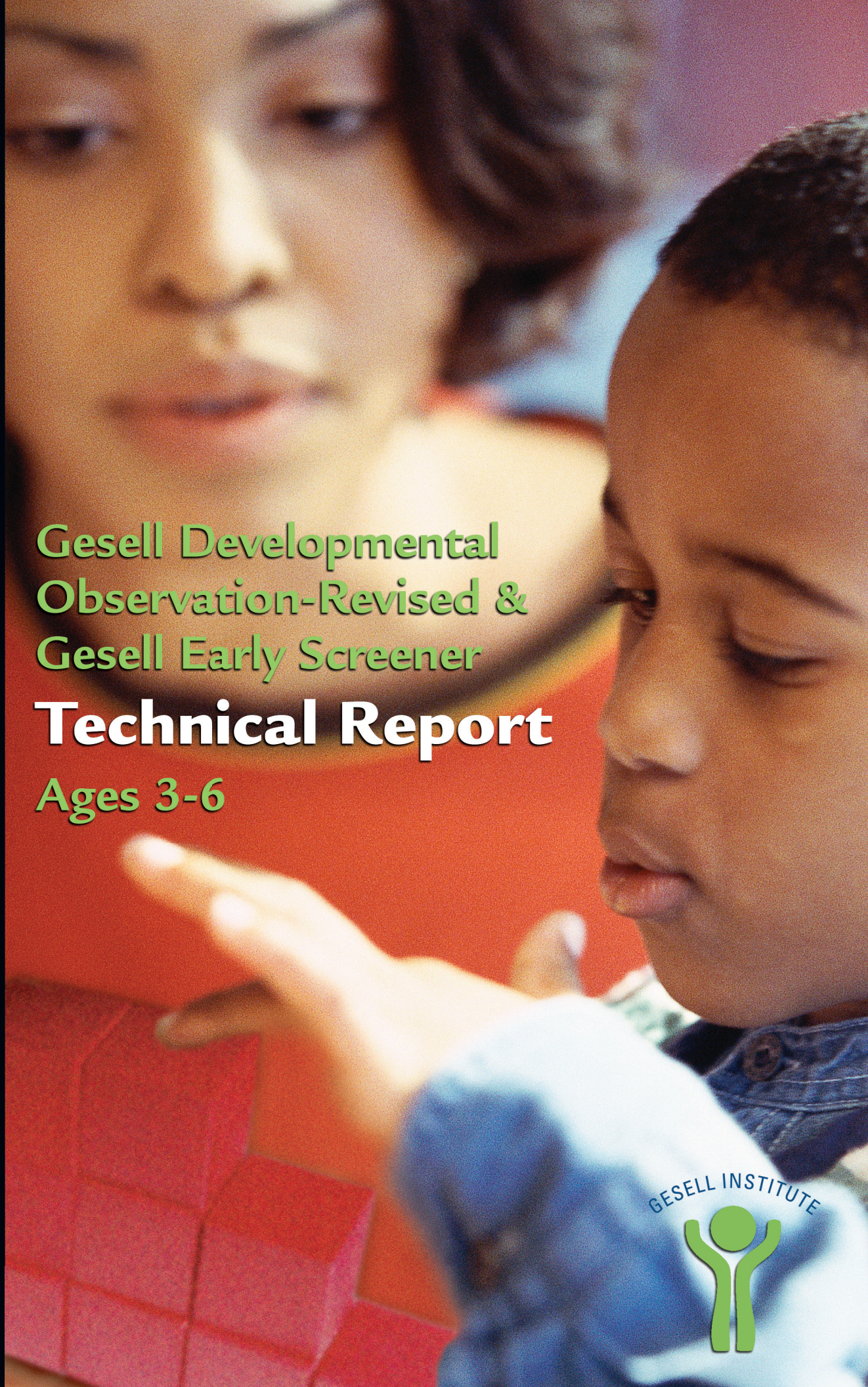


Gesell Developmental Observation-Revised & Gesell Early Screener

Technical Report

Ages 3-6

GESELL INSTITUTE

Gesell Developmental Observation – Revised

and

Gesell Early Screener

Technical Report

Ages 3-6

GESELL INSTITUTE
310 Prospect Street
New Haven, CT 06511
www.gesellinstitute.org

Gesell Institute of Child Development Mission Statement

To promote the principles of child development as the basis for all decision making for young children.

Gesell Institute of Child Development
310 Prospect Street
New Haven, CT 06511
800-369-7709
www.gesellinstitute.org

Table of Contents

ACKNOWLEDGEMENTS

The Gesell Developmental Observation-Revised (GDO-R) and Gesell Early Screener (GES) Technical Report is a product of the efforts of many individuals across the nation. Without the hard work and dedication of these contributing members, this Technical Report would not have come to be. It is with deep gratitude that we at the Institute sincerely thank these individuals for their commitment to this project and the overall mission of the Gesell Institute of Child Development. For their vision and willingness to fund this project, we thank the Gesell Institute Board of Directors. For the qualitative scoring, field testing, and editing, we especially thank Suzy Ferree, Linda Halverson, Marilyn Mansberry, Sallie Wells, and Cheryl Wolfe. Additionally, the Gesell Institute of Child Development would like to extend a special thanks to those schools, organizations, and examiners that volunteered their time and energy to be a part of the national GDO study.

Gesell Institute of Child Development Staff
Marcy Guddemi, PhD, MBA, Executive Director
Andrea Sambrook, MA, Director of Research and Education, GDO Study Coordinator
Sally C. Keller, MAT, Director of National Lecture Staff
Larry R. Joyner, Director of Operations
Jennifer Pelton, MA, Director of Marketing and Communications
Gitta Selva, Director of Programs and Editorial Services

Gesell Institute Board of Directors
Linda Calarco, MA (President)
Judy August MS, CAS (Vice President)
William W. Bakke, MA, JD, LLM (Treasurer)
Barbara M. Stern, MS, MBA (Secretary)
John Cirello, JD
Kathy DeStefano
Gladys Deutsch, MS
Sally Coleman Keller, MAT
Mary L. Pepe, MA
Michelle Rivelli, MD, FAAP
Carole Weisberg, MA
Marcy Guddemi, PhD, MBA (ex-officio)

Supportive Individuals and Former Staff
Candace Azevedo, ECE Consultant
Sha Balizet, PhD, former McREL Staff
Sandra Bishop-Joseph, PhD, Zigler Center
Tiffany Burke, Gesell Institute Consultant
Rhonda Clements, EdD, Manhattanville College
Tynetta Darden, CDF, MPA, former Gesell Institute Board Member
Jane Sikand Edelstein, Gesell Institute Consultant
Ashley Firth, MA, former Gesell Institute Staff
Stefanie Foster-Brown, former Gesell Institute Staff

Jennifer Freeman, Gesell Institute Consultant
Integrated Wellness Group, Dr. Maysa Akbar and Georgette Quackenbush
Tina Mannarino, PhD, New Haven Public Schools
Crista Marchesseault, MA, former Gesell Institute Staff
David Marchesseault, MS, MEd, Consultant
Dorothea Marsden, MEd, ECE Consultant
Shannon Midolo, MA, former Gesell Institute Staff
Richard Ruiter, former Gesell Institute National Lecture Staff
Liz Santy, former Gesell Institute Staff
Joan Turner, Acton Cooperative Preschool
Annette Watert, former Gesell Institute Staff

National Lecture Staff
Erin Akers, MEd
Kiki Ammons, MEd
Suzy Ferree, MS, MA
Linda Pass Halverson
Norman Heimgartner, EdD
Sally C. Keller, MAT
Marilyn Mansberry
Karlen Senseny
Sallie Wells, MS, CFLE
Cheryl Wolfe

Advisory Council
Joan Almon
William Crain, PhD
Stanley C. Finkle, MS
Kathleen E. Fite, EdD
Joe Frost, EdD, LHD
Martha Garcia-Sellers, PhD
Walter S. Gilliam, PhD
Jim Grant
Carla M. Horwitz, MS, EdD
Sharon Lynn Kagan, EdD
Ming Li, MD
Sandra Malmquist, PhD
Tina Mannarino, PhD
Dorothea Marsden, MEd
Linda Mayes, MD
Rae Pica
Rima Shore, PhD
Dorothy Singer, PhD
Jerome Singer, PhD
Fred Volkmar, MD

And a special thank you to Jackie Haines, former Executive Director, Gesell Institute

Psychometric Support
Mid-Continent Research for Education and Learning
Bruce Randel, PhD, Principal Researcher
Susan Shebby, EdM, Research Associate

GDO Study Sites
Acton Cooperative Preschool, Acton, MA and especially Joan Turner
Amberly Elementary, Portage, MI and especially Brian Dobbie
Appleby Elementary, Marathon, NY and especially Ethanie Holl
Applewild Elementary School, Fitchburg, MA and especially Tally Lent
Bluffton Elementary School, Bluffton, OH and especially Nancy Fultz
Bolles School, Jacksonville, FL and especially Leslie Brotman
Charlotte Christian School, Charlotte, NC and especially Cathie Broocks
Charlotte Preparatory School, Charlotte, NC and especially Vicky Wilkison
Christ the King, Lexington, KY and especially Melissa Cashman
Coyote Valley Elementary, Middletown, CA and especially Debby Carter
da Vinci School, Dallas, TX and especially Tracey Mason
Elm Drive Elementary, Millbrook, NY and especially Karen Fitzgerald
Evergreen Country Day School, Evergreen, CO and especially Gray Jessiman
Forsyth Country Day and Messiah Moravian Preschool, Lewisville, NC and especially Judith Kuhn
Grace Brethren School, Simi Valley, CA and especially Marsha McMullen
Greater Portland Christian School, South Portland, ME and especially Lori Libby
Greenhill School, Addison, TX and especially Kim Barnes
Heathwood Hall Episcopal School, Columbia, SC and especially Cheryl Gahagan
International School of Louisiana, Houma, LA and especially Ann Winchell
International School of Minnesota, Eden Prairie, MN and especially Whitney Frisch
Kessler School, Dallas, TX and especially Gayla Dobbins
Laurel Hall School, North Hollywood, CA and especially Patrice Lisy
Living Word Christian Academy, West Haven, CT and especially Karen Schlemmer
Madison Academy, Madison, AL and especially Carmelita Burton
Marlboro Elementary and Middle Hope Elementary, Newburgh, NY and especially Terry Reynolds
Memorial Elementary, Milford, MA and especially Julie Gunduz
Naugatuck Early Childhood Development Center, Naugatuck, CT and especially Cindy Stowell
Naugatuck Valley Community College Child Development Center, Waterbury, CT and especially Marianne Generali
New Haven Head Start Centers, CT (Helene Grant, Hill Central, John Martinez, Truman, and Zigler) and especially Claudia McNeil and the Early Childhood Coordinators
New Haven Public School Readiness Centers, CT (Dwight and Early Childhood Learning Center) and especially Vanetta Lloyd and Cathy Pont
Noah's Ark Preschool, Hastings, MI, and especially Sally Keller
Oakland Christian School, MI and especially Erin Akers
Ravenscroft School, Raleigh, NC, and especially Carol Miedema

Rhonda's Rompers Home Day Care Center, Branford, CT and especially Rhonda Butler
Rosebud Elementary, Rosebud, SD and especially Carrie Reinders
St. Anselm School, Chesterland, OH and especially Joan Agresta
St. Bernadette School, Houma, LA and especially Joy Domangue
St. Maria Goretti Preschool, Scottsdale, AZ and especially Kathleen Bies
Townsend Elementary, Walton, NY and especially Beth Howland
Upland Country Day, Kennett Square, PA and especially Judy Yeatman
Wesleyan School, Norcross, GA and especially Mari Beth King
Westminster Academy and Germantown Methodist Preschool, Memphis, TN and especially Lori Keith
Wichita Collegiate School, Wichita, KS and especially Becky O'Hearn
Wm. H. Frasier School, Fallbrook, CA and especially Kim Kinsman
Zimmerman School, Wahpeton, ND and especially Norma Nosek

Volunteer Examiners and Interns
Ryan Andrews, Eliza Behymer, Jennifer Bowers, Kelsey Buttendorf, Megan DeVizio, Hannah Emerson, Kristie Frumento, Philip Gardner, Dawn Grady, Joe Guddemi, Robyn Jacobs, Victoria Knezek, Sue Landisio, Amy LeBlanc, Lindsay Marcin, Alexander Oji, Ashley Peters, Kristen Provost, Andrea Reinhardt, Meghan Silva, Jessica Simmons, Susan Stevens, Aleesha Young

Pilot Only Data Sites
Bent Tree Child Development Center, Addison, TX and especially Mary Barter
Brevard County Schools,Viera, FL and especially Mona Potter
Cesar Batalla School, Bridgeport, CT and especially Anne Emerson
Glendale Acres Elementary, Fayetteville, NC and especially Donna Vann
Pediatric Therapy Center, CA and especially Pamela Flagg
Timbuktu Academy, Detroit, MI and especially Brenda Parker

CHAPTER I

Gesell Developmental Observation-Revised Technical Report: Data Analyses and Results

FOREWORD

The Gesell Developmental Observation-Revised (GDO-R) Technical Report is the first comprehensive technical publication for the GDO assessment instrument since 1979. The report is intended for both teachers and administrators. Its ultimate purpose is helping to inform curriculum development. It is based on a sample of assessment data for developmental and achievement tasks for children in preschool and Kindergarten. It is anticipated that readers of this Technical Report possess an advanced understanding of appropriate use and application of assessment tools, methods for conducting sound test development, and methodology in statistics and measurement.

PART 1: OVERVIEW

The Role of Developmental Assessment in Early Childhood

The testing of young children with standardized measures is a sensitive process requiring experience and training (Bredekamp & Copple, 2009). In the first seven years of life, a child develops simultaneously in multiple domains. One way to classify the domains of child development is (1) cognitive; (2) language; (3) motor; and (4) social-emotional. Cognition refers to the mental processes involved in gaining knowledge and comprehension, including thinking, knowing, remembering, judging, and problem-solving. These are higher-level functions of the brain and encompass language, imagination, perception, and planning (Lindfors, 2008). Language can be defined as a systematic way to convey meaning using symbols and sounds (Lindfors, 2008). Motor skills (fine and gross) describe how a child utilizes a comprehensive set of physical skills which involve movement in space, balance, and coordinated movement with objects (Clements & Schneider, 2006). Social skills are socially acceptable learned behaviors that enable an individual to interact effectively with others and to avoid or escape negative social interactions with others (Gresham & Elliott, 1990). The core features of emotional development include the ability to identify and understand one's own feelings, to accurately read and comprehend emotional states in others, to manage strong emotions and their expression in a constructive manner, to regulate one's own behavior, to develop empathy for others, and to establish and maintain relationships (National Scientific Council on the Developing Child 2004, 2). "Typical" child development during the early years is fluid and variable and involves many ups-and-downs, including behavior that may look like setbacks or regression. This is often a natural part of the development process (Shonkoff & Phillips, 2000).

As there is variability in development among children of the same age, it is also true that the younger the child, the greater the variability in development, motivation, and distractibility to the task at hand. Consequently, finding developmentally appropriate assessments that accurately

inform educators and parents is challenging (Shepard, Kagan, & Wurtz, 1998). However, as more and more children enter early childhood programs, the need for scientifically based assessments that measure developmental growth has become increasingly important.

The field of child development screening and assessment has an extensive history (Shonkoff & Meisels, 2000). Best practice for how to observe, describe, and evaluate the development of young children continues to evolve over time (Meisels & Atkins-Burnett, 2000). Developmental screening differs from readiness testing, IQ testing, and more comprehensive diagnostic evaluations (Shonkoff & Meisels, 2000). Readiness tests often have different purposes, and are often defined in different ways. This can be problematic for parents, educators, and communities alike if there is not a shared understanding of readiness. The National Education Goals Panel (NEGP) report *Principles and Recommendations for Early Childhood Assessments* (Shepard, Kagan, & Wurtz, 1998) identifies and describes five major purposes for assessing young children. School readiness assessments range widely from evaluation of an individual child, to evaluation of a teacher, program, or school, and typically fall under one of the following purposes:

1) To improve learning by identifying children's strengths and weaknesses on **academic tasks** in order to modify classroom instruction.
2) To identify children with special needs by evaluating general **developmental milestones** in multiple areas, as opposed to whether the child is learning particular concepts covered in a curriculum.
3) To monitor **trends at the school level** over time, so that communities or states may choose to conduct school readiness assessments every few years to provide a snapshot of children as they enter kindergarten.
4) To **evaluate programs** (but not individual children), by showing that a representative group of children from the program has improved.
5) To use for high-stakes accountability to make **decisions about individual children** or teachers. Because few assessment tools for young children meet high standards, the NEGP report (Shepard, Kagan, & Wurtz, 1998) recommends that no child assessments be conducted for high-stakes accountability purposes until third grade.*

*One assessment can never meet all five purposes.

Observational techniques and detailed measurement methodology were first introduced in a formalized way by Dr. Arnold Gesell in 1925. Gesell identified the ages and stages of child development based on his maturationist theory (Gesell, 1925). His contribution to the field provided a new perspective on observing not only the activities mastered by a child at various stages, but the *process* of how a child engaged in the activities. This work guided educators, pediatricians, psychologists, and others in a deeper understanding of the processes which define a child's functioning on the developmental continuum. The ages and stages of typical development established by Gesell provided new momentum to the field of child assessment to help identify children in need of early intervention.

The first national initiative to focus on school readiness programming was Head Start. It was founded in 1965 as part of President Lyndon B. Johnson's War on Poverty, and continues to be

implemented across the nation today. Head Start programming is committed to providing comprehensive health, education, and social services for low-income families through parent involvement and increased home-school communication (Zigler & Styfco, 2010).

In 1985, recognizing the need for a way to measure the quality of educational and developmental services for children, the National Association for the Education of Young Children (NAEYC) started a nationwide, voluntary accreditation program for early childhood education institutions (Bredekamp & Copple, 2009). In addition, NAEYC began publishing national position statements about early childhood education issues including professional development, instructional strategies, multi-culturalism, and assessment. Over the course of the next 20 years, these initiatives served as the primary standard for evaluating developmental assessments and appropriate methodologies (Guddemi, 2003).

The updated Elementary and Secondary Education Act, No Child Left Behind (NCLB), was signed into law in 2002. It included strong language about the role of scientifically-sound instructional methods and assessments. NCLB is premised on four principles: (1) accountability for results; (2) more choices for parents; (3) greater local control and flexibility; and (4) an emphasis on doing what works based on scientific research (Guddemi, 2003). Current assessments are oriented towards this more rigorous evaluative criterion; thus reflecting a significant paradigm shift, and straying from a more traditional focus on the child's *process* of development. These assessments are more intent on achievement or identifying the knowledge of skills deemed necessary to start school. This has created a learning environment in which children are expected to adapt to the curriculum, rather than the curriculum adapting to the developing needs of the child (Almon & Miller, 2011).

One challenge to a heavy focus on achievement is that young children often have uneven and sporadic developmental progress. While they may excel in one area, they may have several others that are latent in growth (Gesell & Ilg, 1940; Piaget & Inhelder, 1969). Therefore, creating a "one-size-fits-all" assessment for what a child must "do" before starting Kindergarten is nearly impossible. However, many schools have set school policy about what children are required to know (National Governors Association, Common Core State Standards, 2010), thus adding stress to teachers who are mandated to regiment the learning pace of children to match school expectations. Memorizing factual information does not mean that the child understands and can use that knowledge in other ways. Thus, "knowing is not understanding" (Elkind, 2012).

Consequently, a more realistic and suitable approach is to assess a child's developmental age, which serves as a comparison to the child's chronological age, and is based on stages of maturation for typically developing children. A child may perform beyond his or her chronological age in one domain, or he or she may be developmentally younger in another. More specifically, all children follow the same "path of development." However, while the stages are the same, the rates at which they are achieved vary among children (Gesell & Ilg, 1940). Usually by the end of third grade, the large discrepancy between children's development across domains is reduced (Suen & French, 2003).

Accurately identifying a child's developmental age aides teachers and administrators in creating a developmentally – and not just chronologically – age appropriate learning environment (Shonkoff & Phillips, 2000). This is exceptionally important, because research has found that when children are put in an academic environment that is beyond their developmental stage, their ability to understand and retain information is adversely impacted, often resulting in decreased focus and increased behavioral problems (Ilg, 1972). Therefore, understanding a child's developmental age enables educators to modify academic expectations, and thus foster a more healthy and appropriate learning environment (Burts et al., 1993; Charlesworth, Hart, Burts & DeWolf, 1993; Hart, Charlesworth, Burts & DeWolf, 1993; Larsen & Robinson, 1989; Marcon, 1992, 1993). The GDO-R is the only assessment that results in a Developmental Age as well as strand Performance Level Ratings (refer to Tables 53-55 for details on Developmental Age and Strand Performance Level Ratings).

Overview of Gesell Developmental Observation-Revised

The 2007 Gesell Developmental Observation (GDO) was the focus of the GDO Study to collect and present updated technical data and validity evidence on 17 of the 19 tasks (as found in ©2007 GDO version), and multiple other measures for children ages 3-6. The ©2007 GDO underwent rigorous review and important improvements (for example, the addition of parent and teacher ratings of child behavior). Revisions (updated interview questions and one action agent item) were made prior to collecting data. This Technical Report contains the data that was collected on the revised GDO, now named the GDO-R.

The GDO Study included a Parent/Guardian Questionnaire (PQ), because parent participation in early childhood assessments improves communication between parents and teachers. Better communication inherently increases the parent's engagement with their child, since parents' attention is drawn towards the child's behavior and learning milestones (Pianta & Kraft-Sayre, 2003). A Teacher Questionnaire (TQ) with ratings of social behavior, emotional development, and adaptive skills was also included to provide additional validity evidence regarding the interpretation of a child's GDO-R results in the field. Multiple studies have supported the predictive validity of teacher ratings on student performance later in school and even into adulthood (Paris & Ayers, 1994; Paulson, Paulson & Meyer, 1991; Wolf & Siu-Runyan, 1996; Valencia, 1990; Zigler, Gilliam & Jones 2006). Other modifications were made as a result of data collection and include improved strand scoring, performance level ratings, and objective scoring rubrics.

The GDO-R is a standardized, performance-based, criterion-referenced instrument that measures a child's behavior through direct observation, and through surveying parents and teachers. The GDO-R is designed for children ages 2 ½ to 9 years old. The 2007 GDO version was used as the basis for the GDO study, including 17 of the 19 tasks for ages 3 to 6. Since examiners who assess children between 7 and 9 years were not the prime users of the GDO, these age bands were not included in the study. Consequently, no data was collected on the Visual III and Right and Left tasks, since these tasks are appropriate only for children age 7-9. New information is provided in the Examiner's Manual for the 2 ½ age band and for older children (6 year 4 months to 9 years old) based on the knowledge of a group of expert examiners.

GDO-R tasks are divided into five strands (related sets of individual tasks), and scores from each strand provide the examiner with the following Performance Level Ratings: Age Appropriate (solid or qualified expectation responses for all or most of the tasks in the strand), Emerging (solid or qualified expectation responses for most or only some tasks in the strand), or Concern, (atypical responses for most tasks in the strand). (For more information on GDO-R Performance Level Ratings, see page 58). The GDO-R can also help determine whether or not a child may need further evaluations to determine appropriate educational planning or remediation in specific domains of development. Refer to Table 1 for the domains measured by the five strands of the GDO-R.

Table 1
Domains Measured by GDO-R Strands

Domain/ Strand	Strand A	Strand B	Strand C	Strand D	Strand E
	Developmental	Letters/ Numbers	Language/ Comprehension	Visual/Spatial Discrimination	Social/ Emotional/ Adaptive
Cognitive	✓	✓	✓	✓	
Language	✓	✓	✓	✓	✓
Motor	✓	✓		✓	✓
Social-Emotional	✓		✓		✓

Note. GDO-R tasks are administered according to age of the child. Refer to Table 2 for a listing of Tasks by Age of Administration.

Table 2

GDO-R Tasks by Age of Administration

Strand Name	Task Name (Number)	2^6	3	3^6	4	4^6	5	5^6	6	7	8	9
Strand A: Developmental	Cubes (Task 1)	●	●	●	●	●	●	●	●			
	Copy Forms (Task 4)	●	●	●	●	●	●	●	●	●	●	●
	Incomplete Man (Task 5)	●	●	●	●	●	●	●	●	●	●	●
	Fine/Gross Motor (Task 19)	●	●	●	●	●	●	●	●			
	Overt Behavior (Task 20)	●	●	●	●	●	●	●	●	●	●	●
Strand B: Letters/Numbers	Writing Name (Task 3)	●	●	●	●	●	●	●	●	●	●	●
	Writing Numbers (Task 3)	●	●	●	●	●	●	●	●	●	●	●
	Identifying Letters (Task 17)			●	●	●	●	●	●			
	Identifying Numbers (Task 17)			●	●	●	●	●	●			
	Counting (Task 18)	●	●	●	●	●	●	●	●			
	One-to-One Correspondence (Task 18)	●	●	●	●	●	●	●	●			
	Conservation (Task 18)	●	●	●	●	●	●	●	●			
	Calculations (Task 18)					●	●	●	●	●	●	●
	Digit Repetition (Task 12)	●	●	●	●	●	●	●	●			
Strand C: Language/ Comprehension	Prepositions (Task 11)	●	●	●	●	●	●	●	●			
	Comprehension Questions (Task 13)	●	●	●	●	●	●	●	●			
	Action Agents (Task 16)	●	●	●	●	●	●	●	●			
	Naming Animals (Task 9)				●	●	●	●	●	●	●	●
	Interests (Task 10)				●	●	●	●	●	●	●	●
	Interview (Task 2)	●	●	●	●	●	●	●	●	●	●	●
Strand D: Visual/Spatial Discrimination	Color Forms (Task 14)	●	●	●	●							
	Three-Hole Form Board (Task 15)	●	●	●	●							
	Visual I (Task 7)				●	●	●	●	●			
	Visual III (Task 8)								●	●	●	●
	Right/Left (Task 6)								●	●	●	●
Strand E: Social /Emotional /Adaptive	Teacher Questionnaire (Task 21)	●	●	●	●	●	●	●	●	●	●	●
	Parent Questionnaire (Task 21)	●	●	●	●	●	●	●	●	●	●	●

Note. Task numbers correspond to Examiner's Manual. The GDO-R contains 19 performance-based tasks, an overall measure of Overt Behavior, as well as Social, Emotional, and Adaptive functioning, for a total of 21 tasks/measures.

Description of GDO-R Materials and Kits

Examiner's Manual

The Examiner's Manual contains an overview and historical background of the GDO-R, and provides the essential characteristics and behaviors of the ages and stages of development, according to original Gesell theory. Complete instructions for administering, recording, scoring (including Decision Trees), and interpreting results are provided for all tasks. In addition, the Manual provides:

- Performance Expectations by Task and Age.
- Strand Scoring Worksheet to calculate strand scores.
- Supplemental scoring trends for responses to tasks for which data was not collected during the GDO Study. These include: Qualitative aspects of Writing Name and Numbers; Incomplete Man Inquiry questions; Physical and verbal overflow for Right and Left Commands and Picture Cards, Visual I, Visual III, and Naming Animals (Type, Number, and Time Sustained).
- Recording Chart for scoring the Social, Emotional, and Adaptive strand.
- Summary Profile Form which incorporates a child's overall performance results to share with parents.

Child Recording Form

The Child Recording Form is a consumable form (one per child) used by the examiner to record responses to each task as well as verbal and physical behaviors observed during the assessment. The Summary Profile Form and the TQ/PQ Recording Chart are included in this form for examiner scoring convenience.

Examiner's Script

The Examiner's Script is a sturdy, spiral-bound booklet which includes the standardized administration script, and the required stimuli cards. The stimuli cards are included in the script for efficient administration, and are purposely sequenced to eliminate possible administration error due to incorrect orientation or sequencing.

Teacher Questionnaire (TQ)

The Teacher Questionnaire collects data from the teacher about a child's self-help skills, self-expression, behavioral observations, and classroom preferences. These ratings provide information on three subscales:

- Social Behavior
- Emotional Development
- Adaptive Skills

Parent/Guardian Questionnaire (PQ)

The Parent/Guardian Questionnaire collects data from parents/guardians about a child's:

- Family background and demographics (race, ethnicity, sex)
- Medical and educational history (birth history, early educational experiences)
- Home environment (siblings, use of technology, exposure to media and literacy opportunities)
- Social, emotional, and adaptive capabilities (self-help skills, choices offered to child, self-expression, behavioral observations)

Manipulatives

The manipulative items required for proper administration of the GDO-R are shown in Table 3.

Table 3
GDO-R Kit Manipulatives

Manipulatives		Items Provided by Examiner
• 10 Red Hardwood Cubes • Large Red Cube and Cylinder • Three-Hole Form Board and Shapes (Green Triangle, Circle, and Square) • Pellets and Bottle	• Beanbag • Right/Left Cards • Visual III Cards • Color Form Cut-Outs	• Pencil • Stopwatch or watch with a second hand • 6 foot length of tape • 20 Pennies

Description of Tasks by Strand

The primary purpose of the Gesell Developmental Observation-Revised (GDO-R) is to observe and evaluate development and academic readiness skills in five areas, or strands:

- The Developmental strand measures perceptual awareness, visual-motor coordination, short-term visual memory, fine and gross motor skills, and overt behavior. Performance in the Developmental strand determines a child's developmental age that may be equal to, older than, or younger than the child's actual chronological age.
- The Letters/Numbers strand measures a child's ability to recognize 26 letters and 12 numbers; to demonstrate one-to-one correspondence; to conserve quantity and perform calculations; and to display short-term auditory memory.
- The Language/Comprehension strand evaluates a child's attention span, and expressive and receptive language.
- The Visual/Spatial Discrimination strand measures visual discrimination of written symbols, manipulative shapes, and left and right orientation (for older children).

- The Social/Emotional/Adaptive strand provides provide information to evaluate the quality of a child's interactions with peers and adults, emotional regulation behaviors, and self-help skills (using the Teacher and Parent/Guardian Questionnaires).

A child's product and process (approach to task) are also evaluated by the GDO-R. Overt behavior is observed for all tasks in the GDO-R, but is scored as part of the Developmental Strand. A child's performance on the tasks for each strand correspond to a child's Performance Level Rating for each strand. Collectively, the Performance Level Ratings on all strands are used to establish an Overall Performance Level Rating, which corresponds to one of three levels: Age Appropriate, Emerging, or Concern.

Historical Validation and Reliability of the GDO-R

In 1925, Dr. Arnold Gesell first published a set of developmental schedules in *The Mental Growth of the Pre-school Child*. The more than 10,000 participants in Dr. Gesell's work (covering ages birth to adolescence) comprised his sample of children from the Yale Child Study Center, and were limited generally to the state of Connecticut and the northeast U.S. Still popular today, these schedules are the basis of the 2007 Gesell assessment known as the Gesell Developmental Observation (GDO).

In 1934, Gesell and colleagues published a revision of the schedules in *Infant Behavior: Its Genesis and Growth*. Gesell and Ilg issued the next revision of the schedules in 1940 in the *First Five Years of Life*. It covered ages 15 months through 6 years of age. In the same year, the Psychological Corporation published the scales commercially. The next revision appeared in Ilg and Ames' *School Readiness* in 1965. This was revised again in 1978 for 3 to 6 year olds. In 1979, Ames, Gillespie, Haines, and Ilg published Gesell Institute's *The Child from One to Six: Evaluating the Behavior of the Preschool Child* with updated technical data. This was the last time a study was conducted using the GDO solely for the purpose of collecting scientific data. A handful of other small-scale studies and dissertation research projects were conducted in the last 30 years for a variety of purposes, mostly surrounding issues of school readiness and transition classrooms.

The purpose of the GDO Study was to provide updated technical data and validity evidence for 17 of the 19 tasks on the ©2007 GDO. In addition, data was collected on social, emotional, and adaptive development by surveying parents and teachers (2 measures), for the purpose of helping to inform classroom instruction for each child by providing scoring at the strand level.

PART 2: VALIDITY AND CONTENT RATIONALE

Validity has no single agreed definition by scientist and statisticians; it generally refers to the extent to which a concept, conclusion, or measurement is substantiated and corresponds accurately to what is in the real world. "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The purpose of test validation is not to validate the test itself but to validate interpretations of the test scores for particular purposes or uses. Validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the assessment process. Every aspect of an assessment may provide evidence in support of its validity (or evidence to the contrary), including design, content specifications, item development, psychometric quality, as well as inferences based on the results. Additional evidence regarding validity can be found in following sections of this report.

The GDO-R was designed and developed to be a standardized, performance-based, criterion-referenced instrument that would produce Performance Level Ratings in five strands, valid for most types of early childhood educational decision-making (refer to Purpose 1 on page 10). Progress can be monitored over years and grades. The performance-based results may be used to analyze the strengths and weaknesses of a child's skills and development in each strand, to plan for further instruction, to strategize for curriculum planning, and to report progress to parents.

Content-Related Validity

Content-related validity is evidenced by uniformity between strand and task content, and the developmental milestones widely accepted to precede instructional content in each area. To ensure such correspondence for the GDO-R, Gesell Institute conducted a comprehensive review of current child development theory, and met with education experts to determine common educational goals and the knowledge and skills emphasized in today's early childhood curricula. The graphic design of the assessment and its manipulative materials reflect the types of activities found in early childhood classrooms and in children's everyday lives. An on-line user survey provided additional information regarding overall assessment effectiveness (addressing such topics as the appropriateness of the criterion for developmental age, ease of administration, and appropriateness for each age). These validation efforts resulted in an assessment that reflects the needs of classroom teachers, children, and parents.

Criterion-Referenced Performance Interpretations

The GDO-R is a developmental assessment that requires training to accurately assess a child's individual performance and qualitative responses to certain tasks, and interpret the results

according to a developmental continuum. As such, the criterion which shaped the scoring rubrics in this revision was based on three sources of information:

- Scientific data collected on a nation-wide sample of 1,287 children. The technical data from the GDO Study provides information about how children across the U.S. have performed on all GDO-R tasks and can be used to compare a child's performance to that of typically developing age matched peers.

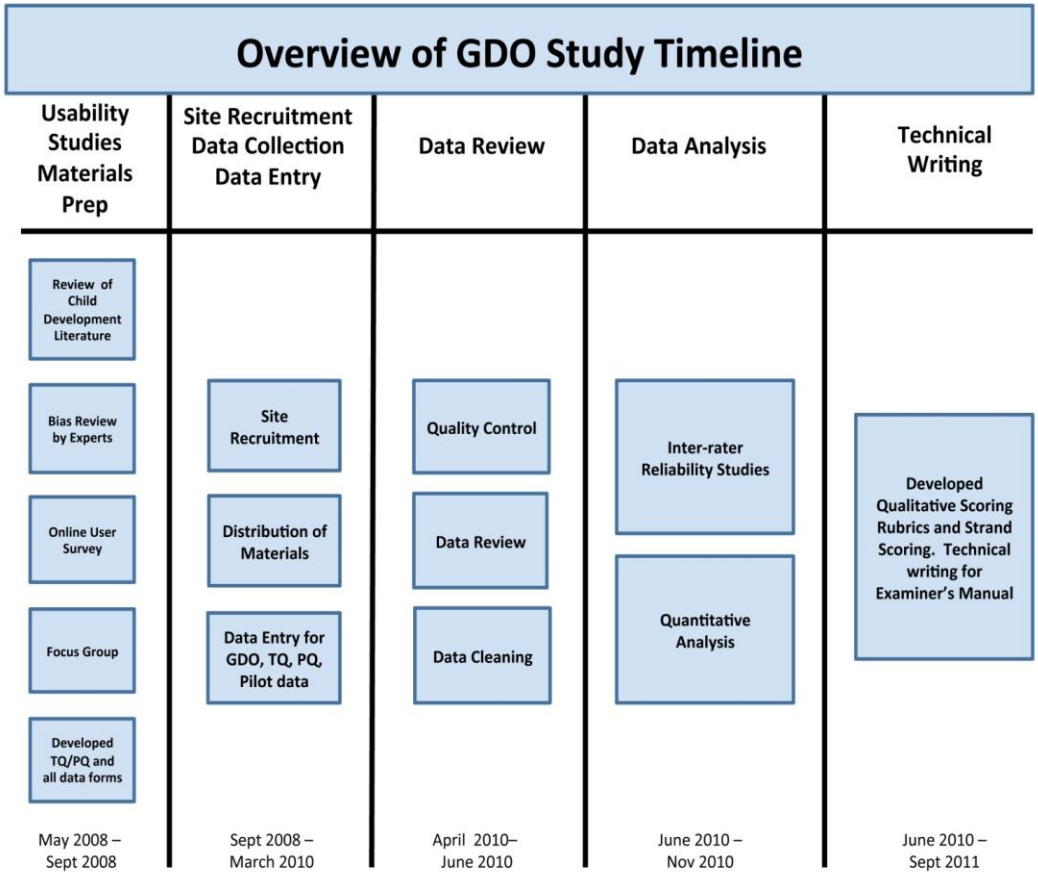
- Knowledge and experience of professionals who teach and work with children in each age band. A panel of nationally recognized experts with extensive experience in the field of child development reviewed the GDO-R performance level definitions as a tool for examiners to confirm a child's overall results on the GDO-R.

- Well-established research findings and theoretical frameworks. Children grow and mature through a series of predictable stages in a sequential order. Their development is dynamic, continuous, and reflects a pace unique to each child.

GDO Study Timeline

The GDO Study was completed over the course of three years. A timeline is presented in Figure 1 to provide context and sequence for the data activities and analyses presented in this Technical Report.

Figure 1: Overview of GDO Study Timeline



Online Survey

Prior to the commencement of the study, an online survey of GDO users was conducted to collect information on how customers used the GDO. A sample of *N*=153 respondents provided feedback. A summary of the questions and quantitative results can be found in Appendix A.

Qualitative feedback from the online user survey on the usefulness of the GDO revealed that some users found the assessment to be time consuming, and containing language and interview questions that were neither culturally sensitive nor up to date. Comments about the layout/organization of the GDO revealed that some users objected to the Child Recording Form having too much paper and blank space, the lack of examiner directions on the recording forms, and lack of consistency between core materials. Feedback on the scoring of the GDO showed that some users found it to lack objective scoring and interpretation rubrics.

Focus Group

In addition, the online survey invited respondents to participate in a focus group. The focus group was held at a Massachusetts preschool and was moderated by a member of the Mid-Continent Research for Education and Learning (McREL) research team. The purpose of the focus group was to gather information on qualitative improvements to the GDO which would be most meaningful to examiners, in addition to the updated technical data which was collected to renew the validity of the instrument. Twenty-nine of the 153 respondents indicated that they wished to participate in a focus group. However, focus group attendance was greatly affected by an unexpected snowstorm and reduced the participation to *N*=3. Attendees were given a $75 honorarium for participating. The focus group questions can be found in Appendix B. Feedback from the focus group can be found in Table 4:

Table 4
Observations Identified by the GDO Focus Group

Observation 1	A child's developmental stage is integral to his or her performance on achievement tests.
Observation 2	Educators in the field are craving validity and reliability data for the GDO.
Observation 3	Parents may be smart and educated, but are primarily misinformed about the ages and stages of child development. Educators feel parents value objective resources for information, such as educational specialists, or Gesell Institute. These are viewed as having more weight than a teacher's input.
Observation 4	Naming Animals is a task that measures response latency ("tempo") which provides information about temperament and has direct application to classroom participation.
Observation 5	A task that measures sequence completion could be helpful with 2-D or 3-D manipulatives.
Observation 6	Using a computer during the assessment would distract the child, require a new means of evaluating grasp (stylus vs. pencil), but could be useful for examiner to gather information, and feed into electronic portfolios.
Observation 7	The GDO-R assessment needs a parent questionnaire measure.
Observation 8	The GDO-R assessment needs a direct measure of children's social behavior, possibly observing a child's response to a card with a story depicted on it as a stimulus.

Procedures for Reducing Bias

The GDO was reviewed by a group of five experts from the fields of early childhood education, special education, physical movement, and test development. Each reviewer was asked to respond to a set of specific questions, in order to identify any biases inherent in the content or methodology of the GDO. The bias review questions can be found in Appendix C. Several GDO tasks raised issues due to a current lack of cultural relevancy. The first item asked the child about his or her most recent birthday celebration, and ability to recall presents that he or she received. This question was omitted from the Interview. The second item, a prompt for the child to name animals, presumed all children have experience of visiting a farm or zoo. This

question was revised to "*Have you ever been to or read a book about the zoo or a farm?*" Also, a new question about watching television was added to the Interview.

In addition, it was recommended that the GDO-R include a measure of emotional regulation and social behavior by surveying the child's teacher and/or parent, and to improve consistency in administration across tasks. A Teacher Questionnaire and Parent/Guardian Questionnaire were developed for the GDO Study, as well as a revised Examiner's Script.

PART 3: RESEARCH STUDY

Overview of Research Studies

Mid-continent Research for Education and Learning (McREL), an independent non-profit corporation employing educational psychometricians, was contracted to advise the Institute on statistical and technical matters over the course of the two-year study, and to conduct final analyses.

This study consisted of several sub-studies to collect both quantitative and qualitative data. Quantitative data (GDO Study) was collected on 17 out of 19 tasks of the GDO (©2007). GDO examiners administered a total of 167 items in one-on-one sessions with children. Data was also collected for each child from the teacher's observation of the child in the classroom (45 items), and from the parent's observation of the child in the home (78 items).

Qualitative data (Gesell Institute National Lecture Staff Review Study) was collected on two developmental tasks on the GDO; the Copy Forms and Incomplete Man. The purpose of the NLS Review Study was three-fold: to collect data on the *qualitative features* of each developmental stage for Copy Forms and Incomplete Man; to establish *inter-rater reliability* for each Copy Form item; and to establish *inter-rater reliability* for assigning a Developmental Age to Copy Forms and Incomplete Man samples.

Procedure for Recruiting Examiners and Data Collection Sites

GDO Study sites were recruited using multiple methods. Initially, a demographic interest form was sent out to current users of the GDO with the intention that interested sites would provide a pool of data for children 2 ½-6 years, from which the research team could select demographically diverse cases for inclusion in the study. This method produced an unexpectedly low number of sites. Data for the study was ultimately procured using the following due diligence approaches across all 50 U.S. states:

- Invited GDO examiners who had been trained within the last 5 years.
- Contacted principals of schools that recently ordered large amounts of GDO materials.
- Appealed to principals of schools that hosted GDO training workshops.

- Promoted participation in the study at 3-day GDO training workshops.
- Procured referrals from Gesell Institute Advisory Council members, Board of Directors, National Lecture Staff, Interns, Website Inquiries, and Home Daycare Networks.
- Promoted participation in the study to all Gesell customers (and schools enrolled in the study) through a monthly electronic *GDO Study Newsletter*.

Each site which enrolled in the study signed an agreement (Appendix D), and secured parental consent forms for each participating child (Appendix E). Each school estimated a number of assessments they could provide, and the expected ages of the children. Since priority was given to securing the target number of valid assessments in each age band, and schools varied on both the quantity and age bands of data they were able to submit, pre-selection of cases at the site level by the research team was not possible.

Trained examiners administered all designated items on the GDO to children within the study age band (see Table 5 later in this report) in individual assessment sessions. A standardized script guided the examiner in the administration of each task. Examiners in the study did not score or determine a Developmental Age. Data was collected from three sources at each school: the child (GDO tasks-GDO), the parent or guardian (Parent/Guardian Questionnaire-PQ), and the child's teacher (Teacher Questionnaire-TQ). The fourth source of data were 70 Pilot questions to be used in the future.

All GDO assessment forms and study materials were provided to each school free of charge, including return shipping and handling. These included Parental Consent Forms, School Agreement Forms, GDO Child Recording Forms, Parent/Guardian Questionnaires, Teacher Questionnaires, Data Collection Forms (including Pilot items), a Study Script, and a Training DVD. The Training DVD was provided to each site to standardize the training for all examiners in the study. A conference call was held between each school's examining team and the Gesell research team to review study protocol, answer questions, and provide support. Assistance was also provided by phone, e-mail, and fax throughout the entire course of the schools' participation in the study. Each site returned hardcopies of all data forms to Gesell Institute for review, validation, cleaning, and data entry. Schools were offered compensation up to $10.00 per child for collecting complete sets of data.

Two teams of examiners were trained in person at Gesell Institute to collect data in southwestern Connecticut. The first, a team of mental health clinicians, was subcontracted in 2008 to collect GDO (and Pilot) data in Head Start classrooms in New Haven, CT. In 2009, a second team of volunteer interns was trained to collect data in School Readiness classrooms in New Haven, CT. The research director observed the interns' initial assessments in the field, and provided feedback and support throughout the course of the data collection process.

All items for each of the designated GDO-R tasks were administered to all children in the seven age bands, whenever practical, using a standardized administration in English. In this way, scientific trends for mastery of items were documented as all children were given an opportunity to try each task, even if the task was developmentally above or below developmental expectations based on their chronological age.

Data Collection Challenges

Significant challenges were identified by data collection sites during the study and in some cases, reduced the number of assessments a school was able to complete. These included:

- Collecting enough GDO assessments for the 2 ½ age band. An insufficient number of sites used the GDO with this age band, and as such, were unable to provide it for the study. This age band was eventually dropped from the study.

- Securing time in the day to administer the GDO-R (40 min), according to study protocol (i.e., that all tasks be administered to all children). These extra items/tasks required for the study are not part of the standard administration of the GDO-R, and require extra time from the examiner and child.

- Not all schools provided teaching coverage for classrooms during testing, especially if the examiner was conducting assessments outside of the regular testing timetable at the site.

- TQ and PQ were new measures, and schools experienced varying responses from parents regarding the sensitivity of the information requested on the PQ. In particular, private school settings in which assessments were conducted as part of the admission procedure were reluctant to use the PQ.

- Some schools were able to participate with the permission of the Principal, while others required the School Board to approve their participation in the study. This delayed the start of data collection for some sites, and eliminated other schools from participating altogether.

- Competing responsibilities at the site such as re-accreditation priorities and curriculum revision, etc. which required investment from all staff in building.

- Some schools had other child assessments ongoing, and thus participation in the GDO Study was limited or eliminated because it would create a hardship for the children.

- Extreme winter storms across the northern part of the U.S. resulted in many school closings and lost days scheduled for data collection.

- Schools reported many cases of real or suspected H1N1 virus which resulted in many children and teachers being absent from school.

Research Sample

The research sample was primarily a sample of convenience drawn from a national population of typically developing children attending schools that administered the GDO. A subset of schools from New Haven, CT also participated in the study. Examiners for these schools were trained at Gesell Institute, since the school did not currently utilize the GDO but wanted to participate in the study. The target sample was an N of 200 children equally distributed in 6 month age bands from 2 years 4 month to six years 3 months. The total target N=1,600 would include 50% males,

50% females, matched to U.S. population proportions for race/ethnicity. Children with special needs were not excluded from the study, although schools that had already identified children as having special needs typically did not include those children in their sample. The research team requested that examiners provide a list of accommodations needed to conduct administrations of the GDO-R with children with special needs. All assessments were conducted in English, and teachers were asked to confirm a child's English language fluency in order to administer a valid GDO-R assessment. All four geographic regions defined by the Census were targeted, but the sample was not intended to wholly match U.S. population proportions by region.

The final sample for analysis for the GDO Study included 1,287 children, age 3 to 6, from 53 geographically diverse sites in 23 states. Data for both studies was organized according to the age bands outlined in Table 5.

Table 5
Chronological Ages of Children Included in Each Age Band

GDO Study Age Band	Chronological ages of children included in each band
3	2 years 9 months and 0 days – 3 years 2 months and 29 days
3^6	3 years 3 months and 0 days – 3 years 8 months and 29 days
4	3 years 9 months and 0 days – 4 years 2 months and 29 days
4^6	4 years 3 months and 0 days – 4 years 8 months and 29 days
5	4 years 9 months and 0 days – 5 years 2 months and 29 days
5^6	5 years 3 months and 0 days – 5 years 8 months and 29 days
6	5 years 9 months and 0 days – 6 years 2 months and 29 days

While the assessment is designed for ages 2 ½ to 9 years, the researchers chose to limit hands-on data collection to the group of children which comprised of the largest number of users of the GDO. The chronological age threshold for each age band is described above in Table 5. The age bands targeted for data collection were 3, 3 ½, 4, 4 ½, 5, 5 ½, and 6. Insufficient data was collected to analyze the 2 ½ age band, and it was subsequently dropped from the study. Since examiners who assess children between 7 and 9 years were not the prime users of the GDO, these age bands were not included in the study. Consequently, no data was collected on the Visual III and Right and Left tasks, since these tasks are appropriate only for children age 7-9. New information is provided in the Examiner's Manual for the 2 ½ age band and for older children (6 year 4 months to 9 years old) based on the knowledge of a group of expert examiners.

Site Sample

The sample included a diverse group of sites in terms of the type of school, region of the country, size, and population served (i.e., ethnicity and percentage eligible for free and reduced lunch). Refer to Figure 2 for sample distribution by state and Table 6 for demographics of participating sites.

Figure 2

States Represented in 2008-2010 GDO Study

Note. Participating states are indicated by a star.

Table 6

Descriptive Statistics for Participating Sites: School Type

School Type	Private	Public
# Sites	33 (62%)	20 (38%)
#GDO Assessments	584 (45%)	703 (55%)

Table 7

Descriptive Statistics for Participating Sites: U.S. Region

Region	Northeast	Midwest	South	West
Participating States	**CT, MA, ME, NY, PA**	**KS, MI, MN, ND, OH, SD**	**AL FL, GA, KY, LA, NC, SC, TN, TX**	**AZ, CA, CO**
# Sites	20 (38%)	9 (17%)	18 (34%)	6 (11%)
# GDO Assessments	672 (52%)	201 (16%)	288 (22%)	126 (10%)

Table 8
Descriptive Statistics for Participating Sites: Ethnicity and SES

Ethnicity	**Mean percent**
African-American	14.8 (34.1)
American Indian	2.9 (.6)
Asian American	4.5 (1.0)
Caucasian not Hispanic	60.7 (42.8)
Hispanic	15.3 (28.1)
Other	1.8 (1.2)
% Eligible for Free/Reduced Lunch	28.2 (35.64)
Enrollment	**N**
Mean # children enrolled	168

Note. Eligibility for Free/Reduced Lunch program is a proxy for lower socioeconomic status. Numbers in parentheses are standard deviations.

Examiner Sample

One hundred and one trained GDO examiners, with a mean of 12 years teaching experience, collected GDO data from children at preschool and elementary schools across the nation. The group of examiners, with a mean of seven years GDO experience, received GDO training in one of two ways: 1) in the last five years through a three-day workshop or 2) a one-day training session at the Institute. Of the examiner sample, 75% were currently teaching, and 25% reported that they were retired, no longer teaching in a classroom, or were volunteer examiners. Approximately eighty-eight percent of the GDO study examiners have a Bachelor's, Master's, or Doctoral degree. Refer to Table 9 for examiner's level of education.

Table 9
Examiner Demographics: Level of Education

Level of Education	**N**	**% sample**
Some College	6	5.94
Associate's Degree	1	.99
Bachelor's Degree	33	32.67
Master's Degree	55	54.46
Doctoral Degree	2	1.98
Missing information	4	3.96
Total	**101**	**100**

Data Review and Validation Procedures

Data was systematically reviewed for completion, accuracy, and any possible serious administration errors prior to being entered into an electronic Survey Gizmo file. Unusable data was put aside. Ten percent of the data entered in each of the four datasets (GDO, Pilot, TQ, and PQ) was checked by a team of interns, and any keystroke or scoring errors were corrected. A subsample of data collected in New Haven, CT by a team of subcontractors was also checked for accuracy and reliability by members of the research team.

Characteristics of unusable GDO-R, TQ, and PQ data:

- Data was beyond the upper limit of the 6 year age band.
- Data was below the lower limit of the 3 year age band.
- Assessments were incorrectly administered in Spanish.
- Assessment was substantially incomplete.

A total of 1,363 GDO assessments were submitted to Gesell Institute for evaluation. After careful review for accuracy, proper administration, and age requirements, a final sample of 1,287 GDO-R assessments were used in the final analysis. Thus, only 5% of the sample could not be used and was deleted from the original sample. Over the course of two years, communication with new sites about data collection/submission procedures improved significantly, and thus further reduced the number of invalid assessments that were submitted by each site.

Sample Descriptive Statistics

The final sample used for analysis was a proportional mix of boys and girls, age 3-6 from ethnically diverse backgrounds. The number of children varied across age bands (Table 10).

Table 10
Overview GDO Study Sample Child Descriptive Statistics by Age Band

	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Number of Students	53	131	186	264	278	221	154
Mean Age	3.08	3.54	4.00	4.52	5.00	5.48	5.97
(Standard Deviation)	(0.13)	(0.12)	(0.15)	(0.14)	(0.15)	(0.15)	(0.14)
Sex (%)							
Male	41.5	56.5	43.5	47.3	51.1	48.0	49.4
Female	54.7	39.7	53.8	50.8	47.5	52.0	50.6
Not Reported	3.8	3.8	2.7	1.9	1.4	0.0	0.0
Ethnicity (%)							
African-American	41.5	48.1	40.9	28.0	17.6	5.9	3.2
American Indian	0.0	0.0	0.5	0.8	2.2	3.6	3.2
Caucasian not Hispanic	26.4	14.5	26.3	44.3	56.8	73.3	80.5
Hispanic	17.0	16.0	14.5	11.4	7.6	4.5	1.3
Multiple Ethnicities	11.3	17.6	14.0	12.1	12.2	10.9	8.4
Other	1.9	2.3	2.1	2.7	2.2	1.4	3.2
Not Reported	1.9	1.5	1.6	0.8	1.4	.5	0.0

Child's Native Language (%)							
English	56.6	39.7	56.5	61.7	79.1	90.0	93.5
Spanish	9.4	13.7	6.5	3.8	3.6	1.8	0.0
Multiple	0.0	1.5	2.2	4.5	1.8	1.8	1.3
Other	0.0	.8	2.7	3.0	2.5	1.8	3.9
Not Reported	34.0	44.3	32.3	26.9	12.9	4.5	1.3

Note. Missing data for child's native language occurred because situations arose which prevented a return of parent questionnaires to Gesell Institute. See the Conclusion section for a report on missing data. In such cases, a shortened form of the PQ (Child Demographic Form) was modified so that basic demographic information could be collected from school records for participating families (sex, ethnicity, DOB, Native Language).

Task and Strand Descriptive Statistics by Age Band

Tables 11 through 17 present raw score descriptive statistics for the 17 GDO-R tasks and/or items, and three subscales from the TQ. Since the Developmental Strand provides a developmental age, and not an overall numeric score, no descriptive statistics are provided at the strand level. Tables 11 through 17 also provide descriptive statistics for the Social Development, Emotional Development, and Adaptive Skills scales from the Teacher Questionnaire. Sample size by strand and task in these tables vary for two reasons. First, when all the items within a task were not administered to the child (due to age of child, or judgment of examiner based on a series of unsuccessful responses), the child's data was excluded from analysis of that task. Second, samples sizes for strands are lower than sample sizes for tasks because the strand sample sizes include only those children who were administered all tasks within the strand.

Tables 11 through 17 provide the number of items by strand and task, sample size (N), maximum possible points, mean raw score (Mean), standard deviation (SD), internal consistency coefficient, and mean p-value for the majority of tasks in each strand. Some tasks/items have been organized into separate tables (34 through 45) because they require different statistical operations to best demonstrate the findings. These are:

1) Tasks which contain continuous items that are scored from 0 to a maximum number (the number of numerals written, the number of those numerals that were in sequential order, the number of animals named, how high the child counted, and the number of seconds for dominant hand pellets). Because the items are scored in terms of a maximum number, means and standard deviations provide better evidence than p-values regarding children's performance.

2) Tasks which contain categories (items from Distinguishing Features and Overt Behavior). These items were analyzed using a frequency distribution, because means and standard deviations are not appropriate.

Internal consistency is typically a measure based on the correlations between different items on the same test (or the same subscale on a larger test). It measures whether several items that propose to measure the same general construct produce similar scores. For example, the Action

Agent task and the Prepositions task are highly correlated, and are both measures of receptive language. Strands or tasks that have more items are expected to have higher internal consistency than strands or tasks with fewer items, since this statistical measure is influenced by the number of items in a strand or task (greater than 1). Internal consistency at the task level was calculated as the internal consistency of all items within the task. Internal consistency at the strand level was calculated as the internal consistency of the task scores. The internal consistency coefficients presented in Tables 11 through 17 are consistent with these expectations, and are acceptable for the intended uses and purposes of the Gesell Developmental Observation-Revised.

The p-value represents the proportion of children who provided the correct responses to the item. For polytomous items (scored with 3 or more score points, e.g., Incomplete Man scored according to level of cues, or Cubes scored according to performance with or without a demonstration), the p-value represents the average proportion of the maximum possible score children received. With few exceptions, Tables 11 through 17 show growth in performance on the GDO tasks from age 3.0 to age 6.0.

The item "Acts Shy" in the Emotional subscale was removed from analysis of TQ data. After consultation with the psychometric team, this item could not be objectively coded as a positive or negative behavior, and was likely related to a child's temperament more than to his or her emotional development. However, the item remains in the TQ and PQ for information, but is not scored.

Interests and Interview tasks are not included in the following tables, since they are scored with a qualitative rubric that augments the Language/Comprehension standardized score.

Table 11

Strand and Tasks Descriptive Statistics Age Band 3.0

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42	45	-	-	-	-	-
Cubes	6	53	12	4.23	1.89	0.36	0.35
Copy Forms	13	53	13	3.7	2.14	0.78	0.28
Incomplete Man	13	53	39	7.3	6.17	0.75	0.19
Fine Motor	1	50	99	25.34	11.46	n/a	0.26
Gross Motor	8	51	22	12.92	3.96	0.69	0.59
Letters/Numbers	71	28	155	10.39	6.55	0.33	0.08
Writing Name	3	51	3	0.08	0.34	0.50	0.03
Identifying Letters	26	50	26	1.18	3.10	0.92	0.05
Counting	1	35	40	5.86	5.82	n/a	0.15
Digit Repetition	12	52	12	1.75	2.12	0.80	0.15
Writing Numbers	2	50	40	0.16	0.55	0.00	0.00
Identifying Numbers	12	50	12	0.64	1.99	0.92	0.05
One-to-One Correspondence (includes Conservation item)	8	53	8	0.98	1.42	0.73	0.12
Calculations	7	44	14	0.55	1.04	0.55	0.04
Language/Comprehension	32	44	91	10.2	8.46	0.74	0.11
Prepositions	5	52	5	2.1	1.35	0.68	0.42
Action Agents	21	51	21	4.84	4.82	0.90	0.23
Naming Animals	1	47	60	1.89	2.19	n/a	0.03
Comprehension Questions	5	51	5	1.37	1.43	0.67	0.27
Visual/Spatial Discrimination	21	47	25	15.19	3.73	0.37	0.61
Color Forms	5	52	5	4.4	1.32	0.87	0.88
Visual I	12	47	12	3.87	2.76	0.73	0.32
Three-Hole Form Board	4	52	8	6.88	1.41	0.75	0.86
Social Development	8	60	40	3.37	0.90	0.91	0.08
Emotional Development	10	59	50	3.48	0.69	0.82	0.07
Adaptive Skills	6	60	30	3.4	0.82	0.83	0.11

Table 12

Strand and Tasks Descriptive Statistics Age Band 3.5

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42	112	-	-	-	-	
Cubes	6	130	11	5.03	2.16	0.55	0.46
Copy Forms	13	131	13	4.47	2.10	0.78	0.34
Incomplete Man	13	130	29	11.74	7.70	0.79	0.40
Fine Motor	1	126	99	24.91	11.43	n/a	0.25
Gross Motor	8	129	22	13.56	4.54	0.77	0.62
Letters/Numbers	71	64	155	18.11	16.31	0.79	0.13
Writing Name	3	122	3	0.27	0.56	0.44	0.09
Identifying Letters	26	125	26	3.41	6.74	0.97	0.13
Counting	1	90	40	6.91	5.51	n/a	0.17
Digit Repetition	12	125	12	2.64	2.33	0.81	0.22
Writing Numbers	2	105	40	0.16	1.00	0.90	0.00
Identifying Numbers	12	122	12	1.6	3.19	0.95	0.13
One-to-One Correspondence (includes Conservation item)	8	128	8	1.42	1.60	0.72	0.18
Calculations	7	107	14	1.24	2.16	0.84	0.09
Language/Comprehension	32	116	91	14.67	8.66	0.72	0.16
Prepositions	5	126	5	2.79	1.52	0.73	0.56
Action Agents	21	129	21	6.8	5.35	0.90	0.32
Naming Animals	1	122	60	2.66	2.31	n/a	0.04
Comprehension Questions	5	127	5	2	1.55	0.70	0.40
Visual/Spatial Discrimination	21	122	25	15.81	3.74	0.46	0.63
Color Forms	5	130	5	4.75	0.88	0.87	0.95
Visual I	12	123	12	3.93	2.82	0.76	0.33
Three-Hole Form Board	4	128	8	7.15	1.53	0.82	0.89
Social Development	8	120	40	3.65	0.73	0.87	0.09
Emotional Development	10	116	50	3.57	0.60	0.74	0.07
Adaptive Skills	6	124	30	3.53	0.87	0.89	0.12

Table 13

Strand and Tasks Descriptive Statistics Age Band 4.0

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42		-	-	-	-	
Cubes	6	186	12	6.26	1.90	0.48	0.52
Copy Forms	13	185	13	4.85	1.96	0.76	0.37
Incomplete Man	13	185	35	16.61	7.38	0.75	0.47
Fine Motor	1	171	99	25.74	11.37	n/a	0.26
Gross Motor	8	172	22	15.47	3.18	0.57	0.70
Letters/Numbers	71	102	155	29.57	19.13	0.73	0.22
Writing Name	3	184	3	0.66	0.76	0.47	0.22
Identifying Letters	26	172	26	5.9	8.25	0.97	0.23
Counting	1	122	40	9.56	6.86	n/a	0.24
Digit Repetition	12	179	12	3.65	2.32	0.80	0.30
Writing Numbers	2	171	40	0.65	1.65	0.83	0.02
Identifying Numbers	12	171	12	3.04	3.98	0.94	0.25
One-to-One Correspondence (includes Conservation item)	8	171	8	2.21	2.07	0.81	0.28
Calculations	7	147	14	1.99	2.43	0.82	0.14
Language/Comprehension	32	159	91	19.65	9.43	0.74	0.22
Prepositions	5	177	5	3.36	1.46	0.73	0.67
Action Agents	21	176	21	9.67	5.39	0.89	0.46
Naming Animals	1	173	60	4.37	3.15	n/a	0.07
Comprehension Questions	5	176	5	2.57	1.63	0.71	0.51
Visual/Spatial Discrimination	21	166	25	17.39	3.06	0.16	0.70
Color Forms	5	175	5	4.94	0.37	0.72	0.99
Visual I	12	170	12	5.09	2.74	0.71	0.42
Three-Hole Form Board	4	176	8	7.43	1.05	0.65	0.93
Social Development	8	153	40	3.97	0.64	0.86	0.10
Emotional Development	10	149	50	3.73	0.59	0.75	0.07
Adaptive Skills	6	147	30	3.92	0.68	0.82	0.13

Table 14

Strand and Tasks Descriptive Statistics Age Band 4.5

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42		-	-	-	-	
Cubes	6	264	12	7.79	2.23	0.61	0.65
Copy Forms	13	261	12	5.91	1.64	0.66	0.49
Incomplete Man	13	264	33	20.91	7.25	0.76	0.63
Fine Motor	1	214	99	25.26	9.97	n/a	0.26
Gross Motor	8	233	21	16.48	3.11	0.56	0.78
Letters/Numbers	71	152	155	56.07	32.56	0.87	0.42
Writing Name	3	260	3	1.42	0.97	0.65	0.47
Identifying Letters	26	232	26	12.04	10.23	0.98	0.46
Counting	1	193	40	16.85	11.43	n/a	0.42
Digit Repetition	12	263	12	5.08	2.76	0.85	0.42
Writing Numbers	2	243	40	4.88	8.13	0.97	0.12
Identifying Numbers	12	230	12	6.3	4.83	0.96	0.53
One-to-One Correspondence (includes Conservation item)	8	231	8	3.82	2.74	0.88	0.48
Calculations	7	208	14	3.6	3.20	0.83	0.26
Language/Comprehension	32	227	91	25.42	9.40	0.76	0.28
Prepositions	5	238	5	4	1.44	0.81	0.80
Action Agents	21	258	21	12.8	5.38	0.90	0.61
Naming Animals	1	262	60	5.98	3.16	n/a	0.10
Comprehension Questions	5	233	5	3.24	1.68	0.75	0.65
Visual/Spatial Discrimination	21	229	25	19.64	3.62	0.31	0.79
Color Forms	5	232	5	4.94	0.45	0.88	0.99
Visual I	12	253	12	7.15	3.22	0.82	0.60
Three-Hole Form Board	4	232	8	7.69	0.82	0.67	0.96
Social Development	8	199	40	4.11	0.72	0.89	0.10
Emotional Development	10	196	50	3.87	0.61	0.78	0.08
Adaptive Skills	6	199	30	4.04	0.76	0.87	0.13

Table 15

Strand and Tasks Descriptive Statistics Age Band 5.0

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42		-	-	-	-	
Cubes	6	278	12	9.02	2.10	0.56	0.75
Copy Forms	13	278	13	6.81	1.73	0.67	0.52
Incomplete Man	13	277	35	23.9	5.52	0.63	0.68
Fine Motor	1	202	99	25.28	10.62	n/a	0.26
Gross Motor	8	232	22	17.4	2.46	0.45	0.79
Letters/Numbers	71	160	155	82.18	33.48	0.85	0.61
Writing Name	3	278	3	2.02	0.86	0.59	0.67
Identifying Letters	26	232	26	16.83	9.67	0.98	0.65
Counting	1	201	40	23.75	11.56	n/a	0.59
Digit Repetition	12	249	12	5.9	2.40	0.81	0.49
Writing Numbers	2	261	40	11.15	10.79	0.97	0.28
Identifying Numbers	12	237	12	8.47	4.08	0.94	0.71
One-to-One Correspondence (includes Conservation item)	8	219	8	5.27	2.53	0.86	0.66
Calculations	7	214	14	5.33	3.63	0.85	0.38
Language/Comprehension	32	226	91	30.91	7.58	0.60	0.34
Prepositions	5	232	5	4.56	0.99	0.76	0.91
Action Agents	21	254	21	14.99	4.12	0.83	0.71
Naming Animals	1	274	60	7.64	3.56	n/a	0.13
Comprehension Questions	5	231	5	3.8	1.40	0.67	0.76
Visual/Spatial Discrimination	21	223	25	21.35	2.93	0.03	0.85
Color Forms	5	225	5	4.98	0.24	0.77	1.00
Visual I	12	248	12	8.55	2.87	0.81	0.71
Three-Hole Form Board	4	225	8	7.91	0.40	0.62	0.99
Social Development	8	198	40	4.07	0.71	0.88	0.10
Emotional Development	10	191	50	3.86	0.61	0.80	0.08
Adaptive Skills	6	198	30	4.05	0.82	0.90	0.14

Table 16

Strand and Tasks Descriptive Statistics Age Band 5.5

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42		-	-	-	-	
Cubes	6	221	12	9.97	1.97	0.53	0.83
Copy Forms	13	220	11	7.52	1.76	0.65	0.68
Incomplete Man	13	221	36	25.68	4.57	0.48	0.71
Fine Motor	1	161	99	24.91	10.63	n/a	0.25
Gross Motor	8	183	22	18.38	2.43	0.53	0.84
Letters/Numbers	71	151	155	105.56	33.66	0.83	0.78
Writing Name	3	220	3	2.3	0.64	0.35	0.77
Identifying Letters	26	184	26	21.6	7.38	0.97	0.83
Counting	1	164	40	29.36	12.36	n/a	0.73
Digit Repetition	12	190	12	6.97	2.39	0.80	0.58
Writing Numbers	2	211	40	18.21	13.61	0.97	0.46
Identifying Numbers	12	187	12	10.33	3.02	0.93	0.86
One-to-One Correspondence (includes Conservation item)	8	179	8	6.37	2.07	0.82	0.80
Calculations	7	179	14	7.22	3.68	0.84	0.52
Language/Comprehension	32	177	91	35.05	5.89	0.45	0.39
Prepositions	5	186	5	4.91	0.47	0.75	0.98
Action Agents	21	189	21	16.53	2.88	0.69	0.79
Naming Animals	1	214	60	9.48	3.94	n/a	0.16
Comprehension Questions	5	183	5	4.11	1.24	0.66	0.82
Visual/Spatial Discrimination	21	182	25	22.97	2.74	0.54	0.92
Color Forms	5	182	5	4.95	0.52	1.00	0.99
Visual I	12	190	12	10.04	2.26	0.79	0.84
Three-Hole Form Board	4	182	8	7.92	0.70	0.96	0.99
Social Development	8	238	40	4.14	0.73	0.88	0.10
Emotional Development	10	236	50	3.87	0.70	0.85	0.08
Adaptive Skills	6	237	30	3.99	0.85	0.90	0.13

Table 17

Strand and Tasks Descriptive Statistics Age Band 6.0

Strand and Tasks	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Developmental	42		-	-	-	-	
Cubes	6	152	12	10.79	1.64	0.52	0.90
Copy Forms	13	153	12	8.12	1.78	0.62	0.68
Incomplete Man	13	154	39	27.53	4.02	0.38	0.71
Fine Motor	1	135	99	21.09	6.52	n/a	0.21
Gross Motor	8	136	22	18.8	2.60	0.63	0.85
Letters/Numbers	71	112	155	127.79	28.16	0.83	0.95
Writing Name	3	154	3	2.61	0.60	0.36	0.87
Identifying Letters	26	139	26	24.07	4.86	0.96	0.93
Counting	1	118	40	34.48	10.09	n/a	0.86
Digit Repetition	12	141	12	7.57	2.51	0.82	0.63
Writing Numbers	2	150	40	29.42	13.35	0.97	0.74
Identifying Numbers	12	140	12	11.46	1.81	0.92	0.96
One-to-One Correspondence (includes Conservation item)	8	135	8	6.9	1.80	0.82	0.86
Calculations	7	137	14	9.57	3.32	0.84	0.68
Language/Comprehension	32	131	91	37.33	5.71	0.52	0.41
Prepositions	5	137	5	4.89	0.55	0.81	0.98
Action Agents	21	140	21	17.61	2.40	0.66	0.84
Naming Animals	1	149	60	10.42	3.58	n/a	0.17
Comprehension Questions	5	134	5	4.15	1.25	0.68	0.83
Visual/Spatial Discrimination	21	136	25	23.88	1.61	0.01	0.96
Color Forms	5	137	5	5	0.00	*	1.00
Visual I	12	142	12	10.92	1.56	0.72	0.91
Three-Hole Form Board	4	137	8	7.96	0.22	0.37	1.00
Social Development	8	244	40	4.26	0.66	0.86	0.11
Emotional Development	10	238	50	4.00	0.67	0.84	0.08
Adaptive Skills	6	243	30	4.13	0.79	0.87	0.14

* Alpha could not be calculated because there was no variance in the item score; all children received the maximum score.

Item P-Values by Task and Age Band

Tables 18 through 33 provide p-values, also known as item difficulties, for the GDO-R items by task and age band. For dichotomously scored items (scored 0 or 1), the p-value represents the proportion of children who provided the correct responses to the item. For polytomous items (scored with 3 or more score points), the p-value represents the average proportion of the maximum possible score children received. Some item responses were missing when children did not provide an answer to an item or were not administered an item because they gave incorrect responses to the number of previous items that met the stop rule. Missing item responses were treated as incorrect for these analyses.

The p-values for items within a task tend to reflect the fact that items gradually increase in difficulty. For this reason, p-values tend to be somewhat higher for earlier items and lower for later items. P-values also reflect the children's increased competency with age; p-values are lower for the younger children and higher for the older children.

For the Copy Forms task, some age 6 children were not administered the first three items when the examiner believed the items were too easy for these children. In this case, the children received a missing score which was treated as incorrect for the analyses, leading to the lower p-values for these items for age 6 children.

Table 18
P-values for Cubes Task by Age Band (polytomous items)

	3.0 (n=53)	3.5 (n=130)	4.0 (n=186)	4.5 (n=264)	5.0 (n=278)	5.5 (n=221)	6.0 (n=152)
Tower	0.86	0.91	0.93	0.94	0.97	0.99	1.00
Train	0.61	0.68	0.83	0.91	0.93	0.96	0.94
Bridge	0.51	0.66	0.87	0.94	0.98	0.99	1.00
Gate	0.07	0.21	0.32	0.64	0.80	0.89	0.93
Steps (6)	0.06	0.05	0.14	0.32	0.49	0.65	0.81
Steps (10)	0.02	0.02	0.05	0.16	0.35	0.52	0.74

Note. In Cubes, examiners were instructed to administer each item, and provide a DEMO if required. Each item was scored using three categories: Successfully completed without DEMO (2), Successfully completed with DEMO (1), Unsuccessful (0).

Table 19
P-values for Writing Name Task by Age Band

	3.0 (n=51)	3.5 (n=122)	4.0 (n=184)	4.5 (n=260)	5.0 (n=278)	5.5 (n=220)	6.0 (n=154)
Letters Only	0.06	0.22	0.50	0.77	0.92	0.98	0.99
First Name	0.02	0.04	0.15	0.53	0.80	0.94	0.97
Last Name	0.00	0.01	0.02	0.12	0.30	0.38	0.66

Table 20

P-values for Copy Forms Task by Age Band

	3.0 (n=53)	3.5 (n=131)	4.0 (n=185)	4.5 (n=261)	5.0 (n=278)	5.5 (n=220)	6.0 (n=153)
Scribble	0.81	0.87	0.90	0.91	0.95	0.91	0.79
Stroke – Vertical	0.75	0.87	0.85	0.92	0.95	0.93	0.82
Stroke – Horizontal	0.74	0.80	0.83	0.91	0.95	0.92	0.83
Circle	0.74	0.85	0.94	0.98	0.99	1.00	0.99
Cross	0.30	0.53	0.67	0.87	0.94	0.97	1.00
Square	0.23	0.25	0.36	0.71	0.83	0.93	0.98
Triangle	0.08	0.15	0.12	0.35	0.55	0.73	0.90
Divided Rectangle	0.02	0.04	0.06	0.12	0.26	0.41	0.65
Diamond – Horizontal	0.02	0.04	0.05	0.05	0.19	0.34	0.48
Diamond – Vertical	0.02	0.04	0.05	0.07	0.19	0.34	0.58
3-D Cylinder	0.00	0.02	0.01	0.01	0.02	0.02	0.07
3-D Cube Face-on	0.00	0.01	0.01	0.01	0.00	0.00	0.01
3-D Cube Point-on	0.00	0.01	0.01	0.00	0.00	0.00	0.01

Note. In Copy Forms, examiners were instructed to administer each item, and to administer demonstrations (DEMOS) if required. However, for the purpose of the study, each item was scored (0, 1) regardless of any DEMO's required.

Table 21

P-values for Incomplete Man Task by Age Band (polytomous items)

	3.0 (n=53)	3.5 (n=130)	4.0 (n=185)	4.5 (n=264)	5.0 (n=277)	5.5 (n=221)	6.0 (n=154)
Arm	0.33	0.57	0.72	0.83	0.92	0.96	0.98
Hand	0.09	0.24	0.50	0.74	0.87	0.94	0.96
Leg	0.53	0.68	0.83	0.93	0.98	0.99	1.00
Foot	0.20	0.36	0.66	0.85	0.94	0.97	0.98
Eyes	0.46	0.50	0.65	0.77	0.87	0.92	0.94
Ear	0.17	0.29	0.43	0.59	0.74	0.83	0.92
Hair	0.19	0.36	0.47	0.65	0.76	0.84	0.91
Neck	0.06	0.11	0.17	0.25	0.31	0.40	0.56
Bowtie	0.00	0.09	0.09	0.21	0.31	0.51	0.68
Knot	0.00	0.03	0.03	0.06	0.12	0.15	0.23
Body Line	0.26	0.38	0.47	0.62	0.78	0.75	0.73
Other – 1	0.12	0.22	0.37	0.34	0.28	0.23	0.22
Other – 2	0.03	0.08	0.14	0.13	0.09	0.08	0.08

Note. The number of body parts included for Incomplete Man task in the study was 13. These include Knot, Other-1, and Other-2 which are not customarily part of the standard GDO-R administration. Thus, the mean number of body parts is relative to a denominator of 13, rather than 10. During data collection, examiners were instructed to use appropriate cueing if required, and to score each body part added into four categories: Body part added Spontaneously (3),

Body part added following a General Cue (2), Body part added following a Specific Cue (1), or Body part not added at all (0).

Table 22
P-values for Visual I Task by Age Band

	3.0 (n=47)	3.5 (n=123)	4.0 (n=170)	4.5 (n=253)	5.0 (n=248)	5.5 (n=190)	6.0 (n=142)
Square with line	0.49	0.37	0.49	0.65	0.72	0.90	0.94
Circle	0.36	0.45	0.52	0.71	0.89	0.95	0.99
E	0.32	0.44	0.47	0.62	0.77	0.89	0.93
Circle over dot	0.43	0.29	0.43	0.61	0.70	0.81	0.94
Triangle over ½ circle	0.34	0.43	0.59	0.77	0.79	0.93	0.96
E backwards 9	0.36	0.35	0.51	0.72	0.85	0.91	0.98
Skip 8 (recognized the skip)	0.06	0.04	0.08	0.21	0.31	0.48	0.58
B	0.09	0.06	0.13	0.28	0.43	0.62	0.75
Arrow	0.30	0.37	0.56	0.74	0.88	0.96	0.97
Circle, square, triangle	0.36	0.46	0.52	0.68	0.82	0.91	0.99
Circle, dot, line, circle, line	0.34	0.25	0.35	0.49	0.60	0.76	0.92
½ circle, square, triangle, circle	0.43	0.41	0.45	0.67	0.79	0.92	0.96

Note. In the Visual I task, the first item was a teaching item (triangle), and "recognizing the skip" was an item scored as part of the total (12).

Table 23
P-values for Prepositions Task by Age Band

	3.0 (n=52)	3.5 (n=126)	4.0 (n=177)	4.5 (238)	5.0 (n=232)	5.5 (n=186)	6.0 (n=137)
On	0.92	0.97	0.98	0.98	0.99	0.99	1.00
Under	0.54	0.67	0.82	0.88	0.97	0.99	0.99
In back of	0.29	0.44	0.66	0.77	0.90	0.99	0.96
In front of	0.15	0.35	0.47	0.68	0.86	0.96	0.96
Beside	0.19	0.35	0.43	0.69	0.84	0.96	0.99

Table 24

P-values for Digit Repetition Task by Age Band

	3.0 (n=52)	3.5 (n=125)	4.0 (n=179)	4.5 (n=263)	5.0 (n=249)	5.5 (n=190)	6.0 (n=141)
6-4-1	0.46	0.62	0.77	0.87	0.96	0.98	0.98
3-5-2	0.38	0.55	0.75	0.84	0.92	0.97	0.99
8-3-7	0.40	0.59	0.75	0.86	0.93	0.97	0.99
4-7-2-9	0.13	0.30	0.41	0.62	0.76	0.87	0.89
3-8-5-2	0.17	0.19	0.34	0.54	0.68	0.79	0.89
7-2-6-1	0.12	0.26	0.42	0.63	0.72	0.85	0.85
2-1-8-5-9	0.02	0.07	0.09	0.22	0.33	0.40	0.60
4-8-3-7-2	0.02	0.02	0.06	0.22	0.28	0.45	0.48
9-6-1-8-3	0.02	0.03	0.05	0.16	0.16	0.27	0.28
2-9-4-8-1-6	0.02	0.00	0.00	0.04	0.04	0.13	0.18
9-6-2-9-3-8	0.00	0.00	0.01	0.06	0.06	0.16	0.23
5-1-7-2-6-9	0.00	0.00	0.01	0.03	0.05	0.13	0.20

Note. Administration was terminated when child unsuccessfully repeated two out of three digit sets in the row.

Table 25

P-values for Comprehension Task by Age Band

	3.0 (n=51)	3.5 (n=127)	4.0 (n=176)	4.5 (n=233)	5.0 (n=231)	5.5 (n=183)	6.0 (n=134)
Hungry	0.41	0.64	0.67	0.70	0.74	0.81	0.80
Sleepy	0.41	0.48	0.56	0.67	0.77	0.79	0.83
Cold	0.25	0.47	0.62	0.70	0.83	0.86	0.84
Lost Something	0.12	0.13	0.26	0.51	0.65	0.79	0.84
Cross Street	0.18	0.28	0.46	0.67	0.81	0.86	0.85

Table 26

P-values for Color Forms Task by Age Band

	3.0 (n=52)	3.5 (n=130)	4.0 (n=175)	4.5 (n=232)	5.0 (n=225)	5.5 (n=182)	6.0 (n=137)
Circle	0.88	0.94	0.98	0.99	1.00	0.99	1.00
Square	0.88	0.92	0.99	0.99	1.00	0.99	1.00
Triangle	0.85	0.97	0.99	0.99	1.00	0.99	1.00
Cross	0.87	0.97	0.99	0.99	1.00	0.99	1.00
Half Moon	0.92	0.95	0.98	0.98	0.99	0.99	1.00

Table 27

P-values for Three-Hole Form Board Task by Age Band

	3.0 (n=52)	3.5 (n=128)	4.0 (n=176)	4.5 (n=232)	5.0 (n=225)	5.5 (n=182)	6.0 (n=137)
Square, Triangle, Circle (presentation 1)	0.96	0.98	0.99	1.00	1.00	0.99	1.00
Circle, Triangle, Square (presentation 2)	0.80	0.89	0.92	0.96	0.99	0.99	1.00
Square, Triangle, Circle (presentation 3)	0.84	0.86	0.89	0.94	0.98	0.99	1.00
Circle, Triangle, Square (presentation 4)	0.85	0.85	0.93	0.96	0.99	0.99	1.00

Note. In the Three-Hole Form Board task, children were given four presentations of the board, each rotated 180 degrees while keeping the board parallel to the table (i.e., board was not flipped over). Each presentation was scored using three categories: Successfully completed (2), Successfully completed with Trial and Error (1), or Unsuccessful (0).

Table 28

P-values for One-to-One Correspondence Task by Age Band

	3.0 (n=53)	3.5 (n=128)	4.0 (n=171)	4.5 (n=231)	5.0 (n=219)	5.5 (n=179)	6.0 (n=135)
4 pennies, count them	0.43	0.55	0.74	0.85	0.94	0.99	0.99
Altogether	0.23	0.25	0.38	0.62	0.79	0.85	0.93
10 pennies, count them	0.13	0.27	0.43	0.59	0.80	0.87	0.87
Altogether	0.08	0.08	0.26	0.46	0.72	0.80	0.85
13 pennies, count them	0.06	0.14	0.18	0.45	0.60	0.80	0.86
Altogether	0.02	0.04	0.10	0.34	0.54	0.74	0.82
20 pennies, count them	0.02	0.06	0.08	0.27	0.45	0.67	0.81
Altogether	0.02	0.02	0.06	0.24	0.43	0.64	0.76

Table 29

P-values for Calculations Task by Age Band

	3.0 (n=44)	3.5 (n=107)	4.0 (n=147)	4.5 (n=208)	5.0 (n=214)	5.5 (n=179)	6.0 (n=137)
2+2	0.08	0.17	0.25	0.44	0.64	0.77	0.87
2+3	0.04	0.13	0.17	0.39	0.53	0.68	0.84
5-2	0.07	0.13	0.21	0.35	0.50	0.63	0.77
7+3	0.01	0.07	0.08	0.19	0.35	0.48	0.70
6-4	0.08	0.11	0.21	0.27	0.34	0.51	0.66
14+3	0.00	0.02	0.03	0.08	0.18	0.30	0.55
16-4	0.00	0.02	0.06	0.09	0.15	0.25	0.41

Note. Children were scored using three categories; Successfully completed without pennies (2), Successfully completed with pennies (1), or Unsuccessful (0).

Table 30

P-values for Action Agents Task by Age Band

	3.0 (n=51)	3.5 (n=129)	4.0 (n=176)	4.5 (n=258)	5.0 (n=254)	5.5 (n=189)	6.0 (n=140)
Sleeps	0.53	0.57	0.68	0.78	0.88	0.88	0.89
Scratches	0.27	0.37	0.56	0.65	0.78	0.89	0.89
Flies	0.37	0.57	0.77	0.84	0.94	0.96	0.98
Bites	0.37	0.48	0.68	0.81	0.89	0.93	0.95
Swims	0.35	0.49	0.72	0.83	0.91	0.96	0.96
Burns	0.29	0.45	0.61	0.70	0.82	0.88	0.92
Cuts	0.33	0.38	0.57	0.74	0.78	0.88	0.96
Blows	0.27	0.33	0.45	0.60	0.75	0.81	0.88
Shoots	0.18	0.41	0.44	0.71	0.81	0.89	0.90
Melts	0.22	0.42	0.56	0.72	0.90	0.90	0.94
Sails	0.14	0.19	0.33	0.57	0.70	0.84	0.93
Boils	0.04	0.10	0.22	0.41	0.41	0.58	0.66
Floats	0.20	0.34	0.47	0.58	0.75	0.78	0.81
Growls	0.12	0.23	0.41	0.62	0.70	0.82	0.91
Stings	0.14	0.22	0.36	0.65	0.79	0.90	0.95
Gallops	0.12	0.16	0.26	0.45	0.60	0.72	0.84
Aches	0.06	0.05	0.09	0.16	0.21	0.32	0.44
Explodes	0.08	0.11	0.20	0.40	0.59	0.74	0.82
Roars	0.31	0.36	0.49	0.70	0.78	0.84	0.91
Mews	0.04	0.09	0.14	0.18	0.30	0.28	0.21
Meows	0.41	0.48	0.66	0.70	0.68	0.72	0.86

Note. The number of Action Agents includes both Mews and Meows as both were tested in the study protocol. Thus, the mean number of Action Agents named correctly is relative to a denominator of 21, rather than 20.

Table 31

P-values for Identifying Letters Task by Age Band

	3.0 (n=50)	3.5 (n=125)	4.0 (n=172)	4.5 (n=229)	5.0 (n=232)	5.5 (n=184)	6.0 (n=139)
A	0.12	0.15	0.34	0.60	0.80	0.89	0.96
B	0.06	0.14	0.30	0.54	0.69	0.86	0.94
C	0.06	0.14	0.31	0.50	0.71	0.87	0.92
D	0.04	0.14	0.20	0.47	0.66	0.82	0.92
E	0.06	0.10	0.24	0.48	0.69	0.83	0.94
F	0.02	0.13	0.20	0.45	0.59	0.84	0.93
G	0.06	0.12	0.20	0.42	0.59	0.79	0.93
H	0.02	0.10	0.20	0.48	0.62	0.84	0.94
I	0.00	0.10	0.13	0.34	0.52	0.64	0.81
J	0.04	0.14	0.20	0.47	0.65	0.83	0.93
K	0.06	0.20	0.21	0.49	0.63	0.82	0.94
L	0.04	0.11	0.19	0.45	0.66	0.82	0.94
M	0.06	0.16	0.24	0.46	0.66	0.82	0.93
N	0.04	0.12	0.19	0.47	0.63	0.81	0.95
O	0.06	0.15	0.34	0.57	0.75	0.90	0.94
P	0.06	0.12	0.23	0.47	0.66	0.88	0.94
Q	0.02	0.16	0.22	0.42	0.60	0.83	0.90
R	0.02	0.13	0.22	0.46	0.65	0.85	0.91
S	0.08	0.11	0.21	0.51	0.70	0.89	0.93
T	0.04	0.14	0.22	0.48	0.66	0.85	0.93
U	0.04	0.07	0.17	0.38	0.56	0.80	0.91
V	0.02	0.08	0.13	0.36	0.56	0.74	0.89
W	0.06	0.16	0.25	0.42	0.64	0.83	0.91
X	0.02	0.17	0.33	0.54	0.72	0.91	0.97
Y	0.04	0.14	0.23	0.46	0.60	0.82	0.93
Z	0.04	0.14	0.23	0.48	0.65	0.83	0.96

Table 32
P-values for Numbers Task by Age Band

	3.0 (n=50)	3.5 (n=122)	4.0 (n=171)	4.5 (n=230)	5.0 (n=237)	5.5 (n=187)	6.0 (n=140)
1	0.10	0.25	0.42	0.69	0.91	0.94	0.99
2	0.12	0.20	0.37	0.62	0.84	0.94	0.99
3	0.06	0.19	0.38	0.64	0.82	0.94	0.98
4	0.06	0.16	0.35	0.63	0.81	0.94	0.97
5	0.06	0.20	0.33	0.61	0.82	0.94	0.97
6	0.02	0.10	0.22	0.50	0.67	0.86	0.93
7	0.04	0.15	0.21	0.55	0.69	0.88	0.96
8	0.06	0.12	0.20	0.52	0.67	0.87	0.95
9	0.04	0.07	0.22	0.46	0.59	0.81	0.94
10	0.04	0.05	0.15	0.43	0.63	0.82	0.94
11	0.02	0.07	0.10	0.38	0.57	0.74	0.92
12	0.02	0.04	0.09	0.27	0.44	0.66	0.92

Table 33
P-values for Motor Task by Age Band

	3.0 (n=51)	3.5 (n=129)	4.0 (n=172)	4.5 (n=233)	5.0 (n=232)	5.5 (n=183)	6.0 (n=136)
Walk on Tiptoe	0.63	0.66	0.82	0.85	0.90	0.96	0.97
Jump in Place	0.88	0.89	0.96	0.96	0.99	0.99	0.98
Stand on One Foot	0.58	0.61	0.69	0.78	0.84	0.90	0.93
Hop on One Foot	0.66	0.74	0.83	0.85	0.93	0.94	0.97
Skip	0.44	0.45	0.58	0.61	0.74	0.79	0.87
Jump	0.69	0.72	0.90	0.93	0.91	0.95	0.96
Beanbag Throw	0.52	0.52	0.50	0.51	0.48	0.50	0.49
Beanbag Catch	0.50	0.52	0.60	0.69	0.71	0.79	0.80

Note. In the Motor tasks, each item was scored according to varying levels of response: Walk on Tiptoe (0-3), Jump in Place (0-2), Stand on One Foot (0-5), Hop on One Foot (0-2), Skip (0-2), Jump (0, 1), Beanbag Throw (0-3), and Beanbag Catch (0-4).

Descriptive Statistics for Continuous Items

In addition to the item p-values presented previously, descriptive statistics are provided in this section for continuous items, items that are scored from 0 to a maximum number. These items include the number of numerals named, the number of those numerals that were in sequential order, the number of animals named, how high the child counted, and the number of seconds for dominant hand pellets. Because the items are scored in terms of a maximum number, means and standard deviations provide better evidence than p-values regarding children's performance. Presented in Table 34 are the sample size (N), maximum score possible (Maximum), mean (Mean), and standard deviation (SD).

Table 34

Descriptive Statistics for Continuous Items

		N	Maximum	Mean	SD
3.0	Number of numerals written	50	20	0.16	0.55
	Number of sequential numbers	50	20	0.00	0.00
	Animals named	47	60	1.89	2.19
	Counting	35	40	5.86	5.82
	Dominant hand pellets (seconds)	50	99	25.34	11.46
3.5	Number of numerals written	105	20	0.12	0.63
	Number of sequential numbers	105	20	0.04	0.39
	Animals named	122	60	2.66	2.31
	Counting	90	40	6.91	5.51
	Pellets (dominant hand)	126	99	24.91	11.43
4.0	Number of numerals written	171	20	0.47	1.04
	Number of sequential numbers	171	20	0.18	0.73
	Animals named	173	60	4.37	3.15
	Counting	122	40	9.56	6.86
	Pellets (dominant hand)	171	99	25.74	11.37
4.5	Number of numerals written	243	20	2.81	4.27
	Number of sequential numbers	243	20	2.07	4.00
	Animals named	262	60	5.98	3.16
	Counting	193	40	16.85	11.43
	Pellets (dominant hand)	214	99	25.26	9.97
5.0	Number of numerals written	261	20	6.15	5.38
	Number of sequential numbers	261	20	5.00	5.57
	Animals named	274	60	7.64	3.56
	Counting	201	40	23.75	11.56
	Pellets (dominant hand)	202	99	25.28	10.62
5.5	Number of numerals written	211	20	9.78	6.52
	Number of sequential numbers	211	20	8.43	7.29
	Animals named	214	60	9.48	3.94
	Counting	164	40	29.36	12.36
	Pellets (dominant hand)	161	99	24.91	10.63
6.0	Number of numerals written	150	20	15.07	6.46
	Number of sequential numbers	150	20	14.35	7.07
	Animals named	149	60	10.42	3.58
	Counting	118	40	34.48	10.09
	Pellets (dominant hand)	135	99	21.09	6.52

Note. In the Naming Animals task, the total number of items in the task was predetermined to be 60 as a baseline for all age groups. For the Counting task, children were allowed to count up to 40.

Frequency Distributions for Distinguishing Features and Overt Behavior Items by Age Band

The Gesell Developmental Observation-Revised includes several items that are assessed using categories, rather than numerical scores. Children's behavior and/or performance were placed into the category by the GDO-R examiner. These categorical items include items from Distinguishing Features and Overt Behaviors on the Child Recording Form. Because these items are scored using categories, mathematical operations (e.g., means and standard deviations) are not appropriate. For this reason, Tables 35 through 45 present the percentage of children, by age band, in each of the respective score categories for the categorical items.

Table 35
Frequency Distribution for Distinguishing Features: Pencil Stroke by Age Band (Percent)

	3.0 (n=39)	3.5 (n=101)	4.0 (n=146)	4.5 (n=227)	5.0 (n=239)	5.5 (n=200)	6.0 (n=131)
Wispy	28.2	29.7	18.5	14.1	10.5	9.0	3.8
Wobbly	66.7	49.5	48.6	35.2	29.7	17.5	14.5
Smooth	5.1	20.8	32.9	50.7	59.8	73.5	81.7

Table 36
Frequency Distribution for Distinguishing Features: Pencil Grasp by Age Band (Percent)

	3.0 (n=37)	3.5 (n=99)	4.0 (n=148)	4.5 (n=227)	5.0 (n=249)	5.5 (n=200)	6.0 (n=139)
Fisted/5 fingers	16.2	13.1	14.9	6.2	5.6	4.0	0.7
Varied	29.7	21.2	10.1	8.4	7.2	3.5	2.9
2-3 fingers - bunch at tip	18.9	13.1	12.2	15.9	25.3	21.0	25.9
5 fingers - bunch at tip	5.4	5.1	4.1	5.3	6.4	7.0	5.8
2-3 finger grasp	21.6	26.3	33.1	47.1	43.4	54.5	53.2
Adult-like	8.1	21.2	25.7	17.2	12.0	10.0	11.5

Table 37
Frequency Distribution for Distinguishing Features: Head Shift by Age Band (Percent)

	3.0 (n=21)	3.5 (n=49)	4.0 (n=75)	4.5 (n=142)	5.0 (n=155)	5.5 (n=141)	6.0 (n=109)
Dominant side	4.8	28.6	36.0	31.0	19.4	10.6	9.2
Non-dominant side	52.4	16.3	30.7	39.4	49.0	55.3	62.4
Alternates	42.9	55.1	33.3	29.6	31.6	34.0	28.4

Table 38

Frequency Distribution for Distinguishing Features: Non-Dominant Hand Posture by Age Band (Percent)

	3.0 (n=23)	3.5 (n=77)	4.0 (n=109)	4.5 (n=196)	5.0 (n=208)	5.5 (n=170)	6.0 (n=118)
Holds paper in place	52.2	58.4	53.2	30.1	25.0	18.2	22.0
Fingers curved holding paper	13.0	23.4	21.1	26.0	22.6	22.4	17.8
Fingers closed/flat on paper	8.7	5.2	4.6	9.2	12.0	11.8	4.2
Fingers flat/little finger or thumb spread	4.3	7.8	8.3	16.3	19.7	17.6	18.6
Fingers spread, flat, sometimes flexed	21.7	5.2	12.8	18.4	20.7	30.0	37.3

Table 39

Frequency Distribution for Distinguishing Features: Eye Movement by Age Band (Percent)

	3.0 (n=20)	3.5 (n=61)	4.0 (n=97)	4.5 (n=167)	5.0 (n=173)	5.5 (n=143)	6.0 (n=113)
Sweeping	0.0	19.7	9.3	4.8	7.5	5.6	3.5
Good tracking	45.0	59.0	56.7	51.5	51.4	65.0	68.1
Restricted	15.0	4.9	12.4	4.2	6.4	4.2	0.0
Constant shift	25.0	9.8	11.3	10.8	15.0	5.6	7.1
Fixate sharply	15.0	6.6	10.3	27.5	18.5	18.2	20.4
Eyes sweep laterally	0.0	0.0	0.0	1.2	1.2	1.4	0.9

Table 40

Frequency Distribution for Overt Behavior: Active, Restless, Much Body Movement by Age Band (Percent)

	3.0 (n=38)	3.5 (n=111)	4.0 (n=154)	4.5 (n=229)	5.0 (n=248)	5.5 (n=200)	6.0 (n=136)
Almost Never	18.4	10.8	16.2	18.8	24.6	25.0	33.1
Rarely	10.5	19.8	17.5	24.0	22.6	22.0	21.3
Sometimes	34.2	17.1	18.8	25.3	19.4	22.0	18.4
Often	5.3	15.3	14.9	9.2	15.3	13.5	11.8
Almost Always	15.8	20.7	17.5	12.7	10.9	12.5	8.8
Not Observed	15.8	16.2	14.9	10.0	7.3	5.0	6.6

Table 41

Frequency Distribution for Overt Behavior: Refused to Complete Tasks by Age Band (Percent)

	3.0 (n=41)	3.5 (n=111)	4.0 (n=153)	4.5 (n=228)	5.0 (n=247)	5.5 (n=199)	6.0 (n=136)
Almost Never	19.5	27.9	28.8	43.9	48.2	61.3	50.7
Rarely	22.0	19.8	20.3	18.0	13.4	15.6	10.3
Sometimes	17.1	16.2	13.7	10.5	9.3	2.0	4.4
Often	17.1	11.7	6.5	6.6	3.6	2.0	0.7
Almost Always	0.0	3.6	3.3	0.9	2.0	0.5	0.7
Not Observed	24.4	20.7	27.5	20.2	23.5	18.6	33.1

Table 42

Frequency Distribution for Overt Behavior: Inattentive, Brief Attention Span by Age Band (Percent)

	3.0 (n=39)	3.5 (n=111)	4.0 (n=150)	4.5 (n=229)	5.0 (n=247)	5.5 (n=200)	6.0 (n=136)
Almost Never	10.3	9.9	21.3	31.0	35.2	44.5	44.1
Rarely	17.9	14.4	22.0	19.7	22.3	19.0	22.8
Sometimes	35.9	34.2	22.0	21.0	13.8	18.0	9.6
Often	10.3	17.1	12.0	10.9	10.9	7.5	7.4
Almost Always	5.1	8.1	9.3	5.2	6.5	4.5	2.9
Not Observed	20.5	16.2	13.3	12.2	11.3	6.5	13.2

Table 43

Frequency Distribution for Overt Behavior: Easily Distracted by Age Band (Percent)

	3.0 (n=39)	3.5 (n=112)	4.0 (n=152)	4.5 (n=231)	5.0 (n=248)	5.5 (n=200)	6.0 (n=136)
Almost Never	5.1	8.9	17.8	27.3	31.9	38.0	44.9
Rarely	20.5	15.2	24.3	26.0	27.0	23.0	20.6
Sometimes	43.6	33.9	21.7	20.8	11.7	20.0	11.8
Often	10.3	16.1	13.2	9.5	12.5	8.5	11.0
Almost Always	5.1	10.7	11.2	7.8	7.7	3.5	2.9
Not Observed	15.4	15.2	11.8	8.7	9.3	7.0	8.8

Table 44

Frequency Distribution for Overt Behavior: Asks Irrelevant Questions by Age Band (Percent)

	3.0 (n=38)	3.5 (n=111)	4.0 (n=152)	4.5 (n=230)	5.0 (n=245)	5.5 (n=200)	6.0 (n=135)
Almost Never	34.2	27.9	31.6	38.7	40.4	46.5	48.9
Rarely	10.5	21.6	16.4	22.6	21.2	21.0	20.7
Sometimes	21.1	23.4	19.7	12.2	15.9	16.0	9.6
Often	10.5	3.6	11.2	7.0	6.9	4.5	3.0
Almost Always	0.0	2.7	3.3	2.2	2.9	1.0	1.5
Not Observed	23.7	20.7	17.8	17.4	12.7	11.0	16.3

Table 45

Frequency for Overt Behavior: Talks When Completing Tasks by Age Band (Percent)

	3.0 (n=40)	3.5 (n=114)	4.0 (n=152)	4.5 (n=231)	5.0 (n=248)	5.5 (n=200)	6.0 (n=136)
Almost Never	22.5	9.6	13.8	15.2	19.0	22.5	23.5
Rarely	5.0	15.8	7.2	10.4	18.5	20.5	19.9
Sometimes	32.5	29.8	30.9	31.2	30.2	23.5	30.1
Often	17.5	21.9	19.7	19.0	13.7	18.0	12.5
Almost Always	12.5	9.6	18.4	17.7	13.7	12.0	11.8
Not Observed	10.0	13.2	9.9	6.5	4.8	3.5	2.2

Inter-correlations Between Tasks by Age Band

Inter-correlations of the Gesell Developmental Observation-Revised were computed to provide information regarding the internal structure of the instrument and the relationships among its various tasks. The inter-correlations (Pearson product moment correlations) presented in Tables 46 through 52 support the score interpretations provided by the Gesell Developmental Observation-Revised. In general, the inter-correlations among the tasks within a strand (e.g., Action Agents and Comprehension) are typically higher than the correlation of tasks between strands (e.g., Comprehension and Fine Motor). The inter-correlations for the younger ages are generally lower than those correlations for the older children, due to the relatively larger measurement error that typically occurs in data from very young children.

A positive correlation of .5 to .9 is indicative of a moderate to strong relationship between two items, so related that one directly implies or is complementary to the other. Correlations between .30-.50 are generally considered to be moderate, and those <.30 are small (Cohen, 1988). Upon reviewing correlations between GDO-R tasks, it became apparent that while some tasks were correlated moderately with others, this was not consistent for the tasks across all age bands. Thus, the GDO-R strands were established using a theoretical framework of child development and criteria derived from early childhood educators on task content and task domain, rather than by correlations between tasks.

Table 46

Inter-correlations Between Tasks: Age Band 3.0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.33																		
3. Incomplete Man	0.40	0.54																	
4. Fine Motor	0.03	0.04	0.19																
5. Gross Motor	0.14	0.18	0.30	-0.17															
6. Writing Name	0.50	0.22	0.33	0.27	0.14														
7. Identifying Letters	0.23	0.15	0.08	-0.09	0.02	-0.05													
8. Counting	0.28	0.26	0.58	0.26	0.13	0.72	0.05												
9. Digit Repetition	0.27	0.22	0.52	0.09	0.24	0.45	0.27	0.61											
10. Writing Numbers	0.41	0.11	0.40	0.20	0.33	0.70	0.05	0.75	0.41										
11. Identifying Numbers	0.20	0.11	0.06	-0.17	0.07	0.09	0.54	0.22	0.49	0.02									
12. One-to-One Correspondence	0.27	0.10	0.08	-0.03	0.32	0.33	0.26	0.35	0.38	0.29	0.52								
13. Calculations	0.23	0.37	0.01	-0.32	0.29	-0.01	-0.03	0.16	0.00	-0.06	0.19	0.41							
14. Prepositions	0.37	0.34	0.36	0.16	0.31	0.38	0.14	0.27	0.34	0.31	0.08	0.19	-0.09						
15. Action Agents	0.45	0.35	0.48	0.38	0.23	0.48	0.31	0.39	0.52	0.43	0.30	0.32	-0.12	0.62					
16. Naming Animals	0.40	0.50	0.43	0.37	0.06	0.38	0.39	0.26	0.33	0.31	0.47	0.31	0.09	0.42	0.75				
17. Comprehension	0.33	0.23	0.39	0.17	0.16	0.36	0.51	0.45	0.47	0.34	0.49	0.44	-0.17	0.46	0.67	0.61			
18. Color Forms	0.22	0.19	0.06	0.02	0.10	0.09	0.18	-0.02	-0.11	0.13	0.15	0.22	0.25	0.24	0.36	0.34	0.23		
19. Visual I	0.31	0.10	-0.01	0.02	0.32	0.28	-0.18	-0.03	0.06	0.16	-0.18	0.12	0.18	0.36	0.05	-0.09	-0.25	0.01	
20. 3-Hole Form Board	0.26	0.37	0.29	0.06	0.44	0.16	0.18	0.03	0.16	0.06	0.16	0.27	0.28	0.18	0.31	0.45	0.27	0.31	0.16

Note. Ns range from 29-53

Table 47

Inter-correlations Between Tasks: Age Band 3.5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.28																		
3. Incomplete Man	0.22	0.40																	
4. Fine Motor	-0.02	-0.19	0.11																
5. Gross Motor	0.25	0.15	0.15	-0.19															
6. Writing Name	0.34	0.27	0.30	0.09	0.18														
7. Identifying Letters	0.24	0.16	0.27	0.09	0.07	0.43													
8. Counting	0.16	0.25	0.36	0.08	-0.06	0.20	0.22												
9. Digit Repetition	0.17	0.25	0.39	-0.09	0.30	0.26	0.27	0.27											
10. Writing Numbers	0.22	0.15	0.25	0.12	0.11	0.52	0.40	0.51	0.07										
11. Identifying Numbers	0.26	0.12	0.36	0.05	0.06	0.36	0.74	0.26	0.21	0.39									
12. One-to-One Correspondence	0.29	0.21	0.34	0.07	0.23	0.39	0.40	0.40	0.37	0.41	0.38								
13. Calculations	0.27	0.18	0.12	0.08	0.10	0.46	0.26	0.14	0.04	0.24	0.21	0.44							
14. Prepositions	0.31	0.18	0.43	0.06	0.27	0.30	0.17	0.28	0.40	0.16	0.18	0.36	0.16						
15. Action Agents	0.30	0.16	0.39	0.10	0.31	0.26	0.20	0.37	0.41	0.17	0.29	0.39	0.23	0.60					
16. Naming Animals	0.26	0.10	0.29	0.15	0.15	0.34	0.18	0.36	0.23	0.43	0.31	0.37	0.37	0.27	0.54				
17. Comprehension	0.25	0.15	0.44	0.21	0.21	0.31	0.22	0.39	0.43	0.23	0.26	0.25	0.17	0.50	0.60	0.34			
18. Color Forms	0.26	0.29	0.27	-0.08	0.26	0.12	0.14	0.13	0.20	0.04	0.13	0.12	0.05	0.28	0.16	0.02	0.22		
19. Visual I	0.26	-0.09	0.10	-0.10	0.27	0.19	0.21	0.11	0.13	0.30	0.28	0.34	0.24	0.17	0.26	0.28	0.11	0.08	
20. 3-Hole Form Board	0.33	0.27	0.35	-0.09	0.33	0.15	0.15	0.24	0.26	0.09	0.05	0.25	0.19	0.26	0.31	0.13	0.33	0.59	0.08

Note. Ns range from 72-131

Table 48

Inter-correlations Between Tasks: Age Band 4.0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.23																		
3. Incomplete Man	0.35	0.33																	
4. Fine Motor	-0.04	-0.18	0.03																
5. Gross Motor	0.18	0.10	0.13	-0.33															
6. Writing Name	0.26	0.27	0.31	-0.06	0.09														
7. Identifying Letters	0.31	0.07	0.30	0.07	0.13	0.40													
8. Counting	0.15	0.02	0.20	0.06	0.18	0.10	0.28												
9. Digit Repetition	0.17	0.15	0.24	0.05	0.27	0.21	0.28	0.27											
10. Writing Numbers	0.13	0.17	0.29	0.06	0.01	0.33	0.29	0.29	0.20										
11. Identifying Numbers	0.33	0.09	0.33	0.05	0.14	0.39	0.77	0.35	0.31	0.35									
12. One-to-One Correspondence	0.26	0.07	0.31	0.02	0.22	0.31	0.48	0.39	0.31	0.24	0.58								
13. Calculations	0.08	-0.08	0.12	-0.10	0.21	0.07	0.18	0.14	0.00	-0.07	0.22	0.29							
14. Prepositions	0.29	0.17	0.42	0.09	0.09	0.26	0.28	0.40	0.32	0.23	0.25	0.25	0.12						
15. Action Agents	0.24	0.10	0.46	0.18	0.13	0.18	0.31	0.37	0.41	0.26	0.32	0.19	0.10	0.61					
16. Naming Animals	0.27	0.20	0.42	0.07	0.13	0.26	0.36	0.29	0.34	0.32	0.37	0.22	0.13	0.39	0.57				
17. Comprehension	0.20	0.07	0.42	0.13	0.18	0.31	0.29	0.35	0.30	0.20	0.33	0.27	0.24	0.50	0.61	0.50			
18. Color Forms	-0.01	0.14	0.05	-0.02	-0.03	0.05	0.08	0.05	0.12	0.06	0.07	0.04	0.12	0.04	0.02	0.07	0.02		
19. Visual I	0.23	0.05	0.21	-0.10	0.28	0.26	0.28	0.20	0.03	0.12	0.28	0.20	0.31	0.30	0.21	0.24	0.27	0.04	
20. 3-Hole Form Board	0.08	0.19	-0.05	-0.07	-0.06	0.04	0.06	-0.05	-0.06	0.03	0.01	0.01	-0.08	-0.01	-0.17	-0.10	-0.10	0.06	0.10

Note. Ns range from 108-186

Table 49

Inter-correlations Between Tasks: Age Band 4.5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.33																		
3. Incomplete Man	0.45	0.33																	
4. Fine Motor	-0.01	-0.21	0.07																
5. Gross Motor	0.17	0.05	0.20	-0.04															
6. Writing Name	0.44	0.44	0.50	0.00	0.13														
7. Identifying Letters	0.42	0.34	0.39	0.03	0.09	0.63													
8. Counting	0.39	0.27	0.25	-0.07	0.08	0.44	0.45												
9. Digit Repetition	0.35	0.17	0.43	-0.04	0.14	0.41	0.44	0.47											
10. Writing Numbers	0.37	0.35	0.32	-0.07	0.09	0.55	0.48	0.49	0.33										
11. Identifying Numbers	0.47	0.34	0.40	-0.04	0.04	0.58	0.79	0.59	0.41	0.50									
12. One-to-One Correspondence	0.46	0.23	0.38	0.02	0.03	0.59	0.57	0.56	0.40	0.39	0.53								
13. Calculations	0.48	0.25	0.29	-0.15	0.11	0.47	0.54	0.41	0.44	0.49	0.54	0.51							
14. Prepositions	0.30	0.17	0.39	0.13	0.14	0.36	0.37	0.37	0.39	0.23	0.47	0.38	0.35						
15. Action Agents	0.43	0.26	0.45	0.08	0.20	0.39	0.50	0.37	0.48	0.28	0.50	0.42	0.43	0.62					
16. Naming Animals	0.32	0.30	0.38	-0.02	0.17	0.41	0.51	0.42	0.38	0.36	0.55	0.35	0.37	0.35	0.47				
17. Comprehension	0.27	0.17	0.32	0.12	0.18	0.24	0.33	0.38	0.28	0.14	0.40	0.27	0.30	0.58	0.60	0.34			
18. Color Forms	0.07	0.05	-0.01	0.03	0.08	0.11	0.04	0.07	0.09	0.02	0.07	0.03	0.12	0.13	0.04	0.00	0.15		
19. Visual I	0.45	0.40	0.37	-0.17	0.14	0.47	0.47	0.49	0.36	0.44	0.54	0.46	0.52	0.38	0.46	0.38	0.34	0.04	
20. 3-Hole Form Board	0.19	0.17	0.13	0.00	0.09	0.19	0.21	0.18	0.20	0.13	0.29	0.17	0.25	0.28	0.15	0.18	0.24	0.49	0.23

Note. Ns range from 171-264

Table 50

Inter-correlations Between Tasks: Age Band 5.0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.31																		
3. Incomplete Man	0.23	0.24																	
4. Fine Motor	0.02	0.01	0.06																
5. Gross Motor	0.11	0.16	0.14	-0.08															
6. Writing Name	0.36	0.26	0.28	0.09	0.22														
7. Identifying Letters	0.37	0.18	0.31	0.10	0.21	0.56													
8. Counting	0.34	0.23	0.33	0.02	0.13	0.30	0.50												
9. Digit Repetition	0.27	0.09	0.26	-0.02	0.09	0.27	0.32	0.36											
10. Writing Numbers	0.32	0.26	0.30	0.05	0.23	0.45	0.51	0.49	0.20										
11. Identifying Numbers	0.39	0.19	0.29	0.04	0.19	0.54	0.74	0.49	0.40	0.54									
12. One-to-One Correspondence	0.35	0.22	0.40	0.04	0.21	0.44	0.56	0.47	0.48	0.51	0.67								
13. Calculations	0.31	0.21	0.25	0.00	0.19	0.34	0.45	0.43	0.27	0.41	0.46	0.51							
14. Prepositions	0.25	-0.01	0.19	0.09	0.09	0.30	0.42	0.24	0.27	0.25	0.41	0.26	0.19						
15. Action Agents	0.29	0.05	0.33	0.08	0.07	0.28	0.47	0.37	0.45	0.24	0.55	0.39	0.43	0.48					
16. Naming Animals	0.29	0.05	0.32	0.05	0.06	0.27	0.37	0.37	0.35	0.21	0.32	0.30	0.25	0.30	0.36				
17. Comprehension	0.16	0.05	0.17	0.06	0.19	0.27	0.24	0.14	0.21	0.17	0.27	0.18	0.23	0.31	0.35	0.15			
18. Color Forms	0.09	0.08	-0.13	0.03	-0.05	-0.04	0.06	0.05	-0.02	-0.08	0.03	0.04	0.01	0.09	0.01	0.08	-0.02		
19. Visual I	0.38	0.25	0.35	0.04	0.16	0.38	0.49	0.44	0.25	0.46	0.54	0.49	0.47	0.39	0.45	0.30	0.26	-0.06	
20. 3-Hole Form Board	0.15	0.03	0.08	0.05	0.02	0.08	0.10	0.13	0.12	0.08	0.17	0.06	-0.09	0.08	0.16	0.12	0.11	-0.02	0.08

Note. Ns range from 171-278

Table 51

Inter-correlations Between Tasks: Age Band 5.5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.37																		
3. Incomplete Man	0.27	0.23																	
4. Fine Motor	-0.17	-0.12	0.08																
5. Gross Motor	0.26	0.14	0.16	-0.22															
6. Writing Name	0.28	0.17	0.27	0.02	0.20														
7. Identifying Letters	0.35	0.16	0.38	-0.09	0.28	0.45													
8. Counting	0.40	0.35	0.24	-0.22	0.16	0.34	0.59												
9. Digit Repetition	0.21	0.17	0.15	-0.16	0.25	0.21	0.19	0.34											
10. Writing Numbers	0.35	0.31	0.21	-0.15	0.19	0.50	0.50	0.54	0.09										
11. Identifying Numbers	0.37	0.23	0.28	-0.22	0.31	0.42	0.75	0.62	0.23	0.52									
12. One-to-One Correspondence	0.36	0.17	0.25	-0.20	0.26	0.35	0.42	0.38	0.18	0.41	0.60								
13. Calculations	0.37	0.34	0.17	-0.25	0.29	0.31	0.41	0.37	0.27	0.50	0.46	0.37							
14. Prepositions	0.38	0.07	0.36	-0.09	0.49	0.26	0.32	0.07	0.29	0.22	0.25	0.28	0.19						
15. Action Agents	0.27	0.11	0.24	-0.24	0.31	0.22	0.25	0.06	0.28	0.20	0.22	0.26	0.28	0.42					
16. Naming Animals	0.27	0.14	0.14	-0.16	0.23	0.16	0.30	0.33	0.26	0.15	0.39	0.26	0.33	0.13	0.34				
17. Comprehension	0.07	-0.08	0.19	0.01	0.22	0.03	0.13	0.08	0.09	-0.03	0.12	0.11	0.05	0.18	0.22	0.11			
18. Color Forms	0.25	0.03	0.10	-0.19	0.43	0.07	0.14	-0.07	0.16	0.15	0.18	0.12	0.18	0.53	0.37	0.15	0.14		
19. Visual I	0.36	0.32	0.39	-0.11	0.28	0.37	0.44	0.41	0.29	0.39	0.53	0.48	0.45	0.40	0.30	0.28	0.13	0.29	
20. 3-Hole Form Board	0.24	0.03	0.17	0.07	0.46	0.18	0.15	-0.05	0.20	0.11	0.20	0.17	0.10	0.64	0.35	0.12	0.17	0.60	0.24

Note. Ns range from 143-221

Table 52

Inter-correlations Between Tasks: Age Band 6.0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Cubes																			
2. Copy Forms	0.37																		
3. Incomplete Man	0.27	0.23																	
4. Fine Motor	-0.17	-0.12	0.08																
5. Gross Motor	0.26	0.14	0.16	-0.22															
6. Writing Name	0.28	0.17	0.27	0.02	0.20														
7. Identifying Letters	0.35	0.16	0.38	-0.09	0.28	0.45													
8. Counting	0.40	0.35	0.24	-0.22	0.16	0.34	0.59												
9. Digit Repetition	0.21	0.17	0.15	-0.16	0.25	0.21	0.19	0.34											
10. Writing Numbers	0.35	0.31	0.21	-0.15	0.19	0.50	0.50	0.54	0.09										
11. Identifying Numbers	0.37	0.23	0.28	-0.22	0.31	0.42	0.75	0.62	0.23	0.52									
12. One-to-One Correspondence	0.36	0.17	0.25	-0.20	0.26	0.35	0.42	0.38	0.18	0.41	0.60								
13. Calculations	0.37	0.34	0.17	-0.25	0.29	0.31	0.41	0.37	0.27	0.50	0.46	0.37							
14. Prepositions	0.38	0.07	0.36	-0.09	0.49	0.26	0.32	0.07	0.29	0.22	0.25	0.28	0.19						
15. Action Agents	0.27	0.11	0.24	-0.24	0.31	0.22	0.25	0.06	0.28	0.20	0.22	0.26	0.28	0.42					
16. Naming Animals	0.27	0.14	0.14	-0.16	0.23	0.16	0.30	0.33	0.26	0.15	0.39	0.26	0.33	0.13	0.34				
17. Comprehension	0.07	-0.08	0.19	0.01	0.22	0.03	0.13	0.08	0.09	-0.03	0.12	0.11	0.05	0.18	0.22	0.11			
18. Color Forms	0.25	0.03	0.10	-0.19	0.43	0.07	0.14	-0.07	0.16	0.15	0.18	0.12	0.18	0.53	0.37	0.15	0.14		
19. Visual I	0.36	0.32	0.39	-0.11	0.28	0.37	0.44	0.41	0.29	0.39	0.53	0.48	0.45	0.40	0.30	0.28	0.13	0.29	
20. 3-Hole Form Board	0.24	0.03	0.17	0.07	0.46	0.18	0.15	-0.05	0.20	0.11	0.20	0.17	0.10	0.64	0.35	0.12	0.17	0.60	0.24

Note. Ns range from 143-221

Performance Level Ratings

Strand A, as previously mentioned, provides Developmental Age, rather than a numerical score. For this reason, the Performance Level Rating for Strand A is a judgment determination by the examiner using the relationship between chronological and developmental age, and the definitions of each performance level. Strand E is measured using a Likert scale, and the child's results on each subscale receives a performance level rating based on a questionnaire response scale (1-2, 3, 4-5), rather than cut scores.

Performance Level Ratings for Strands B, C, and D were determined using the following procedure to establish cut scores. A conceptual definition was composed for each performance level (Age Appropriate, Emerging, Concern), for each age band across all domains of development, and reviewed by a team of early childhood experts. These definitions provided a general expectation of a child's functioning for the three performance levels, and guided the score setting for a target child at each age band and developmental level. The target child was defined as a child whose overall performance minimally met the definition of the performance level. For example, the 4-year-old who scores at the *lowest* end of the range for Age Appropriate, or the 5 ½ year-old who scores at the *lowest* end of the Emerging performance level.

Performance levels were set based on child development theory, opinions of national experts, and GDO-R data. Using the performance level definitions, the research team determined the expected percent of correct tasks across the strand for each age band. In addition, the team reviewed the expected percent of correct items for each *task* for the target child, and then averaged the percent correct across all tasks for that age band. This resulted in a numerical score. Using these individual scores, boundaries were defined between both Concern/Emerging and Emerging/Age Appropriate. Scores above the Emerging threshold were defined as Age Appropriate, and scores above the Concern threshold were defined as Emerging. The percent correct cut score for the target child in each age band was transferred to the frequency distribution of GDO data (impact data) for the appropriate age band and strand. Boundaries were drawn using whole number values just below the cut score, and impact data (the percentage of children within the performance level) was calculated. The impact data was reviewed for reasonableness, continuity across age bands, and relevance for a wider sample of children.

Since development in each domain occurs in a sequential - but not automatic - pace, great care was taken to provide supplemental guidance to the user for interpreting scores on the cusps of a performance level. In particular, Strand D is unique in that it contains the tasks that span a range of ages, only three of which produce the strand score. For some children in certain age bands (3, 3 ½, 4), these tasks may prove either too easy (Color Forms, Three-Hole Form Board), or too difficult (Visual I).

Tables 53 through 55 present the percentages of children in each strand by age band.

Table 53
Percentage of Children in Performance Levels for Letters/Numbers by Age Band

Letters/ Numbers Performance Level Rating		Age 3	Age 3.5	Age 4	Age 4.5	Age 5	Age 5.5	Age 6
Concern	Range	0-2	0-4	0-7	0-18	0-38	0-48	0-64
	% sample	17.9	17.2	12.7	15.1	18.1	14.6	11.6
Emerging	Range	3-8	5-10	8-22	19-40	39-63	49-74	65-85
	% sample	42.9	31.3	41.2	36.2	31.9	33.1	31.3
Age Appropriate	Range	9-100	11-100	23-100	41-100	64-100	75-100	86-100
	% sample	39.3	51.6	46.1	48.7	50.0	52.3	57.1
Total n		28	64	102	152	160	151	112

Table 54
Percentage of Children in Performance Levels for Language/Comprehension by Age Band

Language/ Comprehension Performance Level Rating		Age 3	Age 3.5	Age 4	Age 4.5	Age 5	Age 5.5	Age 6
Concern	Range	0-9	0-16	0-23	0-32	0-46	0-56	0-61
	% sample	15.9	22.4	19.5	18.9	10.6	9.6	13.0
Emerging	Range	10-19	17-36	24-47	33-63	47-66	57-69	62-73
	% sample	27.3	28.4	35.2	36.6	36.7	28.8	35.9
Age Appropriate	Range	20-100	37-100	48-100	64-100	67-100	70-100	74-100
	% sample	56.8	49.1	45.3	44.5	52.7	61.6	51.1
Total n		44	116	159	227	226	177	131

Table 55
Percentage of Children in Performance Levels for Visual/Spatial Discrimination by Age Band

Visual/Spatial Discrimination Performance Rating		Age 3	Age 3.5	Age 4	Age 4.5	Age 5	Age 5.5	Age 6
Concern	Range	0-56	0-61	0-64	0-67	0-76	0-81	0-83
	% sample	14.9	10.7	8.4	7.9	9.4	5.5	5.1
Emerging	Range	57-74	62-74	65-75	68-82	77-90	82-92	84-93
	% sample	42.6	36.9	31.9	31.9	34.5	20.9	5.1
Age Appropriate	Range	75-100	75-100	76-100	83-100	91-100	93-100	94-100
	% sample	42.6	52.5	59.6	60.3	56.1	73.6	89.7
Total n		47	122	166	229	223	182	136

The Social, Emotional, and Adaptive subscales which comprise Strand E are based on data from the Teacher Questionnaire. Each scale has a 5 point rating scale with 1 indicating a low frequency of a behavior, and 5 indicating a high frequency of a behavior. For less desirable behaviors/items, (for example ,"has poor self-control" or "appears sad") the scores are converted so that for these specific items, scores of 1=5, 2=4, 3=3, 4=2, 5=1 to maintain the meaning across all items. However, since each subscale score (1-5) and each item provides its own interpretation, cut scores for these subscales are less meaningful. Thus, the TQ and the PQ offer valuable information about an individual child's development among the three subscales. In addition, they provide context for interpretation of GDO-R results, as opposed to a set of performance levels to be attained. Refer to Table 56 for interpretation of the questionnaire response scale.

Table 56
Interpretation Rubric for TQ/PQ scores

PQ/TQ Score	Interpretation for Social Behavior, Emotional Development, and Adaptive Skills
1-2	Indicates the child has not sufficiently acquired a skill set in the social, emotional, or adaptive domain.
3	Indicates the child has an emerging acquisition of social, emotional, or adaptive capabilities, or may have an inconsistent profile in the domain.
4-5	Indicates consistent behavior at this level, demonstrating near or solid competence in the domain.

Overall Performance Level Definitions

Determining Overall Performance Level on the GDO-R is achieved when a child's Performance Level Ratings for all five strands are collectively determined, evaluated, and then compared to a set of Performance Level Definitions. These definitions were set using generally accepted guidelines for child development in four domains of development: cognitive, language, motor, and social/emotional/adaptive. The definitions were stratified into Age Appropriate, Emerging, or Concern. Then, they were reviewed by two directors of early childhood centers, one physical movement specialist and two members of the National Lecture Staff (NLS). The Performance Level definitions are not intended as scoring rubrics. They represent *general* expectations for each area of development, and aid in confirming the child's Overall Performance Level Rating.

Inter-rater Reliability

Four NLS members participated in the Qualitative Review Study. Three hold a Master's degree in Early Childhood and/or Child Development, and one a Bachelor's in Child Development. Collectively, the Review Study team had over 100 years of experience administering the Gesell Developmental Observation, and conducting Gesell workshops on topics such as school readiness, parent involvement, and child development.

The inter-rater reliability of the Gesell Developmental Observation-Revised was examined to provide evidence regarding the degree to which scores can be reliably assigned and developmental ages can be reliably estimated. The inter-rater reliability study included a sub-sample of children from the Gesell Developmental Observation-Revised technical study. Data consisted of evaluations of Incomplete Man and Copy Forms samples. Inter-rater reliability was examined for assigning developmental age for Incomplete Man, assigning developmental age for Copy Forms, and assigning item scores to the Copy Forms items. Table 57 describes the sample used in the inter-rater reliability study. The sample for Incomplete Man was smaller than the sample for Copy Forms, because some children were rated as unscorable by one or both raters.

Table 57
Inter-rater Reliability Study Sample

	Incomplete Man	Copy Forms	Copy Forms items
Mean age	4.62 years	4.57 years	4.52 years
Number of children by age band			
3.0	15	16	10
3.5	13	18	9
4.0	17	18	11
4.5	21	21	9
5.0	17	18	11
5.5	18	20	10
6.0	21	20	10
Total sample size	122	131	70

The reliability of developmental age as provided by the Gesell Developmental Observation-Revised was estimated by calculating the inter-rater agreement between the developmental ages assigned by two Gesell Institute National Lecture Staff (Rater A and Rater B) for each task (Incomplete Man and Copy Forms). During Phase 1, Rater A and Rater B rated Copy Form samples, while a different team (Rater A and Rater B) rated Incomplete Man samples. In Phase 2, each team rated the opposite task, and inter-rater agreement was calculated for Copy Form items only.

During Phase 1, for both Incomplete Man (IM) and Copy Forms (CF), neither pair of raters had access to the child's chronological age; the raters used only the actual work samples and process sheets of the children in the sample. Inter-rater agreement on developmental age, as measured by the Pearson product moment correlation, was high for both Incomplete Man and Copy Forms (see Table 58). These high correlations provide evidence that developmental age can be reliably assigned by trained raters using the GDO-R.

In addition, each rater was asked to rank order **all** the children in the sample by developmental age. The rank order correlation (Spearman rho) presented in Table 58 provide further evidence of the reliability of developmental ages as assigned by trained GDO-R administrators. The agreement between the two raters' overall developmental age rankings was high for both Copy Forms (.91) and Incomplete Man (.92), showing that raters ranked the children by developmental age very similarly. Finally, to examine the degree to which the developmental age assigned by raters corresponded to the children's actual age (i.e., chronological age), the Pearson product moment correlations between developmental age and chronological age were calculated. Correlations were calculated separately for Rater A and Rater B. These correlations were high (range .78-.82) and in the expected range, providing evidence that the assigned developmental ages corresponded closely, but not exactly, to children's chronological age (see Table 58). Perfect correlations are not expected because of the variation in development between children.

Table 58
Inter-rater Agreement Evidence for Developmental Age

	Incomplete Man	Copy Forms
	Rating Team 1 (A/B) N= 122 IM samples	Rating Team 2 (A/B) N=131 CF samples
Correlation between Rater A and Rater B developmental age	.92	.91
Correlation between Rater A overall developmental age rank and Rater B overall developmental age rank	.93	.93
Correlation between chronological age and Rater A developmental age	.78	.81
Correlation between chronological age and Rater B developmental age	.82	.82

Table 59 presents results of inter-rater reliability for Copy Forms items for Phase 2. Raters used the same children's work samples as were used in the examination of the reliability of developmental age (Phase 1). However, each team of raters that conducted the inter-rater reliability for Incomplete Man during Phase 1 subsequently conducted the inter-rater reliability for Copy Form items for Phase 2, and vice versa. For item inter-rater reliability for individual

Copy Forms items, raters also had access to children's chronological age, because this is the standard scoring practice. Sample sizes varied by item, because some children were rated as unscorable by one or both raters.

The results in Table 59 indicate strong correlation between raters for each Copy Form item. In addition, the means and standard deviations for Rater 1 and Rater 2's scores are very similar. It is important to note that for Cube Face-on and Cube Point-on items, a proportionally large number of children were rated as unscorable by both raters. All children who were scorable received a score of 0, resulting in an inter-rater reliability of 1.00. It should be noted that Cube Face-on and Cube Point-on are some of the most difficult items in the entire Gesell Developmental Observation-Revised. Very few age 6 children in the entire technical study sample received a correct score on these items (see Table 20).

Table 59
Inter-rater Reliability for Copy Forms Items

	Rater 1			Rater 2			r
	N	M	SD	N	M	SD	
Circle	67	.93	.26	69	.90	.30	.71
Cross	66	.68	.47	69	.52	.50	.68
Square	62	.58	.50	65	.42	.50	.44
Triangle	56	.32	.47	56	.52	.50	.66
Divided Rectangle	43	.23	.43	45	.27	.45	.76
Diamond – Horizontal	39	.28	.46	42	.38	.49	.75
Diamond – Vertical	37	.24	.43	39	.33	.48	.81
3-D Cylinder	27	.04	.19	28	.11	.31	.55
3-D Cube Face-on	26	0.00	0.00	27	0.00	0.00	1.00
3-D Cube Point-on	23	0.00	0.00	20	0.00	0.00	1.00

Note. N = sample size. M = mean. SD = standard deviation. r = Pearson product moment correlation coefficient.

CONCLUSIONS

The GDO-R technical study examined the relationship between performance on a set of developmental and academic tasks and the age of the child (for 3-6 year olds), and the results are included in this Technical Report. While no specific comparative analyses were conducted, the quantitative and qualitative results published here appear to mirror the original findings on developmental tasks by Dr. Arnold Gesell on the developmental tasks in Strand A. In addition, new data and validity evidence on other tasks support the GDO-R tasks for renewed use by educators nationwide. No other early childhood assessment purports to measure developmental age nor provides the important distinguishing features of process and product which correspond to learning. Recent research on neurological development confirms that learning interactions which are intentionally designed for the child's stage of development result in optimal learning experiences. Thus, knowing where the child is on the path of development helps teachers and parents plan experiences that result in new brain connections and new learning.

The GDO Study contributes a comprehensive sample of child development data to the educational field at large. Moreover, the data reported in this Technical Report were primarily collected from public schools, providing a heterogeneous sample of children from socioeconomically diverse backgrounds. However, as with any study of its size and scope, it has limitations. The distribution of child ethnicity across the total sample more closely approximated the U.S. Census than did the distribution of child ethnicity in each age band. Thus, interpretation of change across age bands could possibly be attributable to a sample shift as opposed to a definitive age shift. In the case of the Parent/Guardian Questionnaire, the percentage of missing data for child's ethnicity was strongly mitigated by efforts on the part of the school and research team to gather this information from other school records (as reported by parents). However, since the native language of the child was also derived from the PQ, efforts to collect accurate information on a child's native language from the school were less fruitful, and resulted in higher percentages of missing data across age bands. In some sites, it was not possible to administer the PQ due to the nature of the testing at the site (i.e., GDO-R tests were part of admission protocol in private schools, and as such schools did not administer the GDO Study parent questionnaire because it contained questions that could be perceived to affect a child's eligibility for school acceptance (special evaluations, services, level of education of parent). In other sites, bussing of children in urban communities meant that parents did not physically come to the school to return a PQ, or may have been reluctant to share such information with the school administration.

While the examiners were trained carefully on the GDO-R task administration, they did not receive recording and coding rubrics to score the following observations of the child during the assessment: Paper Position, Head Shift, Body Posture, Non-dominant Hand Posture, Eye Movement. This may explain why observational data on these items contain missing cases (examiners did not complete the section of the form). Thus, these items cannot clearly be interpreted which is a shortcoming of this study. However, the Qualitative Review Study and Inter-rater Reliability Study strongly confirm the developmental characteristics of each age band as related to Copy Forms and Incomplete Man tasks. This is very important, not only for providing recent validity evidence for these specific developmental tasks, but for continual improvements to the training of examiners.

SUMMARY

As evidenced by the technical data documented in this report, GDO-R examiners can reliably assess and monitor a child's progress over time, and interpret developmental age and Performance Level Ratings with confidence. It is imperative that learning expectations be aligned with a child's development to promote realistic standards for achievement. A child's developmental age and Performance Level Ratings on the GDO-R enable teachers to more accurately focus their instruction for each individual child and for the entire class. In addition, identifying a child's developmental profile not only promotes more effective communication between administrators and parents to build strong home-school connections, but fosters developmentally appropriate practice in every classroom.

Chapter II

Gesell Early Screener
Technical Report: Data Analyses and Results

FOREWORD

The Gesell Early Screener (GES) Technical Report is the first technical publication for the instrument. The report is intended for both teachers and administrators. Its ultimate purpose is to understand the metrics behind the GES, which identify children who may be in need of further diagnostic evaluation in one or more of the four strands measured. It is based on a sample of assessment data of developmental and achievement tasks for children in preschool and Kindergarten. It is anticipated that readers of this Technical Report possess an advanced understanding of appropriate use and application of assessment tools, methods for conducting sound test development, and methodology in statistics and measurement.

PART 1: OVERVIEW

Role of Developmental Assessment in Early Childhood

Refer to pages 9-12 in Chapter I of the GDO-R/GES Technical Report.

Overview of the Gesell Early Screener

The Gesell Early Screener (GES) is a screening tool that evaluates four domains of a child's development: Cognitive, Language, Motor, and Social/Emotional/Adaptive. The screener assists parents, educators, and other professionals in understanding performance expectations of children in relation to typical growth patterns based on sequential, normative stages of development. Gesell assessments have been used as standardized measures of child growth and development since the introduction of the initial schedules in 1925. Long employed by pediatricians and revised and updated in 2011, Gesell training and assessment materials focus on direct observation to measure a child's neuromotor, language, cognitive, social/emotional, and adaptive development. The GES is designed for use with children age 3-6, and is invaluable for informing educators and parents if a child's development may be "at risk" according to a developmental continuum. The GES can accurately determine when further evaluation or observation may be necessary in one or more domains of development. The Gesell Early Screener consists of the selected tasks listed below. Subsets of these tasks are administered according to the child's age.

Table 1
GES Tasks by Age of Administration

Strand Name	Task Name (Number)	3	3^6	4	4^6	5	5^6	6
Cognitive	Cubes (Task 2)	●	●	●	●	●	●	●
	Copy Forms (Task 3)	●	●	●	●	●	●	●
	Prepositions (Task 4)	●	●	●	●	●	●	●
	One-to-One Correspondence-Four Pennies (Task 5)	●	●	●	●	●	●	●
	One-to-One Correspondence-Ten Pennies (Task 5)				●	●	●	●
	Conservation-Four Pennies (Task 5)	●	●	●	●	●	●	●
	Conservation-Ten Pennies (Task 5)				●	●	●	●
	Identifying Numbers (Task 6)				●	●	●	●
Language/Comprehension	Interview (Task 1)	●	●	●	●	●	●	●
Motor	Fine/Gross Motor (Task 7)	●	●	●	●	●	●	●
Social/ Emotional/ Adaptive	Teacher Questionnaire-TQ (Task 8)	●	●	●	●	●	●	●
	Parent/Guardian Questionnaire-PQ (Task 8)	●	●	●	●	●	●	●

Note. Task numbers correspond to GES Examiner's Manual. The GES contains six performance-based tasks, as well as Social, Emotional, and Adaptive functioning (measured by the TQ and PQ), for a total of eight tasks/measures.

Description of GES Materials and Kits

Examiner's Manual

The Gesell Early Screener Examiner's Manual contains an overview of the GES, a description of each component, and complete instructions for administering, recording, scoring (including Decision Trees), and interpreting results for all tasks. In addition, the Manual also provides:

- Performance Expectations by Task and Age.
- Strand Scoring Worksheet to calculate strand scores.

- Recording Chart for scoring the Social, Emotional, and Adaptive strand.
- Summary Profile Form which incorporates a child's overall performance results to share with parents.

Child Recording Form

The GES Child Recording Form (CRF-S) is a consumable form (one per child) used by the examiner to record responses to each task, as well as verbal and physical behaviors observed during the assessment. The CRF-S also provides a standardized administration script next to each task. The Summary Profile Form and the TQ/PQ Recording Chart are included in this form for examiner scoring convenience.

A checklist on the inside front cover of the CRF-S is provided for the examiner to note any additional observations about physical attributes, speech patterns, and English proficiency. This data is essential to provide an overall picture of the child as an individual, and to note any conditions or circumstances about the administration that may have affected the child's ability to perform the tasks.

Teacher Questionnaire (TQ)

The Teacher Questionnaire collects data from the teacher about a child's self-help skills, self-expression, behavioral observations, and classroom preferences. These ratings provide information on three subscales:

- Social Behavior
- Emotional Development
- Adaptive Skills

Parent/Guardian Questionnaire (PQ)

The Parent/Guardian Questionnaire collects data from parents/guardians about a child's:

- Family background and demographics (race, ethnicity, sex)
- Medical and educational history (birth history, early educational experiences)
- Home environment (siblings, use of technology, exposure to media and literacy opportunities)
- Social, emotional, and adaptive capabilities (self-help skills, choices offered to child, self-expression, behavioral observations)

Manipulatives

The manipulative items required for proper administration of the GES are shown in Table 2.

Table 2
GES Kit Manipulatives

Manipulatives	Items Provided by Examiner
• 10 Red Hardwood Cubes • Copy Forms cards • Numbers card • Beanbag	• Pencil • 6 foot length of tape • 10 Pennies

Description of Tasks by Strand

The primary purpose of the Gesell Early Screener (GES) is to evaluate development and academic readiness skills in four areas, or strands to "flag" a child who may need further evaluation.

- The Cognitive Strand measures visual-motor perception and coordination, and short-term visual memory. In addition, it measures a child's exposure to and proficiency with numbers via one-to-one correspondence and conservation of matter.
- The Language/Comprehension strand evaluates a child's attention span, articulation, and expressive and receptive language.
- The Motor Strand evaluates a child's fine and gross motor skills.
- The Teacher and Parent/Guardian Questionnaires provide information to evaluate the quality of a child's interactions with peers and adults, emotional regulation behaviors, and self-help skills both at home and at school (Social/Emotional/Adaptive strand).

A child's performance on the tasks for each strand is scored and corresponds to one of three Performance Level Ratings for each strand: Age Appropriate, Emerging, or Concern.

Historical Validation and Reliability of the GES

Refer to page 17 in Chapter I of the GDO-R/GES Technical Report.

Development of the Gesell Early Screener

The GES was carefully developed by selecting performance tasks from the GDO-R for the purpose of identifying a child's Performance Level Rating (Age Appropriate, Emerging, or Concern) in four domains in 15 minutes or less. Tasks were selected such that the screening instrument would be easy to administer, objective to score, be reliable and valid when administered by persons with varying levels of expertise, and meet all federal mandates for screening young children. In addition, the instrument was required to provide insight into a child's social/emotional/adaptive development, and meet the needs of early educators in 21st century classrooms.

To this end, GDO-R tasks were sampled from strands A, B, C, D, and E based on a set of criteria that included:

- ease of administration and scoring
- representativeness to classroom expectations and domains of development
- time to administer entire instrument
- ability to meet the needs of the early childhood educational field.

For the Cognitive strand, GES tasks were selected based on strength of correlations between individual task scores, and the total strand score for each age band. The GES Language/Comprehension Strand utilizes the GDO-R Interview task to provide a language sample. Several Motor tasks were not included in the GES Motor strand. The Standing Long Jump was removed because the GDO Study did not include data on distance jumped. In addition, examiner measurement of distance jumped in inches would have introduced an unacceptable level of measurement error. This was not consistent with the criteria for selection of tasks (ease of administration/scoring). The Jump in Place task was also removed because the data from the GDO-R analysis revealed very low variability across age bands (ability for the item to discriminate between performance levels). The Social/Emotional/Adaptive Strand is measured in the same way as the GDO-R, since the Teacher Questionnaire and the Parent/Guardian Questionnaire are complementary measures that stand alone.

PART 2: VALIDITY EVIDENCE AND CONTENT RATIONALE

Relationship of GES to GDO-R

The GES is a quick screening instrument consisting of six tasks selected from the GDO-R. Several of the GDO-R tasks were shortened for use in the GES. The GES is designed to be used with large numbers of children to identify any child who may benefit from further in-depth evaluation. It provides an estimate of a child's performance in four domains, but does not provide a developmental age or inform instruction as results provided by the GDO-R are able to do. The GDO-R provides an in-depth assessment of a child's developmental and academic skills across five strands, and is designed to inform instruction for each child and for periodic reassessment. The GDO-R and the GES are complementary instruments designed to be used together in sequence as part of a wider school-based program of screening, assessment, and intervention services.

Content-Related Validity

Refer to page 18 in Chapter I of the GDO-R/GES Technical Report.

Criterion-Referenced Performance Interpretations

Refer to pages 18-19 in Chapter I of the GDO-R/GES Technical Report.

GDO Study Timeline

Refer to pages 19-20 in Chapter I of the GDO-R/GES Technical Report.

Procedures for Reducing Bias

Refer to pages 21-22 in Chapter I of the GDO-R/GES Technical Report.

Site Sample, Examiner Sample, and Sample Child Descriptive Statistics

Refer to page 24ff. in Chapter I of the GDO-R/GES Technical Report for a discussion on sample demographics.

GES Analytic Method and Sample

Technical data for the Gesell Early Screener (GES) Cognitive Strand are presented below. The technical data are based on the same sample used for the Gesell Developmental Observation-Revised technical study as described in Chaper I of this publication, including 1,287 children, age 3 to 6, from across the United States. Consequently, the results of the GES can be generalized to wider groups of children with similar characteristics in the U.S. Children's total scores on the GES Cognitive Strand and the tasks that make up the GES Cognitive Strand are analyzed and presented by age band. The data presented for the TQ measure is included in Chapter I, and are not duplicated here.

For the GES, several tasks include only a subset of items that are included in the tasks for the GDO-R. The GES Cubes task includes five items: Train, Bridge, Gate, Steps 6, and Steps 10. The GES Copy Forms task includes four items: Circle, Cross, Square, and Triangle. The GES One-to-One Correspondence task and Conservation task include two items (Items 1 and 2) for children ages 3.0, 3.5, and 4.0 years old and four items (Items 1 - 4) for children ages 4.5, 5.0, 5.5, and 6.0 years old. The Prepositions task and the Identifying Numbers task include all the items as administered in the GDO-R.

GES Cognitive Strand and Cognitive Tasks Descriptive Statistics

Tables 3 through 7 display descriptive statistics for the GES Cognitive Strand, and its corresponding tasks by age band. Data for the Cognitive Strand total are presented in their standardized scoring metric. (See Chapter 2, page 51 of the GES Examiner's Manual for explanation of the method for deriving the Cognitive strand score.) Data for each task are presented in their original raw score metric. Within each age band, sample sizes vary by task for two reasons. First, children needed to have a valid response to at least one item to be included in

the task, and have a task level score. Second, children must have a valid score for each cognitive task in order to have a Cognitive strand total.

The tables provide the number of items for the Cognitive strand total and each task, sample size (N), maximum possible points possible, raw score mean (Mean), raw score standard deviation (SD), internal consistency coefficient, and mean p-value. The standard deviations provide a measure of variance. Internal consistencies, given as Cronbach's alpha coefficient, indicate the degree to which the total score is capturing variance on the underlying trait being measured by the respective items. The mean p-values provide an index of the difficulty of the items within the task. With few exceptions, Tables 3 through 9 show growth in performance on the GES Cognitive Strand and tasks from age 3.0 to age 6.0.

Internal consistencies for the Cognitive strand were generally high and in the desirable range. Internal consistency is influenced by the number of items within task, and the variability of scores within the task. Tasks that have more items are expected to have higher internal consistency than tasks with fewer items. The internal consistency coefficients presented in Tables 3 through 9 are consistent with expectations, and are acceptable for the intended uses and purposes of the GES. In a few cases, internal consistencies were lower than optimal, likely because of lack of variability.

Table 3
Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 3.0

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**16**	**51**	**32**	**10.67**	**5.73**	**.71**	**.36**
Cubes	5	53	10	2.51	1.73	.39	.25
Copy Forms	4	52	4	1.37	1.19	.71	.34
Prepositions	5	52	5	2.10	1.35	.70	.42
One-to-One Correspondence 4p, Conservation 4p	2	53	2	0.66	0.81	.68	.33

Note. p=pennies

Table 4
Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 3.5

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**16**	**123**	**32**	**13.98**	**6.05**	**.73**	**.47**
Cubes	5	130	10	3.21	2.02	.57	.32
Copy Forms	4	129	4	1.80	1.19	.69	.45
Prepositions	5	126	5	2.79	1.52	.76	.56
One-to-One Correspondence 4p, Conservation 4p	2	127	2	0.81	0.77	.55	.41

Table 5
Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 4.0

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**16**	**169**	**32**	**17.53**	**5.80**	**.73**	**.58**
Cubes	5	186	10	4.41	1.78	.49	.44
Copy Forms	4	185	4	2.11	1.08	.62	.53
Prepositions	5	177	5	3.36	1.48	.78	.67
One-to-One Correspondence 4p, Conservation 4p	2	171	2	1.12	0.79	.65	.56

Table 6
Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 4.5

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**30**	**216**	**48**	**30.81**	**10.91**	**.93**	**.64**
Cubes	5	263	10	5.91	2.13	.64	.59
Copy Forms	4	260	4	2.95	0.99	.61	.74
Prepositions	5	238	5	4.00	1.44	.89	.80
One-to-One Correspondence 4p/10p, Conservation 4p/10p	4	230	4	2.52	1.45	.84	.63
Identifying Numbers	12	230	12	6.30	4.83	.96	.53

Table 7

Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 5.0

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**30**	**206**	**48**	**38.19**	**8.65**	**.92**	**.80**
Cubes	5	278	10	7.09	2.00	.54	.71
Copy Forms	4	275	4	3.38	0.83	.56	.84
Prepositions	5	232	5	4.56	0.99	.95	.91
One-to-One Correspondence 4p/10p, Conservation 4p/10p	4	219	4	3.26	1.15	.90	.81
Identifying Numbers	12	237	12	4.08	4.08	.96	.71

Table 8

Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 5.5

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**30**	**177**	**48**	**42.78**	**6.26**	**.95**	**.89**
Cubes	5	221	10	8.00	1.97	.57	.80
Copy Forms	4	219	4	3.66	0.60	.50	.91
Prepositions	5	186	5	4.91	0.47	.99	.98
One-to-One Correspondence 4p/10p, Conservation 4p/10p	4	179	4	3.52	0.93	.92	.88
Identifying Numbers	12	187	12	10.33	3.02	.97	.86

Table 9
Gesell Screener Cognitive Strand Total and Task Descriptive Statistics Age Band 6.0

	Number of Items	N	Maximum Points	Mean	SD	Internal Consistency	Mean p-value
Cognitive Strand Total	**30**	**131**	**48**	**45.36**	**4.54**	**.94**	**.94**
Cubes	5	152	10	8.79	1.64	.67	.88
Copy Forms	4	153	4	3.90	0.40	.70	.97
Prepositions	5	137	5	4.89	0.55	.98	.98
One-to-One Correspondence 4p/10p, Conservation 4p/10p	4	135	4	3.65	0.80	.91	.91
Identifying Numbers	12	140	12	11.46	1.81	.98	.95

Inter-correlations for GES Cognitive Strand and Tasks

Inter-correlations for the GES Cognitive Strand were computed to provide information regarding the internal structure of the Strand, and the relationships among its various tasks. A positive correlation of .5 to .9 is indicative of a moderate to strong relationship between two items, so related that one directly implies or is complementary to the other. Correlations between .30 - .50 are generally considered to be moderate, and those <.30 are small (Cohen, 1988). The inter-correlations (Pearson product moment correlations) presented in the Tables 10 through 16 show a strong relationship between each task and the Cognitive Strand total score.

Table 10
Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 3.0

	1	2	3	4
1. Cubes				
2. Copy Forms	0.40			
3. Prepositions	0.39	0.36		
4. One-to-One Correspondence 4p, Conservation 4p	0.11	-0.01	0.20	
5. GES Cognitive Strand total	0.69	0.71	0.80	0.40

Note. Ns range from 51-52

Table 11

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 3.5

	1	2	3	4
1. Cubes				
2. Copy Forms	0.22			
3. Prepositions	0.31	0.21		
4. One-to-One Correspondence 4p, Conservation 4p	0.17	0.20	0.30	
5. GES Cognitive Strand total	0.63	0.66	0.78	0.54

Note. Ns range from 123-129

Table 12

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 4.0

	1	2	3	4
1. Cubes				
2. Copy Forms	0.33			
3. Prepositions	0.29	0.26		
4. One-to-One Correspondence 4p, Conservation 4p	0.27	0.21	0.33	
5. GES Cognitive Strand total	0.65	0.65	0.78	0.59

Note. Ns range from 169-185

Table 13

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 4.5

	1	2	3	4	5
1. Cubes					
2. Copy Forms	0.41				
3. Prepositions	0.30	0.32			
4. One-to-One Correspondence 4p/10p, Conservation 4p/10p	0.43	0.37	0.38		
5. Identifying Numbers	0.48	0.38	0.47	0.49	
6. GES Cognitive Strand total	0.69	0.62	0.69	0.73	0.86

Note. Ns range from 205-258

Table 14

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 5.0

	1	2	3	4	5
1. Cubes					
2. Copy Forms	0.36				
3. Prepositions	0.25	0.15			
4. One-to-One Correspondence 4p/10p, Conservation 4p/10p	0.32	0.36	0.23		
5. Identifying Numbers	0.38	0.31	0.41	0.56	
6. GES Cognitive Strand total	0.65	0.58	0.57	0.73	0.88

Note. Ns range from 193-272

Table 15

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 5.5

	1	2	3	4	5
1. Cubes					
2. Copy Forms	0.38				
3. Prepositions	0.38	0.10			
4. One-to-One Correspondence 4p/10p, Conservation 4p/10p	0.32	0.24	0.37		
5. Identifying Numbers	0.38	0.34	0.25	0.51	
6. GES Cognitive Strand total	0.71	0.50	0.52	0.74	0.84

Note. Ns range from 177-219

Table 16

Inter-correlations between Gesell Screener Cognitive Strand Tasks: Age Band 6.0

	1	2	3	4	5
1. Cubes					
2. Copy Forms	0.15				
3. Prepositions	0.15	0.45			
4. One-to-One Correspondence 4p/10p, Conservation 4p/10p	0.08	0.36	0.16		
5. Identifying Numbers	0.24	0.42	0.66	0.30	
6. GES Cognitive Strand total	0.55	0.62	0.70	0.59	0.82

Note. Ns range from 130-151

GES Performance Level Statistics

Children's scores on the GES were used to calculate and present the percentage of children from the sample in each of the three GES performance levels—Age Appropriate, Emerging, and Concern. These performance level data are based on the Cognitive Strand total scores only.

Table 17 presents the number of children in each age band who had a valid Cognitive Strand total, and the percentage of these children within each performance level. The percentage of children in each performance level are fairly consistent across the age bands, reflecting the use of both criterion-reference and norm-reference information used to set the cut scores defining the performance levels within each age band.

Performance Level Ratings for the Cognitive strand were determined using the following procedures to establish cut scores. Two members of the research team (early childhood professionals) independently set cut score ranges for each age level and performance level based on a target child's expected scores for Age Appropriate, Emerging, or Concern. As a team, the reviewers rigorously examined each set of ranges to produce one set of cut scores that reflected appropriate sensitivity (identification of children in need of further evaluation to determine possible delay), and specificity (guard against incorrect identification of children for further evaluation who are developing normally).

Since development in each domain occurs in a sequential but not automatic pace, great care was taken to provide supplemental guidance to the user for interpreting scores on the upper or lower cusps of a performance level.

Table 17
Percentage of Children in Performance Levels for Gesell Screener Cognitive Strand by Age Band

Cognitive Strand	3.0	3.5	4.0	4.5	5.0	5.5	6.0
	n= 51	n= 123	n= 169	n= 216	n= 206	n= 177	n= 131
Concern	0-5	0-8	0-12	0-21	0-31	0-39	0-43
	15.7%	21.1%	21.3%	23.1%	21.4%	20.3%	19.1%
Emerging	6-9	9-13	13-17	22-32	32-41	40-44	44-46
	29.4%	26.8%	24.3%	26.4%	29.6%	28.8%	26.0%
Age Appropriate	10-30	14-30	18-30	33-48	42-48	45-48	47-48
	54.9%	52.0%	54.4%	50.5%	49.0%	50.8%	55.0%

Inter-rater Reliability Study

The inter-rater reliability of the Gesell Early Screener was examined to provide evidence regarding the degree to which scores on the Cognitive Strand can be reliably assigned. The inter-rater reliability study included a sub-sample of children from the Gesell Developmental Observation-Revised technical study, and a sub-sample of GDO-R tasks which comprise the GES. Data consisted of evaluations of child responses to the six tasks (30 items) in the

Cognitive Strand. Inter-rater reliability was examined for assigning scores for each item on Cubes, Copy Forms, Prepositions, One-to-One Correspondence, Conservation, and Identifying Numbers.

Two raters were used for inter-rater reliability; Rater A was a college graduate intern trained at Gesell Institute on the GES tasks, and Rater B was the GDO examiner for each child originally assessed using the full GDO-R assessment for the GDO Study. Rater A independently scored selected GES items from the original GDO Child Recording Form completed by the GDO examiner. Rater B's scores were taken from GDO examiner's Data Collection Form submitted for the GDO Study. All cases had complete sets of GES task data. Refer to Table 18 for the scope of the inter-rater reliability sample.

Table 18
GES Inter-rater Reliability Study Sample

Age Band		**3.5**	**4.0**	**4.5**	**5.0**	**5.5**	**6.0**
Mean Age (years)	4.74	N	N	N	N	N	N
Cubes		10	10	10	10	10	10
Copy Forms		10	10	10	10	10	10
Prepositions		10	10	10	10	10	10
One-to-One[4]		10	10	10	10	10	10
One-to-One[10]		10	10	10	10	10	10
Conservation[4]		10	10	10	10	10	10
Conservation[10]		10	10	10	10	10	10
Identifying Numbers		10	10	10	10	10	10
Total N		60	60	60	60	60	60

Note. The sample did not include 3 year age band because it lacked complete data for all eight tasks/items. Superscript numbers indicate number of pennies in task.

Two measures of reliability evidence are presented in Table 19. Both Pearson correlations and percent agreement between Rater A and Rater B are very high on each item (.90 and above).

Table 19
Inter-rater Reliability Agreement Evidence

GES Task	GES Item	Correlation between Rater A and Rater B	Agreement between Rater A and Rater B
		r	% agreement
Cubes	**Train**	0.84	95%
	Bridge	0.85	97%
	Gate	0.95	92%
	Steps (6)	0.98	97%
	Steps (10)	0.94	95%
Copy Forms	**Circle**	1.00	98%
	Cross	0.81	95%
	Square	0.88	95%
	Triangle	0.94	97%
Prepositions	**On**	1.00	100%
	Under	1.00	100%
	In back of	0.94	98%
	In front of	0.94	98%
	Beside	1.00	100%
One-to-One	**4 pennies**	0.90	97%
	10 pennies	0.97	98%
Conservation	**4 pennies**	0.89	95%
	10 pennies	0.93	97%
Identifying Numbers	**1**	0.96	98%
	2	1.00	97%
	3	1.00	100%
	4	1.00	95%
	5	1.00	95%
	6	0.97	93%
	7	1.00	95%
	8	1.00	97%
	9	1.00	100%
	10	0.97	93%
	11	1.00	95%
	12	1.00	100%

Table 20 illustrates the means and standard deviations of scores by item for Rater A and Rater B. Both raters are highly consistent with each other.

Table 20
Inter-rater Reliability for All GES Items

GES Task	GES Item	Rater A			Rater B		
		Mean	SD	N	Mean	SD	N
Cubes	Train	1.78	0.58	60	1.82	0.54	60
	Bridge	1.82	0.54	60	1.87	0.43	60
	Gate	1.27	0.88	60	1.28	0.88	60
	Steps (6)	1.00	0.92	60	0.97	0.92	60
	Steps (10)	0.73	0.94	60	0.77	0.93	60
Copy Forms	Circle	2.00	0.00	60	1.97	0.26	60
	Cross	1.70	0.72	60	1.67	0.75	60
	Square	1.47	0.89	60	1.43	0.91	60
	Triangle	1.07	1.01	60	1.00	1.01	60
Prepositions	On	2.00	0.00	60	2.00	0.00	60
	Under	1.87	0.50	60	1.87	0.50	60
	In back of	1.67	0.75	60	1.70	0.72	60
	In front of	1.67	0.75	60	1.63	0.78	60
	Beside	1.60	0.81	60	1.60	0.81	60
One-to-One	4 pennies	1.57	0.83	60	1.63	0.78	60
	10 pennies	1.17	0.99	60	1.13	1.00	60
Conservation	4 pennies	1.40	0.92	60	1.30	0.96	60
	10 pennies	1.17	0.99	60	1.10	1.00	60
Identifying Numbers	1	0.63	0.49	60	0.65	0.48	60
	2	0.60	0.49	60	0.60	0.49	60
	3	0.60	0.49	60	0.60	0.49	60
	4	0.63	0.49	60	0.63	0.49	60
	5	0.58	0.50	60	0.58	0.50	60
	6	0.52	0.50	60	0.53	0.50	60
	7	0.58	0.50	60	0.58	0.50	60
	8	0.53	0.50	60	0.53	0.50	60
	9	0.45	0.50	60	0.45	0.50	60
	10	0.47	0.50	60	0.48	0.50	60
	11	0.48	0.50	60	0.48	0.50	60
	12	0.43	0.50	60	0.43	0.50	60

Note. M = mean. SD = standard deviation. N = sample size. Rater A=Gesell Intern trained to score selected GES tasks. Rater B=individual GDO-R examiners from across the U.S. who scored the same tasks (excerpted from GDO-R Child Recording Form).

CONCLUSIONS

The GES technical study examined the relationship between performance on a set of developmental and academic tasks, and the age of the child (for 3-6 year olds), and the results are included in this Technical Report. The GES is designed to be used with the Teacher Questionnaire (TQ) and the Parent/Guardian Questionnaire (PQ). The GDO Study contributes a comprehensive sample of child development data to the educational field at large. Moreover, the data reported in this Technical Report were primarily collected from public schools, providing a heterogeneous sample of children from socioeconomically diverse backgrounds. As discussed in Chapter I of this publication, the distribution of child ethnicity across the total sample more closely approximated the U.S. Census than did the distribution of child ethnicity in each individual age band. Thus, the reliability and validity evidence presented here for the GES is bound by the same limitations, since the GES data was derived from the larger GDO-R dataset.

Several potential limitations exist in the inter-rater reliability study. The first is that one rater compared her scoring to multiple raters who were GDO examiners themselves during the actual study. Thus, "Rater B" was not one entity, but a group of examiners who had assessed each child in the sample. That is, each child was assessed by an examiner from his or her respective school, and cases were randomly selected from across the entire age sample, but not from one examiner's sample of assessments. This method was seen as preferable to two independent interns rating data recorded by a third, because it best utilized the research staff available. The examiner sample for the GDO-R data was a highly educated group of experienced educators with solid training as GDO examiners. Additionally, the high consistency between raters is possibly a result of high quality recording techniques by examiners garnered over years of training.

The second potential limitation is that since Rater A scores came from the Child Recording Form (CRF), and Rater B scores were taken from the Data Collection Form (DCF) for the GDO Study, it is possible that any transcription errors (from CRF to DCF) were embedded in the inter-rater reliability study. However, when the inter-rater reliability study began, all data had been reviewed for administration errors and transcription errors by the research team, so the likelihood that errors remained in the 4% of the data that comprised the reliability study is low.

SUMMARY

As evidenced by the technical data documented in this report, GES examiners can reliably screen individual children across four domains, interpret Performance Level Ratings (Age Appropriate, Emerging, or Concern) with confidence for each domain, and identify if a child may be in need of further diagnostic evaluation. Because of its narrow scope of items, the GES cannot be used to monitor progress, guide instruction, or determine a developmental age.

APPENDIX

Appendix A: Online Survey

Gesell Developmental Observation (GDO) ©2007 User Survey – Quantitative Results

1. **How is the Gesell Developmental Observation (GDO) used in your school or program? (Check all that apply.)**

	Response Percent	Response Count
Kindergarten screening	**58.2%**	**71**
Developmental screening	54.1%	66
Part of admissions process	35.2%	43
Grade or classroom placement/match	24.6%	30
Other	21.3%	26
	answered question	***122***
	skipped question	***31***

2. **What other assessment or observation tools are used in your school, program, or classroom? (Check all that apply.)**

	Response Percent	Response Count
State screening or framework	9.8%	12
School or district developed	21.3%	26
Brigance (BABS)	8.2%	10
Early Screening Inventory (ESI)	3.3%	4
Bracken (BSRA)	2.5%	3
Peabody (PPVT)	9.8%	12
High Scope (COR)	3.3%	4
Bayley Scales	2.5%	3
Work Sampling System	6.6%	8
Devereux (DECA)	0.8%	1
Creative Curriculum Assessment Tool Kit	5.7%	7
Developmental Indicators for the Assessment of Learning (DIAL)	9.8%	12
None	**32.8%**	**40**
Other	26.2%	32
	answered question	***122***
	skipped question	***31***

3. **The usefulness of the assessment in general:**

	Response Percent	**Response Count**
1 (Poor)	0.0%	0
2 (Fair)	2.6%	3
3 (Not Sure)	15.8%	18
4 (Good)	32.5%	37
5 (Excellent)	**49.1%**	**56**
	answered question	***114***
	skipped question	***39***

4. **The layout/organization of the assessment:**

	Response Percent	**Response Count**
1 (Poor)	0.0%	0
2 (Fair)	4.0%	4
3 (Not Sure)	12.1%	12
4 (Good)	**44.4%**	**44**
5 (Excellent)	39.4%	39
	answered question	***99***
	skipped question	***54***

5. **The usefulness of the Scoring Manual:**

	Response Percent	**Response Count**
1 (Poor)	0.0%	0
2 (Fair)	0.0%	0
3 (Not Sure)	12.4%	11
4 (Good)	33.7%	30
5 (Excellent)	**53.9%**	**48**
	answered question	***89***
	skipped question	***64***

6. **The layout/organization of the Scoring Manual:**

	Response Percent	**Response Count**
1 (Poor)	0.0%	0
2 (Fair)	2.4%	2
3 (Not Sure)	12.2%	10
4 (Good)	36.6%	30
5 (Excellent)	**48.8%**	**40**
	answered question	***82***
	skipped question	***71***

7. The content of the Scoring Manual:

	Response Percent	Response Count
1 (Poor)	0.0%	0
2 (Fair)	0.0%	0
3 (Not Sure)	9.8%	8
4 (Good)	25.6%	21
5 (Excellent)	**64.6%**	**53**
	answered question	***82***
	skipped question	***71***

8. The scoring process:

	Response Percent	Response Count
1 (Poor)	0.0%	0
2 (Fair)	0.0%	0
3 (Not Sure)	12.3%	10
4 (Good)	37.0%	30
5 (Excellent)	**50.6%**	**41**
	answered question	***81***
	skipped question	***72***

9. The profile form:

	Response Percent	Response Count
1 (Poor)	0.0%	0
2 (Fair)	1.3%	1
3 (Not Sure)	13.9%	11
4 (Good)	29.1%	23
5 (Excellent)	**55.7%**	**44**
	answered question	***79***
	skipped question	***74***

10. The 3-day training/GDO workshop:

	Response Percent	Response Count
1 (Poor)	0.0%	0
2 (Fair)	1.3%	1
3 (Not Sure)	13.9%	11
4 (Good)	21.5%`	17
5 (Excellent)	**63.3%**	**50**
	answered question	***79***
	skipped question	***74***

11. Please rank your five top preferred sections of the assessment in the order you find them most useful, where the first column indicates the most useful section, the second column the second most useful section, etc. (Select one per column.)

	Most useful	Second most useful	Third most useful	Fourth most useful	Fifth most useful	Rating Average	Response Count
Cubes	19.6% (10)	25.5% (13)	**37.3% (19)**	15.7% (8)	2.0% (1)	3.45	51
Initial Interview	17.9% (7)	10.3% (4)	20.5% (8)	**30.8% (12)**	20.5% (8)	2.74	39
Paper and Pencil/Copy Forms	**38.5% (20)**	30.8% (16)	21.2% (11)	5.8% (3)	3.8% (2)	3.94	52
Incomplete Man	**33.3% (19)**	**33.3% (19)**	19.3% (11)	8.8% (5)	5.3% (3)	3.81	57
Animals and Interests	8.3% (3)	8.3% (3)	19.4% (7)	19.4% (7)	**44.4% (16)**	2.17	36
Visual I	16.7% (2)	8.3% (1)	16.7% (2)	16.7% (2)	**41.7% (5)**	2.42	12
Visual III	0.0% (0)	0.0% (0)	25.0% (1)	**50.0% (2)**	25.0% (1)	2.00	4
Right/Left	14.3% (1)	14.3% (1)	14.3% (1)	**42.9% (3)**	14.3% (1)	2.71	7
Discriminating Prepositions	0.0% (0)	0.0% (0)	**37.5% (3)**	25.0% (2)	**37.5% (3)**	2.00	8
Digit Repetition	0.0% (0)	0.0% (0)	12.5% (1)	**62.5% (5)**	25.0% (2)	1.88	8
Comprehension Questions	13.0% (3)	21.7% (5)	13.0% (3)	**30.4% (7)**	21.7% (5)	2.74	23
Color Forms	0.0% (0)	11.1% (1)	11.1% (1)	**44.4% (4)**	33.3% (3)	2.00	9
Action Agents	6.3% (1)	6.3% (1)	18.8% (3)	31.3% (5)	**37.5% (6)**	2.13	16
Three-Hole Form Board	0.0% (0)	0.0% (0)	0.0% (0)	33.3% (1)	**66.7% (2)**	1.33	3
Fine Motor	16.7% (3)	11.1% (2)	5.6% (1)	22.2% (4)	**44.4% (8)**	2.33	18
Gross Motor	17.6% (3)	**35.3% (6)**	0.0% (0)	11.8% (2)	**35.3% (6)**	2.88	17
						Answered question	***72***
						Skipped	***81***

12. Which sections do you find LEAST useful, or confusing in some way? (Select up to five sections.)

	Response Percent	**Response Count**
Cubes	2.9%	2
Initial Interview	8.7%	6
Paper and Pencil/Copy Forms	0.0%	0
Incomplete Man	2.9%	2
Animals and Interests	11.6%	8
Visual I	10.1%	7
Visual III	18.8%	13
Right/Left	23.2%	16
Discriminating Prepositions	20.3%	14
Digit Repetition	5.8%	4
Comprehension Questions	2.9%	2
Color Forms	23.2%	16
Action Agents	11.6%	8
Three-Hole Form Board	**27.5%**	**19**
Fine Motor	5.8%	4
Gross Motor	7.2%	5
Other	20.3%	14
	answered question	***69***
	skipped question	***84***

Appendix B: Focus Group Questions

Min	**Topics / Questions**
4.0	Purpose of our chat Expectations and ground rules About me, about McREL OK, let's get started!
5.0	Just to get rolling, use just one word or phrase to describe what "assessment" means to you.
3.0	In your experience, do teachers/administrators think assessment is important? Since NCLB, has your perspective on assessment changed? Tell me yes or no, and why.
4.0	Tell me how you see the difference between a developmental assessment and an academic assessment.
4.0	In your experience, which kind of assessment do teachers and administrators want – an academic assessment or a developmental one? Or a mix?
4.0	And parents that you know, what kind of assessment do they want?
6.0	Talk about the relationship between assessment and curriculum.
4.0	OK, we're going to turn in a different direction, & talk about performance based and observation based assessments. Tell me your opinion – What are the <u>strengths</u> and <u>weaknesses</u> of an early childhood assessment being <u>performance</u> based?
4.0	What about <u>observation-based</u> EC assessment – what are the <u>strengths</u> and <u>weaknesses</u>?
4.0	OK, so as for an assessment being based on observation or performance, should an EC assessment follow just one approach, or a mixture?
8.0	Let's talk about task types. Should an early childhood assessment have manipulatives? How about paper-&-pencil tasks? What about computer tasks?
12.0	So an early childhood assessment, what *should* it assess?
7.0	If you could build an assessment from scratch, what would you consider first?
20.0	Test administration training – how does that relate to assessment quality? In your experience, how do teachers and administrators <u>use</u> assessment results? Tell me about the process of sharing assessment results with the parents. How about the score report?
3.0	OK, so let me see if I can summarize your comments. Is this an adequate summary [. . .] ?
10.0	So, I know it's never possible to cover everything in a chat like this. So let me give you a moment now. What would you like to tell us that we haven't talked about?
3.0	Wrap up

110 min // 1:40 min

Appendix C: Reviewer Questions

Guidelines for Review
2008-09 GDO Renorming and Pilot Study

Thank you for agreeing to review our assessment materials. The GDO (Gesell Developmental Observation) for ages 2 ½ to 6 years of age is a beloved instrument that has been in the field for some time. It measures both developmental tasks and academic/preschool skills. Unfortunately, the norms for this instrument are quite old. While there are many loyal customers who like the GDO just the way it is, in order to comply with new standards for scientifically based materials and to hopefully enter new markets, the Institute is conducting a national normative study on the old GDO. In addition we are also collecting data on some new child items and two new questionnaires—one for parents and one for teachers. These new forms and items will allow us to publish a new, updated and more streamlined and relevant screener. The new screener (name TBD) will be a combination of some of the original Gesell items and some of our new child items. You will be reviewing both the old GDO and three new forms. Included in this packet are:

Script for administration of old items
Old GDO Child Record Form
New Child Pilot Items
New Parent/Guardian Questionnaire
New Teacher Survey

Please review and use the attached form to comment on your review considering possibly the following: Is content the information teachers need/want to know? Is content age appropriate? Is method for soliciting the information appropriate for children? Is method for soliciting the information appropriate for teachers? Is method for soliciting the information appropriate for parents/guardians? In addition, please evaluate for bias (gender, race, age, disability, SES). Any other comment?

Name: _____
Date:_____

Thank you! Please return by Feb. 6, 2009

Appendix D: Site Agreement

June 17, 2009

Name, School, Address, City, State, Zip

Dear _____:

Thank you for agreeing to participate in the 2008-2009 Gesell Developmental Observation National Renorming and Exploratory Study (*2008-2009 GDO Study*). This letter serves as a contractual agreement between Gesell Institute of Human Development and _____ School and includes guidelines for your involvement in the study. **Please note that GDO study guidelines have been slightly modified as of June 1, 2009 and are explained in the *Guidelines for Participation*** which follow.

Your willingness to act as a Data Collection Coordinator for your school is greatly appreciated, as is your interest in and support of our work at Gesell Institute. In order to ensure successful participation in the study, you are being be asked to:

- ☐ Identify one or more individuals/examiners to complete study assessment packets
- ☐ Complete and return the *Contractual Agreement* following this cover letter as soon as possible
- ☐ Train the examiner(s) with materials provided and participate in a conference call for study training and logistics
- ☐ Collect and return a signed *Examiner Demographic and Training Completion Form* for each examiner
- ☐ Collect and return a complete and legible assessment packet (details below) for each child participant

As you know, we are offering compensation for each completed assessment packet that is returned to the Institute, as outlined in the enclosed *Guidelines for Participation*. Please note that there are several options for acceptance of this compensation, including credits toward a future order and/or a donation to Gesell Institute to help defray study costs. Returned assessments must include complete and legible forms in order to qualify for study compensation.

Study forms to be returned to the Institute:

- ☐ *Examiner Demographic and Training Completion Form* (Enclosed)
- ☐ *Data Collection and Compensation Summary Form* (Enclosed)
- ☐ *GDO Child Record* (Original GDO 2007 form)
- ☐ *GDO Data Collection Form* (New form/booklet)
- ☐ *GDO Pilot Child Items* (New form/booklet)
- ☐ *GDO Parent/Guardian Questionnaire* (New form/booklet)
- ☐ *GDO Teacher Questionnaire* (New form)

Please note that for the purposes of the study, the original GDO and child *Profile Sheet* does not need to be scored/completed nor does a Developmental Age need to be determined. If your school is interested in the results of the original 2007 GDO items and the examiner has been fully trained, having attended a 2½-6-Year-Old Gesell Workshop within the last five years, you may make a copy and score/interpret the 2007 portion for your school's assessment needs. However, all original forms, including the new components, must be returned to Gesell Institute when data collection is complete.

Please review the revised *Guidelines for Participation* for additional details, then sign and return one copy of the *Contractual Agreement* as confirmation of your agreement with the terms stated in this letter. If you have any questions, please contact Andrea Sambrook at 203-777-3481 x706 or andrea@gesellinstitute.org. We look forward to our continued work with you!

Sincerely,

Marcia P. Guddemi, PhD, MBA
Executive Director

CONTRACTUAL AGREEMENT BY AND BETWEEN

Gesell Institute of Human Development
310 Prospect Street, New Haven, CT 06511

AND

(School)

As Data Collection Coordinator for my school's participation in the Gesell Developmental Observation National Renorming and Exploratory Study (2008-2009 GDO Study), I agree to the terms as stated in the contractual letter dated June 1, 2009 and the revised *Guidelines for Participation*. I understand that study compensation will only be available for complete and legible assessment forms returned to Gesell Institute in a timely manner, and I agree to select a compensation option on the completed summary form attached to my site's returned packets.

(Print Name)	(School Position/Title)
(Sign Name)	(E-mail Address)
(Date)	(Phone Number)

Please call to schedule your conference training call on 1-800-369-7709. This is an important step to ensure your examiners understand all logistical and methodological procedures for the study and have an opportunity to ask any questions before they begin. Our 2008-2009 GDO Study Team will be present for the training conference call at your convenience.

Thank you!

Guidelines for Participation

Site Administrator/Data Collection Coordinator Responsibilities

Please refer to the cover letter enclosed with these guidelines for specific responsibilities and expectations.

Examiner Requirements

In order to serve as an examiner for the 2008-2009 GDO Study, an individual must meet the following requirements.

- ☐ Watch the training video and become familiar with the new Pilot GDO items and forms.
- ☐ Participate in a training conference call addressing study logistics and questions.
- ☐ Complete and return the *Examiner Demographic and Training Completion Form* (enclosed) as soon as possible prior to testing.
- ☐ Provide a testing environment that is as free from interference as possible.

Training Overview

Training for study examiners will be provided by means of in-person or video training (depending on site location); an accompanying training and administration guide; and a conference call to cover logistics and technical assistance. It is crucial that examiners also take the time to familiarize themselves with the new Pilot forms and study protocol. Examiners must verify that they have completed this training protocol by signing the enclosed training form.

Note: Training for the study consists of an abbreviated GDO training, **covering only how to technically administer** the GDO and the new items. **This training is not in any way a substitute for the full, three-day Gesell Developmental Observation Workshop**. The regular three-day training involves more in-depth professional development, as well as technical training on how to administer, score, and interpret the GDO; and how to apply results to classroom instruction. ***Examiners completing the abbreviated study training will not be qualified GDO examiners for any other purpose than as data collectors for this study.***

Scoring and Interpretation Guidelines

GDO study examiners are asked to administer *both* the original (2007) GDO and the new (2008) Pilot items to each child, recording responses both on the original GDO child record form and on the new data collection forms provided. However, the original GDO *Profile Sheet* does **not** need to be completed, nor does a Developmental Age need to be determined, for the purposes of this

study. If your school is interested in the interpretive results of the original 2007 GDO items and the examiner has been fully trained, having attended a 2½-6-Year-Old Gesell Workshop within the last five years, you may make a copy and score/interpret the 2007 portion for your school's assessment needs. Regardless, all original forms and raw data must be returned to the Institute.

Assessment Packets and Study Compensation Requirements

In order for returned assessment packets to be considered, all of the following complete, legible forms should be returned. While it is certainly most efficient for the GDO study to collect every form for as many children as possible across the country, we recognize that certain factors affect how much time each school has to administer the GDO to an individual child for the purposes of developmental screening.

As of 6/09, slight modifications for data collection have been made by the Institute. Data for GDO child assessments at your school can still be accepted into the study even if they are not accompanied by the other items in the data collection packet. Additionally, the GDO and the Pilot Items may be administered at different times, and even to different children, as each type of data is not linked to any other for the purposes the GDO National Renorming and Exploratory Study (*2008-2009 GDO Study*). Lastly, the data collection window has been extended to October 31, 2009. We hope that this will provide schools with more opportunities to contribute to this important study.

- ☐ Examiner Demographic and Training Completion Form *(required; one for each examiner)*
- ☐ Data Collection and Compensation Summary Form *(required; one for each batch of forms returned)*
- ☐ GDO ©2007 Child Record Form *(required; one for each child)*
- ☐ GDO Data Collection Form *(required; one for each child)*
- ☐ GDO Pilot Child Items (*one for each child)*
- ☐ GDO Parent/Guardian Questionnaire (*one for each child)*
- ☐ GDO Teacher Questionnaire *(one for each child)*

In addition to the free assessment materials provided to each participating site, study compensation will be offered for every completed and legible assessment packet received by Gesell Institute. Compensation will be disbursed as data are received and processed, at a rate of **$2.50 per completed form** (if site is not able to collect all four forms), or **$10 for a completed packet**.

Data forms are to be returned in envelope(s) provided with a completed Compensation Form and sent to: **Gesell Institute of Human Development, Attention: GDO Study Materials, 310 Prospect Street, 2nd Floor, New Haven, CT 06511.**

Technical Assistance Needs

If at any time there are questions regarding the study, please contact Andrea Sambrook at (203) 777-3481 x706 or andrea@gesellinstitute.org.

Examiner Demographic and Training Completion Form

Please provide the information requested below to help inform study results. Be assured that this form will remain entirely confidential. No personal information will be shared with anyone outside of the Institute in any way. Examiner data will only be reported in the final technical report in aggregate across the entire national sample, with no identifying information.

(Name)	(School)
(Position/Title)	(School Street Address)
(Email)	(City) (State) (Zip)
(Phone Number)	(Fax Number)

1. Gender: ☐ Male ☐ Female
2. How would you describe your race and/or ethnicity?
   ☐ Asian ☐ Black/African American ☐ Latino or Hispanic ☐ White ☐ American Indian
   ☐ Hawaiian/Pacific Islander ☐ Multi-racial (Please specify:_____)
   ☐ Other (Please specify: _____)
3. What is your educational background? *(check all that apply)*
   ☐ High school diploma or GED
   ☐ Some College (Please specify area of study: _____)
   ☐ CDA (Child Development Associate) credential
   ☐ Associate's Degree (Please specify area of study:_____)
   ☐ Bachelor's Degree: ☐ Elementary ☐ Early Childhood ☐ Other (Please specify other: _____)
   ☐ Master's Degree: ☐ Elementary ☐ Early Childhood ☐ Other (Please specify other: _____)
   ☐ Doctoral Degree: ☐ Elementary ☐ Early Childhood ☐ Other (Please specify other: _____)
   Other: _____
4. Are you currently a teacher? ☐ Yes ☐ No
5. How many years of teaching experience do you have? _____
6. How many years have you taught children ages 2½ to 6? _____
7. How many years have you been an examiner of the Gesell Developmental Observation? _____
8. Date (including year) of most recent full 2½ - 6 Year Old Gesell Developmental Observation Workshop training: __ /___/ __
9. How often do you administer the GDO? _____
10. When does your school typically administer the GDO? _____

Please verify the completion of the following requirements for study participation.

☐ *I watched the GDO study training video from beginning to end or participated in an in-person study training.*
☐ *I read the Training & Administration Guide, reviewed all study forms, and practiced administering the pilot items.*
Signature: _____ Date: _____

Data Collection and Compensation Summary Form

Please use this form after June 1, 2009

School Name _____ Date _____

Number of Examiners: ____ Number of Examiner Demographic and Training Completion Forms: _____

*Please fill in the chart below and indicate with an **"X"** which forms have been completed for each child. Please double-check to make sure that all forms are complete and included in your return packet.*

***CRF** = Child Record Form (Original); **DCF** = Data Collection Form; **PCI** = Pilot Child Items; **PGQ** = Parent/Guardian Questionnaire; **TQ** = Teacher Questionnaire; **Race/Ethnicity: A**=Asian, **AA**= Black/African American, **H**=Latino or Hispanic, **C**=Caucasian, **AI**=American Indian, **PI**=Hawaiian/Pacific Islander, **MR**=Multi-Racial (specify) **O**= Other (specify)*

Child's Name	Race / Ethnicity	Sex	Native Lang.	DOB	Age	CRF	DCF	PCI	PGQ	TQ
Totals:										

Continue on reverse if more space is needed.

Total Number of Forms:		**If donating an amount to help defray study costs, please select <u>one</u> of the options below as a gesture of our appreciation. Thank you!*
Total Amount Earned :	**x $ 2.50 each**	☐ Option 1: We chose to donate $500 or more and would like a $150 credit toward our next order in the next 12 months.
Less Donation to Gesell Institute*:		☐ Option 2: We chose to donate $300 or more and would like $50 off a workshop registration in the next 12 months.
Total Amount Due:		☐ Option 3: We chose to donate $200 or more and would like $30 off our order of $50 or more in the next 6 months. ☐ Option 4: We chose to donate up to $100 and would like free shipping on our next order in the next 6 months.

If there is a balance due, please select form of compensation: ☐ Mail us a check. ☐ Hold a credit toward our next order.

*Please fill in the chart below and indicate with an **X** which forms have been completed for that child. Please double-check to make sure that all forms are complete and included in your return packet.*

***CRF** = Child Record Form (Original); **DCF** = Data Collection Form; **PCI** = Pilot Child Items; **PGQ** = Parent/Guardian Questionnaire; **TQ** = Teacher Questionnaire; **Race/Ethnicity**=Asian, Black/African American, Latino or Hispanic, White, American Indian, Hawaiian/Pacific Islander, Multi-Racial (specify) or Other (specify)*

Child's Name	Race / Ethnicity	Sex	Native Lang.	DOB	Age	CRF	DCF	PCI	PGQ	TQ

Please photocopy this blank form to use if more space is needed. Thank you!

Appendix E: Parental Consent Forms

Parent/Guardian Information and Consent Letter

Dear Parent or Guardian:

This letter is to inform you that your child's school will be participating in a national research study for which your child may be administered the Gesell Developmental Observation (GDO) in the near future. The purpose of the Gesell Institute's current national study is to collect normative and exploratory data on the Gesell Developmental Observation ©2007 along with some new pilot/test items and questionnaires.

The GDO, based on the original schedules developed by Arnold Gesell, PhD, MD at Yale University in the 1940's and revised many times since at the Institute, is currently used by thousands of preschool and elementary schools across the country to determine a child's stage of development and learning needs. It consists of many familiar performance items—block building; copying shapes and forms; writing name, letters, and numbers; interview questions about the child's personal life; complete a man drawing; language and comprehension questions; etc. This assessment measures a child's physical/neurological growth, language skills, personal-social and adaptive behavior and can help guide understanding of that child's developmental age and associated needs.

A study examiner will submit data from selected child assessments from your child's school to the Institute in order to contribute data to this important national study. If your child is selected for study participation, someone from your child's school will ask you to complete a Parent/Guardian Questionnaire, also to be returned to the Institute. Please be assured that all family and child data will be kept strictly confidential. All names will be removed from study forms and generic ID numbers will be assigned to all study forms and files. Names will not be used in association with the data or in future reports from the Institute.

If you have any questions or concerns about the study at any time, please do not hesitate to contact Gesell Institute at 1-800-369-7709 or gdostudy@gesellinstitute.org. If you do not want your child to participate, please sign the following form and return it to your child's school by _____.

Thank you in advance for your participation in this important endeavor!

Sincerely,

Marcy Guddemi, PhD
Executive Director
Gesell Institute of Human Development

☐ I do not want my child to be administered the Gesell Developmental Observation for this study.

Child's Name: _____

Child's School: _____ Date of Birth: _____

Parent's Name: _____ Phone Number: _____

Parent's Signature: _____ Date: _____

References

Almon, J., & Miller, E. (2011). *The crisis in early education: A research-based case for more play and less pressure.* College Park, MD: Alliance for Childhood.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Ames, L., Gillespie, C., Haines, J., & Ilg, F. (1979). *The child from one to six: Evaluating the behavior of the preschool child*. New York: Harper & Row.

Bredekamp, S., & Copple, C. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.

Burts, D. C., Hart, C. H., Charlesworth, R., DeWolf, D. M., Ray, J., Manuel, K., & Fleege, P. O. (1993). Developmental appropriateness of kindergarten programs and academic outcomes in first grade. *Journal of Research in Childhood Education, 8,* 23-31.

Charlesworth, R., Hart, C. H., Burts, D. C., & DeWolf, M. (1993). The LSU studies: Building a research base for developmentally appropriate practice. In S. Reifel (Ed.), *Perspectives on developmentally appropriate practice in advances in day care and early education, Vol. V* (pp. 3-28). Greenwich, CT: JAI.

Clements, R., & Schneider, S. (2006). *Movement-based learning: Academic concepts and physical activity: Ages three through eight years*. Reston, VA: National Association for Sport and Physical Education.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Elkind, D. (2012). Knowing is not understanding: Fallacies and risks of early academic instruction. *Young Children, 67*(1).

Gesell, A. (1925). *The mental growth of the pre-school child: A psychological outline of normal development from birth to the sixth year, including a system of developmental diagnosis*. New York: Macmillan.

Gesell, A., & Ilg, F. (1940). *The first five years of life*. New York: Harper.

Gesell, A., Thompson, H., & Amatruda, C. S. (1934). *Infant behavior: Its genesis and growth*. New York: McGraw-Hill.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. San Antonio, TX: Pearson Education.

Guddemi, M. P. (2003). The important role of quality assessment in young children ages 3–8. In Wall, J. & Walz, G. (Eds.) *Measuring up: Assessment issues for teachers, counselors, and administrators.* Greensboro, NC: ERIC.

Hart, C. H., Charlesworth, R., Burts, D. C., & DeWolf, M. (1993, March). *The relationship of attendance in developmentally appropriate or inappropriate kindergarten classrooms to first and second grade behavior*. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Ilg, F. (1972). Overplacement: The problem that doesn't have to be. *Teacher, 90*(1), 16-25.

Ilg, F., Ames, L., Haines, J., & Gillespie, C. (1965, 1972, 1978). *School readiness*. New York: Harper & Row.

Larsen, J. M., & Robinson, C. (1989). Later effects of preschool on low-risk children. *Early Childhood Research Quarterly, 4*, 133-144.

Lindfors, J. W. (2008). *Children's language: Connecting reading, writing and talk.* New York: Teacher's College Press.

Marcon, R. A. (1992). Differential effects of three preschool models on inner-city 4-year-olds. *Early Childhood Research Quarterly, 7*, 517-530.

Marcon, R. A. (1993). Socioemotional vs. academic emphasis: Impact on kindergartners' development and achievement. *Early Childhood Development and Care, 96*, 81-91.

Meisels, S. J., & Atkins-Burnett, S. (2000.) The elements of early childhood assessment. In J. P. Shonkoff & S. J. Meisels. (Eds.) *Handbook of early childhood intervention* (2nd ed.), pp. 387–415. New York: Cambridge University Press.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common core state standards. Retrieved March 21, 2012, from http://www.corestandards.org/

National Scientific Council on the Developing Child. (2004). *Children's emotional development is built into the architecture of their brains.* Retrieved March 21, 2012, from http://developingchild.harvard.edu/index.php/download_file/-/view/70/

Paris, S. G., & Ayers, L. R. (1994). *Becoming reflective students and teachers with portfolios and authentic assessment.* Washington, DC: American Psychological Association.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60-63.

Piaget, J., & Inhelder, B. (1969, 2000). *The psychology of the child*. (H. Weaver, Trans.). New York: Basic Books.

Pianta, R. C., & Kraft-Sayre, M. (2003). *Successful kindergarten transition: Your guide to connecting children, families, and schools.* Baltimore: Paul H. Brookes Publishing Co.

Shepard, L., Kagan, S. L., & Wurtz, E. (Eds.). (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

Shonkoff, J. P., & Meisels, S. J. (Eds.). (2000). *Handbook of early childhood intervention* (2nd ed.), pp. 387–415. New York: Cambridge University Press.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early child development.* A report of the National Research Council. Washington DC: National Academies Press.

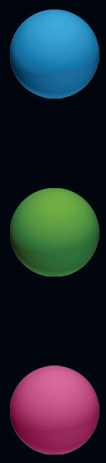
Suen, H. K., & French, J. L. (2003). A history of the development of psychological and educational testing. In C. R. Reynolds & R. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed.), pp. 3-23. New York: Guilford.

Valencia, S. W. (1990). A portfolio approach to classroom reading assessment: The whys, whats and hows. *The Reading Teacher, 43*(4), 338-340.

Wolf, K., & Siu-Runyan, Y. (1996). Portfolio purposes and possibilities. *Journal of Adolescent and Adult Literacy, 40*(1), 30-37.

Zigler, E., Gilliam, W., & Jones, S. (2006). *A vision for universal preschool education.* Cambridge, MA: Cambridge University Press.

Zigler, E., & Styfco, S. (2010). *The hidden history of Head Start*. New York: Oxford University Press.



GESELL INSTITUTE

310 Prospect Street
New Haven, CT 06511

Phone: 203-777-3481
Toll-free: 1-800-369-7709
Fax: 203-776-5001

Email: info@gesellinstitute.org
www.gesellinstitute.org