Outcomes Study

LEVEL OF EVIDENCE
Gold Standard

Teaching Strategies GOLD® Assessment System

Growth Norms Technical Summary

Summary Findings of a Study Conducted by
The Center for Educational Measurement and Evaluation
The University of North Carolina at Charlotte

Teaching Strategies GOLD® Assessment System

Growth Norms Technical Summary

Introduction

Independent researchers found that *Teaching Strategies GOLD*® is a valid and reliable measure of child growth and development. The assessment system was normed so that teachers can use the results to understand how the scores of a particular child compare with the scores of children in a large, nationally representative sample.

Initial research to establish the validity and reliability of *Teaching Strategies GOLD*® was conducted with data collected during the 2010–2011 academic year. To supplement those early findings and to establish growth norms, The Center for Educational Measurement and Evaluation (CEME) at The University of North Carolina at Charlotte conducted additional research. By that time, *Teaching Strategies GOLD*® had been widely

implemented for two full academic years. CEME used a longitudinal study design, tracking groups of children across a single academic year to observe their growth and development directly. The researchers had five goals:

1. further establish validity and reliability evidence for *Teaching Strategies GOLD®*, using the norm sample created for this study;
2. gather evidence that teacher ratings of child development made with *Teaching Strategies GOLD®* can be used to track children's growth over time;
3. determine normed scale scores for each of six developmental areas in *Teaching Strategies GOLD®*, thereby indicating how children of different ages should be expected to score at the beginning of the academic year;
4. create norm tables that show expected growth across the academic year for children of different ages; and
5. create norm tables that show expected *Teaching Strategies GOLD®* scale scores at kindergarten entry.

Growth Norms

Norming is the process of establishing growth trajectories for the children with whom an assessment instrument is used. This is accomplished by first identifying a large sample of children who are representative of the nation's population of similarly aged children and then determining their expected performance according to age. Doing so enables teachers and administrators to assume that the measure (e.g., *Teaching Strategies GOLD®*) will be used equally effectively with children from all parts of the country; children in all types of instructional settings; and children with different backgrounds, racial and ethnic identities, and special needs. Norming enables teachers to compare the assessment results for a particular child to results for children in the norm sample at the beginning of the academic year and at kindergarten entry. Norming also enables teachers to compare a particular child's growth *across* an academic year to the expected growth of children in the norm sample.

Teaching Strategies GOLD® Overview

Teaching Strategies GOLD® is an authentic observation-based assessment system for children from birth through kindergarten. The system may be implemented with any developmentally appropriate curriculum. It blends ongoing observational assessment for all areas of development and learning with performance tasks for selected predictors of school success in the areas of literacy and numeracy. The primary purposes of *Teaching Strategies GOLD*® are to document children's learning over time, inform instruction, and facilitate communication with families and other stakeholders.

Objectives for Development and Learning

Teaching Strategies GOLD® enables educators to focus on and measure the knowledge, skills, and behaviors most predictive of school success. The assessment system includes a total of 38 objectives that are grounded in research and aligned with the *Head Start Child Development and Early Learning Framework* and state early learning standards. Two objectives address English language acquisition, and the other 36 objectives are organized into nine areas of development and learning. These areas are Social–Emotional, Physical, Language, Cognitive, Literacy, Mathematics, Social Studies, Science and Technology, and the Arts.

Progressions of Development and Learning

Teaching Strategies GOLD® presents progressions of development and learning for objectives in the areas of social-emotional, physical, language, and cognitive development and the content areas of literacy, mathematics, and English-language acquisition. Indicators and clear examples enable teachers to rate children's knowledge, skills, and behaviors on a 10-point scale of "Not Yet" to Level 9. Furthermore, with the exception of those for English-language acquisition, the progressions use color-coded bands to show widely held expectations for the performances of children of various ages (birth to 1 year, 1 to 2 years, and 2 to 3 years) and in various classes/grades (preschool 3, pre-K 4, and kindergarten). At a glance, these colored bands show educators and families which skills and behaviors are typical for children of a particular age or class/grade. The bands help teachers manage the complexity of young children's development, which *Teaching Strategies GOLD*® recognizes as being uneven and rapidly changing.

Creating the Norm Sample

Sampling Procedures

The norm sample was taken from the entire population of children whose knowledge, skills, and behaviors were evaluated by teachers who used the *Teaching Strategies GOLD®* assessment system. Children were eligible for admission to the norm sample if their teachers met specific criteria designed to ensure that the ratings were made by users who understand how the measure is intended to assess children's knowledge, skills, and behaviors. The criteria required teachers to (a) use *Teaching Strategies GOLD®* for the entire 2011–2012 school year, (b) successfully complete an interrater reliability check, (c) successfully complete all required in-person *Teaching Strategies GOLD®* training sessions, and (d) successfully complete all *Teaching Strategies GOLD®* online training modules.

Secondly, children were selected for the norm sample only if complete data about their knowledge, skills, and behaviors had been collected across three assessment checkpoints. A window of acceptable assessment dates was set for each checkpoint (fall, winter, and spring) so that the resulting scores could be interpreted as representing a period of approximately 3 months of growth and development between checkpoints. Assessment data recorded during September, October, and November were accepted for the fall checkpoint. Data recorded during December, January, and February were accepted for the winter checkpoint. Finally, data recorded during March, April, May, and June were accepted for the spring checkpoint. These additional admissibility criteria of completeness and timeliness narrowed the candidates for the norm population to approximately 70,000 children.

However, this subpopulation did not represent the exact demographic distribution of children in the United States for race and ethnicity as measured by the 2010 census. Researchers therefore used random sampling to create a norm sample that reflected the U.S. Census Bureau estimates of racial and ethnic diversity of the national population of children ages birth through 5 years. This sampling technique was successful in creating a norm sample that very closely approximated the national population. For example, the percentage of African-American children (13.6 percent) and Hispanic children (25.5 percent) matched the Census Bureau statistics. Four of the remaining five racial and ethnic categories (White, Native-American, Pacific Islander, and Multiracial children) were all

within 0.6 percentage point of the national statistics. Only Asian children were slightly underrepresented (3.2 percent of the *Teaching Strategies GOLD*® norm sample but 4.5 percent of the national population).

Final Norm Sample

Children were included from all regions of the United States, including 34 states, the District of Columbia, and Puerto Rico. Boys comprised 51.3 percent of the norm sample, and girls comprised 48.7 percent. The primary languages spoken in the children's homes were English (77.5 percent), Spanish (17.7 percent), and other languages (4.8 percent). Children with an Individualized Education Program (IEP) or an Individualized Family Service Plan (IFSP) comprised 13.1 percent of the sample. The final *Teaching Strategies GOLD*® norm sample included 54,504 children.

Confirming the Validity and Reliability of *Teaching Strategies GOLD*®

To further confirm that *Teaching Strategies GOLD*® is a valid and reliable tool for measuring early childhood development and learning, researchers conducted new analyses with data for the current norm sample.

Construct Validity: Scale and Item Analysis

Construct validity refers to whether the assessment instrument measures the theoretical constructs (e.g., knowledge, skills, and behaviors) that it is intended to measure. Researchers used a method of analysis known as the Rasch Partial Credit Model to examine the properties of the scale and item scores for each of six developmental areas in *Teaching Strategies GOLD*®: social–emotional, physical, language, cognitive, literacy, and mathematics.

The first step was to determine whether each of the six scale scores measured one and only one developmental area (e.g., social–emotional but *not* language development). This is also referred to as *unidimensionality*. The analysis indicated that, for each of the six developmental areas, the scale scores are unidimensional. That means that they are distinct from one another and acceptably measure only one area in the overall assessment.

Next, researchers analyzed fit statistics, known as infit and outfit mean-square error. The goal of this analysis was to determine whether the specific ratings for each item in the assessment instrument were also unidimensional. If analyses showed that the fit statistics ranged from 0.6 to 1.4, then the specific item was considered to measure effectively only one of the six developmental areas. Every item on the social–emotional, physical, language, cognitive, and mathematics scales of *Teaching Strategies GOLD®* met these criteria. In the literacy scale, the criteria were met for all but one item. These analyses provide strong evidence that the process used to determine the scale and item scores of *Teaching Strategies GOLD®* was applicable to the current norm sample.

Person and Item Reliability

High person and/or item reliability means that there is a high probability of replicating the assessment results obtained by using the instrument. Specifically, person reliability estimates the likelihood of children's performing at the same skill levels across items measuring the same constructs of child development. Similarly, item reliability estimates the likelihood that the instrument's items would follow the same developmental progression if administered to another sample of children with similar skills, knowledge, and behaviors. Person and item reliabilities of .8 or higher are considered strong indicators of reliability. Across the six developmental scales of *Teaching Strategies GOLD®*, person reliabilities ranged from .90 to .97. Item reliabilities were .99 for all six scales.

Internal Consistency Reliability

Internal consistency reliability refers to the consistency of children's responses to all items in each scale of the assessment instrument. The more homogeneous the area measured (e.g., social–emotional, mathematics, etc.), the higher the internal consistency reliability should be. Researchers measured the internal consistency of the items within each of the six developmental scales of *Teaching Strategies GOLD®*. Analyses indicated that internal consistency reliability statistics ranged from .94 (physical) to .97 (language and cognitive), all of which were well above .8. They provide strong evidence of reliability.

Determining Expected Growth With *Teaching Strategies GOLD*®

Norming *Teaching Strategies GOLD*® ensures that the assessment instrument can be used to track the growth over time of children of different ages while also monitoring their readiness for kindergarten. Scale scores created from data for the final norm sample were used to identify expected growth trajectories.

Raw and Scale Scores

The norm sample was divided into 3-month age bands that were based on the ages of the children at the time of the fall assessment checkpoint. Children were placed into 18 age bands. The youngest band included children aged 3 through 5 months, and the oldest band included children aged 54 through 56 months.

Raw scores were derived by summing across the teacher ratings for all of the items in each scale of the assessment instrument. Each item was rated by teachers who used a 10-point scale. The ranges of possible raw scores were as follows: Social–Emotional (0–90), Physical (0–50), Language (0–80), Cognitive (0–100), Literacy (0–120), and Mathematics (0–70). Note that the total number of possible raw score points varied by developmental area because each scale included a different number of items.

Researchers further developed scale scores for each of the six developmental areas by using Item Response Theory, which is a common approach in educational and psychological assessment. Scale scores are generally considered more reliable and meaningful than raw scores. They are interval-level scores that result from the transformation of raw scores, and they are *not* dependent on the number of items rated for each area. Scale scores were calibrated to conform to a normal distribution with a mean of 500 and a standard deviation of 100 across the entire age range. The mean of 500 was associated with children at 36 months of age, which was sensible because that is the intended middle of the age range for *Teaching Strategies GOLD*®.

Expected Growth in Scale Scores

Analysis of the norm sample's scale scores revealed expected growth trends for children evaluated with *Teaching Strategies GOLD®*. As expected, children's *Teaching Strategies GOLD®* scores tended to spread out and have a greater range as they got older. This is supported by looking at the standard deviations. In this analysis, standard deviations for any given scale score tended to be approximately 50 points in each age band for children younger than 30 months at the fall checkpoint. However, standard deviations were approximately 60 points for children aged 30 months or older at the time of the fall checkpoint.

Kindergarten Entry Norms

Increasingly, kindergarten readiness is becoming an area of interest and concern for early childhood educators. Teachers are interested in knowing children's social–emotional skills as well as their literacy and mathematics competencies *before* starting kindergarten. *Teaching Strategies GOLD®* helps teachers thoroughly understand children's development and learning as children enter the kindergarten year. *Teaching Strategies GOLD®* objectives cover all areas of learning, including research-based predictors of school success. Results can inform teacher's approach to classroom instruction and facilitate children's success in kindergarten and the early elementary grades.

Researchers constructed a separate kindergarten norm sample. Consideration for admission to the sample was given to all children who were at least 60 months old at the time of the fall assessment and who were enrolled in a kindergarten classroom in which the teacher used *Teaching Strategies GOLD®*. This initial subgroup of the full norm sample did not match the U.S. Census Bureau statistics for racial and ethnic composition. Researchers therefore used random sampling to match the kindergarten norm sample to the national population as much as possible.

The proportions of White children (52.1 percent), African-American children (13.6 percent), and Hispanic children (25.5 percent) in the kindergarten norm sample were equivalent to those in the national population. Native American children were overrepresented in the kindergarten sample (2.4 percent in the kindergarten sample but 0.9 percent nationally). Hawaiian and Pacific Islander children comprised 0.3 percent of the norm sample and 0.2 percent nationally. Therefore, the final kindergarten norm sample was a relatively nationally representative sample with the exception of over representing Multiracial children and not including any Asian children, who comprise 4.5 percent of the national population.

Final Kindergarten Norm Sample

The final kindergarten norm sample included 4,155 children, of whom 48.2 percent were girls and 51.6 percent were boys. Children with an Individualized Education Program (IEP) comprised 9.4% of the norm sample. English is the primary language spoken in the homes of 68.2% of the children. Spanish is the primary language spoken in the homes of 19.0 percent of the children, and other languages are primary in the homes of 12.8 percent of the children.

Kindergarten Entry Scores

The normed kindergarten scale scores are in the low- to mid-600s for each developmental area. They are appropriate for an assessment instrument with a mean scale score of 500 at age 36 months. One would expect that the knowledge, skills, and behaviors of children who are about to enter kindergarten would be scored *above* the mean on the *Teaching Strategies GOLD®* because the assessment instrument is normed for children from 3 months to 56 months of age.

Conclusion

This summary report serves as a companion piece to the *Technical Manual for the Teaching Strategies GOLD® Assessment System* (Lambert, Kim, Taylor, & McGee, 2010) and the *Teaching Strategies Gold® Assessment System: Growth Norms Technical Report* (Lambert, 2012). The results of this study confirm the validity and reliability of the scale scores created for each developmental area in the assessment instrument. Most importantly, these results demonstrate that the *Teaching Strategies Gold®* raw and scale scores are sensitive to the process of child growth and development. Analyzing the normed scores can help teachers understand children's skills and relative areas of strength and weakness, monitor their expected growth, and compare them to a nationally representative sample.

References

Lambert, R. G. (2012). *Teaching Strategies GOLD® assessment system: Growth norms technical report*. Charlotte, NC: Center for Educational Measurement and Evaluation.

Lambert, R. G, Kim, D. H., Taylor, H., & McGee, J. (2010). *Technical manual for the Teaching Strategies GOLD™ assessment system*. Charlotte, NC: Center for Educational Measurement and Evaluation.

TeachingStrategies.com | 800.637.3652