

BUROS

CENTER FOR TESTING

TEST REVIEWS

Gesell Developmental Observation—Revised

Gesell Developmental Observation–Revised

Purpose

Designed as a comprehensive developmental screening that measures social/emotional/adaptive skills, physical/neurological growth, language skills, and cognitive behaviors such as thinking, memory, perception, attention to task, ability to follow directions, short term visual and auditory memory, cognitive-perceptual thinking, organizational skills, logical mathematical thinking skills, and application of what is learned.

Population

Ages 2.5 through 9 years.

Publication Dates

1964-2012.

Acronym

GDO-R.

Scores

26 tasks: Cubes, Interview, Name and Numbers, Copy Forms, Incomplete Man, Right and Left, Visual I, Visual II, Naming Animals, Interests, Prepositions, Digit Repetition, Comprehension Questions, Color Forms, Three-Hole Form Board, Action Agents, Identifying Letters and Numbers, Numeracy, Counting, One-to-One Correspondence, Conservation, Calculations, Fine Motor, Gross Motor, Overt Behavior, Social/Emotional/Adaptive.

Administration

Individual.

Levels, 2

Ages 2.5 to 6.5; ages 7 to 9.

Price Data, 2013

\$97.80 per Questionnaire bundle including 30 GDO-R forms, 30 Parent/Guardian Questionnaires, and 30 Teacher Questionnaires; \$59.95 per 30 student record forms; \$23.65 per 30 Teacher Questionnaires; \$23.65 per 30 Parent/Guardian Questionnaires; \$30.85 per 30 Spanish language Parent/Guardian Questionnaires for 30 families; \$184.40 per examiner's manual (2011, 230 pages); \$92.65 per examiner's script; price information for technical manual available from publisher.

Foreign Language Edition

Parent/Guardian Questionnaires are available in Spanish.

Time

Untimed; (20-45) minutes.

Comments

5 strands: A (Developmental Tasks), B (Letters/Numbers), C (Language/Comprehension), D (Visual/Spatial Discrimination), E (Social/Emotional/Adaptive); original version of the test was called The Gesell School Readiness Test.

Author

The Gesell Institute of Child Development.

Publisher

Gesell Institute of Child Development.

Cross References

See T5:1085 (4 references) and T4:1035 (13 references); for reviews by Robert H. Bradley and Everett Waters of an earlier version, see 9:438; see also T3:953 (6 references) and T2:1703 (4 references); for excerpted reviews by L. J. Borstelmann and Edith Meyer Taylor, see 7:750 (5 references).

Review of the Gesell Developmental Observation–Revised by THERESA GRAHAM LAUGHLIN, Adjunct Faculty, Nebraska Methodist College of Nursing, Omaha, NE:

DESCRIPTION

The Gesell Developmental Observation–Revised (herein referred to as GDO-R) is a “standardized, performance-based, criterion-referenced instrument” (technical report, p. 12) intended to assess five components of development among children ages 2.5 to 9 years. The five components or “strands” include Developmental, Letters/Numbers, Language/Comprehension, Visual/Spatial Discrimination, and Social Behavior, Emotional Development and Adaptive Skills. Each strand consists of different tasks. For example, Development includes 5 tasks: Cubes, Incomplete Man, Copy Forms, and Fine and Gross Motor. Other strands include more tasks. For example, Letters/Numbers consists of 9 tasks. Developmental, Letters/Numbers, Language/Comprehension, and Visual/Spatial Discrimination are assessed using the same 26 tasks included in the 2007 GDO. Emotional Development/Adaptive Skills is based on a newly created Parent/Guardian Questionnaire and Teacher Questionnaire. It is intended to evaluate children’s developmental and academic performance on a “developmental continuum” (examiner’s manual, p. 2). In addition, information is used to assist educators in their curriculum development and to identify children who may need academic or social intervention.

The GDO-R should be administered individually by a trained examiner and is estimated to take between 20 and 45 minutes, depending on the age of the child. Scoring is estimated to take between 15–20 minutes. It is recommended that examiners complete a 3-day workshop offered by the Gesell Institute to learn test administration and scoring protocols. Given the complexity of administration and scoring and reliance on examiner judgment, extensive training is necessary to ensure standard administration and interpretation. An Examiner’s Script is provided to ensure standard administration. In addition, for each task, the examiner’s manual provides information regarding the rationale, the materials needed, administration guidelines with decision trees, protocol for recording child behavior, and tips for scoring, including detailed scoring rubrics. The examiner’s manual provides guidelines regarding starting points. Decision Trees are utilized to determine end points. Three forms are included in the GDO-R: the Child Recording Form, the Parent/Guardian Questionnaire, and the Teacher Questionnaire.

The Child Recording Form is a comprehensive recording instrument that provides space for recording and coding child behavior. Because the GDO-R relies on examiner observation, the Child Recording

Form includes space for process and qualitative information. In addition, the Child Recording Form contains the Summary Profile, which is used to score each strand and to record Developmental Age and Overall Performance Level. For each task, rubrics are provided to determine the child's Developmental Age. Strand scores, based on the subset of tasks within each strand, are calculated using the Strand Scoring Worksheet provided in the examiner's manual. Scores are converted to Performance Level Ranges using cut scores. For each strand, children's performance is summarized in terms of the following three categories: Age Appropriate, Emerging, and Concern. Finally, Overall Performance Level is determined by the examiner's "collective impression of the child" (examiner's manual, p. 57) and should be consistent with the Performance Level Definitions provided in the examiner's manual. Overall Performance Level can be categorized as Age Appropriate, Emerging, and Concern. Although performance is dependent on child performance on task, significance weight is also placed on verbal and nonverbal behavior observed by the examiner.

The Parent/Guardian Questionnaire and the Teacher Questionnaire should be completed within 2 weeks of administration of the GDO-R and are intended to offer insight into different characteristics of the child, including social engagement and temperament. Three components are assessed in the questionnaires: social behavior, emotional development, and adaptive skills. The recording chart for the two questionnaires is included in the Child Recording Form. Guidelines for scoring are provided in the examiner's manual. Cut scores are used to interpret the questionnaires and are categorized as Age Appropriate, Emerging, or Concern.

DEVELOPMENT

The GDO-R is based on a number of developmental principles, most notably the ideas of Arnold Gesell and his maturational developmental theory. The *Mental Growth of the Pre-School Child* (Gesell, 1925) outlined developmental milestones in motor, adaptive, language, and personal social behavior that were used to develop the tasks used in the GDO-R. The examiner's manual summarizes Gesell's work, outlines typical behavior of children across the age groups included in the GDO-R, and provides age-related implications for administering the GDO-R. Although it is clear that the GDO-R is based on developmental principles, no information is provided regarding specific item development. In part, this is due to the fact that the items used in the GDO-R are the same as those that were developed for earlier versions of the test.

The Parent/Guardian and the Teacher Questionnaire are based on analysis of other unnamed questionnaires. According to the examiner's manual, the questionnaires reflect "current thinking on child behavior continuums, best practice in early childhood classrooms, and current cultural influences" (p. 57). However, no specific information is given regarding item development or analysis.

TECHNICAL

Scant technical information is found in the examiner's manual although there is a chapter titled "Standardization, Reliability, and Validity." A separate technical report on the GDO-R provides more in-depth information regarding the standardization study completed. Prior to the standardization study, an online survey was conducted to gather information from customers using the GDO. In general, respondents (N = 153) indicated that the GDO was time-consuming and not culturally sensitive. Feedback on the Child Recording Form led the test authors to make adjustments in the form prior to the GDO study. Respondents' comments on the scoring and rubrics led the test authors to establish strand scoring and update rubrics in the GDO-R. Finally, a review of the GDO by five experts in the early childhood field prompted changes in two items on the Interview task and the addition of the Teacher and Parent/Guardian Questionnaire.

The GDO study included 1,287 children ages 3–6 years, recruited from 53 schools (62% private; 38% public) in 23 U.S. states. Of the total participants, 45% were from private schools, and 55% were from public schools. No standardization information or data were gathered for children under age 3 or for children ages 7–9 years. The sample was divided into seven age bands, in 6-month increments. The actual number of children within each age band differed widely, ranging from 53 for 3-year-olds to 278 for 5-year-olds. Information regarding ethnicity and gender distribution across the age groups tested is provided. Ethnic distribution differed significantly across the age groups. For example, over 40% of the sample of 3-, 3.5-, and 4-year-olds were African American, whereas less than 6% of the sample of 5.5- and 6-year-olds were African American. No information regarding participant socioeconomic status is provided. Discrepant information regarding the training of the examiners who administered the GDO study is provided in the examiner's manual and the technical report. The examiner's manual indicates that test administrators either completed a 3-day GDO workshop or viewed a training DVD. The technical report makes no mention of the training video. Scoring of the GDO-R was completed by Gesell Institute National Lecture Staff. This discrepancy is important given the reliance on examiner judgment.

Reliability

Reliability was based on internal consistency coefficients for each task within the strands for each age band. Internal consistency coefficients are generally higher for tasks with more items and, for the most part, are within an acceptable range. Internal consistency coefficients for the strands were lower than the values reported for the tasks. For example, the internal consistency coefficients for the Visual/Spatial Discrimination Strand ranged from .01 for 6-year-olds to .54 for 5.5-year-olds. The internal consistency coefficients for the three components included in the Teacher Questionnaire were high. However, no data were provided for the Parent/Guardian Questionnaire.

Descriptive information regarding child performance and p-values across the various tasks was provided. Child performance increased, as expected, across the various age bands. Older children were more successful on tasks than younger children. Two tasks that are usually administered to children older than 6 years were not included in the study.

No test-retest evaluation was performed. However, interrater reliability was assessed on subsamples of 122 and 132 children for two tasks: The Incomplete Man and the Copy Forms items, respectively. Four individuals with experience administering the GDO produced high interrater reliability for assigning Developmental Age on these two tasks. No interrater reliability was provided for any of the other items.

Validity

Evidence of content-related validity was based on a comprehensive review of child development, discussions with experts in the field, and online surveys of past users of the GDO-R. Construct validity evidence was based on intercorrelations performed between tasks by age band. Moderate to small to no correlation was found between tasks. Moreover, it was clear that items within a strand were often unrelated. The test authors stated that the strands were based, then, on a theoretical framework rather than on correlations between tasks. No other indicators of validity were provided. Moreover, the GDO-R was not compared with other preschool assessments nor tied to any prediction of future behavior.

COMMENTARY AND SUMMARY

The GDO-R is intended to measure a wide array of significant cognitive, language, motor, and social-emotional components of development in children ages 2.5 to 9 years of age. The assessment relies on examiner observation of children performing a variety of tasks and from parent/guardian and teacher questionnaires. The cornerstone of this measure is its history in the tradition of Arnold Gesell and his followers.

Unfortunately, the utility of the current version of the GDO-R is compromised by a number of factors raised in the standardization study conducted. First, the study only includes children between the ages of 3–6 years of age. Without data for children ages 7–9 years of age, it is difficult to confirm reliability and validation for this age group. Moreover, clearly there is a relationship between socioeconomic status and the achievement of many of the milestones measured in the GDO-R. Without any specific data on SES and oversampling of children enrolled in Head Start, it is possible that some of the data gathered might be skewed, which in turn may adversely affect the cut scores generated and interpretation of data.

In addition, one of the new components to the GDO-R is the inclusion of the Teacher and Parent/Guardian questionnaires. Although descriptive data are presented from the Teacher

Questionnaire, no data are presented from the Parent/Guardian questionnaire. Although the test authors provide a variety of explanations for the lack of data, they offer no suggestions on how to improve return rates for future users of the GDO-R.

Finally, although many of the tasks included in the GDO-R are very useful in providing a picture of preschool development, it requires a well-skilled examiner to administer and score the test. It is unlikely that many teachers have the resources or time to devote 3 days for intensive training: a significant investment of time and energy to provide teachers information to inform their curriculum and classroom instruction. It is unlikely that the measure is useful at the classroom level. In sum, the current standardization study does not provide enough information to assess the validity and reliability and utility of the GDO-R for the entire age group for which this measure is intended.

REVIEWER'S REFERENCE

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.

Cite this review

Graham Laughlin, T. (2014). [Test review of Gesell Developmental Observation–Revised]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook*. Retrieved from <http://marketplace.unl.edu/buros/>

REVIEW 2 OF 2

Review of the Gesell Developmental Observation–Revised by TIMOTHY R. KONOLD, Professor of Research, Statistics, and Evaluation, and KATHAN SHUKLA, Doctoral Student in Research, Statistics, and Evaluation, University of Virginia, Charlottesville, VA:

DESCRIPTION

The Gesell Developmental Observation–Revised (GDO-R) assessment system is designed to provide an assessment of children's developmental progress for purposes of guiding instruction. This revised version of the GDO incorporates word revisions to some tasks, newly developed observation forms, and a technical manual.

The GDO-R consists of an individually administered direct assessment as well as informant-based parent and teacher observation forms. The direct assessment measures both developmental and academic (i.e., Letters/Numbers, Language/Comprehension, and Visual/Spatial Discrimination) domains through the measurement of 26 tasks, and the parent-teacher observation forms assess home and classroom Social/Emotional/Adaptive experiences of the child. Child level score reports are available in the form of summary profiles that reveal patterns of strengths and weaknesses across tasks, as well as an overall performance level rating.

The direct assessment is designed to be individually administered to children between 2 years 6 months to 9 years of age. Users of the GDO-R are required to complete a 3-day workshop on the assessment system that covers developmental theory, an overview of the GDO-R, and administration and scoring instructions. Administration is said to take approximately 20–45 minutes for trained examiners, and those with experience with the instrument can expect to complete the scoring profile form within 15–20 minutes. GDO-R materials include separate teacher and parent/guardian questionnaires, a child recording form, an examiner’s manual, an examiner’s script (i.e., standardized instructions for administration), manipulatives, and a technical report (available upon request) that describes the psychometric characteristics of the GDO-R.

DEVELOPMENT

Arnold Gesell’s maturational-developmental theory served as the primary framework for development of the GDO-R. Gesell’s early childhood theory outlines general stages of development, and provides the framework for reasonable growth expectations. Following Gesell’s work, general developmental characteristics that are typical of children in the areas of motor behavior, language behavior, personal-social behavior, and learning behavior are described in the examiner’s manual for 6-month periods that span the ages of 2 years 6 months to 9 years. These characteristics provide foundation for the test blueprint that guided the GDO-R item development.

TECHNICAL

Standardization

The GDO-R was administered to a sample of 1,287 children (55% from public schools) between 3 and 6 years of age by 101 examiners within 53 schools (38% public) that were located within 23 different states across the U.S. Although the GDO-R is advertised for use with children between 2 years 6 months through 9 years of age, the standardization sample did not include children between 7 and 9 years of age, or between 2 years 6 months to 3 years of age. The tested sample is described as being

composed of typically developing children from schools that normally administer the GDO. Ethnic diversity was present in the total sample of tested children. However, some racial groups appear to be underrepresented across age groups. At the same time, there were approximately equal numbers of boys and girls across the included age bands. All examiners had either participated in the required 3-day workshop necessary to become trained in the administration and scoring of the GDO-R, or were presented with a training DVD within 5 years of testing.

The standardization sample was employed to provide users with various descriptive statistics across the GDO-R tasks and strands. Separate tables are provided for 6-month increments ranging from 3 to 6 years of age. Summary tabled descriptive statistics include the number of items in each task or strand, the number of tested children, maximum number of obtainable points, means and standard deviations, and estimates of internal consistency and difficulty values. No data are provided for children 2–2.5 years of age or 7–9 years of age.

Reliability

Several investigations into the reliability of scores are presented in the form of internal consistency measures at both task level and strand level. A review of the technical report reveals that there is wide variability in these estimates across both tasks and strands. In many instances, internal consistency estimates are within reasonable expectations (i.e., $> .80$). At the same time, a number of age-level strand and task estimates are well below expectations (i.e., $< .50$), which the test authors attribute to fewer numbers of items. Users should be mindful of how these errors of measurement might impact interpretation of these measures.

Interrater reliability evidence is presented using correlation coefficients between developmental ages assigned by the pairs of raters on a subsample (approximately 10%) of children for two tasks (i.e., Incomplete Man and Copy Forms). Results suggest high interrater agreement (coefficients of .92 and .91, respectively) between developmental ages assigned by two raters. No information on interrater reliability is provided for other tasks on the GDO-R. In addition, estimates of test-retest reliability are not provided.

Validity

It is widely recognized that validity refers to the accumulation of evidence to support the interpretation of test scores in the context of their purpose. The primary form of validity evidence provided for the GDO-R is in the form of content validity, which was appropriately based on review of child development theory and consultations with educational experts for determining contemporary educational goals, and

knowledge and skills highlighted in early childhood curricula.

Overall, the intercorrelations between tasks of different strands is higher than within strand tasks. There are some moderate to high correlations (.30–.77), but many extremely low correlations are also found. For example, correlations between the Fine Motor task and all other tasks in its strand is less than .22 for all ages, with one exception (age 4 with Gross Motor, $r = .33$). Correlational estimates between tasks within strands are smaller in general at younger ages. For example, at age 3, the correlation between Visual I and Color Forms is .01, and between Visual I and 3-Hole Form Board is .16. The test authors attribute this outcome “to the relatively larger measurement error that typically occurs in data from very young children” (technical report, p. 50). However, even at age 6, correlations between Comprehension and all other tasks of its strand are less than .22.

COMMENTARY

Development of the GDO-R appears to be based on a strong theoretical foundation that served as a blueprint for its development. Further, much work appears to have gone into establishing a system of scoring that would be useful to the end user. At the same time, several psychometric characteristics of the instrument are less than desirable or unknown from the information provided in the GDO-R’s user’s manual and technical manual. Beyond a strong basis to support the content validity of the instrument, there is little evidence provided to support the GDO-R’s associations with external measures of current or future academic skills (i.e., concurrent and predictive validity), and no empirical evidence is provided to link the tasks with the strands that they purportedly measure. Users would benefit from factor analytic work to provide additional support for the construct validity of this instrument. For example, it is unclear why letter identification and writing are in the same strand with numbers but separated from language/comprehension. There are limited data for some age groups for which the GDO-R is intended (e.g., 3 years and 7–9 years of age). The need to complete a 3-day workshop prior to administering and interpreting the instrument may be considered a limitation to some users and may place the GDO-R at a disadvantage to other similar instruments that are more easily accessible to educators.

Given the technical and other limitations, this instrument at best may serve as supporting evidence for a child’s development, and more psychometric evidence is needed to support its use as a sole source for use in high stakes decision making.

SUMMARY

The GDO-R is a multidimensional assessment tool that allows for the measurement of children’s developmental, academic, and social-emotional-adaptive skills. It appears to be a carefully constructed, individually administered assessment tool that aligns with a respected theory of children’s development.

This revision incorporates some word changes to tasks, informant-based reports, and an optional technical report. It is designed to provide insight into strengths and weaknesses that may be useful for guiding future instruction, and to serve as a communication tool for teachers and parents. Users would benefit from additional psychometric studies that evaluate additional forms of reliability (e.g., test-retest) and validity (e.g., concurrent, predictive, dimensionality) to provide users with greater confidence that the scores resulting from use of this assessment with children are indeed related to their intended uses.

Cite this review

Konold, T. R., & Shukla, K. (2014). [Test review of Gesell Developmental Observation–Revised]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook*. Retrieved from <http://marketplace.unl.edu/buros/>