Outcomes Study

LEVEL OF EVIDENCE
Gold Standard

Teaching Strategies GOLD® Assessment System

Technical Summary 2013

A Summary of Research Examining the Reliability and Validity of *Teaching Strategies GOLD®*

Teaching Strategies GOLD® Assessment System

Technical Summary 2013

Introduction

Selecting a Meaningful Assessment

When selecting an assessment to use with young children, the most important consideration is the *validity* and *reliability* of the measure. To ensure that *Teaching Strategies GOLD®* is both valid and reliable, the Center for Educational Measurement and Evaluation (CEME) conducted extensive research with thousands of children and teachers over the 2012/2013 school year. This document is a summary of the results obtained from that research.

Teaching Strategies GOLD® Overview

Teaching Strategies GOLD® is an authentic, observational 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. *Teaching Strategies GOLD*® is inclusive of children with disabilities, children who are English-language learners, and children who demonstrate competencies beyond typical developmental expectations.

Using *Teaching Strategies GOLD*®

The primary purpose of *Teaching Strategies GOLD*® is to document children's learning over time, inform instruction, identify at-risk children, and facilitate communication with families and stakeholders. It is important to remember that *Teaching Strategies GOLD*® is not intended as a screening or diagnostic measure, an achievement test, or a program evaluation tool. The information obtained should be used as just one part of a larger system of data collection for decision making.

Objectives for Development and Learning

Teaching Strategies GOLD® focuses educators' time by measuring the knowledge, skills, and behaviors most predictive of school success. The tool has a total of 38 objectives, including 2 objectives related specifically to English language acquisition. Thirty-six objectives are organized into nine areas of development and content-area learning:

- Social–Emotional
- Physical
- Language
- Cognitive
- Literacy
- Mathematics
- Science and Technology
- The Arts

The Norm Sample

Determining the Sample

When determining the validity and reliability of an early childhood assessment, it is extremely important to identify a large sample of children who are representative of the nation's population of similarly aged children. By doing so, teachers and administrators can assume that the assessment will be equally effective with children from all parts of the country, in all types of instructional settings, and with different backgrounds, races, ethnicities, and special needs.

The study behind the 2012/2013 technical report was based on a nationally representative norm sample of 18,000 children. It contained children from all 50 states, the District of Columbia, and Puerto Rico. The Center for Educational Measurement and Evaluation determined the norm sample from a total of 933,000 children who had scores available using *Teaching Strategies GOLD*® over the 2012/2013 school year. The norm sample contained 3,000 children in each of the six age or class/grade cohorts: birth to 1, 1 to 2, 2 to 3, 3 or preschool, 4 or prekindergarten, and kindergarten.

Final Sample

A total of 18,000 children were retained in the final sample used to evaluate the validity and reliability of *Teaching Strategies GOLD*®. This sample also provided enough data to provide norms for children birth to 1, 1 to 2, 2 to 3, preschool, prekindergarten, and kindergarten children. Overall, the final sample used in this research was large, broad, and highly representative of young children in the United States.

Scale

Teaching Strategies GOLD® items for the social–emotional, physical, language, cognitive, literacy, and math areas are measured on a 0 to 9 scale. These developmental progressions represent the knowledge, skills, and behaviors of children from birth through kindergarten and increase in difficulty as the scale increases. The results show strong statistical evidence that the items within each scale work very well together to measure a single underlying construct or domain of development. Additionally, the items within each scale yield information that fits the statistical model that was used to develop the scoring strategy that is used to create scale scores.

Use of *Teaching Strategies GOLD*®

Teaching Strategies GOLD® has more than 45,000 teachers completing assessment data each year. There is strong statistical evidence that teachers are able to use the rating scale to place children accurately along a continuum of growth and development. When the items within each domain of development are arranged from the easier objectives for children to master to the most difficult objectives for children to master, the hierarchy that is created matches very well with what developmental theory indicates.

The range of items, with varying degrees of difficulty, indicates that teachers understand the developmental trajectory that children will follow, and can use each area and the objectives and dimensions within *Teaching Strategies GOLD*® to help all children progress.

The statistical results indicate that the developmental progressions within *Teaching Strategies GOLD*® produce high reliability as measured by multiple indexes. These high reliability statistics were not only found for the overall norm sample, but extend to each age cohort: birth to 1, 1 to 2, 2 to 3, 3 or preschool, 4 or prekindergarten, and kindergarten.

Teaching Strategies GOLD® users have four opportunities per year to complete checkpoint data. High reliability statistics were found across fall, winter, and spring assessments for all scales and age or class/grades.

Norms

The norm sample was created using raw scores and summing across the teacher ratings for all the items within each scale of the assessment. Each item is rated using a 10-point scale, and the range of possible raw scores is as follows: social-emotional (0–90), physical (0–50), oral language (0–80), cognitive (0–100), literacy (0–120), and mathematics (0–70). Note that the total number of possible raw score points varies by developmental domain since each scale includes a different number of items.

Researchers further developed scale scores for each of the six developmental domains using a common approach in educational and psychological assessment known as Item Response Theory. 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, are *not* dependent on the number of items rated for each domain, and are 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 is associated with children at 36 months of age, which is sensible as it is the intended middle age range for *Teaching Strategies GOLD*®.

The new analyses yielded norm tables across all six areas of development and learning. Each norm table includes expected scores for children across 24 different 3-month age bands from 0–2 months through 69–71 months. The norm tables include expected scores for fall, winter, and spring assessments and expected growth from fall to spring. The norm tables clearly demonstrate that teachers can use the *Teaching Strategies GOLD*® assessment system to track growth across that academic year for children of different ages or classes/grades. Children in the 0–35 month age bands can be expected to make about 50–60 scale score points of growth across the academic year. Children in the 36–71 month age bands can be expected to make about 70–100 scale score points of growth across the academic year.

Conclusion

The *Teaching Strategies GOLD*® assessment system continues to yield highly valid and reliable results. Teachers' ability to use the rating scale to place children accurately along a continuum of growth and development ensures that *Teaching Strategies GOLD*® produces reliable data. These results further validate the magnitude of teacher effects when administering the assessment and for instructional planning, communicating with families, and monitoring overall child progress. Further evidence of the high reliability and validity of *Teaching Strategies GOLD*®, along with the updated norms, provides users with the assurance that their data are meaningful and should be used for helping all children progress.

TeachingStrategies.com | 800.637.3652