Performance Based Assessments to Identify Gifted and Talented Students

C. June Maker, Professor Emerita, University of Arizona, Tucson, AZ, USA

Ketty Sarouphim, Associate Professor of Psychology, Lebanese American University, Beirut, Lebanon

Randy Pease, Teacher Educator, University of Arizona, Tucson AZ, USA

Abdulnasser Alhusaini, Assistant Professor, Department of Special Education, University of Jeddah, Jeddah, Saudi Arabia

Performance based assessments often are overlooked as methods for identifying gifted and talented students because they are perceived as being less valid or reliable measures of ability. However, certain performance assessments have been found to be more reliable and valid measures of ability, especially for students from underserved and underrepresented groups such as children whose primary language is not English, those from cultures other than the dominant one, those with difficulties, and those from low-income groups (Maker, 2005; Romanoff, Algozzine, & Nielson, 2009; Sarouphim, 2001; 2002; 2004; 2009; Sarouphim & Maker, 2010).

In studies of **reliability**, Griffiths (1996) found that expert DISCOVER observers agreed more frequently and consistently (r =.92 to 1.00) than novice DISCOVER observers (.47 to .92). However, across experience levels, observers agreed 95% of the time on the highest rating of problem solving, the category used for placement in special programs (Griffiths, 1996). Kassymov (2000) found that agreement between DISCOVER personnel with varying levels of experience was 81% overall and 100% on the highest rating. Agreements between DISCOVER personnel and school district teams was 85%, and was 82% among the members of the district’s team.

In studies of **construct validity**, Sarouphim (2000, 2002, 2004) studied the psychometric properties of the DISCOVER assessment and the behavior checklist. She repeated the same study with the four different forms of the assessment: K-2, 3-5, 6-8, and 9-12, and these results are reported in three publications. The results were consistent across studies—written linguistic and oral linguistic activities have the highest relationship, but still seem to tap different aspects of linguistic ability; logical mathematical and spatial analytical ratings were related, and most other correlations ranged from .00 to .25 except one correlation between oral linguistic and math at grade four (.39). As predicted by the theory, the highest correlations were between the oral and written linguistic activities. She found expected significant correlations between the spatial analytical and math activities and unexpected significant correlations between the spatial analytical and oral linguistic activities in younger children. In general, she found low correlations among activities, showing that the abilities being assessed are indeed separate, as predicted by the theories on which the assessment is based. She found no gender bias in identification of gifted students. She also found that student behavior observed corresponded to abilities the activities were designed to measure. The behavior most frequently observed in students, regardless of the rating they received was “follows through to completion,” indicating that the activities are engaging to students of all age and ability levels (Sarouphim, 1997).

In studies of **concurrent validity**, Sarouphim (1999a) found that DISCOVER assessment results for spatial, logical-mathematical, and linguistic abilities were consistent with independent ratings made by both classroom teachers and a researcher who observed students in the classroom. Lori (1997) found a significant correlation between storytelling ability and personal intelligences in Bahraini children. Sarouphim (1999b) also studied the relationship between scores on the Raven in kindergarten, second, fourth, and fifth grade students. In this group of predominately Navajo and Mexican American students, she found the highest correlations with spatial artistic (r = .58, p<.01), spatial analytical (r = .39, p<.01) and math (r = .35, p.<01) as one would expect, since the Raven is a measure of non-verbal logical reasoning. Lower correlations appeared between Raven scores and oral (r = .20, ns) and written linguistic (r = .093, ns) DISCOVER ratings. In a study of 55 gifted private school Caucasian students (without correcting for restricted range of IQ) Stevens (2000) found significant correlations between DISCOVER spatial artistic (r = .373, p.<.01) and written linguistic (r = .34, p<.05) activities and Full-Scale IQ. He found significant correlations between spatial artistic (r = .27, p<.05) and written linguistic (r = .388, p<.01) activities and verbal IQ. The only significant correlation with performance IQ was the spatial artistic activity (r = .369, p<.01).

**Predictive validity** was evaluated in several studies. Sak and Maker (2003) found that the DISCOVER assessment, when administered in Kindergarten, was a significant predictor of achievement as measured by grades, state assessments, and Stanford 9 achievement when students were in grades 3 and 6. In a unique study of stability of scores, one researcher found that the results of the performance assessment were more stable across several years than the scores of a well-known IQ test and a test of creativity (Alhusaini & Maker, 2018). In another study of predictive validity, Romanoff (1999) examined a problem-solving assessment (PSA) containing three of the DISCOVER assessment activities. In her study, the PSA was used with students referred by their teachers as showing promise; the students referred, tested, and selected were compared with those referred, tested and not selected. Scores in reading and math, on North Carolina end of grade tests, averaged across grades 3, 4, and 5 were significantly higher for all students identified as gifted (M=84.03 for gifted and 58.37 for non-gifted). She also found that the differences between gifted African American and Caucasian students identified as gifted were not as great as those between non-gifted African American and Caucasian students. Most recently, when DISCOVER was compared with a well-known non-verbal test of intelligence, the Raven, the performance assessment administered at the beginning of grade 2 was a significant predictor of American Indian students’ achievement at the end of grade 3. The non-verbal test also was a significant predictor, but the performance assessment explained 43.9% of the variance in students’ achievement (*R2 = .439; F = 5,55; p = .000; N = 79*) while the non-verbal IQ test explained only 19% [*R2 = .195, F (1, 49) = 11.84, p = .001; N = 65*] of the variance (Tan and Maker, 2015; Erdimez and Maker, 2015).

The STEM assessments, developed in a grant from the National Science Foundation in cooperation with the Bio5 Institute, college of pharmacy, college of science, and superintendent of a school in the Navajo Nation, were compared with conventional methods used to select students for special programs (grade point average [GPA], teacher recommendation, and self-statements). Students identified by both methods were invited to participate in a special internship program in laboratories of scientists on the campus of a R1 university in the Southwestern USA. Existing methods limited the diversity of students identified as exceptionally talented and invited to participate in the program. Significant differences were found between students identified by the new methods and existing methods in GPA, ethnicity, and parent level of education. Ethnicity differences may have been due to the ethnic makeup of the partner schools where the assessments were conducted, but GPA differences and parent level of education cannot be attributed to the location of schools. The factor structure of the new assessments and their intercorrelations was evidence for their construct validity (Maker, 2017); other researchers and practitioners are encouraged to implement the new assessments and conduct studies on their usefulness and psychometric properties. Studies are ongoing.

The new DISCOVER 456 Assessments of young children, developed in cooperation with the Ministry of Education in the United Arab Emirates and International Namaa Association of Research and Development, have been field tested in both the USA and the UAE with approximately 1100 children of all three ages in public and private schools in both Arabic and English. Their reliability and validity have not yet been studied, but a grant proposal has been submitted for a long-term study in a public school in a low-income area in Tucson, AZ.

**References**

Alhusaini, A. & Maker, C. J. (2018). Who is gifted?The stability of scores on the DISCOVER assessment and the Raven Progressive Matrices in Diné gifted children. *Turkish Journal of Giftedness and Education. 8*(2), 114-142.

Erdimez, O. & Maker, C. J. (2015). *The Predictive Validity of the DISCOVER Performance-Based Assessment to Identify the Academic Achievement of Diné Students.* Unpublished manuscript, University of Arizona, Tucson, AZ, USA.

Griffiths, S. (1997). *The comparative validity of assessments based on different theories for the purpose of identifying gifted ethnic minority students*. Unpublished doctoral dissertation. University of Arizona.

Kassymov, A. (2000). *Researchers and implementers: Comparison of inter-rater reliability*. Unpublished manuscript. University of Arizona.

Lori, A. (1997). Storytelling and personal traits: Investigating the relationship between children’s storytelling ability and their interpersonal and intrapersonal traits. *Gifted Education International, 13*, 57- 66.

Maker, C. J. (2005). *The DISCOVER Project: Improving assessment and curriculum for diverse gifted learners*. Senior Scholars Series Monograph. Storrs, CT: National Research Center on the Gifted and Talented.

Maker, C. J. (2017). *Identifying exceptional talent in Science, Technology, Engineering, and Mathematics (STEM): Increasing diversity and assessing creative problem solving*. Manuscript in preparation, Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

Romanoff, B.S., Algozzine, B. & Nielson, A.B. (2009). Achievement of African American and Caucasian students referred and placed or not placed in gifted programs. *Journal for the Education of the Gifted*, *33*(2), 156-175.

Sak, U. & Maker, C. J. (2003). The Long-Term Predictive Validity of a Performance-Based Assessment Used to Identify Gifted CLD Students. *Proceedings of the 15th Biennial World Conference of the World Council for Gifted and Talented Students*. Adelaide, Australia: World Council for Gifted and Talented Students.

Sarouphim, K. (1999a). DISCOVER: A promising alternative assessment for the identification of gifted minorities. *Gifted Child Quarterly, 43*(4), 244-251.

Sarouphim, K. (1999b). Discovering multiple intelligences through a performance-based assessment: Consistency with independent ratings. *Exceptional Children, 65*(2), 151-161.

Sarouphim, K. M. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly*, *45*, 130-138.

Sarouphim, K. M. (2002). DISCOVER in high school: Identifying gifted Hispanic and Native American students. *The Journal of Secondary Gifted Education, 14*, 30-38.

Sarouphim, K. M. (2004). DISCOVER in middle school: Identifying gifted minority students. *The Journal of Secondary Gifted Education, 10*, 61-69.

Sarouphim, K. M. (2009). Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted Lebanese students: ls DISCOVER effective? *Journal for the Education of the Gifted, 33*(2), 275-295.

Sarouphim, K. & Maker, C.J. (2010). Ethnic and gender differences in identifying gifted students: A multicultural analysis. *International Education*, *39* (2), 42-56.

Tan, S. & Maker, C.J. (2015). *The Predictive Validity of the Raven Progressive Matrices for Identifying Academic Achievement of Diné Children.* Unpublished manuscript, University of Arizona, Tucson, AZ, USA.

For more information contact

junemaker@hotmail.com