Article

Arnold Gesell's Developmental Assessment Revalidation Substantiates Child-Oriented Curriculum

Marcy Guddemi[1], Andrea Sambrook[1], Sallie Wells[1], Bruce Randel[2], Kathleen Fite[3], Gitta Selva[1], and Kelly Gagnon[1]

Abstract

Educators and parents are increasingly concerned about effects of high-stakes testing on children who may not be developmentally ready to perform tasks according to rigorous standards of today's kindergartens. In response to this issue, and to provide new psychometric data for the Gesell Developmental Observation (GDO), Gesell Institute conducted a nationwide study with nearly 1,300 children aged 3 to 6 years. Results confirm that children are able to perform developmental items according to a sequential trajectory of increasing difficulty, relative to their chronological age in 6-month intervals, and that performance mastery on items does not occur at the same time for all children of the same age. Results support the continued use of the revised GDO, now named the Gesell Developmental Observation–Revised (GDO-R), as an instrument to determine a child's developmental level along a continuous path of growth and learning. Also discussed is the importance of establishing effective and appropriate academic goals based on a child's developmental assessment results.

Introduction

Promoting academic expectations that are inappropriately aligned with a child's development creates unrealistic levels of achievement for young children. As the problematic No Child Left Behind legislation saw many children indeed left "behind," American educators and policy makers advocated for more consistent and rigorous standardized learning benchmarks for younger children, resulting in the Common Core State Standards for Kindergarten to Twelfth Grade in 2010. While these standards represent a widespread initiative to try to close the achievement gap and provide accountability through test scores, many early childhood advocates question their appropriateness and whether or not they are achievable.

Many policy and decision makers appear to be obsessed with the educational idea that sooner is better when it comes to learning core knowledge. Jean Piaget (1896-1980), an influential Swiss child psychologist, referred to the American fixation that children should learn "sooner and faster" as "the American disease" (Guddemi & Zigler, 2011). Teaching academic tasks to children at earlier ages will *not* result in greater learning for the vast majority of children, due to the developmental trajectory of child development and individual differences among children. Furthermore, increased testing will not help or hasten the process. In fact, testing is very

unreliable with young children under age eight. Experts have determined there is a 40% correlation between intelligence tests administered prior to kindergarten and results on achievement tests in third grade (Kim & Suen, 2003). However, informed parents, educators, and advocates of young children seek effective ways to establish academic goals for young children which correspond appropriately to developmental level, including social, emotional, and adaptive capacities.

A tool that addresses this need is the Gesell Developmental Observation–Revised (GDO-R) which is an instrument that evaluates a child's performance on a series of developmental and academic tasks in relation to the sequential ages and stages of typical child development in the cognitive, motor, language, and social/emotional/adaptive domains. The results of the GDO-R reveal a child's overall Developmental Age and Performance Level Ratings in each of the four domains at a point in time. Developmental age is an age in

---

[1]Gesell Institute of Child Development, New Haven, CT, USA
[2]Century Analytics, Centennial, CO, USA
[3]Texas State University, San Marcos, USA

Corresponding Author:
Marcy Guddemi, Gesell Institute of Child Development, 310 Prospect St., New Haven, CT 06501, USA.
Email: office@gesellinstitute.org

years and half years which best describes a child's collective behavior and performances on a developmental scale. A Developmental Age may differ from the child's chronological age, being lower, higher, or the same. Knowing each child's Developmental Age enables educators to customize developmentally appropriate academic experiences and expectations to best meet the learning needs of every child.

Arnold Gesell, PhD and MD, developed an assessment of human development, identifying the ages and stages of child development based on his maturationist theory (Gesell, 1925). He published the original Gesell assessment, known today as the GDO-R. It was updated in 1940 and 1965. 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 for the GDO. In 2011, after a nationwide study of 3- to 6-year-old children, the newly revalidated and revised GDO-R was published.

This article presents the psychometric results for each of the tasks on the GDO-R, indicating the typical ages at which specific developmental and academic tasks, needed for realistic, developmentally appropriate success in kindergarten, are mastered. Reliability and validity evidence are also reported to support the continued use of the GDO-R as a developmental assessment for children aged 3 to 6 years. Three tasks are discussed to illustrate the need to establish effective and appropriate academic goals based on a child's developmental assessment results.

Literature Review

Experts have stressed the significance of the kindergarten year as it relates to the child's development and the child's ability to succeed within the school environment. Kindergarten sets the tone for learning and future school success (Black, 2008; Guddemi & Zigler, 2011). Embarking on new learning creates numerous opportunities for the development of the child not only in areas of cognitive, social, emotional, and physical growth, but also as an individual within a community. Life-long, vital skills are acquired through the learning opportunities presented within the kindergarten environment.

With an increased emphasis placed on rigorous new standards and accountability, educators and parents are faced with new challenges relating to school readiness and the kindergarten curriculum. Schools play an important role in readiness; however, various schools have different expectations regarding readiness. A child may be considered prepared for one school environment and not prepared for another based on that particular school's expectations for readiness (Maxwell & Clifford, 2004). Parents and educators are concerned due to the increase of pressures and demands within early learning environments. The academic expectations of today's kindergarten are set similar to the achievement levels of first grades 20 years ago (Almon & Miller, 2011; Miller & Almon, 2009).

Although there is much discussion related to the readiness of incoming kindergarten children, it is a school's responsibility to educate children who are legally of age to attend school. Most states require children to attend school by a certain age regardless of their readiness or skill level. On the flip-side, there is also a need for schools to be *ready* for the child. It is widely accepted that school readiness is multidimensional (Ewing Marion Kauffman Foundation, 2002; National Education Goals Panel, 1997) and encompasses the following areas: physical well-being and motor development, social and emotional development, language development, approaches to learning, and cognition and general knowledge (National Education Goals Panel, 1997). Therefore, to be ready for the child, all of these areas must be addressed by the school. Furthermore, it is essential that schools, communities, and families acknowledge gaps in each child's educational abilities that can occur based not only on individual differences in normal development but also on such factors as birth weight, nutrition, television viewing, parent–child ratio, children's exposure to language and literacy, and parental involvement and participation in the child's well-being.

Despite a national focus on early childhood education, current research suggests that educational gaps continue to exist and that achievement gaps occur prior to the beginning of elementary school (Langham, 2009). It has been suggested that high-quality early education in combination with high-quality kindergarten through third-grade programs plays a critical role in attempting to close educational gaps and potentially contributes to enhancing the child's development, school readiness, and future school success. Long-term effects of a quality pre-kindergarten experience can affect grade retention, placement, special education, and school dropout rates (Barnett, 1993; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Mead, 2008). Unfortunately, not all children have the opportunity to participate in a *high-quality* pre-kindergarten program.

A consistent characteristic of high-quality pre-kindergarten and kindergarten programs is the reliance on developmentally appropriate practices (DAPs) for each child based on each child's needs. DAP is defined by the National Association for the Education of Young Children (NAEYC) as knowing where a child is developmentally, providing unique experiences based on his or her stage of development that are both challenging and achievable, and possessing knowledge about how young children learn. The organization supports educators in "promoting young children's optimal learning and development" (Copple & Bredekamp, 2009, p. 16). NAEYC's DAP encourages educators to provide learning opportunities that will enhance all areas of a child's development and to understand that a child's development follows a well-documented, sequential order. DAP is based on being aware that each child develops at his or her own unique rate and that learning opportunities need to be challenging, but within the child's ability. Understanding a child's development is key to setting expectations that are

appropriate and to planning curriculum that meets the child's needs and abilities. Early educators need to adhere to methods and practices of teaching that foster a child's development with learning being concentrated in all areas of development—cognitive, social, emotional, language, and physical (Kagan & Reid, 2009).

To plan DAP, assessing where a child is on the path of development is essential in determining what experiences a child is ready for. This use of readiness assessments should never exclude children from learning opportunities, but rather help determine how and what educational and learning experiences should be developed and/or modified to meet the child's developmental level (Gullo, 2005). Through documentation and assessment, an educator is better able to understand the child. These instruments offer insight into a child's development and his ability to learn, making learning visible to the educator (Seitz, 2008). As learning is multidimensional, it is important for an assessment to be used as a tool to help educators better understand children, their development, and how they learn (Tomlinson, 2008).

About the GDO-R

The GDO-R is a standardized, performance-based, criterion-referenced developmental assessment tool. It is designed for children from 2½ to 9 years of age and is used to inform educators and parents about a child's progress on developmental continuums. This information helps to set appropriate expectations for performance as well as instruction for children based on their developmental stage or level. When combined with the Parent/Guardian Questionnaire (PQ) and the Teacher Questionnaire (TQ), the GDO-R functions as a comprehensive assessment system. The GDO-R can also help determine whether or not a child may need further diagnostic evaluations to suggest appropriate planning or remediation in specific areas of development.

The purpose of the GDO Study was to provide updated technical data and reliability evidence for 17 of the 19 original tasks on the ©2007 GDO (see Table 1); two tasks that were intended for assessing older children, Right and Left and Visual III, were omitted. Another purpose of the GDO Study was to define Overt Behavior (Task 20) and to strengthen the social/emotional/adaptive domain (Task 21). The criterion that shaped this study, and subsequent outcomes, was based on three sources of information:

- *Scientific data collected on a nationwide sample of nearly 1,300 children*. This technical data provide information about how children across the United States performed on all GDO-R tasks, and it 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.

Method

The GDO study consisted of several sub-studies designed to collect both quantitative and qualitative data. Quantitative data were collected on children in seven age bands (spanning ages 3-6 years) and is reported here. Examiners administered a total of 167 items in one-on-one sessions with children. Data were 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 at home (78 items).

Qualitative data were collected on two developmental tasks on the GDO, the Copy Forms and Incomplete Man, as part of the Gesell Institute National Lecture Staff (NLS) Review Study. The purpose of the NLS Review Study was threefold: 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.

GDO Study Timeline

The GDO Study was completed over the course of 3 years following American Educational Research Association (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) standards. A timeline is presented in Figure 1 to provide context and sequence for the data activities and analyses of the GDO Study. The steps in the timeline are as follows.

Review of Child Development Literature

A comprehensive review of child development literature on observation, assessment methodology, and assessment instruments for children ages 2½ to 9 was undertaken. The GDO instrument was evaluated as a whole based on appropriateness of developmental tasks to early educational goals for young children, culturally sensitive measurement approaches, and consistency with developmental capabilities of children between the ages of 2½ to 9.

Content Validity

After a careful review of the literature and with the results of a Bias Review, the research team felt the items on the GDO

Table 1. GDO-R Tasks.

Task No.	Task name		Refer to Table(s)
1	Cubes	This set of items requires the child to reproduce block structures built by the examiner: The Tower, the Train, the Bridge, the Gate, Steps with 6 cubes, and Steps with 10 cubes. The ability to reproduce the structures successfully and the approach to the item used by the child provides information about horizontal and visual perception, fine motor coordination, attention span, spatial judgment, and short-term memory.	8
2	Interview	A child's responses to the series of questions related to home life (such as his or her favorite story or TV program) revealed expressive and receptive language skills, as well as the ability to recall everyday experiences. Responses provide a glimpse of the child's cognitive organizational skills, ability to stay on task, and ability to follow directions. While these are important GDO-R tasks, study data for Interview and Interests (Task 10) were not analyzed in aggregate since the examiner evaluated individual language samples across the entire assessment session, and scored the domain using a qualitative rubric.	
3	Name and Numbers	This set of items requires the child to first write his or her name and as many numerals (up to 20) as he or she can. Both this task and Task 4 (Copy Forms) evaluate a child's competence in integrating visual information with motor abilities, visual tracking skills, and discrimination abilities. The size, shape, and organization of the products drawn indicate maturity in fine motor ability, organizational skills, awareness of detail, visual perceptions, ability to execute angles, and overall eye–hand coordination.	9, 10
4	Copy Forms	The child is asked to copy a Circle, Cross, Square, Triangle, Divided Rectangle, Vertical and Horizontal Diamonds, and ultimately 3-dimensional shapes (cube and cylinder) according to his or her age and demonstrated ability. Some age 6 children were not administered the first three items (Scribble, Horizontal, and Vertical Stroke) when the examiner believed the items were too easy for these children. In this case, the children received a missing score that was treated as incorrect for the analyses, leading to the lower p values for these items for age 6 children.	11
5	Incomplete Man	This task requires a child to add missing symmetrical body parts to a given drawing. It measures fine motor skill, perceptual awareness, balance, symmetry, and spontaneous task completion.	12
6	Right and Left	Not included in the GDO study. This task is intended for children aged 6 and older.	
7	Visual I	This visual discrimination task requires a child to match symbols presented one at a time on a card to the corresponding symbol on a worksheet. This task measures a child's competence in left-to-right directionality, visual discrimination, ability to sustain attention, to find one's place repeatedly, and to carry out directions.	13
8	Visual III	Not included in the GDO study. This task is intended for children aged 6 and older.	
9	Naming Animals	The child is encouraged to name all the animals he can think of. Responses provide information about a child's level of expressive and receptive language, retrieval skills, and cognitive organization processes. Recall, ability to conceptualize, attention to task, and classification skills are also observed in this 60 second timed task.	14
10	Interests	See description in Task 2: Interview	
11	Prepositions	This item assesses the child's understanding of specific prepositional phrases and his or her ability to apply them to a corresponding action (placing place a cube on, under, in back of, in front of, and beside a chair).	15
12	Digit Repetition	This task requires the child to repeat a series of digits with increasing length. It measures auditory and short-term memory, as well as listening ability. As age increases, children's ability to repeat increasingly longer digit sets increase as well.	16
13	Comprehension Questions	During the Comprehension Question task, the child is asked "What must you do when you are hungry, sleepy, cold, have lost something, or cross the street?" A child's performance in this area measures cognitive processes related to problem solving, personal experience, and knowledge and understanding of specific words and phrases.	17

(continued)

Table 1. (continued)

Task No.	Task name		Refer to Table(s)
14	Color Forms	This task measures visual discrimination by asking a child to place cut-out shapes on a corresponding board. This task is designed to better differentiate performance at younger ages.	18
15	Three-Hole Form Board	This item uses puzzle-like materials to measure a child's visual discrimination, depth perception, and spatial perceptual accuracy in a variety of orientations. This task is expected to better differentiate performance at younger ages.	19
16	Action Agents	This item requires the child to generate a word (noun) that could produce the action suggested; e.g., "what cries or what runs?" Language comprehension skills are measured on this item. It requires a relatively long period of sustained attention to the task.	20
17	Identifying Letters and Numbers	(a) *Identifying Letters*: This task requires a child to identify random capital letters. This task is dependent on prior exposure and knowledge of the alphabet. (b) *Identifying Numbers*: This task requires a child to identify random numerals 1 to 12 by name. This task is dependent on exposure to and knowledge of numerals.	21, 22
18	Numeracy (Counting, One-to-One Correspondence, Conservation, and Calculations)	(a) *Counting*: Counting reveals the child's experience with and ability to remember numbers in a sequence. Children were allowed to count up to 40. (b) and (c) *One-to-One Correspondence and Conservation*: One-to-one correspondence evaluates the child's understanding that each item is represented only once by a number name. Children learn to count with one-to-one correspondence before they learn to conserve the same number of items; e.g., the child when asked, "how many altogether?" must know that the number of pennies that he or she just counted is still the same number. (d) *Calculations*: The calculation task demonstrates a child's ability to compute simple mathematical problems. They may use the pennies to help figure out the answer.	23, 24, 25
19	Motor (Fine and Gross)	(a) *Fine Motor: Pellets*: The fine motor task determines hand-eye coordination skills as the child drops one pellet at a time into a small jar, using first the dominant hand and then the non-dominant hand. *Pencil Stroke*: A solid pencil stroke is not a solid expectation until the child is 5.5 years of age. *Pencil Grasp*: The 2-3 finger grasp is the preferred method of holding a pencil as the child's age increases. (b) *Gross Motor*: Large motor activities like tip-toeing, hopping, balancing on one foot, jumping up and down, broad jump, skipping, catching and throwing offer additional information concerning large motor skills, hand-eye coordination, and visual perception.	26, 27, 28, 29
20	Overt Behavior	Not reported here.	
21	Social, Emotional, and Adaptive	The GDO-R utilizes three subscales of the Teacher and Parent/Guardian Questionnaires to measure social interactions with adults and peers, ability to self regulate and cope with transitions, and self-help skills in daily life.	39

Note. GDO-R = Gesell Developmental Observation–Revised.

instrument remained relevant with the exception of two items. The GDO Study was designed to collect updated technical and baseline data using the ©2007 GDO as the basis of the investigation to renew the reliability and validity evidence for the instrument. In addition, the PQ and the TQ were developed to address the social, emotional, and adaptive development of the child.

Procedures for Reducing Bias

The five experts who reviewed the GDO for bias also evaluated content and age appropriateness for each task. The team was selected 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, to identify any biases inherent in the content or methodology of the GDO. The bias review questions can be found in the full Technical Report (Gesell Institute of Child Development, 2012) at www.gesellinstitute.org.

Several GDO tasks raised issues due to a current lack of cultural relevancy. One item within the Interview task 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 Naming Animals because it failed to allow for the breadth and complexity of birthday celebrations in different cultures and by families with lower socioeconomic circumstances. Another item in the Interview task prompts the child to name animals, presuming all
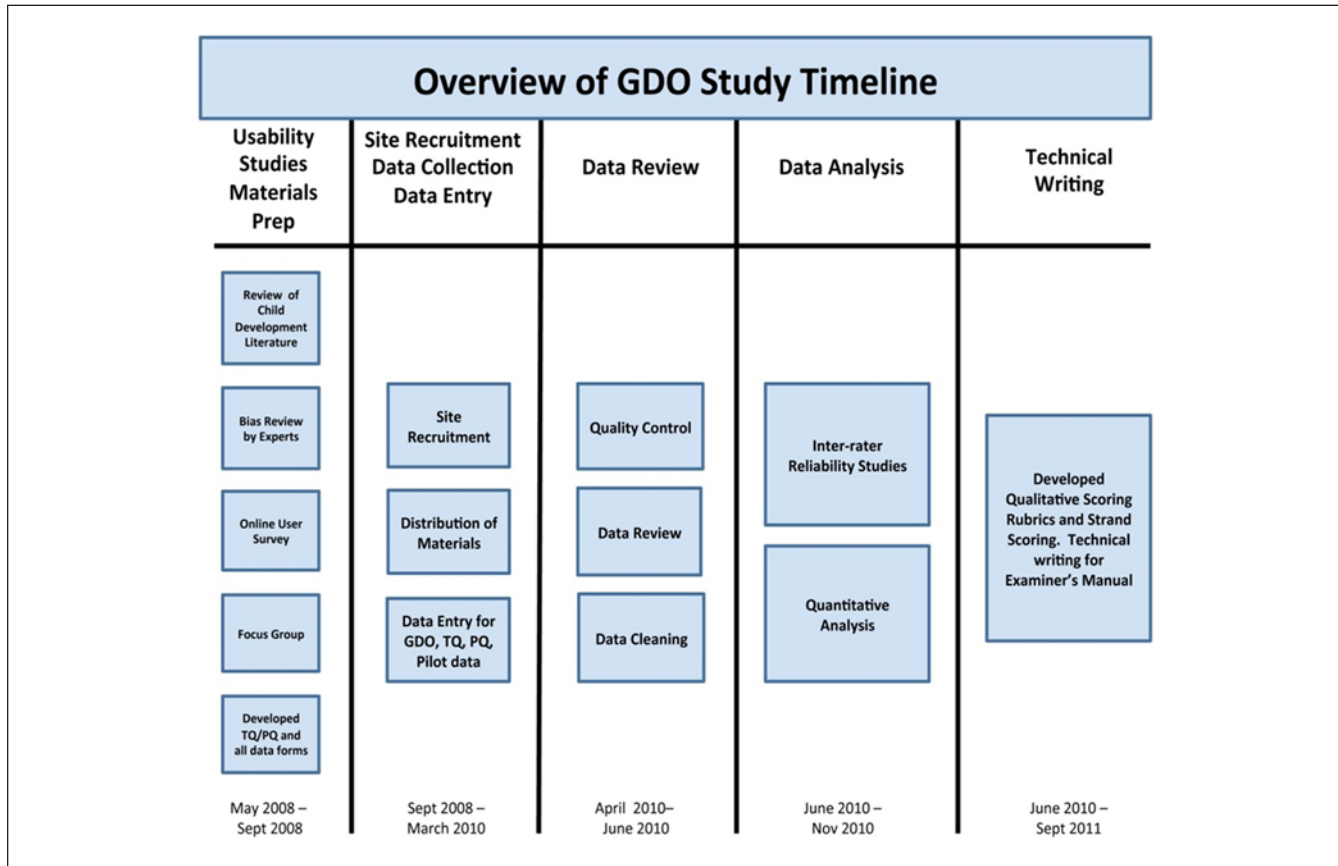
Figure 1. Overview of GDO study timeline.
Note. GDO = Gesell Developmental Observation.

children have had the 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 to elicit more exchange with the child for a language evaluation.

Online User Survey

Prior to the commencement of the study, an online survey of GDO users was conducted to collect information on how customers use the GDO. A sample of $N = 153$ respondents provided feedback. A summary of the questions and quantitative results can be found in the full Technical Report (2012) Appendix A.

Focus Group

A focus group held at a Massachusetts preschool 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 that would be most meaningful to examiners, in addition to the updated technical data, which were collected to renew the validity of the instrument.

TQ and PQ

It was recommended by all reviewers that the GDO-R include a measure of emotional regulation and social behavior by surveying the child's teacher and parent or guardian. After reviewing the literature on emotional, social, and adaptive behaviors and examining multiple existing parent and teacher questionnaires, the TQ and PQ were developed for the GDO Study.

Procedures for Recruiting Examiners and Distributing Materials

Each site that enrolled in the study signed an agreement and secured parental consent forms for each participating child (Gesell Institute of Child Development, 2012).

Trained examiners administered all designated items on the GDO to children within the study age band (see Table 2) 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 were collected from three sources at each school: the child (GDO tasks-GDO), the parent or guardian (PQ), and the child's teacher (TQ).

Table 2. 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 to 3 years 2 months and 29 days
3[6]	3 years 3 months and 0 days to 3 years 8 months and 29 days
4	3 years 9 months and 0 days to 4 years 2 months and 29 days
4[6]	4 years 3 months and 0 days to 4 years 8 months and 29 days
5	4 years 9 months and 0 days to 5 years 2 months and 29 days
5[6]	5 years 3 months and 0 days to 5 years 8 months and 29 days
6	5 years 9 months and 0 days to 6 years 2 months and 29 days

Note. GDO = Gesell Developmental Observation

All GDO assessment forms and study materials were provided to each school free of charge, including return shipping and handling. A Training DVD was given 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 offer support. Assistance was also supplied 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.

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, as the schools did not currently utilize the GDO but wanted to participate in the study.

The final sample for analysis for the GDO Study included 1,287 children from 53 geographically diverse sites in 23 states. While the assessment is designed for ages 2½ to 9 years, the researchers chose to limit hands-on data collection to the group of children that comprised the largest number of users of the GDO. The chronological age threshold for each age band is described in Table 2. The age bands targeted for data collection were 3, 3½, 4, 4½, 5, 5½, and 6.

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 Tables 3 to 5 for demographics of participating sites.

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 7 years GDO experience, received GDO training in one of two ways: (1) in the last 5 years through a 3-day workshop or (2) a 1-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 88% of the GDO study examiners have a bachelor's, master's, or doctoral degree. Refer to Table 6 for examiner's level of education.

Data Validation and Entry Procedures

Data were systematically reviewed for completion, accuracy, and any possible serious administration errors prior to being entered into an electronic survey gizmo file. Unusable data were put aside. Ten percent of the data entered in each of the data sets (GDO, 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.

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 assessments were used in the final analysis. Thus, 5% of the sample could not be used and was deleted from the original sample. Over the course of 2 years, communication with new sites about data collection/submission procedures improved significantly, and thus further reduced the number of invalid assessments (i.e., child was too young or old for the study, etc.) that were submitted by each site.

Results

Sample Descriptive Statistics

The final sample used for analysis was a proportional mix of boys and girls in each age band (3.0-6.0 years) from ethnically diverse backgrounds. The number of children in each age band varied (see Table 7).

Task Descriptive Statistics by Age Band

The p value represents the proportion of children who provided the correct responses to the item (scored with 0, 1). For
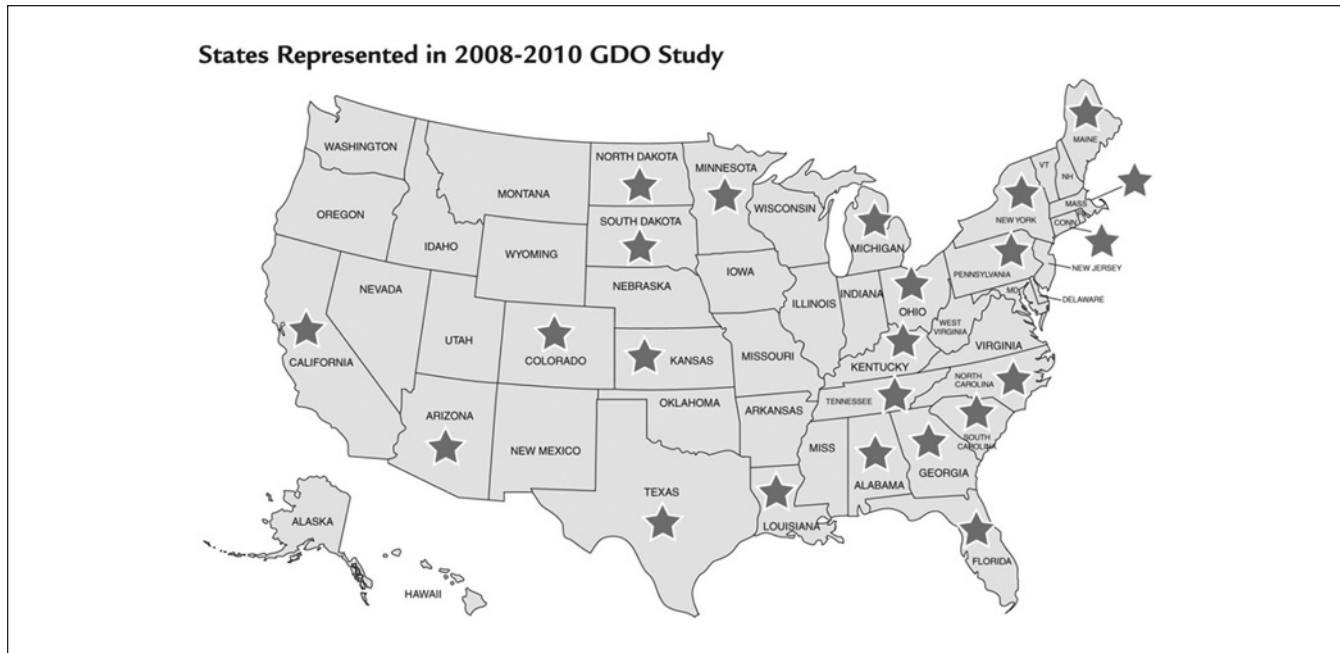
Figure 2. Sample distribution by state.
Note. Participating states are indicated by a star.

Table 3. Descriptive Statistics for Participating Sites: School Type.

School type	Private	Public
# Sites	33 (62%)	20 (38%)
# GDO-R assessments	584 (45%)	703 (55%)

Note. GDO-R = Gesell Developmental Observation–Revised.

Table 4. 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-R assessments	672 (52%)	201 (16%)	288 (22%)	126 (10%)

Note. GDO-R = Gesell Developmental Observation–Revised.

Table 5. Descriptive Statistics for Participating Sites: Ethnicity and Socioeconomic Status.

	M %
Ethnicity	
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. Numbers in parentheses are standard deviations. Eligibility for free/reduced lunch program is used as a representative variable in the sample for lower socioeconomic level.

polytomous items (scored with 3 or more score points; for example, 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 8 through 29 show growth in performance on the GDO tasks across age bands.

Some tasks/items have been organized into separate tables because they require different statistical operations to best demonstrate the findings. These are as follows:

1. *Tasks that contain continuous items that are scored from 0 to a maximum number*: the number of numerals written, and the number of those numerals that were in sequential order (Table 10); the number of animals named (Table 14); how high the child counted (Table 23); and the number of seconds for dominant hand pellets (Table 26). Also included are Social/Emotional/Adaptive Items (Table 30). 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.

Table 6. Examiner Demographics: Level of Education.

Level of education	*n*	% sample
Some college	6	5.94
Associate's degree	1	0.99
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

Table 7. Overview of Gesell Developmental Observation Study Sample Child Descriptive Statistics by Age Band.

	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Number of children	53	131	186	264	278	221	154
M age	3.08	3.54	4.00	4.52	5.00	5.48	5.97
SD	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 the child's native language occurred when Parent/Guardian Questionnaires (PQs) were not returned. In such cases, a shortened form of the PQ (Child Demographic Form) was modified so that basic demographic information could still be collected from school records for participating families (sex, ethnicity, DOB, Native language). Missing data for ethnicity are within the acceptable limit. Missing data for native language do not affect findings because all children were administered the assessment in English and had to be deemed fluent by the teachers who administered the assessment.

2. *Tasks that contain categories*: items from Distinguishing Features and Overt Behavior. These items were analyzed using a frequency distribution, because means and standard deviations were not appropriate. These are not reported here.

Item p *Values by Task and Age Band*

Tables 8 through 29 provide results for the GDO-R items by task and age band. The tasks are described in order of test administration. 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 children's increased competency with age; *p* values are lower for the younger children and higher for the older children. Dr. Gesell utilized two levels for identifying developmental competence, successful or not successful, on items and tasks at a 51% benchmark. The results presented here utilize *three* Performance Level Expectations for each GDO-R task, rather than two. The criteria set for each expectation were established using a developmental framework of growth and learning in which a child acquires the capacity to succeed at more difficult items after having mastered the less complex items which precede it. They are shaded accordingly:

- Solid Expectation (SE)—over 70% of the children could complete the task (dark gray shading)
- Qualified Expectation (QE)—50% to 69.9% could complete the task (light gray shading)
- Not Yet Expected—under 50% of the children could complete the task (no shading)

The Performance Level Expectations for each task reflect responses of a large group of children of the same age in the sample that were able to complete the task independently, *without* demonstrations or cues from the examiner. Typical GDO-R administration allows for demonstrations and cues for some items because it allows the examiner to differentiate between developmental levels and also reveals the child's approach to the task, especially when they are challenged by an item or task. In this article, we report on Solid Expectation, over 70% of the children could complete the task.

Content-Related Validity

Content-related validity is evidenced by uniformity between 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 online user survey provided additional information regarding overall assessment effectiveness (addressing such topics as the appropriateness of the criteria 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.

Table 8. *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 9. *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 10. Descriptive Statistics for Writing Numbers Task by Age Band.

	Age band						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Number of numerals written							
n	50	105	171	243	261	211	150
Maximum	20	20	20	20	20	20	20
M	0.16	0.12	0.47	2.81	6.15	9.78	15.07
SD	0.55	0.63	1.04	4.27	5.38	6.52	6.46
Number of sequential numbers							
n	50	105	171	243	261	211	150
Maximum	20	20	20	20	20	20	20
M	0	0.04	0.18	2.07	5	8.43	14.35
SD	0	0.39	0.73	4	5.57	7.29	7.07

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 Qualitative Review Study team held over one hundred years of experience administering the GDO and conducting Gesell workshops on topics such as school readiness, parent involvement, and child development.

Inter-rater reliability of the GDO-Revised provides evidence regarding the degree to which Developmental Age can be reliably assigned. The inter-rater reliability study included a subsample of children's performance on the Incomplete Man and Copy Forms tasks. Table 31 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 unable to score by one or both raters.

Inter-rater reliability was calculated by comparing the agreement between the developmental ages assigned by Rater A and Rater B of Team 1 for each task. During Phase 1, Rater A and Rater B of Team 1 rated Copy Forms, while Rater A and Rater B of Team 2 rated Incomplete Man. Inter-rater agreement for assigning overall Developmental Age was calculated for Copy Forms and Incomplete Man samples. Inter-rater agreement was also calculated for each individual Copy Form item in Phase 2.

During Phase 1, for both Incomplete Man and Copy Forms, neither team 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 32). 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 32 provides 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

Table 11. *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 oint-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 required.

Table 12. *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)
Eyes	0.46	0.50	0.65	0.77	0.87	0.92	0.94
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
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
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
Body line	0.26	0.38	0.47	0.62	0.78	0.75	0.73
Bowtie	0.00	0.09	0.09	0.21	0.31	0.51	0.68
Neck	0.06	0.11	0.17	0.25	0.31	0.40	0.56
Knot	0.00	0.03	0.03	0.06	0.12	0.15	0.23
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
Mean number of body parts added (out of a total of 39 possible points)	7.3 (6.17)	11.74 (7.7)	16.61 (7.38)	20.91 (7.25)	23.9 (5.52)	25.68 (4.57)	27.53 (4.02)

Note. The number of body parts included for Incomplete Man task in the study was 13. These include Knot, Other-1, and Other-2 that are not customarily part of the standard Gesell Developmental Observation–Revised (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).

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 32). Perfect correlations are not expected because of the variation in development between children.

In Phase 2, Table 33 presents results of inter-rater reliability for Copy Forms items. 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 of 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 unable to score by one or both raters.

Table 13. *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
Mean *p* value	.32	.33	.42	.60	.71	.84	.91
Mean number of items correct	3.87 (2.76)	3.93 (2.82)	5.09 2.74	7.15 (3.22)	8.55 (2.87)	10.04 (2.26)	10.92 (1.56)

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). Standard deviations are indicated in parentheses.

Table 14. Number of Animals Named by Age Band.

	Age band						
Naming Animals	3	3.5	4	4.5	5	5.5	6
n	47	122	173	262	274	214	149
Maximum	60	60	60	60	60	60	60
M	1.89	2.66	4.37	5.98	7.64	9.48	10.42
SD	2.19	2.31	3.15	3.16	3.56	3.94	3.58

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.

Table 15. *p* Values for Prepositions Task by Age Band.

	3.0 (n = 52)	3.5 (n = 126)	4.0 (n = 177)	4.5 (n = 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

The results in Table 33 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 unable to score by both raters. All children who could be scored received a score of 0, resulting in an inter-rater reliability of 1.00. Cube Face-on and Cube Point-on are some of the most difficult items in the entire GDO-Revised. Very few 6-year-old children in the entire study sample received a correct score on these items.

Table 16. *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 17. *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
Cold	0.25	0.47	0.62	0.70	0.83	0.86	0.84
Sleepy	0.41	0.48	0.56	0.67	0.77	0.79	0.83
Cross street	0.18	0.28	0.46	0.67	0.81	0.86	0.85
Lost something	0.12	0.13	0.26	0.51	0.65	0.79	0.84

Table 18. *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 19. *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 20. *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
Mean *p* value	.23	.32	.46	.61	.71	.79	.84
Mean number of action items correct	4.84 (4.82)	6.8 (5.35)	9.67 (5.39)	12.8 (5.38)	14.99 (4.12)	16.53 (2.88)	17.61 (2.4)

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.

Limitations

The GDO Study contributes a comprehensive sample of child development data to the educational field at large. As with any study of its size and scope, it has limitations. The distribution of child ethnicity across the total sample more closely resembled 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 PQ, 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 testing at the site (i.e., GDO-R tests were part of admission protocol in private schools. These schools did not administer the GDO Study PQ 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.

Table 21. 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
Mean p value	.05	.13	.23	.46	.65	.83	.93
Mean number of letters identified	1.18 (3.10)	3.41 (6.74)	5.9 (8.25)	12.04 (10.23)	16.83 (9.67)	21.6 (7.38)	24.07 (4.86)

Table 22. p Values for Identifying 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
Mean p value	.05	.13	.25	.53	.71	.86	.96
Mean number of numerals identified	.64 (1.99)	1.6 (3.19)	3.04 (3.98)	6.3 (4.83)	8.47 (4.08)	10.33 (3.02)	11.46 (1.81)

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 all tasks in the assessment: Paper Position, Head Shift, Body Posture, Non-Dominant Hand Posture, and Eye Movement. This may explain why observational data on these items contain missing cases (examiners did not complete the section of the form). Thus, a shortcoming of this study is that these items cannot clearly be interpreted. However, the Qualitative Review Study and Inter-rater

Table 23. Descriptive Statistics for Counting Task.

Counting	Age band						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
n	35	90	122	193	201	164	118
Maximum	40	40	40	40	40	40	40
M	5.86	6.91	9.56	16.85	23.75	29.36	34.48
SD	5.82	5.51	6.86	11.43	11.56	12.36	10.09

Table 24. p Values for One-to-One Correspondence and Conservation Tasks 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 25. 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 26. Descriptive Statistics for Pellets Task.

Pellets (dominant hand)	Age band						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
n	50	126	171	214	202	161	135
Maximum	99	99	99	99	99	99	99
M	25.34	24.91	25.74	25.26	25.28	24.91	21.09
SD	11.46	11.43	11.37	9.97	10.62	10.63	6.52

Reliability Study strongly confirm the developmental characteristics of each age band as related to the Copy Forms and Incomplete Man tasks. This is very important because it provides recent validity evidence for these specific developmental tasks and allows for continual improvements to the training of examiners.

Table 27. Frequency Distribution for Distinguishing Features: Pencil Stroke by Age Band (%).

	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 28. Frequency Distribution for Distinguishing Features: Pencil Grasp by Age Band (%).

	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

Implications

The most valuable implication of this research is that the GDO-R has renewed reliability and validity evidence to support its continued use as a developmental instrument to evaluate growth and development of children aged 3 to 6 years and to inform instruction for developmentally appropriate activities. The results from this study also support the original findings for developmental tasks as originally published by Arnold Gesell (Gesell, 1925). Children are developing and reaching the major developmental milestones at about the same time as they did when Dr. Gesell first started collecting date over a century ago.

A few of the important implications of the research for educators nationwide include the following:

1. Perceiving oblique lines is a prerequisite to letter formation and writing—two essential expectations in the kindergarten curriculum of today. Building the Gate (Task 1: Cubes) and copying the Triangle (Task 4: Copy Forms) require that the child not only perceive the oblique angle of the cube or the form, but is able to reproduce the structure in 3-D or on paper. The GDO study documents that this developmental capacity is solid only by age 5 (Task 1: Cubes–Gate) and 5.5 (Task 4: Copy Forms–Triangle). Educators must be alert to both variations in chronological age and developmental level to properly balance the pace and sequence of daily learning activities for each child.

2. Children correctly identify letters in the alphabet in a graduated process that is affected by age, experience,

Table 29. *p* Values for Motor Items 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)
Jump in Place	0.88	0.89	0.96	0.96	0.99	0.99	0.98
Jump	0.69	0.72	0.90	0.93	0.91	0.95	0.96
Walk on Tiptoe	0.63	0.66	0.82	0.85	0.90	0.96	0.97
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
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).

Table 30. Mean Scores of Social/Emotional/Adaptive Items.

	Age band						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Social	3.37 (.90)	3.65 (.73)	3.97 (.64)	4.11 (.72)	4.07 (.71)	4.14 (.73)	4.26 (.66)
Emotional	3.48 (.69)	3.57 (.60)	3.73 (.59)	3.87 (.61)	3.86 (.61)	3.87 (.70)	4.00 (.67)
Adaptive	3.40 (.82)	3.53 (.87)	3.92 (.68)	4.04 (.76)	4.05 (.82)	3.99 (.85)	4.13 (.79)

Note. Standard deviations are in parentheses.

Table 31. Inter-Rater Reliability Study Sample.

	Incomplete Man	Copy Forms	Copy Forms items
M 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

Table 32. Inter-Rater Agreement Evidence for Developmental Age.

	IM	CF
	Rating Team 1 (A/B)	Rating Team 2 (A/B)
	n = 122 IM samples	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

Note. IM = Incomplete Man; CF = Copy Forms.

and exposure to the printed word. As such, the average 4.5 year old can successfully identify approximately 12 letters of the alphabet while a year later, at 5.5, they can identify 21 to 22 letters. Educators who attempt to teach writing letters before the age of 5.5 (when most children can perceive and execute the oblique lines of letters) are doing their young students a disservice, which may result in a child internalizing failed attempts at writing before his or her developmental capacity for the task exists. Taking the time to understand how developmental level can be leveraged for teaching will benefit both children and teachers.

3. Educators who are able to recognize when a child is beginning to conserve 10 or more items will likely find that the child can also begin to succeed at simple calculations which have final answers less than 5 (beginning around 5.5 years and solid expectation by 6). Until a child can conserve item sets of 13 to 20, his or her success at calculations will likely remain the product of memorization or chance, as opposed to concepts of true numeracy.

Table 33. Inter-Rater Reliability for Copy Forms Items.

	Rater 1			Rater 2			
	n	*M*	*SD*	*n*	*M*	*SD*	*r*
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	.00	.00	27	.00	.00	1.00
3-D cube point-on	23	.00	.00	20	.00	.00	1.00

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

Conclusion

The results of this study, based on a culturally and socioeconomically diverse sample of children 3 to 6 years of age in seven age bands, provide evidence that children's performance on developmental and academic tasks, as measured by the GDO-Revised, occurs in a sequential progression of mastery which increases with age. In addition, the results provide evidence that not all children of the same chronological age arrive at each developmental level for the same tasks at the same time. Thus, there exists variation in performance on developmental and academic tasks between children of the same age. Future research should include a more intensive analysis of the data by weighing variables such as child ethnicity, geography, and socioeconomic level to pursue stability in the findings.

It is essential that educators, policy makers, and parents understand the significance of developmental level when setting standards for all children. Because children in kindergarten are at various chronological ages and develop at varying rates, having the same set of standards and expectations for *all* children at a given time is both inappropriate and potentially harmful for children.

Utilizing standardized, performance-based instruments to understand a child's developmental level, cultural and social influences, and individual interests allows for appropriate expectations, relevant goals for learning, and proper accountability in the educational system. Educators can utilize each child's unique developmental profile to plan curriculum that respects the developmental level and potential of the child by using robust observational methods coupled with comprehensive developmental assessment tools.

The results of the GDO Study presented here strongly support the GDO-R as a reliable and valid developmental measurement tool, confirm the essential role that a child's developmental level plays in his or her success for learning today, and suggest that having the same expectation for all children at the same time is inappropriate if not impossible.

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

Barnett, W. S. (1993). Benefit-cost analysis of preschool education: Findings from a 25-year follow up. *American Journal of Orthopsychiatry*, *63*(4), 25-50.

Black, S. (2008). Early education, later success. *American School Board Journal*, *196*(9), 61-63.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science*, *6*, 42-57.

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

Ewing Marion Kauffman Foundation. (2002). *Set for success: Building a strong foundation for school based on the social-emotional development of young children*. Kansas City, MO: Author.

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, NY: Macmillan.

Gesell Institute of Child Development. (2012). *Gesell developmental observation-revised and Gesell early screener technical report ages 3-6*. New Haven, CT: Author. Available from http://www.gesellinstitute.org

Guddemi, M., & Zigler, E. (2011). *Community early childhood LEADership e-kit* [CD-ROM]. New Haven, CT: Gesell Institute of Child Development.

Gullo, D. (2005). *Understanding assessment and evaluation in early childhood education*. New York, NY: Teachers College Press.

Kagan, S., & Reid, J. (2009). Invest in early childhood education. *Phi Delta Kappan*, *90*, 572-576.

Kim, J., & Suen, H. K. (2003). Predicting children's academic achievement from early assessment scores: A validity generalization study. *Early Childhood Research Quarterly*, *18*, 547-566.

Langham, B. A. (2009). The achievement gap: What early childhood educators need to know. *Texas Child Care Quarterly*. Retrieved from http://www.childcarequarterly.com/fall09_story2a.html

Maxwell, K., & Clifford, R. (2004). *Research in review: School readiness assessment*. Washington, DC: National Association for the Education of Young Children.

Mead, S. (2008). Find success in early childhood education. *American School Board Journal*, *195*(11), 25-29.

Miller, E., & Almon, J. (2009). *Crisis in the kindergarten: Why children need to play in school*. College Park, MD: Alliance for Childhood.

National Education Goals Panel. (1997). *The national education goals report: Building a nation of learners*. Washington, DC: U.S. Government Printing Office.

Seitz, H. (2008). The power of documentation in the early childhood classroom. *Young Children*, *63*(2), 88-92.

Tomlinson, C. A. (2008). Learning to love assessment. *Educational Leadership*, *65*(4), 8-13.

Author Biographies

Marcy Guddemi, PhD, MBA is the Executive Director of Gesell Institute of Child Development, a nonprofit child advocacy organization on the Yale campus. She is a well-known for her work on play and children's assessments.

Andrea Sambrook, MA, is the Owner of Annette's Preschool in Hinesburg, VT. She was the Research Director for the Gesell GDO Study.

Sallie Wells, MA, is the Head of School at The Clariden School in Southlake, TX. She is also a member of the National Lecture Staff at Gesell Institute of Child Development and an Adjunct Professor at Texas Woman's University.

Bruce Randel, PhD, has extensive experience and expertise in educational measurement, assessment, and psychometrics. He has been the technical lead for state-wide assessment programs for NCLB, nationally normed achievement tests, nationally normed aptitude tests, and early childhood assessments including the GDO-Revised. He owns Century Analytics, Inc. in Denver, CO.

Kathleen Fite, EdD, has worked as a writer, consultant, researcher, educator, and leader at many levels. She is a Gesell International Ambassador and serves on the Gesell Institute Advisory Council. She is a Distinguished Alumni and professor of education at Texas State University.

Gitta Selva studied communications at New York University. She has been working in the field of education for many years, and copublished several articles about a service-learning project at Queensland University of Technology in Australia. Gitta is a staff writer at Gesell Institute of Child Development.

Kelly Gagnon recently earned a Bachelor of Arts degree in psychology and French from Hobart and William Smith Colleges. She plans to pursue further education in the field of early childhood development. She is currently a research intern at the Gesell Institute.