and classify the objects by two attributes simultaneously (color and shape).

**Pilot study.**

*Overall pilot results.*

The assessment tool was subject to a pilot study carried out at a preschool center from September 6, 2012 to September 10, 2012 with six children aged between 3 years 9 months and 4 years 11 months. The original instrument used in the pilot test measured children’s sorting and classifying skills based on one attribute (color, shape, size), two attributes (color and shape, shape and size, color and size) simultaneously, and three attributes (color, shape, size) simultaneously. Results of this pilot test revealed that one of the six children tested did not understand well the concept of sorting and classifying. Most of the time s/he needed the assessor to demonstrate an activity, after which, s/he was able to sort and classify the objects correctly. It was also found that initial contact for her was important. In line with this observation, it seemed that some children were not able to grasp the wordings used in the instrument. Instructions and directions seemed unclear to them. Other results revealed that during the first level of the instrument, for the most part, children grouped only one set of objects and did not include all objects based on a particular attribute. They were also found to randomly sort the objects and mix them up, and these responses/sorting decisions seemed to occur due to fatigue. All of the six children were found to be more familiar with color than the other two dimensions, shape and size. Children did not perform well on the second level of the instrument, when they had to sort by two attributes. Children also showed attitudes of uncertainty in regards to their responses and sorting strategies. These were indicated by either the questions they were asking to the assessor or by their body gestures.
Changes made to the initial instrument.

Based on the results and observations of the pilot study, the instrument was modified to focus on sorting and classifying skills with respect to one attribute (color, shape, size) and two attributes (color and shape) simultaneously, thus dividing the instrument into two levels instead of three. The reason for excluding the third level of the instrument, which measured the child’s sorting and classifying ability by three attributes, was that a large amount of objects was necessary to accommodate sorting by three attributes at the same time. In addition, pilot results indicated that at this level children showed emotions of fatigue and confusion as they randomly grouped objects. When asked to sort and classify by three attributes simultaneously, they seemed to diverge from the activity, thus losing interest in the assessment. Pilot results also revealed that children seemed to be confused when they had several options of grouping objects by two attributes simultaneously. The majority of children were unable to sort and classify by two attributes when size and shape were involved. Therefore, it was decided to leave out sorting and classifying by color and size, and size and shape, and to keep the only option sorting by color and shape simultaneously.

On the whole, based on information from the pilot study, three types of information were added to the instrument: (1) more detailed scoring possibilities in terms of the child’s responses or decisions regarding sorting and classifying strategies, (2) an additional column indicating the attribute the child selected or the assessor modeled (for clarity), and (3) inclusion of graphics (color, shape, and size of foam objects) in the instrument illustrating the child’s performance of the sorting and classifying task, so that the assessor could circle the exact graphics selected by the child.

Description of the final instrument.
The first level of the instrument (Appendix B, Table B1) is comprised of three iterations in which the child’s ability to sort by a single attribute is assessed. Initially, the child is asked to determine which objects should go together. This open-ended question assesses whether or not the child can abstract an attribute to use as a rule for sorting the items. Application of a rule is a key concept in algebraic thinking. If the child is not able to respond to the open-ended question, the assessor models by grouping three items according to color. The child is then asked what else could go into the assessor’s group. Since many young children are incomplete in their responses to sorting and classifying tasks, the assessor provides from one to two prompts when needed to encourage the child to complete groupings. Once the child has completed his or her response to sorting by the first attribute, the same procedure is followed for the remaining two attributes.

The second level of the instrument (Appendix B, Table B1), which measures the child’s sorting and classifying ability based on two attributes simultaneously, differs from the first level in terms of the process. The assessor starts this level by excluding all red objects and all heart-shaped objects, thus giving rise to four groups of objects to be sorted correctly (blue circles; blue squares; yellow circles; yellow squares). The assessor measures the child’s ability to sort and classify by color and shape simultaneously.

A third type of assessment is included after each sorting sequence. Children are asked an open-ended question to explain how they decided which objects to group together. Scoring is based upon the accuracy and inclusiveness of the response.

**Scoring of the final instrument.**

For level one (sorting by one attribute), the child’s raw score is determined by counting the number of objects the child is able to sort based on each attribute (color, shape, and size).
Objects sorted by the child without modeling or prompting ("Self") are recorded separately from objects sorted after assessor modeling ("Model") or prompting ("One prompt," "Two prompts"):  

Total raw score = Self + Model + One prompt + Two prompts  

It was determined that children should receive more points for items they can sort independent of receiving a model or various prompts, because this denotes a higher level of reasoning. For this reason, a weighting system was derived and applied to the equation above. Thus, the adjusted total score on each section of the first level of the ACBM is determined by the following equation:  

\[ A \text{ (Self)} + B \text{ (Model)} + C \text{ (One prompt)} + D \text{ (Two prompts)} \]  

where \( A, B, C, D \) are the weightings that demarcate scores based on level of assistance needed. The final weightings applied to the equation were \( A=3, B=2, C=2, \) and \( D=1 \).  

**Determination of weightings.**  

In order to determine the best weights to apply to the ACBM score levels, various weightings (based on a trial-and-error method) were applied to the original equation and analyzed graphically. The assumption was that the best weights should spread out the scores through the four components ("Self," "Model," "One prompt," "Two prompts") on each section. To achieve this goal, in a plot where weightings of 1/1/1/1 ("Self"/"Model"/"One prompt"/"Two prompts") fall on the y-axis and the weighting scheme under trial falls on the x-axis, high scores through the “Self” component should stretch out to the right, thus creating a curvilinear fit. The subsequent sections explain how plots were handled to determine the weights that would best spread out the scores through the four components. In terms of weighting, each level of the assessment was considered separately because the first level yields a maximum raw score of 27,
whereas the second level yields a maximum raw score of 12. Below is a description of the way weightings were derived and applied to create adjusted scores.

The total number of participants in this study was 120, which resulted in 360 scores for the three sections (color, shape, and size) of the first level of assessment. When plotting graphs with weightings of 1/1/1/1 (Total Score = Self + Model + One prompt + Two prompts) versus weighted scores (Adjusted Total score = A (Self) + B (Model) + C (One prompt) + D (Two prompts), the y-axis has a range of 0 to 27, representing the minimum and maximum scores that can be attained on any of the first three sections. Due to the hierarchical nature of the scoring, each assessment section was assigned into a raw group based on the highest number of correct responses. For example, in the case of an assessment section in which the child sorted 0 objects on the “Self” level, 3 objects on the “Model” level, 12 objects on the “One prompt” level, and 9 objects on the “Two prompts” level, the assessment section would be assigned to the “One prompt” group.

Based on this grouping system, when an analysis of scatterplots that compared various weightings against un-weighted data was conducted, the “Model” group and the “One prompt” group were found to overlap. This raised the question of whether the “Model” group and the “One prompt” group actually represent different levels. By design of the ACBM, if a child obtained any non-zero score on the “Self” level, the “Model” level was skipped; the assessment moves to the “Model” level only if the score on the “Self” level is 0. Seen from this perspective, modeling was considered as a different kind of prompting for this study. Thus, the “Model” group and the “One prompt” group were considered similar and labeled as “Model/One prompt” group. Based on this argument, weightings of 3/2/2/1 were plotted graphically against weightings of 1/1/1/1 (Appendix C, Figure C1) and have been applied to the equation Adjusted total score =
\( A \text{ (Self)} + B \text{ (Model)} + C \text{ (One prompt)} + D \text{ (Two prompts)}, \) where \( A=3, B=2, C=2, \text{ and } D=1. \)

The horizontal spreading of the scores are “cleaner,” and the three components, “Self,” “Model/One prompt,” and “Two prompts” are well demarcated. Among the various combinations of weightings that were tested, the 3/2/2/1 weighting was found to be the best. Therefore, the adjusted total score at the one-attribute level is determined by the following equation:

\[
\text{Total score} = 3 \text{ (Self)} + 2 \text{ (Model)} + 2 \text{ (One Prompt)} + 1 \text{ (Two Prompts)},
\]

where \( 3, 2, 2, \text{ and } 1 \) are the weightings.

A similar procedure was considered for the second level of the ACBM in order to determine the best weights to be applied to score the three components at this level, that is, “Self,” “Model,” and “One prompt.” The weightings selected were 3/2/1. The scores among these three groups were well spread when the equation with these weightings was plotted against the equation with weightings of 1/1/1 graphically (Appendix C, Figure C2). Therefore, the adjusted total score at the two-attribute level is determined by the following equation:

\[
\text{Total score} = 3 \text{ (Self)} + 2 \text{ (Model)} + 1 \text{ (One Prompt)}, \text{ where } 3, 2, \text{ and } 1 \text{ are the weightings}
\]

In contrast to the hierarchical scoring system of a child’s sorting and classifying skills on the ACBM, the question pertaining to “How did you decide which objects to put in your bowls?” which was asked to the child at the end of each sorting and classifying activity, followed a relatively straightforward scoring system. The responses were categorized as follows: A raw score of 4 was attributed if a child gave a completely accurate response to the question (e.g. “Because reds, blues, yellows are together”); a raw score of 3 was attributed if a child gave a partially accurate response to the question (e.g. “I put different colors together”); a raw score of 2 was attributed if a child gave a general response to the question (e.g. “They are the same”); a raw
score of 1 was attributed if a child pointed to some correct objects (e.g. Child pointed to red objects); and a raw score of 0 was attributed if a child either gave a random/wrong response(s) (e.g. “I just wanted to”) to the question or if there was no response. However, due to the large difference in the range of possible scores for sorting and classifying skills as compared to the range of scores for providing an explanation, weights were applied to these raw scores (0 - 4) for this question. The process of applying a specific weight to the raw scores was based on the highest score a child could obtain for sorting and classifying on a section. At the one-attribute level, the highest score that could be obtained was 81 (with weightings 3/2/2/1) and at the two-attribute level, the highest score that could be obtained was 36 (with weightings 3/2/1).

Therefore, the weightings that were chosen to score the question “How did you decide which objects to put in your bowls?” were 20 at the one-attribute level and 9 at the two-attribute level, thus yielding a maximum adjusted score of 80 and 36 respectively. Table 3 below shows how responses to the question were scored with respect to the five distinct categories (0 – 4) and their weightings (0 – 80 at the one-attribute level and 0 – 36 at the two-attribute level).

Table 3

<table>
<thead>
<tr>
<th>Description</th>
<th>Scoring</th>
<th>Weighted scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely Accurate</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>One-attribute level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-attribute level</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Partially Accurate</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>One-attribute level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-attribute level</td>
<td></td>
<td>27</td>
</tr>
</tbody>
</table>
General: Not wrong 2
One-attribute level 40
Two-attribute level 18

Points to some correct objects 1
One-attribute level 20
Two-attribute level 9

Random responses; wrong responses; 0
No response; “I don’t know”
One-attribute level 0
Two-attribute level 0

Data Management

Data file structure.

The main study consisted of data coming from three different sources: demographic information, such as participants’ age, gender, socioeconomic status, race/ethnicity; qualitative data from participants’ responses to assessor’s questions regarding how they decided to group the objects; and quantitative data from scoring of the ACBM. Children’s demographic information, that is, age, gender, socioeconomic status (based on Head Start enrolment), and race/ethnicity of children, which were obtained from the school’s records upon receipt of parental consent, were the only identifiers that were made accessible to the principal investigator. Demographic information was entered into an Excel file. Assessment data, which included the participant’s study identification number (used instead of the child’s name on the study questionnaire), the date of assessment, the assessor’s initials, the child care center at which the participant was enrolled, the time taken to complete the assessment, qualitative data from the
participants’ comments, and the participant’s scores on each item, were entered into a separate Excel file.

Data quality.

Data coding.

Four types of data required coding: participant’s name, gender, ethnicity, and socioeconomic level. Each name was given a separate identification number starting from 001. Gender was coded as 0 for male participants and 1 for female participants. Ethnicity, as obtained from the school’s files, were self-reported by families. The different classifications of ethnicities were collapsed into four categories (White, Black, Asian, and Hispanic) according to the National Institutes of Health (2001) guidelines and have been given a numeric code of 1 to 4 respectively. Socioeconomic level was coded as 0 for the children who were not externally funded and as 1 for children who were externally funded.

Data entry.

Each participating child was assigned an identification number, which was then associated with their assessment data, date of birth, age, gender, socioeconomic status, time taken, and race/ethnicity in the Excel file with the appropriate coding.

Data checking.

Assessment data were entered into an Excel file. In order to ensure accuracy of data, data were checked manually after input into the file. Any data errors were corrected accordingly.

Data cleaning.

Missing data were not applicable to the main study because the assessment tool was administered and scored by the assessor himself. However, if missing data arose at the individual
student level with respect to the individual item response, that is, when a child did not respond to a specific item, that particular item was scored as 0.

Confidentiality of data.

Information about participants has been and will be kept private. Demographic information was obtained through the children’s school file upon receipt of parental consent. No identifiers other than those listed above were made accessible to the research team.

A study ID number was assigned to each child and was used instead of a child’s name on the study questionnaire. Data were kept on campus in the faculty researcher’s locked office cabinets. Signed consent documents and a master list of children’s names and respective ID numbers were kept in a separate location from the research forms in the faculty researcher’s campus office. Only the research team had access to the research data. Any research data that were saved on a computer were password-protected. After completion of data analysis and review, the master list of names and study ID numbers will be destroyed by shredding. Raw data entered in the password-protected computer will be kept for five years after the study is closed, after which it will be deleted. Signed documents will also be kept for five years after the study is closed. If the study is published, no children’s data will be identified by name, classroom, teacher, or school.

Data and Assumption Checks

Data checking procedures were conducted using SPSS, version 21.0, to verify data quality and to check univariate and multivariate assumptions. The dataset was checked for the following: missing data, normality, linearity, and outliers.

Missing data.
Missing data (demographic or assessment data) were not applicable to the study. The child care centers from where the study was conducted provided all the demographic information that was required for this study. Since the instrument administration and scoring were done by the assessors themselves, there were no missing assessment data. The only type of missing data arose at the individual student level in regards to the individual item response, that is, when a child either did not respond to a specific task or question or when s/he indicated that s/he did not know the answer. In such cases, as per the ACBM’s assessment and scoring guide (Appendix B, Tables B2 and B3), a score of 0 was attributed to corresponding tasks or questions.

**Normality checks.**

When a frequency distribution follows a normal distribution, the assumption of normality is met. A normal distribution is symmetrical, with the highest frequency found in the middle, and frequencies decreasing when moving toward each extreme (Gravetter & Wallnau, 2009). For a normal distribution, the skewness and kurtosis values are 0 (Tabachnick & Fidell, 2007). According to Waigandt (2004), it is a common belief that skewness and kurtosis values within the limits of ± 1.0 give an appropriate indicator of normality. In other words, skewness and kurtosis values that are within the range of ± 1.0 are considered acceptable in terms of assessing normality.

Normality checks for the main study data were assessed both statistically (skewness and kurtosis values, Kolmogorov-Smirnov test and Shapiro-Wilk test) and visually (histogram) by using the software, Statistical Package for the Social Sciences (SPSS), version, 21.0. Descriptive statistics included the mean, standard deviation, skewness, and kurtosis (Appendix D, Table D1).

*Skewness and kurtosis.*
Both the skewness and kurtosis values for two scales “Color” and “Color and Shape” were within acceptable limits of ±1.0. The values for the scale “Color” were -0.107 (skewness) and -0.929 (kurtosis); the values for the scale “Color and Shape” were 0.787 (skewness) and -0.832 (kurtosis). While the skewness values for the scales “Shape” and “Size” were within acceptable limits of 0.264 and 0.459 respectively, the kurtosis values were slightly above the limit at -1.094 (“Shape”) and -1.211 (“Size”). The skewness and kurtosis values for the scale “Explanation” were above the limit at 1.813 and 3.642, respectively. Finally, the skewness and kurtosis values for the total measure (Total ACBM) were barely within acceptable limits at 0.925 and -0.990 respectively.

**Kolmogorov-Smirnov test.**

Kolmogorov-Smirnov (K-S) test values were calculated for the five scales of the ACBM, as well as for the total measure (Appendix D, Table D2). K-S values for the scales “Color,” “Size,” “Color and Shape,” and “Explanation” were all significant at \( p = .000 \), indicating deviation from normality. On the other hand, K-S values for the scale “Shape” and the total measure were not significant at \( p = .121 \) and \( p = .408 \) respectively, indicating a normal distribution.

**Shapiro-Wilk test.**

Shapiro-Wilk values were also calculated for the five scales of the ACBM, as well as for the total measure (Appendix D, Table D3). The Shapiro-Wilk test values for all five scales “Color,” “Shape,” “Size,” “Color and Shape,” “Explanation” were significant at \( p = .000 \), indicating deviation from normality.

**Histograms.**
In addition to the K-S and Shapiro-Wilk normality tests, histograms were produced for all five scales: “Color,” “Shape,” “Size,” “Color and Shape,” and “Explanation” as well as for the total measure of the ACBM (Appendix E, Figure E1). These are further discussed below.

With highest frequencies at scores of 0 and 45, the scale “Color” showed a bi-modal distribution instead of a normal distribution. The highest peak (36.7% of the children) was found slightly to the right-hand side of the histogram. The second highest peak was found to the complete left, that is, at a score of 0, indicating that some children (21.7%) were unable to sort and classify by color at all. The peak that was found completely to the right of the distribution indicated that some children (10%) were able to obtain the highest possible score when they sorted and classified the objects by color.

With highest frequencies at scores of 0, 50, and 80, instead of being normally distributed, the scale “Shape” demonstrated a multi-modal distribution. The highest peak was found to the complete left of the distribution at a score of 0, indicating that some children (20.8%) were unable to apply the concept of sorting and classifying by shape at all. On the other hand, the second highest peak, which was slightly above the mean, was found to the right of the distribution, indicating that some children (19.2%) had the ability to sort the objects by shape. As compared to the scale “Color,” there were more children (14.2%) who were able to attain the highest score on the scale “Shape.” This is represented by the third highest peak, which is found at the complete right of the distribution.

With highest frequencies at the scores of 0 and 50, the scale “Size” followed a bi-modal distribution rather than being normally distributed. The highest peak of the distribution, which was found at the complete left at a score of 0, indicated that part of the sample was unable to apply the concept of sorting and classifying by size at all (36.7%). The second highest peak was
found above the mean, on the right of the distribution, indicating that some children (25.8%) were partly successful in sorting and classifying the objects by size. Finally, the last frequency bar, found on the complete right of the distribution, showed that only one child was able to sort and classify all the 27 objects by size.

With highest frequencies at scores of 0, 20, and 36, instead of following a normal distribution, the scale “Color and Shape” showed a multi-modal distribution. The highest peak, which was located to the complete left of the distribution at a score of 0, indicated that some children (45%) were unable to apply the concept of sorting and classifying by these two attributes simultaneously at all. On the complete right of the distribution was found the second highest peak, indicating that there were some children (16%) who demonstrated high ability in sorting and classifying items by color and shape simultaneously, by obtaining the maximum score.

The mean score on the scale “Explanation” was 34.14 and the standard deviation was 51.395. About 88% of the scores were found to the left of the distribution. These data showed that when the question “How did you decide which objects should go together?” was asked, most children in this sample were not successful in giving correct responses. More than 50% of children were either not responding to this question, or saying “I don’t know,” or were giving completely wrong responses. About 12% of scores were found in the long tail of the distribution, indicating that some children were able to respond to the question above. With a skewness and kurtosis value of 1.813 and 3.642 respectively, the scale “Explanation” represents a positively skewed distribution instead of following a normal distribution.

As far as the total measure is concerned, the overall mean score and standard deviation of this sample were 140.26 and 100.388 respectively. The histogram shows that about 69% of the
scores obtained by this sample of children were close to the mean of the distribution. The two highest peaks are located to the complete left of the distribution, indicating that about 17% of the sample obtained low scores on the ACBM. On the other hand, about 14% of children obtained total scores above the mean. As shown by the histogram, the total ACBM does not represent a normal distribution, but rather shows a positively skewed distribution.

To summarize, none of the scales on the ACBM, as well as the total measure, were normally distributed. While the distributions for the scales “Color,” “Shape,” “Size,” and “Color and Shape” were either bi-modal or multi-modal, the scale “Explanation” and the total measure were positively skewed. As Tabachnick and Fidell (2007) indicate, a distribution with a positive kurtosis value indicates a flat distribution with many cases in the tails. This is represented by the histogram for the scale “Explanation” and the total measure.

**Data transformation.**

Due to problems of normality with the data, various transformations were performed with the data. The square root transformation was successful in producing a normal distribution for the total scale.

**Square root transformation.**

A square root transformation is commonly applied to count data. When a square root transformation is applied, the square root of every value is taken. In the case of the square root of numbers above 1, these always become smaller (e.g. \(\sqrt{9} = 3\)). The square root of number 0 remains the same. In general, a square root transformation is used for reducing right skewness of a distribution. For instance, if a distribution is moderately positively skewed, Tabachnick and Fidell (2007) suggest that a square root transformation be used. In doing so, the data may be converted to a normal distribution. In regards to this study, the fact that the total measure of the
ACBM includes count data and is a moderately positively skewed distribution, a square root transformation reduced the difference between the two tails of the distribution by pulling the both sides toward the middle, thus normalizing the data.

The transformed histogram for the total measure indicates a normal distribution (Appendix E, Figure E1) with a mean score of 10.95 and a standard deviation of 4.529 and the highest peak close to the mean. Skewness and kurtosis values of -0.121 and -0.230 respectively also indicate a normal distribution for the total scale (Appendix D, Table D4). The percentage of transformed scores that fall under the curve is 68.3%, indicating that these scores were close to the mean of the distribution. In this regard, as Tabachnick and Fidell (2007) note, for a distribution to be normal, approximately 68% of the data should fall under the curve. Therefore, the transformed total measure clearly fits a normal distribution.

Visually, none of the five transformed scales of the ACBM met normality requirements. They still showed bi-modal, multi-modal, or skewed distributions. Data have not been reported because there is not much difference between the raw and the transformed histograms. Statistically, most of the skewness and kurtosis values for the five scales of the ACBM were within acceptable limits of $\pm 1.0$ for the transformed data, with the exception of the kurtosis values of the scales “Size,” and “Color and Shape,” which were at -1.631 and -1.584 respectively. Descriptive statistics of the five transformed scales and total measure include the mean, standard deviation, skewness, and kurtosis (Appendix D, Table D4).

The K-S normality test values (Appendix D, Table D5) showed that the transformed scales were significant ($p = .000$), thus indicating that they were still non-normally distributed, with the exception of the total ACBM, which was non-significant at $p = .686$ respectively, thus indicating that they met normality requirements.
Furthermore, the Shapiro-Wilk normality test values (Appendix D, Table D6) also showed that the five transformed scales were significant \( (p = .000) \), thus indicating that they were still non-normal even after the square root transformation. However, the total measure was non-significant \( (p = .648) \), thus indicating a normal distribution.

**Linearity.**

In order to examine the degree of correlation (a) among the five scales on the ACBM (Appendix D, Table D7), and (b) between age and scores on the five scales of the ACBM as well as the total measure (Appendix D, Table D8), linearity was examined by means of the Pearson correlation coefficient, \( r \), and \( R^2 \) values. Scatterplots (Appendix E, Figure E2) were also generated to compare children’s age in months to the total score of each of the five scales on the ACBM as well as the total score for the entire measure. \( R^2 \) values were examined to describe the percentage of variance explained by the model.

**Correlations among the scales on the ACBM.**

Correlations between each pair of scales were examined. Weak to moderate linear relationships among scales were obtained, as determined by the Pearson correlation coefficient, \( r \), with values ranging from .260 to .460 (Appendix D, Table D7). On the whole, these data indicate that the assumption of linearity was met for all scales.

**Correlations between age in months and scale scores on the ACBM.**

The scores for the two scales “Color” and “Shape” showed relatively low correlations with age \( (r = .215, p = .018; \text{ and } r = .225, p = .014 \text{ respectively}) \), but were both statistically significant. This means that an increase in age is significantly related to an increase in scores on these two scales. As indicated by the \( R^2 \) values, 4.6% and 5.1% of these respective scales was explained by age. The correlation between age and the scale “Size” was also low \( (r = .061, p = \)
but was not statistically significant, implying that this scale was not explained by age. On the other hand, the two remaining scales, namely, “Color and Shape” and “Explanation,” as well as the total score on the ACBM, showed moderate correlations with age ($r = .311, p = .001$; $r = .310, p = .001$; and $r = .330, p = .000$ respectively). These correlations were all statistically significant. This means that an increase in age is significantly related to an increase in scores on the scales “Color and Shape” and “Explanation” as well as for the total score of the ACBM. As indicated by the $R^2$ values, 9.7% and 9.6% of these two respective scales were explained by age in months, and 10.9% of the total score on the ACBM was explained by age in months.

In addition to determining the degree to which age was correlated with each of the scale on the ACBM by means of the Pearson correlation coefficient, $r$, scatterplots for each of these scales (Appendix E, Figure E2) were also generated. Both the Pearson correlation coefficient and scatterplots were helpful in determining the relationship between age and scores on each of the scales. In particular, scatterplots were helpful in determining the strength and direction of the relationship between these variables visually. How strong or weak the relationship between age and the scales depended largely on the extent to which the dots were oriented toward the direction of the fit line.

In reference to the scatterplots produced for the two scales “Color” and “Shape,” weak positive relationships were shown between age and each of the scales, that is, age was a weak predictor of children’s sorting and classifying ability by “Color” and “Shape.” A closer look at the scatterplot for these two scales showed that most of the dots were not oriented toward the direction of the fit line. Instead, the plots showed that, irrespective of age, children were obtaining low to high scores on these two scales.
As far as the scatterplot for the scale “Size” is concerned, the relationship between age and this scale was not significant, that is, age was not a significant predictor of preschool children’s ability to sort by size. The slope of the fit line, which was very close to 0, was almost flat, implying that there was no relationship between age and the scale “Size.” In addition, many dots were not oriented toward the direction of the fit line. Furthermore, regardless of age, many children were obtaining low to high scores when they were sorting objects by size. For instance, as shown by the scatterplot, many students of varying ages obtained a score of 0.

As opposed to the three scales discussed above, the scatterplots for the two scales “Color and Shape” and “Explanation” showed moderate and positive linear tendencies. The dots had a tendency to be more oriented toward the direction of the fit line, thus showing a stronger positive relationship between age and these two variables. However, irrespective of age, some children were still obtaining low to high scores on these scales. This situation was slightly more pronounced on the scale “Color and Shape” rather than the scale “Explanation,” that is, the scatterplot for “Explanation” showed more scatter.

Finally, the scatterplot for the ACBM total score showed a moderate and positive linear trend with a good deal of scatter. Most of the dots were oriented toward the direction of the fit line, thus showing a moderate positive relationship between these two variables, age and total score of the ACBM. An important point to note is that, regardless of age, children obtained scores as low as 0 (lowest) and scores as high as 80 (highest) when they sorted objects by one attribute; and scores as low as 0 and scores as high as 36 when they sorted objects by two attributes simultaneously.

**Outliers.**
An outlier is a score or measurement that illustrates an atypical exception to a general pattern by deviating largely from other scores or measurements in a group; it should be considered as a special case and handled cautiously (Fraenkel & Wallen, 2009). This particular score or measurement can be in the form of a univariate outlier (an extreme value on one variable) or a multivariate outlier (an odd combination of scores on more than one variable) (Tabachnick & Fidell, 2007).

**Univariate outlier check.**

A check for univariate outliers was performed by calculating the values for each of the five scales of the ACBM, using the formula $M \pm 4 \times SD$, since the sample size was greater than 80 (Pan, 2011). All values fell within the calculated range, indicating an absence of univariate outliers (Appendix D, Table D9).

**Multivariate outlier check.**

A check for multivariate outliers was performed by calculating Mahalanobis’ Distance, $\chi^2_{0.001}(5)$ critical = 20.515 for the five scales on the ACBM. All values fell within the calculated range, indicating no multivariate outliers.

**Data Analysis Plan**

The two research questions of this study were investigated by means of psychometrics and Confirmatory Factor Analysis (CFA). As Switzer, Wisniewski, Belle, Dew, and Schultz (1999) put forward, psychometrics is concerned with determination of instrument quality by measuring variations of constructs within the context of instrument design. As such, psychometric theory forms the basis of instrument development (Rust & Golombok, 2009; Waltz, Strickland, & Lenz, 2005). Two concepts that are considered important in psychometric theory are reliability and validity (White, 2011).
Analysis of first research question.

Since this study is driven by theory, a Confirmatory Factor Analysis (CFA) was performed to answer the first research question: “What is the contribution of each sub-construct to the latent construct of sorting and classifying?” CFA is a statistical method used to specify and estimate one or more acceptable factor structure models, where each model suggests a set of latent variables, also known as factors, to explain co-variances among a set of observed variables (Doll, Raghunathan, Lim, & Gupta, 1995). Specifically, CFA was used to measure the level of contribution of each sub-construct to the latent construct of sorting and classifying on the ACBM. In other words, the CFA was used to examine the relationships between the scales on the ACBM and the latent construct of sorting and classifying.

The first step of the analysis involved the creation of several hypothesized models, which were based on logic, theory, and previous studies. After running a CFA with these models, the most appropriate model was considered to be the best and final hypothesized model for this study (Appendix E, Figures E3 and E4). This final hypothesized model consisted of five sub-constructs as measured by the five scales of the ACBM. It was also assessed with the transformed data. Therefore, two models were generated. These two models were assessed with the software Analysis of Moment Structures (AMOS) version 21.0 (Arbuckle, 2012) to: (a) measure the degree to which each of the five scales (“Color,” “Shape,” “Size,” “Color and Shape,” “Explanation”) on the ACBM contributed to the latent construct of sorting and classifying, and (b) assess the extent to which the model fit the observed data.

The degree to which each of the five scales contribute to the latent construct was established by examining factor loadings, also known as beta weights (Kline, 2000) or regression coefficients (Hox & Bechger, 2007) or regression weights. Factor loadings, which represent the
correlations of the variable with the factor (Kline, 2000), were examined for statistical significance at \( \alpha = .001 \) as well as practical significance.

As far as assessment of model fit is concerned, there are many statistical tests that can be examined to determine an overall fit of the model. However, there is no universally acceptable and established statistic to assess whether the model is a good fit for the observed data or not (Doll et al., 1995). As such, several measures of fit were considered. As Tabachnick and Fidell (2007) state, assessment of model fit is an active area of research, where new fit indices appear to be developed daily. According to them, the two most commonly reported fit indices are the Comparative Fit Index (Bentler, 1990), and the Root Mean Square Error of Approximation (Browne & Cudeck, 1992). In addition to these two fit indices, Thompson (2004) notes that currently, the other two most frequently considered fit indices to assess the goodness of fit between the hypothesized and measured models are the chi-square \((\chi^2)\) statistical significance test and the Normed Fit Index (Bentler & Bonnett, 1980). As suggested by Tabachnick and Fidell, model fit can also be assessed by considering the ratio of the \(\chi^2\) to the degrees of freedom. Therefore, these five fit indices were considered for this study. These are described below:

- \(\chi^2\) statistical significance: Test of the model’s ability to reproduce the sample variance/covariance matrix (Doll et al., 1995).
- Normed Fit Index (NFI): The NFI, as put forward by Thompson (2004), compares the \(\chi^2\) for the estimated model against the \(\chi^2\) for the independence model, assuming the complete independence of the measured variables.
- Comparative Fit Index (CFI): The CFI, like the NFI, also assesses model fit with respect to a tested and baseline model (Thompson, 2004).
• Root Mean Square Error of Approximation (RMSEA): Tabachnick and Fidell (2007) refer to RMSEA as an estimation of the lack of fit in a model as compared to a perfect (saturated) model. In other words, RMSEA also determines how well model parameters will do at reproducing population co-variances (Thompson, 2004).

• Ratio of $\chi^2$ to degrees of freedom: describes the level of efficiency of competing models in accounting for the data (Doll et al., 1995).

**Analysis of second research question.**

Reliability and validity assessments were carried out to address the second research question: “What is the technical adequacy of the ACBM in measuring preschool children’s sorting and classifying skills?” Technical adequacy is a conventional term that provides a description of the levels of reliability and validity of a particular assessment (Hampton, 2011). The technical adequacy of the ACBM was examined by means of reliability assessments, namely, internal consistency reliability (by Cronbach’s alpha), construct reliability, and inter-rater reliability; and validity of assessments, namely, content validity (was addressed in stages), construct validity (was established by the CFA), and internal validity. Potential threats to internal validity, which were history, maturation, and location, were controlled.

**Reliability assessments.**

As Fraenkel and Wallen (2009) put forward, reliability is attributed to the consistency of scores or answers from the administration of one instrument to another, and from a collection of items to one another. In other words, an assessment tool is considered reliable if it can provide consistent results. As far as the ACBM is concerned, its reliability was measured by internal consistency reliability, construct reliability, and inter-rater reliability.

**Internal consistency reliability.**
The first reliability assessment that was conducted was internal consistency reliability. It was examined by the calculation of coefficient alpha (Cronbach, 1951), also known as Cronbach’s Alpha. In fact, Cronbach’s alpha is regarded as the best reliability assessment as far as internal consistency reliability is concerned (Kline, 2000). Along the same lines, White (2011) notes that Cronbach’s alpha is the most commonly reported measurement of internal consistency for quantitative measurement instruments and is based on correlations between items. In other words, internal consistency refers to how items within a scale are strongly related to the latent variable under study (DeVellis, 2003). In regards to the present study, internal consistency was assessed by calculating Cronbach’s Alpha for all the items of the ACBM using SPSS 21.0.

**Construct reliability.**

Construct reliability (Fornell & Larcker, 1981) was the next form of reliability that was examined. The purpose of calculating construct reliability is to determine the consistency of the construct validity indicator (Hamdan, Badrullah, & Shahid, 2011). In other words, construct validity was calculated to determine to what extent the indicators on the ACBM are internally consistent. The construct reliability for both the raw model and the square root transformed model was reported. As suggested by Hamdan et al. (2011) the formula to calculate construct validity is:

Construct reliability = Square of Total Standardized Loading/Square of Total Standardized Loading + measurement error

where measurement error = 1 - (standardized loading)^2

**Inter-rater reliability.**

Instruments that make use of direct observation are sensitive to observer differences, thus requiring researchers to report the degree of scoring agreement (Fraenkel & Wallen, 2009). This
form of reliability was assessed in the following way: training that included explanations and
discussions of procedures involved in the study (Fraenkel & Wallen, 2009) was provided by the
principal investigator to the other data collector. Fraenkel and Wallen (2009) put forward that a
scoring agreement of at least 80% should be attained. In the case of this study, 25% of subjects
were randomly chosen and independent scoring was performed for this sample by the assessors.
The scores were then compared to account for inter-rater reliability.

*Test-retest reliability.*

Initially, another reliability assessment that was planned was the test-retest reliability. As
Fraenkel and Wallen (2009) note, the test-retest method requires the same instrument to be
administered a second time to the same group of participants after that a certain period of time
has elapsed. In order to demonstrate reliability, there should be at least a 3-month gap between
the two administrations of the test and the sample size should include at least 100 participants
(Kline, 2000). However, the duration of time between the administrations of the test affects the
reliability coefficient in the sense that with a longer interval of time, the reliability coefficient
will tend to be lower because changes in participants are most likely to have occurred before they
take the test a second time (Fraenkel & Wallen, 2009). For this reason, Fraenkel and Wallen
(2009) suggest the importance of choosing an appropriate time interval. In regards to this study,
to assess test-retest reliability while simultaneously considering an appropriate time interval
between the two administrations of the test, as well as fulfilling the conditions of enrolling the
minimum number of 100 participants, proved to be difficult in the following way: first, one child
care center was closed for the summer period, and this resulted in loss of nearly 50% of subjects;
and second, due to the busy summer period schedule for the other child care centers, the staff did
not get an opportunity to redistribute parent permission slips for the retest. As a matter of fact, 7
months had elapsed and it was decided not to conduct the test a second time.

Validity assessments.

Validity is the degree to which an instrument measures the intended attribute and not
some other attribute (Waltz et al., 2005). In other words, the validity of an instrument is referred
to as the degree to which results from it enable researchers to make guaranteed deductions
regarding the characteristics of the individuals under study (Fraenkel & Wallen, 2009). The
validity of the ACBM was determined through content validity, construct validity, and internal
validity. However, criterion validity was not examined due to the reasons provided below. A
description of each of the different types of validity as they pertain to this study is provided
below.

Content validity.

This type of validity designates the judgments on the content and rational format of an
assessment tool as it is to be used in a particular study (Fraenkel & Wallen, 2009). Scale
development of the ACBM was addressed in stages. The first stage was comprised of three steps:
first, the literature was searched for a conceptual definition of sorting and classifying; second,
instrument administration procedures, instructions, and test questions used at the early childhood
mathematics level were researched in the literature; and third, an operational definition of the
latent construct was derived, followed by the creation of the scales, test questions, and
procedures of administration of the ACBM. The second stage consisted of two steps: first, after
development of the initial instrument, it was pilot tested to identify any possible discrepancies;
and second, after the pilot test, the instrument was modified and presented to four experts for
feedback. These people have one or more of the following characteristics: experience in teaching
mathematics and developing mathematics curriculum at the preschool level, knowledgeable in mathematics related to early childhood education, and experience in using statistical and assessment measures. There were no suggestions for change from these people, and therefore the instrument was considered as valid in terms of content validity.

*Construct validity.*

As Thorndike and Thorndike-Christ (2010) indicate, in the field of psychology and education, the term construct is referred to as an unobservable trait that is literally created by the researcher to summarize or to explain regularities or relations in observed behavior. This form of validity characterizes the nature of the psychological construct or attributes being measured by the instrument (Fraenkel & Wallen, 2009). In other words, construct validity refers to the extent to which the entire evidence obtained is consistent with theoretical anticipation. The construct validity of the ACBM was assessed by conducting a CFA to examine the relationships between the scales on the instrument and the latent construct sorting and classifying.

*Internal validity.*

A study is said to have internal validity when any relationship found between two or more variables is distinct and directly related to these variables instead of being the result of something else (Fraenkel & Wallen, 2009). Researchers term the possible hypotheses that can explain the lack of internal validity in the outcomes of a study as threats to internal validity (Fraenkel & Wallen, 2009). The potential threats to internal validity that could have hindered the process of this study were assumed to be history, maturation, and location. Ways to minimize these potential threats to the internal validity of the study were established. These potential threats, as well as ways to minimize them, are described below.

*Potential threats to internal validity.*
The major potential threats that could have hindered the process of this study were assumed to be history, maturation, and location.

The history threat materializes when, one or more unexpected situations occur during the progression of a study, thus affecting the subjects’ responses (Fraenkel & Wallen, 2009). History was considered a potential threat to the internal validity of the study because during the course of the study, children might have been exposed to and taught about sorting and classifying activities as part of their regular curriculum.

The maturation threat occurs when, modifications during the study may be due to factors related to time that has elapsed rather than the study itself. As far as the present study is concerned, maturation was considered to be a potential threat because very young children mature quite rapidly, which could be reflected in their sorting and classifying skills.

The location threat occurs when data is collected at different locations, which might affect results. Location was identified as a potential threat to the internal validity of this study because data was collected at four different child care centers and thus four different locations.

Minimizing potential threats to internal validity.

Several ways were established to minimize the potential threats to internal validity. Below are the ways these threats have been addressed and minimized:

History threat: Teachers were asked to allow children to play with any concrete objects that they wished to during the course of the study, but were also asked not to interact with the children from a sorting and classifying perspective during the process of the study.

Maturation threat: Since the study took place during school hours and spanned a short period of time (6 weeks), the maturation threat was minimized.
Location threat: Since data was collected from four different centers, the location threat was addressed within the centers. A location for data collection that involved few distractions was selected at each center; this location remained the same for all participants of a particular center.

*Criterion validity.*

This type of validity is referred to as the relationship between the scores obtained with a particular measure and the scores obtained with one or more measures that assess the same criteria (Fraenkel & Wallen, 2009). In other words, Fraenkel and Wallen (2009) refer to this form of validity as the degree to which information obtained with one instrument is in agreement with information obtained by using other independent instruments. The qualities desired in a criterion measure include relevance, freedom from bias, reliability, and availability (Thorndike & Thorndike-Christ, 2010). Due to the lack of assessment measures for this young age group and the difficult problem of identifying or creating an acceptable criterion measure (Thorndike & Thorndike-Christ, 2010), criterion validity was not examined.
Chapter 3: Results

First Research Question: What is the contribution of each sub-construct to the latent construct on the ACBM?

A CFA was performed to answer the first research question. In order to assess the contribution of the five scales on the ACBM, the hypothesized model (Appendix E, Figures E3 and E4), before and after transformation, was evaluated through AMOS 21.0. Based on the modification indices suggested in the AMOS output, co-variances were added to the models. The results of the square root transformed model did not differ much from the results of the raw model in terms of factor loadings, squared variances, and fit indices. The following sections illustrate both models with their results.

Evaluation of model before square root transformation.

All goodness of fit indices that were examined indicated a good fit for the model. The $\chi^2$ test was not significant, $\chi^2 (2, N = 120) = 0.882, p = .643$. The ratio of the $\chi^2$ to the degrees of freedom was 0.441. The values for NFI, CFI, and RMSEA were .992, 1.000, and .000 respectively. All standardized regression coefficients (Appendix E, Figure E5) showed statistical significance at the .001 level (2-tailed). Since the standardized regression coefficients for the CFA ranged from .43 to .69, they demonstrated practical significance ($\beta > .3$). The standardized regression coefficients and their corresponding $R^2$ values were as follows: “Color,” $\beta = .43, R^2 = .18$; “Shape,” $\beta = .59, R^2 = .34$; “Size,” $\beta = .69, R^2 = .48$; “Color and Shape,” $\beta = .58, R^2 = .34$; “Explanation,” $\beta = .62, R^2 = .38$, where $R^2$ represents the percentage of variance in the indicator variable that is explained by the factor (Tabachnick & Fidell, 2007). In regards to this study, the $R^2$ values represent the percentage of variance in the sub-constructs that is explained by the latent construct.
Evaluation of model after square root transformation.

All goodness of fit indices that were examined after the square root transformation also indicated a good fit for the model. The $\chi^2$ test was not significant, $\chi^2 (2, N = 120) = 0.749, p = .688$. The ratio of the $\chi^2$ to the degrees of freedom was 0.375. The values for NFI, CFI, and RMSEA were .992, 1.000, and .000 respectively. All standardized regression coefficients (Appendix E, Figure E6) showed statistical significance at the .001 level (2-tailed). Since the standardized regression coefficients for the CFA ranged from .44 to .63, they demonstrated practical significance ($\beta > .3$). The standardized regression coefficients and their corresponding $R^2$ values were as follows: “Color,” $\beta = .44, R^2 = .19$; “Shape,” $\beta = .57, R^2 = .33$; “Size,” $\beta = .63, R^2 = .39$; “Color and Shape,” $\beta = .62, R^2 = .38$; “Explanation,” $\beta = .55, R^2 = .30$, where $R^2$ represents the percentage of variance in the indicator variable that is explained by the factor (Tabachnick & Fidell, 2007). In the context of this study, the $R^2$ values represent the percentage of variance in the sub-constructs that is explained by the latent construct.

Comparison of models.

Table 4 below shows a comparison between the model before the square root transformation and the model after the square root transformation in terms of factor loadings, standardized regression weights ($\beta$), and squared multiple correlations ($R^2$):
Table 4

*Comparison of models before and after transformation*

<table>
<thead>
<tr>
<th>Scale/Model</th>
<th>Standardized regression weights, $\beta$</th>
<th>Squared multiple correlations, $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before transformation</td>
<td>After transformation</td>
</tr>
<tr>
<td>Color</td>
<td>.43</td>
<td>.44</td>
</tr>
<tr>
<td>Shape</td>
<td>.59</td>
<td>.57</td>
</tr>
<tr>
<td>Size</td>
<td>.69</td>
<td>.63</td>
</tr>
<tr>
<td>Color and Shape</td>
<td>.58</td>
<td>.62</td>
</tr>
<tr>
<td>Explanation</td>
<td>.62</td>
<td>.55</td>
</tr>
</tbody>
</table>

Table 5 below shows a comparison between the goodness of fit indices before the square root transformation and the goodness of fit indices after the square root transformation:

Table 5

*Comparison of goodness of fit indices before and after transformation*

<table>
<thead>
<tr>
<th>Fit indices</th>
<th>Before transformation</th>
<th>After transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2 (2, N = 120)$</td>
<td>.882 ($p = .643$)</td>
<td>.749 ($p = .688$)</td>
</tr>
<tr>
<td>Ratio of the $\chi^2$ to the degrees of freedom</td>
<td>.441</td>
<td>.375</td>
</tr>
<tr>
<td>NFI</td>
<td>.992</td>
<td>.992</td>
</tr>
<tr>
<td>CFI</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>
Second Research Question: What is the technical adequacy of the ACBM in measuring preschool children’s sorting and classifying skills?

**Reliability.**

Reliability was assessed in terms of internal consistency reliability, construct reliability, and inter-rater reliability. These are described below:

**Internal consistency reliability.**

The internal consistency reliability of the ACBM was assessed through the calculation of Cronbach’s Alpha for all the items on the ACBM using SPSS, version 21.0. The resulting Cronbach’s Alpha was .730.

**Construct reliability.**

The construct reliability was calculated for both the raw model and the square root transformed model by using the formula mentioned previously in the data analysis plan. The resulting construct reliability value for the raw data was .721 and the resulting construct reliability value for the square root transformed data and was .700.

**Inter-rater reliability.**

As far as inter-rater reliability is concerned, out of 120 assessments, 30 (25%) were co-scored and a 100% inter-rater agreement was attained.

**Validity.**

The validity of the assessment tool was assessed in terms of content validity, construct validity, and internal validity. The content validity and internal validity of the assessment tool were discussed in the data analysis plan. The construct validity of the ACBM is discussed below.

**Construct validity.**
After performing a CFA with both the raw and square root transformed data, the factor loadings of each scale onto the latent construct, for both models, were examined to assess the construct validity of the ACBM. In other words, a CFA was performed to examine the relationships between the scales on the instrument and the latent construct sorting and classifying. The factor loadings for each scale of the raw model were as follows: $\beta = .43$ for “Color”; $\beta = .59$ for “Shape”; $\beta = .69$ for “Size”; $\beta = .58$ for “Color and Shape”; and $\beta = .62$ for “Explanation.” The factor loadings for each scale of the square root transformed model were as follows: $\beta = .44$ for “Color”; $\beta = .57$ for “Shape”; $\beta = .63$ for “Size”; $\beta = .62$ for “Color and Shape”; and $\beta = .55$ for “Explanation.” Since all these standardized regression coefficients for both models were above .30, the ACBM showed practical significance as well as statistical significance at the .001 level (2-tailed).
Chapter 4: Discussion

The purpose of this research study was to design and validate an assessment tool entitled the Algebra Curriculum Based Measure (ACBM) with the intent of measuring preschool children’s sorting and classifying skills. Specifically, it was hypothesized that the five sub-constructs, namely, color, shape, size, color and shape, and explanation, would contribute significantly to the latent construct sorting and classifying, thus making the ACBM a significant indicator of preschool children’s sorting and classifying skills. Two research questions pertained to this study. The first research question examined the level of contribution of each of the five sub-constructs to the latent construct. The second research question examined the technical adequacy of the ACBM.

Major Findings

Two analyses were run with AMOS 21.0, one with the raw data and one with the square root transformed data, thus giving rise to two models. The CFA of both models showed similar results in the sense that all five scales significantly contributed to the latent construct, thus demonstrating that these five scales are significant indicators of preschool children’s sorting and classifying skills. Furthermore, a good fit between the corresponding hypothesized and observed models was obtained. Factor loadings demonstrated statistical as well as practical significance. In regards to the second question, which examined the technical adequacy of the assessment tool, results indicated that the ACBM is a reliable and valid measure of sorting and classifying skills for this sample of preschool children.

First research question.

A CFA was performed with both the raw data and the square root transformed data to evaluate the model fit between the hypothesized and observed models. Specifically, the CFA was
used to answer the first research question and to test the guiding hypothesis of the study. The addition of three co-variances based on the modification indices suggested in the AMOS output significantly improved both models to provide good-fitting models. Results about the model fits are as follows:

- **χ² statistical significance**: As put forward by Tabachnick and Fidell (2007), a non-significant χ² is desired, thus indicating no significant difference between the hypothesized and observed relationships. The chi-square for both models was not significant. For the raw model, the results were as follows: χ²(2, N = 120) = .882, p = .643; and for the square root transformed model, the results were as follows: χ²(2, N = 120) = 0.749, p = .688.

- **Normed Fit Index (NFI)**: High NFI values of .95 or more are considered to reasonably fit the model (Thompson, 2004). Similarly, Tabachnick and Fidell (2007) stress that NFI values greater than .95 indicate a model of good fit. For both the raw model and the square root transformed model, the NFI was .992, indicating a good model fit.

- **Comparative Fit Index (CFI)**: As Thompson (2004) indicated, values close to 1.000 are preferred, with high CFI values of .95 or more considered to reasonably fit a model. Similarly, Hu and Bentler (1999) put forward that CFI values of .95 or greater indicate a good model fit. The CFI value for both the raw model and the square root transformed model was 1.000, which indicates a good model fit for the data.

- **Root Mean Square Error of Approximation (RMSEA)**: A RMSEA value of .06 or less generally represents a reasonable model fit (Browne & Cudeck, 1992; Tabachnick & Fidell, 2007; Thompson, 2004). Both the raw model and the square root transformed model reached a RMSEA value of .000, which indicates a good model fit for the data.
• Ratio of $\chi^2$ to degrees of freedom: A ratio of the $\chi^2$ to the degrees of freedom of less than 2 indicates a good-fitting model (Tabachnick & Fidell, 2007). The ratio of the $\chi^2$ to the degrees of freedom for the raw model was .441 and the ratio of the $\chi^2$ to the degrees of freedom for the square root transformed model was .375. In both cases, these values indicate a good fit between the hypothesized and measured models.

After examining model fit, factor loadings of both models were examined for statistical and practical significance. Factor loadings greater than .30 can be regarded as significant (Kline, 2000) and a factor loading of over .30 is considered to be the minimum level of practical significance (Hair, Black, Babin, Anderson & Tatham, 2006). Therefore, since factor loadings for the raw model ranged from .43 to .69, and factor loadings for the square root transformed model ranged from .44 to .63, they all demonstrated practical significance ($\beta > .3$). In addition, the greater the loading is, the more the variable is considered to be a “pure measure of the factor” (Tabachnick & Fidell, 2007, p. 649). In regards to the two models, factor loadings ranging from .43 to .69 and 44 to .63 were obtained, thus indicating acceptable factor loadings that also demonstrated statistical significance at $\alpha = .001$ (2-tailed).

**Second research question.**

As far as the second research question is concerned, the technical adequacy of the ACBM was examined. Various reliability and validity assessments were conducted to examine the extent to which the ACBM is a reliable and valid assessment tool in measuring preschool children’s sorting and classifying skills. In terms of reliability assessments, internal consistency reliability, construct reliability, and inter-rater reliability were conducted. The validity of the assessment tool was assessed by content validity, construct validity, and internal validity. These different types of reliability and validity, as applicable to the ACBM, are discussed below.
Reliability.

Internal consistency reliability.

As calculated from the items on the ACBM, the resulting Cronbach’s Alpha was .730. As DeVellis (2003) suggests, a Cronbach’s alpha value below .60 is unacceptable; between .60 and .65 is undesirable; between .65 and .70 is minimally acceptable; between .70 and .80 is respectable; between .80 and .90 is very good; and higher than .90 needs a decrease in the length of the scale. Therefore, a Cronbach’s alpha of .730 on the ACBM is considered to be a respectable value for reliability.

Construct reliability.

Hair et al. (2006) put forward that the desired value for construct reliability should be .70 or higher. The resulting construct reliability value for the raw model was .721 and the resulting construct reliability value for the square root transformed model and was .700.

Inter-rater reliability.

Inter-rater reliability was checked for all assessors to determine the percentage of inter-rater agreement between the assessors on co-scored assessments. In this regard, Fraenkel and Wallen (2009) indicated that a scoring agreement of at least 80% should be attained. As far as this study is concerned, out of 120 assessments, 30 (25%) were co-scored and an inter-rater agreement of 100% was attained.

Validity.

Content validity.

Important aspects of content validity consist of (a) the format of the assessment tool, including language appropriateness and clear directions; and (b) having individuals who are knowledgeable about the content area and research context review the assessment tool (Fraenkel
& Wallen, 2009). Fraenkel and Wallen (2009) suggest the following procedure of addressing content validity: The researcher starts by writing a definition of what s/he intends to measure. Judges, who might be colleagues who are knowledgeable about the research context, can be asked to review and provide feedback on the items included in the instrument and on the content areas that should be included or excluded from it (DeVellis, 2003).

Content validity was addressed in stages, as described in the data analysis plan: theoretical definition of latent construct; literature search for instructions, and test questions used at the early childhood mathematics level; creation of the scales, test questions, and procedures of administration of the ACBM; pilot testing; revision of the assessment tool; and finally presentation of the assessment tool to professionals for feedback. Since there were no suggestions for change from these people, the ACBM was considered as valid in terms of content validity.

*Construct validity.*

Construct validity was established by the CFA, where acceptable factor loadings were obtained for both models. For the raw model, the factor loadings ranged from .43 to .69 and for the square root transformed model, the factor loadings ranged from .44 to .63. All these factor loadings demonstrated practical significance since they were all greater than .30 (Hair et al., 2006) as well as statistical significance at $\alpha = .001$ (2-tailed).

*Internal validity.*

The three potential threats to internal validity, namely, history, maturation, and location were controlled by the researcher. Since the strategies to address and minimize these threats during the course of the study were successful, the ACBM is considered internally valid.
Based on the results obtained, the hypothesis was supported by this research. This was indicated by the following: a good model fit, as demonstrated by fit indices examined for both the raw and the transformed model; factor loadings demonstrating statistical significance at $\alpha = .001$ (2-tailed) as well as practical significance for both models; and by the various reliability and validity assessments conducted. Therefore, the ACBM can be considered as a reliable and valid measure of the sorting and classifying skills for this sample of preschool children.

**Other Findings**

The statistical results showed that the ACBM proved to be a valid and reliable tool in measuring preschool children’s sorting and classifying skills. Both levels of the assessment tool, that is, sorting and classifying objects by one attribute and sorting and classifying objects by two attributes simultaneously, indicated that children at the ages of 3 to 5 years encountered some level of difficulty. In particular, sorting and classifying by one attribute was found to be challenging to many children because they were initially provided with no directions regarding how to sort and classify the objects as well as what attributes to use. Indeed, prior research showed that 4-year-olds need explicit sorting instruction to perform activities because most often they rely upon naming and visualizing objects that should be remembered (Baker-Ward et al., 1984). Other researchers (Inhelder & Piaget, 1964; Mandler & Stephens, 1967; Vygotsky, 1934, 1962) reported that when no specific direction for classification tasks were provided to 2- to 5-year-old children, they had the tendency to sort objects by switching from one attribute to another attribute, or they might also select another attribute found around them (Denney & Acito, 1974). A more recent study by Deák et al. (2002) demonstrated that 75% of 3-year-old children who received instructions about classifying objects by function were successful in doing so and 97% who received instructions about sorting by shape did so correctly. However, the authors
found that half of the 4-year-olds who did not receive specific instructions were able to match objects by shape. This finding was closely related to the results of the present study, which indicated that 46% of 4-year-olds were able to sort and classify by shape independently. A second finding from Deák et al.’s (2002) study that is closely related to the findings of the present study is that irrespective of attribute, 75% of 4-year-olds could sort independently as compared to 70% for the present study.

Regardless of age, on the ACBM, children were more inclined to sort and classify objects by shape (35 children; 29.2% of sample), followed by color (28 children; 23.3% of sample), and finally by size (6 children; 5% of sample) when they were not provided with any specific instructions. On the whole, regardless of the choice of attributes on the first level of the ACBM where the child is required to sort and classify the objects by one attribute without receiving instructions (that is, with no modeling or prompts), 57.5% of the children were able to sort and classify the objects based on the first attribute they selected; 18.3% of children could do so based on the second attribute they selected regardless of whether they succeeded in sorting and classifying the objects by their first chosen attribute; and 4.2% of children could do so based on the third attribute they selected regardless of whether they succeeded in sorting and classifying the objects by their first and second chosen attributes. These percentages indicate that as the assessment progresses, it becomes more challenging for the children to identify a different attribute.

Compared to the first level of the assessment tool, at the second level (sorting by two attributes simultaneously) children in this study were given clear and specific instructions regarding how to sort and classify the objects by two particular attributes simultaneously. This activity was also found to be challenging to some children despite clear instructions on how to
sort and classify the objects, which were relatively few in number. The majority of the children needed a model to get started on the activity at this level. Results showed that while only 27.5% (33 children) of the sample was successful in starting the task without receiving help from the assessor, 72.5% (87 children) needed a model to start the sorting and classifying activity. This result is aligned with previous research (Inhelder & Piaget, 1964; Mandler & Stephens, 1967; Vygotsky, 1934, 1962) which revealed that it is only after the age of 8 that children are able to sort and classify objects independently by several attributes simultaneously.

As mentioned above, some children in this study demonstrated the ability to sort and classify by one attribute and two attributes simultaneously without assistance from the assessor. On the first level of the assessment tool, where the child was initially asked to sort and classify the objects by an attribute of his/her choice, findings showed that all three attributes, color, shape, and size, were chosen: 23.3% (28 children) for color; 29.2% (35 children) for shape; and 5% (6 children) for size. In general, for the entire assessment at this first level, while 28.3% (34 children) were able to sort and classify by color without a model, 71.7% (86 children) needed a model to sort by color. While 40% (48 children) were able to sort and classify by shape without a model, 60% (72 children) needed a model to sort by shape. While 11.7% (14 children) were able to sort and classify by size without a model, 88.3% (106 children) needed a model to sort by size.

In this regard, as previous literature (Deák et al., 2002) revealed, children at the preschool age, when not obtaining specific instructions, tend to sort and classify objects by shape rather than color or size.

When children were asked to sort and classify the objects by an attribute that was different from their original choice, results showed that 15% (18 children) were able to abstract a second attribute to use to group the objects and at least begin sorting without a model. If the first
attribute chosen was shape, the second attribute chosen was usually color; conversely, if the first attribute chosen was color, the second attribute chosen was usually shape. Finally, when asked to sort by a third attribute, out of the 18 children who were able to abstract a second attribute, only one child was able to abstract a third attribute to use to group the objects and at least begin sorting without a model. Regarding the second level of the ACBM, where the child was required to sort and classify the objects by color and shape simultaneously, 27.5% (33 children) were able to do so without assistance from the assessor and 72.5% (87 children) needed a modeling activity to start the sorting and classifying task.

Children’s sorting and classifying skills were also determined by the question posed to them after each sorting and classifying activity: “How did you decide which objects to put in your bowls?” The inclusion of this question after each sorting and classifying activity was intended to measure the extent to which preschool children can explain their sorting and classifying strategies, that is, how they created the different sets based on particular attribute(s). Results showed that the majority of children were unsuccessful in responding to this question, regardless of the level on the ACBM, which indicates that while many children can perform sorting tasks, particularly with some instructions or a model, they cannot explain how they performed the task. As Piaget (1971) indicates, many young children are able to sort and classify objects correctly, but may not always be aware of the attributes being used (e.g. color, shape), and this may be wrongly considered as a lack of knowledge, or inability to sort and classify (Halpenny & Petterson, 2013). Piaget further notes that during the pre-operational stage of cognitive development, children between the ages of 2 and 7 do not have a clear understanding of concrete logic, that is, they usually think in a non-logical way and thus cannot organize information mentally. In other words, children can perform some mental operations, but do not
appear to be familiar with the strategies involved because they cannot explain them (Hetherington & Parke, 2002).

**Limitations**

Two limitations pertain to this study. First, as mentioned in the data analysis plan, because of some unforeseen circumstances, it was not possible to collect data again with the same participants within an appropriate time frame to account for test-retest reliability. Second, the sample for this study comprises about twice the number of children from middle- to upper-income families as compared to the number of children from lower-income families. Since the percentage of families whose income is below poverty level for Hamilton county for the year 2012 is about 16% (Hamilton County, Ohio Demographics, 2013), this indicates that the sample for this study under-represents children who come from families with income above the poverty line and over-represents children who come from families with income below the poverty line. Furthermore, to be considered as a fully validated instrument, the sample assessed on the ACBM should come from various geographical areas.

**Implications**

Within the algebra strand, it is expected that all PreK-2 students should "…sort, classify and order objects by size, number and other properties" (NCTM, 2000, p. 90). Further, NCTM indicates that algebraic thinking in young children is a crucial element that contributes to the development of other strands of mathematics. Young children’s sorting and classifying skills significantly promote the development of other mathematical areas, such as patterning, data analysis, problem solving, number sense, and the algebra strand in general.

Children in this sample were found to be more able to sort and classify objects by color and shape than by size and by two attributes simultaneously. Furthermore, most of the children were unable to explain how they sorted and classified the objects according to particular
attribute(s). Therefore, it is recommended that teachers put more emphasis on developing preschool children’s awareness of other attributes, with an awareness that sorting by two attributes simultaneously is challenging for most preschool children. Teachers may wish to encourage children to think about how they decided to sort and classify objects based on one or more specific attributes. By eliciting the basis of criteria children are using when sorting and classifying objects into groups, which is a fundamental component in the instruction of algebra, teachers prompt them to think that two different situations or objects can have the same attribute, and thus be grouped accordingly. Asking these types of questions to children is a way for teachers to help children understand and be aware that sorting and classifying could be described as the creation of distinct groups based on the same and/or different attribute(s), thus promoting their algebraic thinking and reasoning skills. This procedure is an early introduction to the concept of algebra to children, thus integrating at least part of the algebra strand in early childhood mathematics.

On the whole, the results of this study indicate that children’s ability to sort and classify can be measured on the ACBM and that some aspects of sorting and classifying are harder than others. In this regard, it is recommended that teachers implement scaffolding by modeling or prompting in their instruction. By initially helping children in sorting and classification tasks and by gradually reducing the amount of help as children feel more comfortable with these types of tasks, teachers are in a better position to promote children’s performance in sorting and classifying by one and two attributes simultaneously.

**Considerations for Future Research**

Several considerations for future research emerged from this study. First, since test-retest reliability could not be performed due to previously mentioned circumstances, it would be
critical to find if there is a significant correlation between the initial testing and the retest. Therefore, future use of the ACBM could include retesting at least 100 participants after a 3-month interval. Second, parallel versions of the ACBM could be developed and tested to provide teachers with repeated probes for ongoing measurement purposes. Third, different objects, as well as different attributes, could be used. This would provide an idea of whether differences in object form and object attribute have an effect on children’s sorting and classifying skills. Fourth, the ACBM could be used with children attending different child care programs as well as different geographical regions. Effectiveness of programs could be determined by examining children’s performance on the assessment. In addition, teachers using the ACBM could also determine if the tool informs their curricular activities.

Conclusion

This research demonstrates that a curriculum-based measure such as the ACBM can accurately measure the development of sorting and classifying skills in young children. Confirmatory factor analysis demonstrated that the scales, “Color,” “Shape,” “Size,” “Color and Shape,” and “Explanation” on the ACBM were able to measure the latent construct of sorting and classifying, implying that the tool was a reliable and valid measure of preschool children’s sorting and classifying ability. The ACBM can now be used for formative and summative assessment of children, for program evaluation, and for further research related to children’s emergent understanding of sorting and classifying.
References


*Psychological Science, 17*(8), 665-669.


Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3303449)


### Table A1

*Table of Specifications for the Algebra Curriculum Based Measure*

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Content</th>
<th>Composition of sets/explanation</th>
<th>Total number of objects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sorts by color</strong></td>
<td>Creation of three sets based on color</td>
<td>9 red objects 9 blue objects 9 yellow objects</td>
<td>27</td>
</tr>
<tr>
<td><strong>Sorts by shape</strong></td>
<td>Creation of three sets based on shape</td>
<td>9 circles 9 squares 9 hearts</td>
<td>27</td>
</tr>
<tr>
<td><strong>Sorts by size</strong></td>
<td>Creation of three sets based on size</td>
<td>9 big objects 9 medium objects 9 small objects</td>
<td>27</td>
</tr>
<tr>
<td><strong>Sorts by color and shape simultaneously</strong></td>
<td>Creation of four sets based on color and shape simultaneously</td>
<td>Blue: 3 circles and 3 squares Yellow: 3 circles and 3 squares</td>
<td>12</td>
</tr>
<tr>
<td><strong>Explains sorting and classifying strategies</strong></td>
<td>Question: “How did you decide which objects to put in your bowls?”</td>
<td>Question posed after each of the four objectives above: sorting by color; shape; size; color and shape simultaneously</td>
<td>N/A</td>
</tr>
</tbody>
</table>
## Appendix B

**Algebra Curriculum Based Measure**

Table B1

<table>
<thead>
<tr>
<th>Student ID:</th>
<th>Center:</th>
<th>Date:</th>
<th>Assessor:</th>
<th>Classroom:</th>
<th>Time taken:</th>
<th>Comments</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Instructions</th>
<th>Child response/Decisions</th>
<th>Attribute selected</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Circle as appropriate (A, B, C, D, E) and proceed as indicated.</td>
<td>circle one</td>
<td>Circle child’s responses.</td>
<td></td>
</tr>
</tbody>
</table>

### 1.a 1st attribute

Assign a name.

*Assessor says: “We’re going to play a game with these objects. Find the ones that you think should go together. You can put them in this bowl.”*

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Instructions</th>
<th>Child response/Decisions</th>
<th>Attribute selected</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.a</td>
<td></td>
<td>A. Child sorts all 3 groups</td>
<td>Color</td>
<td>R ○ • • ■ ■ ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B. Child sorts 1 set (e.g. red)</td>
<td>Shape</td>
<td>B ● ● ● ■ ■ ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C. Child sorts part of set /makes some errors (e.g. 4 out of 9 red or RRRRYBB)</td>
<td>Size</td>
<td>Y ● ● ● ■ ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D. Child randomly groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E. No response from child</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 1.b Modeling

Assign a name.

*Assessor says: “Look, child’s name, I could put red objects together in this bowl.”*

*Assessor places the large red circle, medium red square, and small red heart in the bowl.

*Assessor says: “What else”*

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Instructions</th>
<th>Child response/Decisions</th>
<th>Attribute selected</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.b</td>
<td></td>
<td>A. Child sorts all 3 groups</td>
<td>Color</td>
<td>R ○ ● ● ■ ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B. Child sorts 1 set (e.g. red)</td>
<td>Shape</td>
<td>B ● ● ● ■ ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C. Child sorts part of set /makes some errors (e.g. 4 out of 9 red or RRRRYBB)</td>
<td>Size</td>
<td>Y ● ● ● ■ ■ ■</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D. Child randomly groups</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 1.c Options B or C

<table>
<thead>
<tr>
<th>A. Child sorts all 3 groups</th>
<th>B. Child has sorted 1 set (e.g. red)</th>
<th>C. Child has sorted some objects (e.g. 4 out of 9 reds)</th>
<th>D. Some errors</th>
<th>E. Random</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assessor points to the remaining objects and says: “What about these? Do any of them go together? You can sort them into these bowls.” Put a circle around objects that go into the 2nd bowl and a square around objects that go into the 3rd bowl. <em>Then go to 1.d.</em></td>
<td>Assessor points to the remaining objects and says: “Are there any more that should go in your bowl?” Put an X through newly selected objects. <em>Then give prompt for item B and score.</em></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Performance of task**

| N/A | R ● ● ● [square] ● ● ● ● ● | R ● ● ● [square] ● ● ● ● ● | N/A | N/A | N/A |
| N/A | B ● ● ● [square] ● ● ● ● ● | B ● ● ● [square] ● ● ● ● ● |               |               |               |
| N/A | Y ● ● ● [square] ● ● ● ● ● | Y ● ● ● [square] ● ● ● ● ● |               |               |               |

### 1.d

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Child’s Response</th>
<th>Additional Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor says: “How did you decide which objects to put in your bowls?”</td>
<td>Record exactly what child says, or as closely as possible</td>
<td></td>
</tr>
</tbody>
</table>

### Item

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Child Response/Decisions</th>
<th>Attribute selected circle one</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle as appropriate (A, B, C, D, E) and proceed as indicated.</td>
<td>Circle child’s responses.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 2.a 2nd attribute

**Assessor says:** “You sorted all these objects by attribute in cell 1a. Is there another way to sort them? What is a different way to put these objects together?”

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| A. | Child sorts all 3 groups  
**Go to 2.d.** |
| B. | Child sorts 1 set (e.g. circles)  
**Go to 2.c.** |
| C. | Child sorts part of set/makes some errors (e.g. 4 out of 9 circles/5 circles, 2 hearts, 1 square)  
**Go to 2.c.** |
| D. | Child randomly groups  
**Go to 2.b below** |
| E. | No response from child  
**Go to 2.b below** |

#### Color

<table>
<thead>
<tr>
<th>R</th>
<th>○</th>
<th>●</th>
<th>■</th>
<th>◆</th>
<th>▼</th>
<th>▼</th>
</tr>
</thead>
</table>

#### Shape

<table>
<thead>
<tr>
<th>B</th>
<th>○</th>
<th>●</th>
<th>■</th>
<th>◆</th>
<th>▼</th>
<th>▼</th>
</tr>
</thead>
</table>

#### Size

<table>
<thead>
<tr>
<th>Y</th>
<th>○</th>
<th>●</th>
<th>■</th>
<th>◆</th>
<th>▼</th>
<th>▼</th>
</tr>
</thead>
</table>

### 2.b Modeling

**Assessor says:** “Look, child’s name. I could put circles together in this bowl.”  
Assessor places the large red circle, medium blue circle, and small red circle in the bowl.  
Assessor says: “What else should go here?”

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| A. | Child sorts all 3 groups  
**Go to 2.d.** |
| B. | Child sorts 1 set (e.g. circles)  
**Go to 2.c.** |
| C. | Child sorts part of set/makes some errors (e.g. 4 out of 9 circles/5 circles, 2 hearts, 1 square)  
**Go to 2.c.** |
| D. | Child randomly groups  
**Go to 2.d.** |
| E. | No response from child  
**Go to 2.b below** |

#### Shape

<table>
<thead>
<tr>
<th>R</th>
<th>○</th>
<th>●</th>
<th>■</th>
<th>◆</th>
<th>▼</th>
<th>▼</th>
</tr>
</thead>
</table>

### 2.c Options B or C

A. Child sorts all 3 groups  
**Go to 2.d.**

B. Child has sorted 1 set (e.g. circles)  
**Go to 2.c.**

C. Child has sorted some objects (e.g. 4 out of 9 circles)

D. Some errors

E. Random

| Prompt | Assessor points to the remaining objects and says: “What about these? Do any of them go together? You can sort them into these bowls.”  
Put a circle around objects that go into the 2nd bowl and a square around objects that go into the 3rd bowl. **Then go to 1.d.** |
|---|---|
| Assessor points to the remaining objects and says: “Are there any more that should go in your bowl?”  
Put an X through newly selected objects.  
**Then give prompt for item B and score.** |
<p>| E. Random | N/A |
| N/A | N/A |</p>
<table>
<thead>
<tr>
<th>Performance of task</th>
<th>N/A</th>
<th>R ● ● ● ● □ □ □ □ ● ● ●</th>
<th>R ● ● ● ● □ □ □ □ ● ● ●</th>
<th>N/A</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B ● ● ● ● □ □ □ □ ● ● ●</td>
<td>B ● ● ● ● □ □ □ □ ● ● ●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y ● ● ● ● □ □ □ □ ● ● ●</td>
<td>Y ● ● ● ● □ □ □ □ ● ● ●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2.d</th>
<th>Instructions</th>
<th>Child’s Response</th>
<th>Additional Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options A - C</td>
<td>Assessor says: “How did you decide which objects to put in your bowls?”</td>
<td>Record exactly what child says, or as closely as possible.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Instructions</th>
<th>Child Response/Decisions</th>
<th>Attribute selected</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.a 3rd attribute</td>
<td>Assessor says: “You sorted all these objects by attributes in cell 1a and 2a. Is there still another way to sort them? What is a different way to put these objects together?”</td>
<td>A. Child sorts all 3 groups Go to 3.d. B. Child sorts 1 set (e.g. big) Go to 3.c. C. Child sorts part of set /makes some errors (e.g. 4 out of 9 big/4 big, 1 medium, 2 small ) Go to 3.c. D. Child randomly groups Go to 3.b. below. E. No response from child Go to 3.b. below</td>
<td>Color</td>
<td>R ● ● ● ● □ □ □ □ ● ● ●</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Shape</td>
<td>B ● ● ● ● □ □ □ □ ● ● ●</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Size</td>
<td>Y ● ● ● ● □ □ □ □ ● ● ●</td>
<td></td>
</tr>
</tbody>
</table>
### 3.b Modeling

- **Assessor says:** “Look, child’s name, I could put big objects together in this bowl.”
- Assessor places the large red circle, large blue square, and large yellow heart in the bowl.
- Assessor says: “What else should go here?”

**Options**

- **A. Child sorts all 3 groups**
  - Go to 3.d.

- **B. Child sorts 1 set (e.g. big)**
  - Go to 3.c.

- **C. Child sorts part of set /makes some errors (e.g. 4 out of 9 big/4 big, 1 medium, 2 small)**
  - Go to 3.c.

- **D. Child randomly groups**
  - Go to 3.d.

- **E. No response from child**
  - Go to 4.a. and ask child to sort by color and shape

### 3.c Options B or C

<table>
<thead>
<tr>
<th>Prompt</th>
<th>N/A</th>
<th>A. Child sorts all 3 groups</th>
<th>B. Child has sorted 1 set (e.g. big)</th>
<th>C. Child has sorted some objects (e.g. 4 out of 9 big)</th>
<th>D. Some errors</th>
<th>E. Random</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assesssor points to the remaining objects and says: “What about these? Do any of them go together? You can sort them into these bowls.”</td>
<td>N/A</td>
<td>Assessor points to the remaining objects and says: “Are there any more that should go in your bowl?”</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Put a circle around objects that go into the 2nd bowl and a square around objects that go into the 3rd bowl. Then go to 3.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Then give prompt for item B and score.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.d

**Instructions**

Record exactly what child says, or as closely as possible

**Additional Comments**

---

122
### Options

A - C

<table>
<thead>
<tr>
<th>Assessor says: “How did you decide which objects to put in your bowls?”</th>
</tr>
</thead>
</table>

### Item

#### 4.a Sorting by 2 attributes - color and shape

- **Instructions**: Assessor removes the red items, and all the hearts, and puts out 4 small bowls. Assessor says: “Let’s try something else. Can you find all the objects that have the same color and the same shape?”

<table>
<thead>
<tr>
<th>Child response/Decisions</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Child sorts all 4 groups correctly; Go to 4.d</td>
<td>B ○ ● ● ■ ■</td>
<td></td>
</tr>
<tr>
<td>B. Child sorts 2 groups correctly; Go to 4.c</td>
<td>Y ● ● ● ■ ■</td>
<td></td>
</tr>
<tr>
<td>C. Child sorts 1 group correctly; Go to 4.c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Child sorts by only one attribute (color or shape) Go to 4.b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Child randomly groups objects; Go to 4.b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. No response from child; Go to 4.b</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 4.b Modeling

- **Instructions**: Assessor says: “Look, child’s name, I could put blue circles together. They are the same color and the same shape.” Assessor places the large and small blue circles in the bowl. Assessor says: “What other objects are the same color and the same shape? You can put them in these bowls.”

<table>
<thead>
<tr>
<th>Child response/Decisions</th>
<th>Performance of task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Child sorts all 4 sets correctly; Go to 4.d</td>
<td>B ○ ○ ● ■ ■</td>
<td></td>
</tr>
<tr>
<td>B. Child sorts 2 sets correctly; Go to 4.c</td>
<td>Y ● ● ● ■ ■</td>
<td></td>
</tr>
<tr>
<td>C. Child sorts 1 set correctly; Go to 4.c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Child sorts by only one attribute (color or shape) Go to 4.d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Child randomly groups objects; Go to 4.d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. No response from child; End assessment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.c Options (A, B, C, D)

#### Options

<table>
<thead>
<tr>
<th>A. Child sorts all 4 groups</th>
<th>B. Child sorts 2 sets correctly</th>
<th>C. Child sorts 1 group correctly</th>
<th>D. Random grouping</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor points to the remaining objects and says:</td>
<td>Assessor points to the remaining objects and says:</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Performance of task</td>
<td>N/A</td>
<td>B ● ● ● ■ ■ ■</td>
<td>B ● ● ● ■ ■ ■</td>
<td>N/A</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----</td>
<td>----------------</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>Y ● ● ● ■ ■ ■</td>
<td>Y ● ● ● ■ ■ ■</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.d</th>
<th>Instructions</th>
<th>Child’s Response</th>
<th>Additional Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options A - E</td>
<td>Assessor says: “How did you decide which objects to put in your bowls?”</td>
<td>Record exactly what child says, or as closely as possible</td>
<td></td>
</tr>
</tbody>
</table>
Table B2
Assessment and Scoring Guide

<table>
<thead>
<tr>
<th>Child’s trajectory</th>
<th>Options</th>
<th>Child’s responses</th>
<th>Example</th>
<th>Instructions</th>
<th>Scoring (number of objects correctly selected)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Child creates 3 sets</td>
<td>All 9 red objects together + all 9 blue objects together + all 9 yellow objects together</td>
<td>Go to 1.d.</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Child creates 1 set</td>
<td>(1) All 9 red objects together (2) All 9 circles together (3) All 9 big objects together</td>
<td>(a) Go to 1.cB* [count number of additional objects to complete group(s)] (b) Then go to 1.d.</td>
<td>9; 11-27</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Child creates part of set (minimum 4 objects)</td>
<td>(1) 4 out of 9 red objects (2) 4 out of 9 circles (3) 4 out of 9 medium objects</td>
<td>(a) Go to 1.cC** – and 1.cB if needed- [count number of additional objects to complete group(s)] (b) Then go to 1.d.</td>
<td>4-27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>1 big red circle + 1 medium blue square + 1 small yellow heart</td>
<td>Go to 1.b</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------</td>
<td>---------------------------------------------------------------</td>
<td>-----------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>No response from child</td>
<td>N/A</td>
<td>Go to 1.b</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Child creates 3 sets</td>
<td>All 6 red objects together + all 9 blue objects together + all 9 yellow objects together</td>
<td>Go to 1.d.</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>(1) All 6 red objects together</td>
<td>(a) Go to 1.cB*** [count number of additional objects to complete group(s)]</td>
<td>6; 8-24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) All 9 blue objects together</td>
<td>(b) Then go to 1.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) All 9 yellow objects together</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>(1) Add 1 red object</td>
<td>(a) Go to 1.cC***** – and 1.cB if needed- [count number of additional objects to complete group(s)]</td>
<td>1-24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) Add 2**** blue objects</td>
<td>(b) Then go to 1.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) Add 4 yellow objects</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model (e.g. 3 red objects modeled by assessor)**
### Sorting by Two Attributes

<table>
<thead>
<tr>
<th>Child’s trajectory</th>
<th>Options</th>
<th>Child’s responses</th>
<th>Example</th>
<th>Instructions</th>
<th>Scoring (number of objects correctly selected)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self</strong></td>
<td>A</td>
<td>Child creates 4 sets correctly (3 objects in each set)</td>
<td>1 set of blue squares + 1 set of blue circles + 1 set of yellow squares + 1 set of yellow circles</td>
<td>Go to 4.d</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Child creates 2 sets correctly (3 objects in each set)</td>
<td>1 set of blue squares + 1 set of blue circles</td>
<td>(a) Go to 4.cB. (count number of objects)</td>
<td>6; 8-12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b) Then go to 4.d.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Child creates 1 set correctly (3 objects in the set)</td>
<td>1 set of blue squares</td>
<td>(a) Go to 4.cC. (count number of objects)</td>
<td>3; 5-12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b) Then go to 4.d.</td>
<td></td>
</tr>
</tbody>
</table>
| Model (1 big blue circle and 1 small blue circle modeled by assessor) | D | Child creates set(s) based on only one attribute | (1) Sorting by color only (e.g. blue objects)  
(2) Sorting by shape only (e.g. circles) | Go to 4.b | 0 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>Child randomly creates sets of objects</td>
<td>1 big blue circle + 1 small yellow square</td>
<td>Go to 4.b</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>No response from child</td>
<td>N/A</td>
<td>Go to 4.b</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Child creates 4 sets correctly</td>
<td>1 set of blue circles (1 object) + 1 set of blue squares (3 objects) + 1 set of yellow squares (3 objects) + 1 set of yellow circles (3 objects)</td>
<td>Go to 4.d.</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>
| B | Child creates 2 sets correctly | 1 set of blue circles (1 object) + 1 set of blue squares (3 objects) | (a) Go to 4.cB. (count number of objects)  
(b) Then go to 4.d. | 4; 6-10 |
| C | Child creates 1 set correctly | (1) 1 set of blue circles (1 object)  
(2) 1 set of yellow squares (3 objects) | (a) Go to 4.cC. (count number of objects)  
(b) Then go to 4.d. | 1; 3-10 |
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| D | Child creates sets on only one attribute | (1) Sorting by color only (e.g. blue objects)
(2) Sorting by shape only (e.g. circles) | Go to 4.d. | 0 |
| E | Child randomly creates sets of objects | 1 big blue square + 1 small yellow circle | Go to 4.d. | 0 |
| F | No response from child | End assessment | N/A | 0 |

* Number of additional objects should pertain to the selected attribute and should include a minimum of 2 objects of one subset of the same attribute (e.g. a group of 9 reds followed by 2 blues after prompt)

** A total number of 9 objects selected by the child does not always represent a group of a single subset of an attribute but may also represent the score obtained when a child sorts part of set. For example, the child is given a raw score of 9 IF there is at least a group of 4 objects of a single subset of an attribute included in his/her set, in addition to at least 2 objects in other subsets of the same attribute (e.g. 4 reds, 3 blues, 2 yellows OR 5 reds, 4 blues, 1 yellow)

***after completing sorting of modeled set, child should add at least 2 objects of another set of the same modeled attribute (e.g. after completing sorting of modeling of red color, child groups 2 blue objects)

****regardless of whether child adds objects to modeled set, a score of 2 is attributed if the child groups 2 objects of another set of the same modeled attribute (e.g. after modeling red color, child groups 2 blue objects)

*****number of additional objects should pertain to the selected attribute and should include a minimum of 1 object of one subset of the same attribute (e.g. a group of 4 yellows followed by 1 yellow)
Table B3

**Scoring Guide for the Question “How did you decide which objects to put in your bowls?”**

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely Accurate</td>
<td>“Because reds, blues, yellows are together”</td>
<td>4</td>
</tr>
<tr>
<td>Partially Accurate</td>
<td>“I put different colors together”</td>
<td>3</td>
</tr>
<tr>
<td>General: Not wrong</td>
<td>“They are the same”</td>
<td>2</td>
</tr>
<tr>
<td>Points to some correct objects</td>
<td>Child points to red objects</td>
<td>1</td>
</tr>
<tr>
<td>Random/wrong responses</td>
<td>“I just wanted to”</td>
<td>0</td>
</tr>
<tr>
<td>No response</td>
<td>N/A</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix C

Graphs used to determine scoring on the ACBM

Figure C1. Adjusted scores (3221) against raw scores (1111) at the one-attribute level
Figure C2. Adjusted scores (3211) against raw scores (1111) at the two-attribute level
Appendix D

Tables

Table D1

*Descriptive Statistics for the Scales of the ACBM and the Total Measure Before Data Transformation*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>36.93</td>
<td>25.445</td>
<td>-0.107</td>
<td>-0.929</td>
</tr>
<tr>
<td>Shape</td>
<td>35.06</td>
<td>27.392</td>
<td>0.264</td>
<td>-1.094</td>
</tr>
<tr>
<td>Size</td>
<td>22.52</td>
<td>22.380</td>
<td>0.459</td>
<td>-1.211</td>
</tr>
<tr>
<td>Color and shape</td>
<td>11.62</td>
<td>13.450</td>
<td>0.787</td>
<td>-0.832</td>
</tr>
<tr>
<td>Explanation</td>
<td>34.14</td>
<td>51.395</td>
<td>1.813</td>
<td>3.642</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>140.26</td>
<td>100.388</td>
<td>0.925</td>
<td>-0.990</td>
</tr>
</tbody>
</table>
Table D2

*Kolmogorov-Smirnov Values for the Scales of the ACBM and the Total Measure Before Data Transformation*

<table>
<thead>
<tr>
<th>Scale</th>
<th>K-S statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>2.300</td>
<td>.000</td>
</tr>
<tr>
<td>Shape</td>
<td>1.183</td>
<td>.121</td>
</tr>
<tr>
<td>Size</td>
<td>2.295</td>
<td>.000</td>
</tr>
<tr>
<td>Color and shape</td>
<td>2.806</td>
<td>.000</td>
</tr>
<tr>
<td>Explanation</td>
<td>3.616</td>
<td>.000</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>.889</td>
<td>.408</td>
</tr>
</tbody>
</table>
Table D3

*Shapiro-Wilk Values for the Scales of the ACBM and the Total Measure Before Data Transformation*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Shapiro-Wilk statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>.884</td>
<td>.000</td>
</tr>
<tr>
<td>Shape</td>
<td>.906</td>
<td>.000</td>
</tr>
<tr>
<td>Size</td>
<td>.834</td>
<td>.000</td>
</tr>
<tr>
<td>Color and shape</td>
<td>.785</td>
<td>.000</td>
</tr>
<tr>
<td>Explanation</td>
<td>.709</td>
<td>.000</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>.938</td>
<td>.000</td>
</tr>
</tbody>
</table>
Table D4

*Descriptive Statistics for the Scales of the ACBM and the Total Measure After Data Transformation*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>5.26</td>
<td>3.061</td>
<td>-0.836</td>
<td>-0.748</td>
</tr>
<tr>
<td>Shape</td>
<td>5.06</td>
<td>3.085</td>
<td>-0.543</td>
<td>-0.961</td>
</tr>
<tr>
<td>Size</td>
<td>3.63</td>
<td>3.068</td>
<td>-0.071</td>
<td>-1.631</td>
</tr>
<tr>
<td>Color and shape</td>
<td>2.43</td>
<td>2.403</td>
<td>0.231</td>
<td>-1.584</td>
</tr>
<tr>
<td>Explanation</td>
<td>3.63</td>
<td>4.599</td>
<td>0.743</td>
<td>-0.869</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>10.95</td>
<td>4.529</td>
<td>-0.121</td>
<td>-0.230</td>
</tr>
</tbody>
</table>
Table D5

*Kolmogorov-Smirnov Values for the Scales of the ACBM and the Total Measure After Data Transformation*

<table>
<thead>
<tr>
<th>Scale</th>
<th>K-S statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>2.729</td>
<td>.000</td>
</tr>
<tr>
<td>Shape</td>
<td>1.730</td>
<td>.005</td>
</tr>
<tr>
<td>Size</td>
<td>2.721</td>
<td>.000</td>
</tr>
<tr>
<td>Color and shape</td>
<td>3.217</td>
<td>.000</td>
</tr>
<tr>
<td>Explanation</td>
<td>4.034</td>
<td>.000</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>.715</td>
<td>.686</td>
</tr>
</tbody>
</table>
Table D6

Shapiro-Wilk Values for the Scales of the ACBM and the Total Measure After Data Transformation

<table>
<thead>
<tr>
<th>Scale</th>
<th>Shapiro-Wilk statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>.800</td>
<td>.000</td>
</tr>
<tr>
<td>Shape</td>
<td>.876</td>
<td>.000</td>
</tr>
<tr>
<td>Size</td>
<td>.825</td>
<td>.000</td>
</tr>
<tr>
<td>Color and shape</td>
<td>.802</td>
<td>.000</td>
</tr>
<tr>
<td>Explanation</td>
<td>.749</td>
<td>.000</td>
</tr>
<tr>
<td>Total ACBM</td>
<td>.991</td>
<td>.648</td>
</tr>
</tbody>
</table>
Table D7

Correlations Among the Scales on the ACBM

<table>
<thead>
<tr>
<th>Scale</th>
<th>Color</th>
<th>Shape</th>
<th>Size</th>
<th>Color and shape</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>1</td>
<td>.286</td>
<td>.263</td>
<td>.260</td>
<td>.460</td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td>1</td>
<td>.410</td>
<td>.314</td>
<td>.321</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td>1</td>
<td>.416</td>
<td>.385</td>
</tr>
<tr>
<td>Color and shape</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>.364</td>
</tr>
<tr>
<td>Explanation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

All Correlations are significant at the 0.01 level (2-tailed)
Table D8

*Correlations Between Age in Months and Scores on the ACBM*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Pearson, r</th>
<th>Significance</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>.215*</td>
<td>.018</td>
<td>.046</td>
</tr>
<tr>
<td>Shape</td>
<td>.225*</td>
<td>.014</td>
<td>.051</td>
</tr>
<tr>
<td>Size</td>
<td>.061</td>
<td>.509</td>
<td>.004</td>
</tr>
<tr>
<td>Color and shape</td>
<td>.311**</td>
<td>.001</td>
<td>.097</td>
</tr>
<tr>
<td>Explanations</td>
<td>.310**</td>
<td>.001</td>
<td>.096</td>
</tr>
<tr>
<td>Total measure</td>
<td>.330**</td>
<td>.000</td>
<td>.109</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)
**Table D9**

*Test for Univariate Outliers on the Scales of the ACBM*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Mean ± 4SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>36.93</td>
<td>25.445</td>
<td>(-64.85, 138.71)</td>
</tr>
<tr>
<td>Shape</td>
<td>35.06</td>
<td>27.392</td>
<td>(-74.508, 144.628)</td>
</tr>
<tr>
<td>Size</td>
<td>22.52</td>
<td>22.380</td>
<td>(-67, 112.04)</td>
</tr>
<tr>
<td>Color and shape</td>
<td>11.62</td>
<td>13.450</td>
<td>(-42.18, 65.42)</td>
</tr>
<tr>
<td>Explanation</td>
<td>34.14</td>
<td>51.395</td>
<td>(-171.44, 239.72)</td>
</tr>
</tbody>
</table>
Appendix E

Data Figures
Figure E1. Histograms representing total scores on the ACBM
Figure E2. Scatterplots comparing age in months to scores on the ACBM
Figure E3. Hypothesized CFA model for the ACBM before transformation
Figure E4. Hypothesized CFA model for the ACBM after transformation
Figure E5. Measured model with standardized factor loadings of the scales of the ACBM before transformation
Figure E6. Measured model with standardized factor loadings of the scales of the ACBM after transformation