Identifying Exceptional Talent in Science, Technology, Engineering, and Mathematics (STEM):
Increasing Diversity and Assessing Creative Problem Solving

C. June Maker, Ph.D., Litt.D.

The University of Arizona

Maker, C. J. (2020). Identifying exceptional talent in science, technology, engineering, and mathematics: Increasing diversity and assessing creative problem-solving. *Journal of Advanced Academics*. 31(3), 161–210. DOI: 10.1177/1932202X20918203

Author Note

C. June Maker, Professor Emerita, Department of Disability and Psychoeducational Studies, University of Arizona.

Correspondence concerning this manuscript should be addressed to C. June Maker, Department of Disability and Psychoeducational Studies, University of Arizona. E-mail: junemaker@hotmail.com

This research was supported by the National Science Foundation, Grant #1321190, Cultivating Diverse Talent in STEM; PI, Uwe Hilgert; Co-PIs C. June Maker, Frans Tax, and Martha Lindsey, University of Arizona; Harold Begay, Tuba City Public Schools

Abstract

In the Cultivating Diverse Talent in STEM Project, funded by the National Science Foundation in the USA, new assessments were developed, field tested, used to identify students with exceptional talent in STEM, and compared with existing methods (Grade Point Average [GPA], letters of recommendation, self-statements). Students identified by both methods participated in an internship program in laboratories of scientists on the campus of an R1 university in the Southwest. Existing methods limited the diversity of students identified. Significant differences were found between students identified by the new methods (M2) and existing methods (M1) in GPA, ethnicity, and parent level of education. Ethnicity differences may be due to the ethnic makeup of the partner schools, but differences in GPA and parent level of education cannot be attributed to the location of schools. Although GPAs of M1 students were significantly higher (3.71) than those of M2 students (3.07), the M2 students scored higher on all the performance assessments of creative problem solving and at similar levels on concept maps and mathematical problem solving. Studies on the usefulness and psychometric properties of the new assessments are needed with different groups and in different contexts.

Keywords: creative problem solving, performance-based assessments, concept maps, STEM, identification of exceptional talent

Identifying Exceptional Talent in Science, Technology, Engineering, and Mathematics

(STEM): Increasing Diversity and Valuing Creativity

In the past two decades, scientists, policy-makers, and the general public have been concerned with the state of Science, Technology, Engineering, and Math (STEM) education in the United States, recommending reforms to enhance the ability to produce a workforce that is literate in scientific, numeric, and technological areas; calling for a concerted effort to stop importing STEM talent from other countries and to develop a generation of STEM innovators (National Science Board [NSB], 2010). STEM innovators are defined as “individuals who have developed the expertise to become leading STEM professionals and perhaps the creators of significant breakthroughs or advances in scientific and technological understanding” (p. vii). The National Science Board members recommended policy changes in three main areas: (a) “provide opportunities for excellence”, (b) “cast a wide net to identify *all* types of talents and to nurture potential in *all* demographics of students”, and (c) “foster a supportive ecosystem” (NSB, 2010, pp. 2 & 3). Thus, assessments designed to identify future STEM innovators must include measures of varied abilities needed to be innovators, not just conventional, easy-to-administer tests, and they must be appropriate for use with students from all cultures, language groups, and socioeconomic levels. Identification of potential talent must be accompanied by opportunities and programs that are challenging and matched to the levels of ability and interests of the students.

The Importance of Developing New Assessments

When designing assessments to identify future STEM innovators, test developers must create methods to find not only those who can learn and develop expertise in STEM areas, but also those who have the potential to make significant breakthroughs or advances in

understanding. In an extensive review of research on talent identification and development, Subotnik and colleagues (Subotnik, Olszewski-Kubilius, & Worrell, 2011) concluded that the most important variables associated with outstanding achievements in both productivity and performance are “general and domain-specific ability, creativity, motivation and mindset, task commitment, passion, interest, opportunity, and chance.” (p. 4) Although many scholars would agree this is an important list of variables, less agreement exists about how to measure them. In the sections that follow, each of these variables and the methods for measuring them are reviewed.

General and Domain-Specific Ability. General abilities usually are measured by intelligence tests and overall Grade Point Average (GPA). These measures often are criticized because students of color (Miller, 2004; Plucker, Burroughs, & Song, 2010) and students from low socioeconomic status (SES) levels (National Science Board [NSB], 2010; Smith, Cech, Metz, Huntoon, & Moyer, 2014; Rojas, 2015) have lower average scores than students from middle and high SES levels. Domain-specific ability most often is measured by achievement tests in specific domains and by grades in specific subjects (e.g., math, English, life science, physics). Measurement of domain-specific abilities as a method of selection also has been criticized because of the achievement gap between students of color and White students and the achievement gap between students from low income groups and those from high and middle-income groups. Miller (2004), for instance, found that African Americans, Hispanics, and American Indians were “severely underrepresented” among the top 1%, 5%, and 10% on almost every measure of achievement: grades, GPA, class rank, and standardized test scores. Plucker and colleagues (2010) found similar results using National Assessment of Educational Progress (NAEP) and state achievement test results.

In recent years, an important paradigm shift has been occurring that has important implications for identification and nurturing of talent (Dai & Chen, 2013; Subotnik, Tai, Rickoff, & Alamarode, 2009). For many years, intelligence was viewed as being fixed and stable; but much research has shown that not only does IQ change in individuals (Dai & Chen, 2013; Sternberg, 2005), but it also changes in large populations, as evidenced by the “Flynn Effect:” massive IQ gains worldwide during the 20th century (Flynn, 1987, 2007; Neisser, 1998). These gains are not consistent across domains of ability, and in some later studies in different cultures, have reached a plateau or reversed (Teasdale & Owen, 2008). However, these studies taken together demonstrate that intelligence is not fixed or stable. An alternative view of intelligence, an emerging paradigm, is that of intelligence as developing expertise (Ericcson, Nandagopal, & Roring, 2005; Sternberg, 1999; Subotnik, et al., 2009). From this perspective, any test or assessment simply reflects the level of development of a particular form of expertise.

Another problem with tests of achievement and similar measures of domain-specific abilities is the ceiling effect for students at the highest levels, which has led to the use of out-of-level testing such as that employed in the Study of Mathematically Precocious Youth (SMPY) to select students for special opportunities through their talent search programs (Lubinski & Benbow, 2006). Another solution (Sternberg, 1999), consistent with the definition of STEM Innovators by the National Science Board (2010) and with the emerging paradigm, is to consider intelligence (general ability) and domain specific ability as *developing expertise within domains*. Intelligence and achievement tests measure a limited aspect of developing expertise, defined as “...the ongoing process of the acquisition and consolidation of a set of skills needed for a high level of mastery in one or more domains of life performance.” (Sternberg, p. 359) In studies comparing novices and experts across different domains, researchers (Bransford, Brown &

Cocking, 2000; Chi, Feltovich, & Glaser, 1981; Glaser & Chi, 1988; Dogusoy-Taylan & Cagiltay, 2014) have found that expertise in a domain increased sensitivity to patterns of meaningful information and experts' knowledge was organized around core concepts, which helped them establish meaningful relationships between concepts. From this perspective, rather than simply testing the amount of knowledge possessed by students, one must ask them to show the relationships they perceive among concepts. A method often used in science education is concept mapping, developed by Novak (Novak & Gowin, 1984) and further extended by Ruiz-Primo and colleagues (2007; Ruiz-Primo & Shavelson, 1996; Ruiz-Primo, Shavelson, Li, & Schultz, 2001).

When concept mapping is used, students are given a list of concepts related to a topic and are asked to “map” them, showing which ones are connected, which ones are at higher levels than others, and which ones are connected to others in different sections of a map. Unlike “mind-mapping”, students must write words on the lines connecting different concepts and they must make a hierarchical map. The words on the lines explain the relationship between two concepts and the hierarchies identified by students show how they have organized the concepts. This approach to assessment of domain-specific knowledge has several advantages: (a) assessing the students' levels of development of expertise in a domain, (b) having no ceiling effect, (c) assessing higher levels of thinking rather than memory and comprehension (Austin & Shore, 1995; Erdimez, Tan & Zimmerman, 2017; İngeç, 2009; Vanides, Yin, Tomita, & Ruiz-Primo, 2005), and (d) offering an alternative to the usual multiple-choice standardized tests or grades for assessing domain-specific ability and understanding (Tan, Erdimez, & Zimmerman, 2017; Zimmerman, Maker, Gomez- Arizaga, & Pease, 2011; Maker & Zimmerman, 2019).

Creativity. Assessment of creativity, another variable associated with outstanding achievements (Subotnik, et al., 2011) is even more controversial than assessment of domain-general and domain-specific ability, with some researchers favoring the use of self-report inventories (c.f., An & Runco, 2016), others focusing on divergent thinking tests (c.f., Torrance, 2008), others using integrative approaches (c.f. Jellen & Urban, 1989; Urban & Jellen, 1996) and others favoring inclusion of both divergent-exploratory and convergent-integrative creative abilities (c.f., Barbot, Besancon, & Lubart, 2016; Lubart, Zenasi, & Barbot, 2013). Criticisms of creativity assessment are related mainly to the scoring procedures: (a) a strong relationship has been found between uniqueness and fluency (Silvia, 2015; Torrance, 2008); (b) uniqueness scores are strongly sample-dependent (Silvia, 2015); and (c) divergent thinking tasks resemble traditional verbal fluency tasks too closely (Nusbaum & Silvia, 2011). A different approach, the Consensual Assessment Technique (CAT), was developed by Amabile (1982). Judges provide subjective ratings of the creativity of each response, and if multiple judges are employed, the method has been shown to be a reliable and valid method for assessing uniqueness without being confounded with fluency scores (Silvia, Martin, & Nusbaum, 2009; Silvia, Winterstein, Willse, Barona, Cram, Hess, Martinez, & Richard, 2008). In addition to these different approaches to creativity assessment and the criticisms of the scoring procedures, creativity assessment research includes two long-standing debates: the relationship between creativity and intelligence (c.f., Getzels & Jackson, 1958; Runco, 2007; Runco & Albert, 1986; Torrance, 1977; Wallach, 1971) and the domain-specificity versus domain-generality of creativity (c.f., An & Runco, 2016; Baer, 1998, 2015; Barbot, et al., 2016; Hong & Milgram, 2010; Plucker, 1999). An extensive discussion of these debates is beyond the scope of this paper, but the main aspects of the debates are reviewed.

Creativity and intelligence. Out of the debate about whether creativity was the same as or different from intelligence and the relationships between the two constructs (c.f., Kaufman, Plucker, & Baer, 2008; Runco, 2007; Sawyer, 2006; Weisberg, 2006) came the belief that a threshold exists: intelligence and creativity are related only up to a certain IQ level, usually believed to be 120 (Getzels & Jackson, 1962; Guilford & Christensen, 1973; Simonton, 1994). However, studies of the threshold theory have not yielded consistent results, nor are they viewed as being conclusive (c.f., Runco & Albert, 1986), which may result from the use of different measures of both constructs, gender, age, and SES in the samples. These inconsistencies prompted Kim (2005) to conduct a meta-analysis of the results of several studies. She found that, in 21 studies including 45,880 participants, the average correlation across different measures of both creativity and intelligence was small ($r = .174$), but differed significantly according to age and different creativity tests. In elementary students, the correlation was .086; for the middle school group, .210; for the high school group, .261; and for adults, .205.

Going beyond the debate about the distinction between creativity and intelligence and the threshold theory, and building on the research of Kim (2005), Karwowski and colleagues (Karwowski, Dul, Gralewski, Jauk, Jankowska, Gajda...& Benedek, 2016) report results of eight studies involving 12,225 participants, in which they concluded that intelligence is a *necessary but not sufficient* condition for creative behaviors and accomplishments: different aspects of creativity related to intelligence differently, and very few participants in the studies with high creativity and achievement had low intelligence.

In Silvia's (2015) review of the history of thought on how intelligence and creativity are related, he suggests that statistical advances (e.g., Silvia, 2008; Hong & Milgram, 2010); advances in assessment of creativity, especially uniqueness scoring; emphasis on executive

processes (Gilhooly, Fioratou, Anthony, & Wynn, 2007; Nusbaum & Silvia, 2011); and the emerging neuroscience of creativity (Beaty, 2015; Beaty, Benedek, Wilkins, Jauk, Fink, Silvia, Hodges, Koschutnig, & Neubauer, 2014) have provided support for the view that intelligence and creativity are strongly linked. Silvia and others (Nusbaum & Sylvia; Weisberg, 2015) do not suggest that these constructs are the same, but that they are more closely related than many theorists and researchers previously have thought. By studying the deeper connections between intelligence and creativity and viewing intelligence and creativity as "...families of processes and functions that the mind can do..." (Silvia, 2015, p. 604), researchers and practitioners can see a larger conceptual picture: "The mind can do a lot of things, and the processes that it uses for solving closed-ended problems are similar to those used for making, judging, and playing with ideas." (Silvia, 2015, p. 604)

Consistent with this perspective, Sak and Maker (2006), in a study of mathematical problem solving using the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) assessment, concluded that in elementary age students two standard deviations above the mean of domain-specific knowledge was a threshold for a one standard deviation above the mean of creativity in math. The DISCOVER math assessment (Bahar & Maker, 2011; Bahar & Maker, 2020; Maker, 1993, 2005), created in 1992 from a perspective similar to that of Silvia (2015), includes solving known problems with known operations and correct answers, accuracy in the solving of other types of problems (domain-specific knowledge and skills), and solving problems with multiple correct answers by applying mathematical concepts and operations in many and novel ways (domain-specific and domain-general creative processes) (e.g., write as many problems as you can with the answer of 10). The

DISCOVER assessment framework is described in more depth in a later section of this article and the math assessment in more depth in another publication (Bahar & Maker, 2020).

Domain-Specificity versus domain-generality of creativity. Although the literature contains many debates over the years, recent consensus (c. f., Amabile, 1996, 2013; An & Runco, 2016; Barbot, et al., 2016; Barbot & Tinio, 2015; Hong & Milgram, 2010; Plucker & Zabelina, 2009) seems to be that (a) creativity has both domain-general and domain-specific components, (b) these two components of creativity are related, and (c) task characteristics and motivation play a part in the relationship between the two components. Amabile (1996) initially proposed that creativity consists of domain-specific knowledge and skills, creativity-relevant processes, and task motivation; but later acknowledged the importance of both positive and negative effects of environment on the development and expression of creativity (2013). Teasing out these relationships has become the focus of many researchers rather than debating whether creativity is specific to domains or general across domains (Barbot, et al., 2016; Lubart, et al., 2013). For example, Lubart and colleagues pointed out the importance of task demands, describing each individual as having a unique profile of resources, and the degree to which one is creative depends on the fit between her or his profile of cognitive (i.e., divergent thinking, analytic thinking, mental flexibility, associative thinking, selective combination) and conative (i.e., tolerance of ambiguity, risk taking, openness, intuitive thinking, motivation to create) resources and the demands of the task. Creative potential is the confluence of these distinct, but interrelated, resources and involves more than a simple sum of each person's level, mainly due to their interactive effects, with the environment serving as a catalyst for person-centered resources to be activated. Evaluation of creative potential can be production-based (focused on the product) or resource-based (focused on the cognitive and conative resources of the individual (Lubart, et

al., 2013). This definition is consistent with the varied views of other researchers and comprehensive in that it pulls together significant research in the field.

Motivation and mindset. Renzulli (1978) was perhaps the first scholar in the field of education for the gifted to highlight the importance of a combination of traits as necessary for exceptional talent to be evident. His three-ring conception included above average ability, creativity, and task commitment, three of the variables identified by Subotnik and colleagues (2011). He defined task commitment as "...a refined or focused form of motivation...." (Renzulli, 1986, p. 69) This form of motivation is important in achievement in both elementary school (Curby, Rudasill, Rimm-Kaufman, & Konold, 2008) and in math and science (Benbow & Arjand, 1990; Collins, 2018). Other support comes from the research on "deliberate practice" (Daley, 1999; Eriksson, 1996; 2015; Ericsson, Krampe, & Tesch-Romer, 1993; Eriksson, Nandagopal, & Roring, 2005). Without a high level of task commitment, people will not invest the time (10,000 hours as proposed by Ericsson) and energy needed to become experts or to master the skills needed for high-level productivity or performance. They must believe they *control* the development of their talents and abilities through *effort* on their part (Deschamps & Beauvois, 1994; Brown & Weiner, 1984; Ericsson, Prietula, & Cokely, 2007; Maker, 1978; Siegle, 2013; Siegle & McCoach, 2005; Siegle, McCoach, & Roberts, 2017; Weiner, 1986; Whitmore & Maker, 1985) because talents and abilities are dynamic and malleable (Dweck, 2012). Those interested in becoming experts or mastering certain skills also must believe they are capable of overcoming obstacles, learning from mistakes, and recovering from setbacks: "growth mindset" as described by Dweck (2006).

From another perspective, Feist (2006a) found that members of the National Academy of Science knew they wanted to be scientists at an early age: 25% by age 14, 50% by age 18, and

75% by age 20. In his study of Westinghouse Talent Search finalists, he found "...the only precocity variable that reached the .05 level of significance with lifetime productivity was age that one first conducted formal research." (p. 70) Results of research on motivation and practice, when combined, demonstrate their importance in high-level achievement. One could also argue that the motivation needed for long-term task commitment comes from an individual's passion for certain activities, another trait Subotnik and colleagues (2011) found to be important to outstanding productivity and performance. For gifted students, interest is highly correlated with performance in a variety of talent domains (Siegle, Rubenstein, Pollard, & Romey, 2010).

Opportunities. One cannot separate opportunities from motivation. Motivation, or the desire and willingness to do something, cannot develop or be expressed without opportunities in the areas of interest. Those opportunities must be appropriately challenging (Gallagher, Harradine, & Coleman, 1997; Maker & Schiever, 2010; Rimm, Siegle, & Davis, 2018) and in students' "zone of proximal development" (Vygotsky, 1978); in other words, matched to the abilities and interests of the students (Pease, Vuke, Maker, & Muammar, 2019; Wang & Eccles, 2013). To spend the 10,000 hours of practice proposed as necessary for the development of expertise, teachers, mentors, programs, and other opportunities for this practice must be provided, accompanied by specific feedback to improve performance (Hattie, 2009). In a recent study related to this discussion, Wai, Lubinski, Benbow, and Steiger (2010) used the concept of "educational dose", consisting of both acceleration and enrichment opportunities in STEM, to study its effects on accomplishments in STEM. In one study, they followed a large sample of students who were identified in the Study of Mathematically Precocious Youth (SMPY) talent search to determine their levels of achievement and productivity 25 years later. In a second study, they included participants in graduate programs in STEM. The combined results of these

studies led to the conclusion that “The high-dose groups consistently earned a greater proportion of the STEM outcomes [STEM PhD, STEM publications, STEM tenure, STEM patent, STEM occupation] than the low-dose groups, with the exception of STEM patents for the total sample and STEM patents and publications for the graduate student women.” (Wai, et al., p.869)

Integrating Intelligence, Creativity, Motivation, and Opportunities through Problem Solving

In 1993, Maker concluded that educators and psychologists have made artificial distinctions between the constructs of intelligence and creativity because of the tasks included in the tests used to measure them. This perspective is consistent with the recent conclusions of researchers: (a) intelligence and creativity are not as different as once assumed (Karwowski, et al., 2016; Nusbaum & Silvia, 2011; Silvia, 2015), (b) motivation and mindset are important individual differences (Dweck, 2012; Ericsson, et al., 2007; Renzulli, 1978), and (c) individuals have a unique profile of resources and the degree to which one is creative depends on the fit between her or his profile of resources and the demands of the task (Lubart, et al., 2013).

Using the framework proposed by Getzels and Csikszentmihalyi (1967; 1976), who described problem situations (tasks) as being on a continuum of open-endedness depending on how much information was given to the problem solver, how much was known (specified) by the presenter of the problem, and how much was unknown (unspecified) to either or both (Table 1), Maker analyzed tests of intelligence, achievement, and creativity. She found that tests of intelligence and achievement included mostly “closed” problems, which were well-defined with known methods and right answers (Types I and II in Getzels and Csikszentmihalyi’s continuum). At the other end of the continuum, tests of creativity included open-ended problems, problems that were either well-defined or not defined with multiple appropriate methods and many

possible solutions. No intelligence, achievement, or creativity test was identified that included the most open-ended type (Type III in Getzels and Csikszentmihalyi's continuum and Type VI in Maker and Schiever's [1993, 2010] modified continuum), the type Getzels and Csikszentmihalyi (1967; 1976) found to distinguish creative artists and scientists from those who were not as creative.

Sternberg (1999) also emphasized the solving of varied types of problems as essential in his theory of intelligence as developing expertise. He suggested that expertise includes certain metacomponents of thinking: "recognition of problems, definition of problems, formulation of strategies to solve problems, representation of information, allocation of resources, and monitoring and evaluation of problem solutions." (p. 361) In his view, problem solving needs to be defined more broadly than it is usually defined in intelligence and achievement tests; it needs to include the solving of open-ended problems as well as monitoring the solutions as they are applied in practical situations.

In science education, as a result of extensive experience and research with the Westinghouse Science Talent Search (now called the Intel Talent Search), Brandwein (1992, 1995) distinguished *general giftedness* from *science proneness* and science proneness from *science talent*. Giftedness (measured by test of general ability) was recognized as high achievement in the general scholastic program. Science *prone* students "displayed a keen and driving interest, as well as high achievement, in science generally, and in addition, demonstrated skill in the laid-out laboratory, then in guided discovery aided by clues (problem-doing), and then in solving the more sophisticated puzzles that require deferred solution." (1992, p. 122) Using the problem continuum (Table 1), these would be the closed and semi-open problems. Science *talent* was defined as "...ability to plan and complete an investigation involving a

problem without a known solution (attested by a scientist in the particular field of study)”. (1992, p. 123) In the problem continuum, these would be the open-ended problems. More recently, Feist (2006a) found that the age at which one first conducted original research was the only precocity variable that predicted lifetime productivity. STEM innovators would be those with science *talent*, as evidenced by their ability and interest in “problem finding”, not just science *proneness*, as evidenced by their interest in solving well-defined problems.

Using a problem solving focus and a framework initially proposed by Getzels and Csikszentmihalyi (1967; 1976) and modified by Maker and Schiever (Maker, 1993, Maker, Alhusaini, Pease, Zimmerman, & Alamiri, 2015; Maker & Schiever, 2010) enables integration of the variables involved in the expression of talent (Subotnik, et al., 2011) by focusing on each student’s ability to solve various types of problems in STEM domains rather than on labeling a student as intelligent, creative, or even “exceptionally talented”. The assessments designed through this project can help to identify which types of problems students are already capable of solving, which types they are developing the expertise to solve, which ones are challenging for them, and how these abilities and challenges are different in different domains (Pease, et al., 2019). Figure 1 shows how these components are related. Through these assessments one can observe (a) domain-specific knowledge and abilities and domain-general abilities through students’ performance in solving closed problems (Types I and II), (b) domain-specific and domain general creativity through students’ performance when solving open-ended problems (Types V and VI), and (c) the interplay between domain-specific and domain-general knowledge and abilities and domain general and domain specific creativity through the solving of semi-open problems, and (d) task motivation through observing performance in all types of problems across

domains (Figure 1). To determine opportunities, an examination of family, school, community, and economic factors is essential.

Cultural, Ethnic, Linguistic, and Economic Diversity

Another policy change recommended by the National Science Board (NSB; 2010) is to identify all types of talent and nurture potential in all demographics of students. Conventional methods for identifying exceptional talent severely limit the diversity of students recognized as exceptionally talented and served in special programs for gifted students. As reflected in Office for Civil Rights data and State Department of Education statistics for the state in which the project was located, in the year 2006, White students made up 47.04 % of the overall student population and 64.65 % of the students served in programs for exceptionally talented students. African American students made up 5.13% of the student population and 2.88% in programs for exceptionally talented students. Hispanic students made up 39.67% of the student population and 23.21% of students in programs for exceptionally talented students. American Indian students made up 5.64% of the student population and 3.23% of students in programs for exceptionally talented students (Snyder & Dillow, 2012; AZED, 2006). Nationally, the picture is better, but certain groups remain overrepresented while others remain underrepresented: White, 57.7% of the school-aged population and 67.7% of programs for talented students; Black, 14.8% of the school population and 9.1% of students in programs for talented students; Hispanic, 19.7% and 12.8%; American Indian/Alaskan Native 0.9% and 0.9% (Snyder & Dillow).

Examination of these statistics shows that American Indian students are represented proportionally across the country, but not in the state in which this project was located. Perhaps the reason is that the percentages in the state are higher in general, and students are mostly on reservations where fewer programs exist and the income level is lower than the state and national

average. One cannot ignore the fact that economic differences are important determiners of academic opportunities (Miller & Kimmel, 2012; National Science Board, 2010), and may be more important than cultural differences for the American Indian students across the country, especially those on reservations. Of all ethnic groups, American Indians have the highest rate of poverty, 25.9% (Rogers, 2009; Smith, et al., 2014). Students from families in high and middle socioeconomic status (SES) groups are more likely to have opportunities to discover their talents, develop their interests, and participate in special programs, thus developing their skills and confidence in STEM areas (Archer, DeWitt, Osborne, Dillon, Willis, & Wong, 2012).

Some of the reasons for the lack of diversity in special programs for exceptionally talented students are (a) heavy reliance on standardized quantitative measures of achievement and aptitude (Joseph & Ford, 2006; Miller, 2004; Smith, et al., 2014; Plucker, et al., 2010), (b) using only one method of assessment (Ivcevic & Kaufman, 2013; Joseph & Ford, 2006; Kaufman, Baer, & Gentile, 2004; Ortiz, 2002; Silvia, 2015; Sternberg, 2010; Stemler, Grigoreko, Jarvin, & Sternberg, 2006) (c) using conventional assessment methods rather than authentic, performance-based assessments (Ford, 1998; 2006; Joseph & Ford, 2006; Luria, O'Brien, & Kaufman, 2016; Van Tassel-Baska, 2002), (d) using limiting definitions of giftedness (Dai & Chen, 2013; Maker, 2005; Sternberg, 1999; Subotnik, Tai, Rickoff, & Alamarode, 2009), and (e) using identification procedures heavily influenced by teacher perceptions such as grades and teacher recommendations (Clasen, Middleton, & Connell, 1994; Maker, 1996; McCoach & Siegle, 2003).

Often, in attempts to increase diversity, selection committees or individuals use the SAT, grades, and teacher recommendations, but apply weighting criteria or use other methods to select students from underrepresented groups who have lower scores and grades than White male

students (Smith & Garrison, 2005). However, this practice has been the subject of a number of lawsuits in which White students have claimed that students of color with lower scores were given discriminatory preference (Margulies, 2002; Olivas, 1999). Indeed, Maker (1996) argued that not only is this practice discriminatory to the White students, but it also sends a message to students of color that they are somehow inferior.

Fortunately, promising practices exist and have been used successfully to achieve a more equitable balance of students from diverse ethnic and economic groups in special programs without lowering cutoff scores or applying weighting formulas to equalize scores (Glover, 1976; Ivcevic & Kaufman, 2013; Kaufman, et al., 2004; Kaufman, 2006; Luria, et al., 2016; Sternberg, 2006; Stemler, et al., 2006; Torrance, 1971; 1973; Troiano & Bracken, 1983). A number of researchers have found that, unlike measures of ability and achievement, scores of students from diverse cultural groups and low income groups on various measures of creativity are not significantly different from those of high income and mainstream cultural groups (Glover, 1976; Ivcevic & Kaufman, 2013; Kaufman, 2006; Kaufman, et al., 2004; Torrance, 1971). In fact, researchers have found that cognitive flexibility, open-mindedness, and imagination, traits important in creativity, are higher in African American students than in White students (Jenkins, 2005), that bilingual students tend to demonstrate higher creativity than monolingual students (Ghonsooly & Showqui, 2012; Kharkhurin, 2012), and that American Indian students are more likely to solve problems creatively as a group and to generate creative stories (DeVries & ShiresGolon, 2011).

In addition to the fact that measures of creativity do not seem to have the same biases as measures of intelligence, inclusion of creativity and/or creative problem solving in definitions and methods used to select students for special programs has the additional advantage of being

more aligned with the research on characteristics needed for talent development and innovation (Luria, et al., 2016; Maker, 1993, 1996, 2005; Sternberg, 2010; Sternberg & Coffin, 2010). In fact, Sternberg and his colleagues (Sternberg, 2010; Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006) found that measures based on his theories (including analytical, creative, and practical abilities; intelligence as developing expertise) predict college success more accurately than conventional admissions tests while significantly reducing differences by ethnicity. For example, in the Kaleidoscope Project, applications to university programs from African American students and Hispanic American students increased significantly, admission of African American students increased by 30%, and admission of Hispanic American students increased by 15% (Sternberg, 2010).

This approach to identification has characterized the DISCOVER Projects since their beginnings in 1992, with components to assess both domain-specific knowledge and abilities in spatial, linguistic, and mathematical domains and creative problem solving within and across domains (Maker, 1992, 1993, 1996, 2001, 2005; Sarouphim & Maker, 2010). Across different grade levels, different forms of the assessment, different ethnic and economic groups, and at different times, no significant differences have been found in the percentages of students from different economic, cultural, and gender groups identified using the DISCOVER assessments (Maker, 2005; Nielson, 1994; Sarouphim, 2001, 2002; Sarouphim & Maker, 2010). In one school district, for instance, the DISCOVER assessments were used in designated schools in low-income, high-minority schools because of the problem of underrepresentation of students of color. Using conventional methods (teacher nomination and IQ tests), prior to the use of DISCOVER, in a school district with an overall population of 58% Hispanic students, only 26% of the students served in programs for gifted students were Hispanic; although only 37% of

students in the schools were White, the percentage of students served in programs for the gifted who were White was 72%; and for African American students, 3% in the school population and 1.3% served in programs for the gifted. The balance changed in the targeted schools to more equitable percentages after DISCOVER was incorporated into the identification process: Hispanic, 73.5% in the eight schools and 71.5% served in special programs in the schools; White, 21% in the schools and 26.5% in the special programs; African American, 4.2% in the schools and 2.0% in special programs. In a school district with a high percentage of African American students, the balance was changed to reflect the ethnic balance in the school district when an assessment modified from DISCOVER was used as an identification tool (Reid, Udall, Romanoff, & Algozzine, 1999; Romanoff, Algozzine, & Nielson, 2009). The conceptual framework for these and the new assessments that were created using the same framework are presented in the following sections.

Conceptual Framework: Integrating Intelligence, Creativity, Problem Solving, and Motivation

The first, and most important aspect of the conceptual framework for creating the assessments was a definition of ability that could serve as the unifying theme for all assessments. Gardner (1983, p.60) provided an excellent theme as he defined intelligence as “...a set of skills of problem solving enabling the individual to resolve genuine problems or difficulties...and ... the potential for finding or creating problems—thereby laying the ground work for the acquisition of new knowledge.” This definition is consistent with the definition of STEM innovators (NSB, 2010), as it includes a clear relationship to the kinds of problems STEM innovators face, the need for a knowledge base, problem solving as well as problem finding, and the potential for creative breakthroughs.

Defining Exceptional Talent and Creating Instruments

Using a conceptual framework based on Gardner's (1983) and Sternberg's (1999) theories of intelligence and analysis of tests of creativity and intelligence, Maker (1993) proposed a different definition of giftedness or exceptional talent: "...the key element in giftedness or intellectual competence is the ability to solve the most complex problems in the most efficient, effective, or economical ways" and in addition, "...are capable of solving simple problems in the most efficient, effective, or economical ways." (p. 71) Later, teams of observers added "elegant" (i.e., pleasingly ingenious and simple) to the list of adjectives describing solutions to problems as an alternative to more common words such as novel, creative, or unique. Using this framework, Maker and her colleagues (Maker, 1993, 1996, 2005), with funding from various federal sources, developed a performance-based assessment, Discovering Intellectual Strengths and Capabilities through Observing Varied Ethnic Responses (DISCOVER). Over a period of more than 25 years, Maker and her colleagues found that the assessments based on this framework not only resulted in recognition of students with creative problem-solving ability, but also changed the balance of underrepresented students in programs for exceptionally talented students in ways that reflected the ethnicity distribution of students in their communities (Maker, 2005; Nielson, 1994; Sarouphim, 2001, 2002; Sarouphim & Maker, 2010).

Not only were students from diverse and traditionally underrepresented groups identified through DISCOVER shown to be exceptionally talented, but they were successful in programs for gifted students or in regular education programs (Reid, et al., 1999; Romanoff, et al., 2009; Sak & Maker, 2003; Erdimez & Maker, 2015). For example, Erdimez and Maker (2015) found that the combined scores on the components of the DISCOVER assessment administered at the beginning of grade 3 explained 43.9% of the variance in American Indian students' overall

achievement at the end of grade 4 while an often-used non-verbal test of intelligence explained only 19.5% of the variance in achievement in the same group of students (Tan & Maker, 2015). Sak and Maker (2003), in two predictive validity studies, found that the DISCOVER assessment administered in kindergarten was a significant predictor of student grades and achievement in grade 3 and grade 6 in schools in low-income areas with high percentages of Hispanic students. The areas of student strength identified in kindergarten corresponded to the areas in which they excelled in later years. Sarouphim (2002) found that the high school form of the DISCOVER assessment showed no gender or ethnic biases in her study involving Hispanic, American Indian, and White students. Results such as these lead to the conclusion that an instrument based on the same principles could have the potential not only to identify exceptionally talented students in STEM, but also to be appropriate for identifying the talents of students from underrepresented groups rather than being biased against them.

For this research, a definition of exceptional talent in STEM was developed based on the conceptual framework presented here, research and practice in psychology, sociology, general education, and science education: Exceptional talent in STEM has two essential components: *(a) a highly integrated and interconnected knowledge structure; and (b) the ability and willingness to solve a variety of types of problems, from well-structured and known to ill-structured and novel, in science, technology, engineering, and mathematics in the most effective, efficient, elegant, or economical ways.*

Expertise: An Integrated and Interconnected Knowledge Structure

To assess the first component of exceptional talent, Sternberg's (1999) theory of intelligence as developing expertise, which included both recognition of the knowledge structure of experts and an emphasis on problem solving, was the framework for developing the

assessments. Sternberg defined intelligence as “developing expertise” and described it as the ongoing process of acquiring and consolidating of a set of skills needed for mastery in one or more domains. Studies have shown that as expertise in a domain grows, knowledge becomes increasingly interconnected (Chi, et al., 1981; Glaser & Bassok, 1989; Shavelson, 1972; Shavelson, Carey, & Webb, 1990). The most knowledgeable individuals in a domain have a highly integrated conceptual structure organized around central concepts (Ruiz-Primo, et al., 2001). Thus, to solve problems in any domain, and particularly in science, knowledge and conceptual understanding in specific domains (content), problem solving skills (process), and the ability to apply knowledge and understanding to novel situations (application) are essential (Shavelson, et al., 1990). These three areas are roughly similar to the three types of intelligence (i.e., analytical, creative, and practical) combined with the elements of expertise (i.e., metacognitive skills, learning skills, thinking skills, knowledge, and motivation) outlined by Sternberg in both his Theory of Successful Intelligence (1997, 2005) and his view of intelligence as developing expertise (1999). Concept mapping, a way to assess the conceptual structure of students’ knowledge base, was incorporated into the new DISCOVER assessments in STEM (Maker & Zimmerman, 2019) as a way to assess the first component of the definition of exceptional talent.

The Importance of Creativity and Problem Solving

Innovation in any area is the process of creating new products, new ideas, new methods, and new ways of thinking. It involves transformation, change, upheaval, and breakthrough (NSB, 2010; Sawyer, 2006; Weisberg, 2006). For these reasons, to assess abilities in the second component of the definition of exceptional talent, performance-based assessments of creative problem solving were developed. Creative problem solving is essential to innovation in STEM,

and it needs to include the solving of a variety of types of problems, both those involving general and domain-relevant skills and abilities (closed and semi-open), general and specific creativity-relevant processes (semi-open and open-ended), and task motivation (observed during the assessment process (Amabile, 1996, 2013; Renzulli, 1978; Sternberg, 2000, 2005) (Figure 1).

Most conventional measures used to identify exceptional talent in STEM and other areas do not include creative thinking and problem solving, perhaps because these traits are seen as being difficult to measure. Combining the views of Gardner (1983, 1999), Maker and Anuruthwong (2003), and Sternberg (1999), problem solving can be conceptualized as a domain-general ability while the knowledge and skills needed for each domain are domain-specific skills. In Sternberg's theory (2000, 2005), one of the metacomponents is problem solving; core abilities such as memory, metacognition, and reasoning cut across different domains of intelligence, while others are specific to domains.

In a similar vein, social psychologist Amabile (1996, 2013) developed and has found support for a componential model of creativity, which includes (a) domain-relevant skills that lead to exceptional performance in a specific domain (such as math or science); (b) creativity-relevant processes that cut across domains of creative performance; and (c) task motivation, the interest in or attitudes toward a specific task. In the field of education for gifted and talented students, Renzulli's (1978) definition of giftedness consisted of three interacting clusters corresponding to those identified by Amabile, Gardner, and Sternberg: above average ability, creativity, and task commitment.

Motivation

In any discussion of problem solving, development of expertise, and abilities, motivation must not be ignored. An individual's passion for solving certain types of problems as well as her

or his willingness to spend hours developing the expertise needed to function at high levels is a major determinant of success (Amabile, 1996, 2013; Feist, 2006a, 2006b; Gallagher & Gallagher, 2013; Renzulli, 1978; Sternberg, 1999; Subotnik, et al, 2011). Maker and Anuruthwong (2003), as a result of years of research with and observations of children and youth, have described this motivation by noting that the question, situation, challenge, or discrepancy in a situation of interest to the problem solver provides the spark that initiates the action of abilities. They use the metaphor of a prism, describing the “challenge” as the white light coming into one side of the prism and the activated or observed human abilities as the varicolored lights coming out of another side of the prism. Similarly, in Sternberg’s (1999) developing expertise model, the driving force is motivation. Without it, nothing happens. If motivation is present, the other elements, metacognitive skills, learning skills, thinking skills, knowledge, and context, interact as the novice “...works toward expertise through deliberate practice.” (p. 364)

Although the long-term motivation and task commitment described by Sternberg (1999) and Renzulli (1978) cannot be seen during a performance-based assessment, task motivation (Barbot, et al., 2016; Amabile, 1996, 2013) can be observed. When the DISCOVER assessments were being developed, observers were asked to share their perceptions of the performance of the students in their groups, which often included statements about motivation. If, for example, an observer said “She was highly motivated”, the observer was asked to list the behaviors he or she saw that led to the conclusion that a particular student was motivated. Over the years, the behaviors were compiled and are now included on the forms observers use to record their observations during an assessment: shows involvement in task (e.g., focuses on own work rather than that of others, not easily distracted), works continuously, persists on tasks or activities,

increases in motivation or enjoyment as problems increase in open-endedness, follows through to completion, does not want to quit even when others have finished, shows non-verbal enjoyment of task or activity (e.g., smiling, laughing, playing), and verbalizes enjoyment of task.

Defining and Assessing Problem Solving

With problem solving as a key element of the assessment, an operational definition was necessary as a guide to developing tasks. Drawing upon earlier work, the research team used the continuum of problems from the research of Getzels and Csikszentmihalyi (1967, 1976). In Csikszentmihalyi's early research, the ability (and willingness) to structure an open-ended, ill-structured problem, or "problem-finding", as it was later labeled, was the single trait that most accurately predicted the later creative achievements of artists and scientists. His results were similar to those of later researchers, Brandwein (1992, 1995) and Feist (2006a), in their studies of science talent search winners and accomplished scientists.

Csikszentmihalyi's research on problem finding had a significant effect on the field of education for talented students, leading to the development of teaching approaches in which problem-finding was valued over the solving of already-defined problems or problems with known solutions (Gallagher, Stepien, & Rosenthal, 1992; Gallagher & Gallagher, 2013; Maker & Nielson, 1995; Maker & Schiever, 2010). However, this same emphasis was not incorporated into the identification process, with the use of tests consisting of only well-defined problems that had known methods for solving them and one right answer (IQ and achievement tests). Assessments to identify exceptional talent must be aligned with the programs and curricula designed to cultivate them. More specifically, these new assessments must include methods to recognize students who can identify problems, design their own methods, and create new

solutions rather than simply remembering facts and methods or implementing solutions designed by others, the types of items included in standardized tests of abilities and achievement.

Using Getzels and Csikszentmihalyi's (1967; 1976) framework of problem types, Maker and Schiever (Maker & Schiever, 2010; Schiever & Maker, 1997) developed a modified continuum consisting of five, and later, six problem types that could be used to design both assessments of problem solving and curricula to develop it. In their continuum (Table 1), the first two types are closed, the second two are semi-open, and the third two are open-ended. The closed types are characterized by clearly defined problems, specified methods, and right answers; the semi-open types are characterized by clearly defined problems, a range of possible methods, and either a right answer or a range of possible answers. The open-ended types have either a clearly defined problem or an undefined problem, an unlimited number of appropriate methods, and an unlimited number of possible solutions. This continuum has been a guide to developing tasks for the DISCOVER assessments since 1992. Each assessment in the new STEM battery included at least one closed problem (e.g., mechanical/technical) or an aspect of scoring that was based on accuracy (e.g., life science), one semi-open problem, and one open-ended problem (Table 2). For example, in the assessment of mechanical-technical abilities (closely related to physics and engineering), the closed problem was to make a gear box given a set of materials and a diagram, the semi-open problem was to make one of two vehicles based on a picture of the final product and a picture of the materials needed to make it, but with no instructions telling how to put the pieces together to make the vehicle; and the open-ended problem was to make a machine that moved with the remote and motors that was different from the vehicle and of the student's own creation (Alfaiz, Pease, & Maker, 2019).

Tasks that Resemble Real-Life Problems

Another aspect of assessment noted by Milgram and Hong (1993) as important in the assessment of creativity within domains is the relationship between the tasks presented to students and the problems people solve in real-life situations. The problems usually presented to be solved in domain-general creative thinking measures are very different from the kinds of problems people face in real life. For instance, many tests include “unusual uses” of common objects or making different drawings using figural stimuli. In real-life, people must solve problems in a life or career context. For these reasons, Hong and Milgram developed creativity assessments in which respondents were given opportunities to use domain-specific creative thinking ability in a wide variety of real-life situations. This may be an important reason why they found that the scores on their tests and indicators of creativity were good predictors of creative adult accomplishments across several domains (Milgram & Hong, 1993, 1999).

In the example from the mechanical-technical assessment in the physical science domain (Table 2), the tasks presented are similar to the kinds of problems engineers will face as they learn new techniques and apply their ideas to develop new designs. Another example is that in the life science performance assessment, students are presented with two separate tasks. In the first task, students are invited to choose either flowers or insects, and then to make as many groups as they can based on the similarities they observe. In this task, they are using the skills of life scientists, who must observe carefully, make inferences about relationships, and give names and descriptions to their categorizations of phenomena. In the second task, students are directed to make an ecosystem of their own. They are to indicate interdependencies, interactions, and connections among the natural phenomena (Zimmerman, Maker, & Alfaiz, 2019). Studying and recognizing these interdependencies and interactions are life-long tasks of scientists.

First and Second Order Knowledge

Another aspect of the conceptual framework was a consideration of the kind of knowledge being assessed. When designing the performance-based components of the assessment, practical intelligence (Sternberg, 1997, 2005), skills, and knowledge gained from experience was the focus, called “first-order” knowledge (Gardner, 1992) because it is developed through experiences rather than during formal schooling. “Second-order” knowledge (Gardner) is the knowledge developed in school. These two types of knowledge are also roughly equivalent to the fluid (first-order) and crystallized (second-order) abilities identified by Cattell (1957). Fluid ability is involved in solving new problems, using logic in new situations, and identifying patterns. Crystallized ability is using learned or acquired knowledge. Both types of knowledge and ability are essential.

Depending on their environments, the cultural values of the community, and the schools they attend, students are exposed to different knowledge and experiences. In fact, a number of studies have shown that the level of a family’s economic resources may have a greater impact on students’ academic achievement (second-order knowledge) (Miller & Kimmel, 2012; NSB, 2010) than cultural and ethnic factors (Archer, et al., 2012; Xie, Fang, & Kimberlee, 2015). Often, the discrepancies result from both the quality of education provided in schools and the families’ lack of resources available to give their children opportunities for enrichment experiences in STEM and other areas. Thus, items designed to measure second-order knowledge and crystallized abilities often are measures of exposure to information rather than measures of the ability to learn the information (Maker, 2005). All domains in the original DISCOVER assessments and in the new STEM battery include a measure of mainly second-order knowledge (e.g., concept maps in life science and physics and mathematical problem solving) and a measure

of mainly first-order knowledge (e.g., performance assessment of life science, mechanical-technical, and spatial analytical abilities).

Purpose and Research Questions

The purpose of this study was to evaluate the effectiveness of existing instruments and others created specifically for the Cultivating Diverse Talent in STEM (CDTIS) project based on the new definition of exceptional talent in STEM over a four-year period. Several studies were conducted as a part of the CDTIS project to identify and cultivate exceptional talent in STEM. The project was a collaboration between (a) faculty members from three colleges (Pharmacy, Science, and Education) and a biomedical research institute at an R1 University in the Southwestern USA and (b) educators from three public schools, one community-controlled school funded by the Bureau of Indian Affairs, and one charter school in the same state. For this study, two research questions were the focus of the inquiry:

1. What were the differences in characteristics of students identified as exceptionally talented in STEM through measures of (a) academic success (i.e., teacher recommendation, GPA) and ability to present themselves in writing (i.e., personal statements) and (b) creative problem solving (i.e., performance-based assessment and math) and sophisticated conceptual understanding of STEM (i.e., concept mapping)?
2. Did the use of either of the identification methods (e.g., the existing process used in the past to select students for the internship experiences [M1] or the new instruments [M2]) limit the diversity of the characteristics of students selected for the program?

Method

During the CDTIS research project, faculty members at the R1 university partnered with educators from five schools with high percentages of students from culturally diverse groups and

low socioeconomic groups usually underrepresented in special programs for talented students to develop, field test, and implement four new instruments to identify high school students for an internship program on campus. In addition, one of the original DISCOVER assessments (spatial analytical) was added to the battery of measures used to assess the students and one instrument (mathematics problem solving) was modified to include a greater variety of skills. The new instruments were developed by a team of educators (a researcher with extensive experience developing assessments, a post-doctoral scholar with a degree in education of gifted students and experience teaching science, graduate students in education of the gifted, and science teachers) and scientists based on the new definition of exceptional talent in STEM. Identified students were placed in laboratories of scientists on campus, and were provided with support and special opportunities to present their research to the campus community at the conclusion of the internship.

Students from varied ethnic and socioeconomic backgrounds who applied for admission to the internship program, Keep Engaging Youth in Science (KEYS), also were involved in the research. Both the new instruments (M2) and existing procedures (M1) for selecting students (overall grade point average [GPA], teacher recommendations, and self-statements) were used to choose the most promising students to participate in the internship program.

Developing the New Assessments

The process of development followed by the research team to create all the instruments was the *assessment square* proposed by Ruiz-Primo and colleagues (Ruiz-Primo, 2003, 2007; Shavelson, Ruiz-Primo, Li, & Ayala, 2002). The four corners are the *construct*, *observation*, *assessment*, and *interpretation*. Guiding development of the assessments were certain underlying principles. Each task was designed to (a) assess the ability to solve either closed, semi-open, or

open-ended problems (Table 1, Table 3) based on the continuum modified from the research of Getzels and Csikszentmihalyi (1967, 1976) by Maker and Schiever (Maker & Schiever, 2010; Schiever & Maker, 1991, 1997), (b) assess core competencies outlined in the theoretical frameworks (Table 3), (c) be engaging and developmentally appropriate, (d) be as closely related as possible to the types of problems people would be expected to solve in real life in each domain, (e) assess either experiential or school-based knowledge (described as first- and second-order knowledge by Gardner [1992]), and (f) elicit observable behaviors in students that are indicators of different levels of ability and motivation. Using the underlying theories and principles as a basis, constructs were identified for each assessment and each task within the assessments. Constructs included (a) core capacities and abilities considered essential in the domains being assessed; and (b) elements of analytical, creative, and practical intelligence.

From these principles and constructs, trial instruments and instructions for administering the assessments were designed. Next, research team members solved all problems themselves to identify any discrepancies and to understand the processes involved; they tested the items on friends, their children, and students in their classes (observation). At this time, research team members began to note what problem-solving processes were needed to solve the problems, whether the tasks were engaging and interesting to themselves and others, and the kinds of behaviors that could be observed to distinguish different levels of performance and motivation (observation). Next, these items were tested on larger populations, especially those in the target groups of students to be assessed when the instruments were completed (assessment). The processes of observation continued: noting problem solving behaviors, student engagement and interest, and behaviors that could be observed to distinguish performance levels. Finally, the research team developed scoring systems that could be used to reliably distinguish levels of

performance on all items and in all sections of the assessments (interpretation). This process was cyclical in that after all phases, revisions were made and the new items were tested in the same way as the initial versions of the items.

When the team was satisfied that the instruments were ready for trial on a larger scale, they field tested them in the partner schools with students in grade 12 (assessment). During the field-testing, lists of student behaviors to observe, rubrics for assigning ratings to products, and scoring systems were developed (interpretation). The process was iterative as the research team returned to earlier steps to refine the assessments. Based on field tests, instruments and observer instructions were again revised, and after field testing and additional revisions as needed, instruments were administered to the students in grade 11 at the same partner schools. Development of each of these instruments and an in-depth description of their current form can be found in other publications (Alfaiz, et al., 2019; Bahar & Maker, 2020; Maker, 2020; Zimmerman, et al., 2019; Maker & Zimmerman, 2019).

Participants

Participants included high school juniors (11th grade) in two groups. One group (M2) was selected from the partner schools. The first year, three schools on an American Indian reservation participated and in the second year, two schools participated from the American Indian reservation. One partner school was located in an urban, low-income area. Students in the M2 group were selected using the six measures: three performance-based measures of creative problem solving, two measures of sophisticated conceptual understanding using concept maps, and one assessment of application of knowledge and creative application of concepts in math. The other group of students (M1) consisted of students from a variety of types of schools in the same state who applied for and were accepted to the summer internship program at the

university. They came from different schools, urban and rural, and from varied ethnic and economic backgrounds, and were selected using the existing procedures: overall GPA, teacher recommendations, and student self-statements. From the pool of students selected for the Keeping Engaged Youth in Science (KEYS) internship, the research team chose those who were juniors, and from the juniors, attempted to choose participants who came from areas of the state or school districts with demographics (ethnicity and SES levels) similar to those of the partner schools. They also had to agree to participate in the M2 assessments before, during, and after the internship program. Because no American Indian students had participated in the KEYS program prior to the CDTIS project and teachers were not aware of the program, very few American Indian students were available for selection through the existing procedures. In Table 3, the number of students selected using the new assessments (M2) and their ethnic backgrounds are presented along with the number of students and their ethnicities who were selected using the existing methods (M1) and who agreed to participate.

Instruments

In each domain to be assessed, two types of measures were developed and tested, corresponding to the two aspects of the definition of exceptional talent. One was a performance-based measure of problem solving with a particular emphasis on the solving of semi-open and open-ended problems (e.g., life science, mechanical-technical, and spatial analytical); and the other was a measure of domain-specific skills and creative application of knowledge (e.g., concept maps in life science and physics and mathematical problem solving). The definition of each domain was derived from the conceptual framework presented in an earlier section of this article and drawing upon the research and theories of Maker (1993), Maker and Anuruthwong (2003), Gardner (1983, 1999), Sternberg (1985, 1997, 1999) and Amabile (1996, 2013).

For the domain-specific knowledge assessment in life science and physical science, concept maps were developed. Concept mapping is a way to assess students' understanding of the complexity of concepts and their interrelationships, which is essentially a way to determine the extent to which their knowledge base approximates the "...highly integrated knowledge structure..." (Ruiz-Primo, Shavelson, & Schultz, 1997, p. 2) of experts in a domain. In math, instead of concept maps, a paper-and-pencil assessment of knowledge, application of knowledge and heuristics, and conceptual understanding of mathematical processes was modified from an existing instrument considered to be too narrowly focused on computational mathematics. The core competencies, problem types, and tasks for each assessment are summarized in Table 3. Each assessment is described in depth in other publications (Alfaiz, et al., 2019; Maker, 2020; Bahar & Maker, 2020; Maker & Zimmerman, 2019; Zimmerman, et al., 2019).

Data Collection and Scoring

In most cases, all six assessments, which included life science performance-based assessment (Zimmerman, et al., 2019), mechanical-technical performance-based assessment (Alfaiz, et al., 2019), spatial analytical performance-based assessment (Maker, 2020), concept maps (Maker & Zimmerman, 2019) and math (Bahar & Maker, 2020) were administered over a two-day period at each school. Students gathered in the library or another large room at tables for the performance-based assessments, one observer for each 3 to 5 students; and for the paper and pencil assessments, they gathered in a large room with one research team member administering the assessment and two to four others assisting as monitors.

Performance-Based assessments. The performance-based problem-solving assessments (life science, physical science, and spatial analytical) were administered to students in small groups of not more than five students by trained observers who were part of the research team or

graduate students in special education for talented students. Observers gave instructions, distributed materials, interviewed students and recorded their responses, and as a group (along with other observers) scored responses and assigned ratings. Scoring systems and rubrics were designed for each assessment during the field testing and revised as needed before being used for decision-making in this study.

After the assessments were completed, observers met to review the students' performance. First, they listened to the audio records of student interviews and transcribed their responses, completed their notes about each student in the group, and assigned tentative scores and ratings. When all observers had completed this process, they met to discuss the students' performance. They reviewed each student's performance and reached group consensus about the scores to assign.

Concept maps and math assessments. The three paper-and-pencil assessments were administered in large groups by one or two members of the research team. Depending on the size of the group, two to four monitors walked around the room, checking students' maps and their responses to math problems to make certain they were following instructions, and to answer students' questions. They did not suggest answers or comment on students' responses.

Ratings and scoring procedures. After all scoring of the paper-and-pencil assessments of school-based knowledge was completed and rubrics were applied to the results of the performance-based assessments, every student was assigned a rating on each of the six assessments. The first (lowest) level was *Unknown*, signifying that the student had not demonstrated exceptional abilities in that domain so the level of ability was not known. The second level was *Maybe*, signifying that the student had shown some evidence of competence, but her or his performance was generally below the average of the students assessed. The third

level was *Probably*, signifying that the student's performance, while being above the average of those in the school, was not outstanding, but the student probably had strengths in this domain. The fourth level was *Definitely*, signifying that the student definitely performed at an exceptional level in this domain. The fifth level, *Wow*, was reserved for those students whose performance was exceptionally high; they performed beyond even the highest expectations of the observers.

During the development of the DISCOVER assessments, a decision was made to use words to denote levels of ability rather than numbers, and to recognize, by using unknown for the lowest level, that if a student does not demonstrate abilities, the reasons are not clear. Perhaps the student was having difficulties that day, and was unable to perform. The tasks may also not be engaging to some students. In some of the areas, ratings were based on a rubric specific to the assessment. In other areas, such as the concept maps and math assessments, after scoring was complete, the research team members assigned overall ratings of unknown, maybe, probably, definitely, and wow using Jenks Natural Breaks system (Jenks, 1967); a method in which all scores are listed from highest to lowest, and the same rating is assigned for scores that are grouped together and are separated from other groups by at least 3 to 5 points. Specific scoring procedures, including rubrics, points assigned, and criteria for decision-making are included in articles about each assessment (Alfaiz, et al., 2019; Bahar & Maker, 2020; Maker, 2020; Maker & Zimmerman, 2019; Zimmerman, et al., 2019).

Selection of M1 Students

Students in the M1 group were selected by an admission committee from the College of Pharmacy, the Bio5 Institute, and the College of Science. The committee considered the combination of each student's GPA, letter(s) of recommendation, and self-statement. The cut-off for GPA on the program website was 3.2, so if a student with a lower GPA was considered, the

decision was based on the strength of the teacher recommendation(s) and the student self-statement. After initial acceptance, students were interviewed to determine their specific interests to make certain placements were available that matched their interest areas. Further selection of students to participate in the study was based on the similarity of the demographics of their schools or communities (location in the state, ethnicity, SES) to the demographics of the partner schools and the students' willingness to participate in the M2 assessments before, during, and after the internship program.

Selection of M2 Students

After a day of testing, the team gathered to analyze the results of the performance-based assessments. All members of the team reviewed all information collected about the students and made decisions about the ratings to give for each student's performance. This is an important procedure to note because observer bias is less likely to be a factor in decision-making when multiple observers discuss and assign ratings (Griffiths, 1996; Kassymov, 2000; Maker, 2005). Students at the large high school in the urban area, predominantly Hispanic, were compared to other students at that school, and students at the smaller high schools with predominantly American Indian students were compared with each other. Initially, students at each of the two schools with American Indian students were compared with only the students at their school, but later, these results were combined because of the small numbers at one school. Comparison of students at similar schools is an important factor to note as it lessens the effect of differences in opportunities to learn information taught in school. The research team selected students for the internship program in the following order: first were the students who had ratings of definitely or wow in all areas assessed, next were those who had ratings of definitely or wow in 4 or 5 areas, and next were those who had definitely or wow in 3 areas. If more placements were available,

the team considered the patterns of ratings on performance assessments and assessments of second-order knowledge. Similar to the M1 student selection, after initial acceptance, students were interviewed to determine their specific interests to make certain placements were available that matched their interest areas.

Schools

The new assessments to select M2 students were conducted in the project's partner schools: four high schools located in the Southwestern United States (Table 3). The number of students in these schools ranged from 94 to 2245, and more than 71% received free or reduced-price lunches. In two schools, all students were American Indian; in one school, 97% were American Indian and others were White; and in the school in the urban area, the makeup of the school was the following: Hispanic, 83.4%; American Indian, 5.0%; Asian, 0.7%; Black, 3.7%; White, 7.0%; Pacific Islander, 0.3%. All partner schools were located in poverty areas with an unemployment rate that ranged from 7.7% to 45.8%. A limited number of academic programs were offered in the larger schools, one on the American Indian reservation and one in the urban area, such as Honors and Advanced Placement (AP) classes and programs for gifted students. However, the smaller schools on the American Indian reservation did not have these programs.

The students selected by conventional methods (M1) came from a variety of schools and types of schools: most were from public schools in urban areas, but students also came from charter schools, a science academy, a special public high school for high ability/high achieving students, and one school in a high income area of the community with a significant percentage of Hispanic students. Two students came from a public school on the Mexico border, which was predominantly Hispanic, two students came from the partner school in the urban area the second

year, and two students came from the partner schools in rural areas that were predominantly American Indian, one each year of the project.

Data Analysis

To answer the two research questions, the research team conducted Chi-square analyses of the differences between the students identified by both methods in the following areas: gender, ethnicity, primary language identified, and highest educational level of at least one parent. Then, they conducted a *t*-test of the differences in overall GPA.

Results

Research Question 1

Gender. In Table 4, the results of the Chi-square analysis for gender are shown. The Chi-square statistic was .322 and the p-value (2-sided) was .756. The result was not statistically significant, showing that significant differences were not found in gender of the students identified by the two methods.

Ethnicity. In Table 5, the results of the Chi-square analysis of ethnicity by group are shown. The Chi-square statistic was 14.371 and p-value (2-sided) was .006. The ethnic balance of the two groups was significantly different.

Language. In Table 6, the results of the Chi-square analysis of language by group are shown. The Chi-square statistic was 2.163 and the p-value (2-sided) was .539. The groups were not significantly different in primary language.

Highest level of parent education. In Table 7, the results of the Chi-square analysis of highest level of education of at least one parent, using four levels (e.g., middle school, high school, a bachelor's degree, graduate degree [MA or PhD]) are shown. The Chi-square statistic was 5.326 and p-value was .149. The differences were not statistically significant. A second Chi-

square analysis was conducted because of the observed differences between the highest and lowest two groups (high school and graduate degree), which would have the greatest impact on expectations for achievement and the number and type of opportunities provided for students. For this new analysis, the Chi-square statistic was 4.073 and the p-value was .044, which is significant. If only the highest (MA and PhD) and lowest levels (high school) were considered, the groups were significantly different, and if all four levels were considered, the groups were not significantly different in the highest level of parent education, mainly because the groups had the same number with parents who had bachelor's degrees.

Overall grade point average. A *t*-test was applied to find differences in overall achievement as measured by unweighted Grade Point Average (GPA) between students identified by conventional methods (M1 Students) and those identified by the new methods (M2 Students). The Mean of M1 students was 3.93 and the Standard Deviation was 0.20. The Mean of M2 students was 3.07 and the Standard Deviation was 0.59. The *t*-test statistic was 6.43 and p-value was .000. The groups were significantly different in overall GPA.

Research Question 2

Students identified by the conventional methods (M1 Students) and the students identified by the new methods (M2 Students) were not significantly different in gender or primary language (Tables 4 and 6). When the differences between four levels of educational attainment of parents were considered (Table 7), the groups were not significantly different; however, when only the lowest and highest levels of parent educational attainment were considered, the groups were significantly different. The M1 and M2 groups were significantly different in ethnicity (Table 5) and overall achievement (GPA). Thus, the traditional method

limited the diversity of students identified in three areas: highest level of educational attainment of parents, ethnicity, and overall grade point average.

Discussion

Limitations

Before discussing the results of the analyses, some general limitations need to be noted because they have a potential impact on the results. Perhaps the most significant limitation was the difficulty securing parent permission for students to participate in the assessments. Although the team was prepared to assess all 11th grade students at each partner school, this was not possible. For parents unfamiliar with legal requirements for research, the length and complexity of the forms was a barrier. The forms were provided in English and the predominant languages of parents; however, many adults do not read fluently in their native languages, and some do not read at all in their native languages. Many American Indian languages, for instance, have not been written or have been written only recently, so only those who attended bilingual schools with an emphasis on writing in Native languages were able to read the materials in their home languages. At School D, for example, to get even a small percentage of parents to sign consent forms, a community liaison drove to the homes of students in an attempt to get parent or guardian signatures on the forms.

The pool of students assessed was limited by the lack of parent consent. At School A, out of approximately 300 11th graders, only 75 participated in Year 1 and 100 in Year 2. At School B, out of 60, 21 participated in Year 1 and 25 in Year 2. At School C, all except one or two students participated. Many were old enough to sign the student assent forms without parent permission. At School D, out of 150 students, 74 participated in Year 1 and the school declined to participate in Year 2 because obtaining parent permission was so expensive and time-

consuming. The fact that the groups did not contain the entire available population may limit the generalizability of the results.

Another limitation was the restriction on collection of data from students in the partner schools who were assessed but not offered placement in the special internship program. Because these specific students' demographic characteristics could not be described, the research team was unable to determine the degree to which the students assessed were representative of the population of each school. Considering the demographics of the schools, however, one can conclude that the students who were selected by the new methods were definitely from low-income, culturally diverse groups usually underrepresented in programs for exceptionally talented students in STEM. The generalizability of the findings is affected by this limitation.

In general, the small sample size was a limitation, and was due to both the level of funding and the research approval process. Only a few schools could be included in the research project; they were purposefully selected because of their high percentages of the two cultural groups most often underrepresented in programs for exceptionally talented students in the state. Administrators of the schools and school districts also were those who were willing to participate in the project. Thus, the generalizability of the findings is limited, and others are encouraged to use the assessments with larger populations to determine their validity as tools for identifying exceptionally talented students in STEM from underrepresented groups.

Differences Between Groups (Questions 1 and 2)

The characteristics of students identified as exceptionally talented with traditional methods such as teacher recommendations, GPA, and student self-statements (M1) were compared with students identified using the new assessments of problem solving and abilities in STEM developed during the research project (M2): performance-based assessments, concept

maps, and a written assessment of problem solving in math. No statistically significant differences were found in gender or primary language. However, the groups were significantly different in highest level of parent education, ethnicity, and overall grade point average (GPA).

Overall grade point average (GPA). The most important difference between the two groups was in their GPAs. The differences were statistically different at $p = .000$. In the M1 group, GPAs ranged from 3.71 to 4.0 with an average of 3.93, while the M2 students' GPAs ranged from 2.11 to 4.00 with an average of 3.07. In many programs for exceptionally talented students, participants are screened out based on their overall grade point average. The special internship program is no exception. In program materials, the stated requirement is a GPA of 3.2 on a 4.0 grading scale. Of the students in the M2 group, 8 out of 23 would not have met the requirement of 3.2 on a 4.0 grading scale, and most likely would not have applied to the program. Perhaps more importantly, examination of the GPAs of students identified using the conventional methods shows that the lowest GPA of students accepted was 3.71. Using this real decision-making cutoff, 18 out of 23 would not have been selected.

Another important fact to consider in this discussion is that many exceptionally talented students have uneven profiles. They can demonstrate high levels of competence in one area with low levels or even lack of competence in another area (Winner, 2000a, 2000b; Newman & Sternberg, 2004). Similarly, research on underachievement has shown that underachievement is more common than expected in students with exceptional talent when achievement is defined by conventional measures such as overall grade point average (McCoach & Siegle, 2003).

Another factor to consider when thinking of these results is that the students who do well on assessments such as those designed during this project, which include open-ended problem solving requiring creative thinking, may not get high teacher recommendations or have high

GPA's. Research on teacher perceptions of creative students has shown that teachers have a negative view of students who are creative, and that this view often affects their assignment of grades (Westby & Dawson, 1995) even though measures of creative thinking and creative activities such as involvement in active research may be more valid predictors of productivity and performance in STEM areas than intelligence and school grades (Feist, 2006a, 2006b; Hong & Milgram, 2010; Milgram & Hong, 1993). Students who receive high grades and high recommendations may be better at "problem-doing" in planned laboratory experiences than in "problem-solving" in which they must identify a researchable problem, design appropriate scientific methods, and persist until they have reached an acceptable solution (Brandwein, 1992, 1995; Maker, 1993, 1996, 2005). Together, these results suggest that, as McCoach and Siegle (2003) conclude from their study of achievers and underachievers, some highly talented high school students may not value the goals teachers have set for them, and are not motivated to achieve those goals.

M1 and M2 student performance on the assessments created and used in this project provide further support for the use of methods other than GPA for selecting high school students for special STEM projects. Researchers (Miller, 2004; Plucker, et al., 2010) have found that students of color are "severely underrepresented" among the top 1%, 5%, and 10% on almost every measure of achievement: Grades, GPA, class rank, and standardized test scores such as state achievement and National Assessment of Educational Progress [NAEP] results. This was not the case with the new assessments created for the CDTIS project. Students of color and students from low-income groups were well-represented at all levels (Figures 1 and 2). Although one might expect, based on the research of Miller and Plucker and colleagues, that scores of the M2 students on the measures of domain-specific and domain general knowledge and skills

(concept maps and mathematical problem solving) consisting of mainly knowledge gained in school (second-order knowledge) would be significantly lower than scores of M1 students, this was not the case. The scores were very similar (Figure 1) on physics concept maps (Maker and Zimmerman, 2019) and math (Bahar & Maker, 2020), and the average scores of M2 students were higher than scores of the M1 students on the life science concept maps (Maker and Zimmerman, 2019). Figure 1 shows a comparison of the Unweighted GPAs (UW GPA) of the two groups and their scores on concept maps and mathematical problem solving. A clear and substantial difference was found between the two groups on GPAs but not on the new assessments.

On the performance-based measures of creative problem solving in life science (Zimmerman, et al., 2019), mechanical-technical (Alfaiz, et al., 2019), and spatial analytical (Maker, 2020) domains (Figure 2), the ratings of students in the M2 group were higher than the ratings of students in the M1 group in all areas. The average of M1 students on the life science performance-based assessment was 3.35 and the average for M2 students was 3.50; while the average of M1 students on the mechanical-technical performance-based assessment was 3.00 and the average of M2 students was 3.27. The greatest difference was on the spatial analytical performance-based assessment: 3.25 for M1 students and 3.64 for M2 students. Thus, one can conclude that the new measures were culturally responsive measures of the knowledge, skills, and abilities needed for future STEM innovators.

Gender. Although no significant gender differences were found between M1 and M2 groups, an interesting fact is that across the two groups, more girls were selected (26) than boys (17). Reasons for these differences could include factors such as these: (a) because the internships were offered during the summer before students' last year of high school, boys may

have been more likely to be expected to work or more likely to want to work, so they did not apply or did not choose to accept if selected; and (b) more girls than boys applied to the program through the usual method or agreed to participate in the assessment, so the pool of female applicants was greater than the pool of male applicants.

Primary language. Of the 43 who were selected for the special program for exceptionally talented students, only 8 identified a primary language other than English. In the partner schools, information about primary language is not available to researchers unless it is collected directly from students. The Institutional Review Board did not allow the research team to collect information about the primary languages of all students who were assessed. This information was collected on only those who were selected and chose to participate in the internship program. Although a few more M2 students than M1 students identified a primary language other than English, the differences were not significant. The reasons for the lack of significant differences are unclear, but one possibility is that the pool of students whose parents signed the consent forms were predominantly native English speakers. Perhaps the English-speaking parents were more likely to sign the forms and also were more familiar with legal requirements for research projects. In the school with a large percentage of Spanish-speaking students, forms were provided in Spanish, but not in other languages spoken by the students.

From another point of view, because of the significant differences in ethnicity of M1 and M2 students in the two major language groups of the state, the following information is important: of the American Indian students in the M2 group, 1 out of 13 (7.6%) identified Navajo as her or his native language, higher than the state average of 3%; of the five Hispanic students in the M2 group, three (60%) identified Spanish as her or his native language, much higher than the state average of 12%. In the state where the project was located, the drop-out rate

of English Language Learners (ELLs) is significant: only 18% of ELLs in the state graduate, while across the country, the graduation rate is 63% (Sanchez, 2017). A related factor, perhaps a reason for the low graduation rate, is that policies in the state limit the participation of English Language Learners in regular classes. They are placed in instructional blocks that keep them from earning credits in other classes, including those required for graduation (Jung, 2017). Thus, many of these students may not have participated in the new assessments, both because their teachers did not understand that a high level of reading or oral language proficiency was not required, especially for the performance-based assessments, and the students were not encouraged by their parents to participate in the assessments. Comparison of numbers of students from the two main cultural groups in the state and their identified native languages shows that a higher percentage of M2 students than M1 students identified a native language other than English; however, the differences were not statistically significant and the reasons for the differences or lack of statistical significance in the differences are not clear.

Highest level of parent education. The research team initially identified four levels of parent education and collected information about all four levels: middle school, high school, bachelor's degree, and master's or doctoral degree. The number of parents with a bachelor's degree was the same (5), but the differences between the two groups were greater when the lowest level with more than one student, high school (8 M1 and 15 M2), and the highest level, graduate degree (6 M1 and 3 M2) were considered (Table 7). For these reasons, the Chi-square analysis was run a second time, including only high school and graduate degree. These differences were statistically significant.

To help understand these differences, state averages for all adults are compared with the levels of attainment of parents of M1 and M2 students. In the state where the project was located,

approximately 25% of all individuals age 25 and older have a high school degree or equivalent. Of the students in the study, 40% in the M1 group and 65.2% in the M2 group had parents with only a high school degree. One could conclude that the new assessments, compared with the usual methods, enabled selection of a higher percentage of students whose parents have the lowest level of educational attainment. Further, in the state, 10% of adults age 25 and older have a graduate or professional degree. Of the students in the study, 30% of M1 and 13% of M2 students had parents with a graduate or professional degree. The pattern of attainment in the M1 group was much higher than the state average for all ethnic groups, and the pattern of attainment in the M2 group was aligned with the demographics of the state. Based on these statistics, one could conclude that the new assessments enabled students whose parents were not highly educated to be recognized as exceptionally talented in STEM.

Ethnicity. The students identified through the usual methods (M1) and the students identified using the new methods (M2) were significantly different in ethnicity (Table 5). These results need to be interpreted from several points of view. First, identifying a higher percentage of American Indian students in the M2 group can be expected because of the ethnic makeup of the partner schools involved. Three of the schools had 97 to 100% American Indian students, while the fourth school had 83.4% Hispanic students. Thus, one would expect that in the M2 group, selected from partner schools, high percentages of Hispanic and American Indian students would be identified.

Another way of looking at the results is to consider the selection process as consisting of recruitment, assessment, and selection. For M1 students, the process of recruitment consisted of visiting schools, describing the program, distributing application forms, and encouraging students to apply. In many cases, students who had attended the program the previous year

assisted by encouraging others to apply. Prior to the project, no partner schools with high percentages of American Indian students were included in the recruitment process. Thus, they did not know about the program and did not apply. During the project, however, these schools were included in the recruitment process. Discussions with the teachers and administrators during the project revealed that many students did not apply through the M1 process because they did not think they would be accepted. In addition, many of the students who were recommended for acceptance to the program through the M2 selection process were reluctant to complete the required application forms until they were shown their profiles and told how their performance compared with the performance of others. In short, many of the students were uncertain of their abilities and lacked the academic self-confidence needed to seek out advanced experiences and apply for admission to a special program in STEM. This result is consistent with research showing that African American, Hispanic, and American Indian students' lower academic self-perceptions and lower self-expectations affect their achievement and their willingness to engage in special programs and opportunities (McCoach & Siegle, 2003). A helpful conclusion to make from these results and the potential reasons for them is that the use of performance-based assessments and other similar measures are important for students from low-income and culturally diverse groups, and both the students and their parents need to be given the results in a format they can understand (Pease, et al., 2019). The students need encouragement and assistance in identifying potential opportunities (Subotnik, et al., 2011; Olszewski-Kubilius, Subotnik, & Worrell, 2015) to develop their abilities.

Comparison with other research. These results are similar to the findings of researchers who compared students selected with the DISCOVER performance-based assessments of creative problem solving with students selected by conventional methods such as IQ and

achievement. The percentages of students selected using DISCOVER and a performance-based assessment modified from DISCOVER paralleled the percentages of individuals with certain demographic characteristics in the populations of low-income, highly diverse students in the schools in several areas, including the following: race/ethnicity (Nielson, 1994; Reid, et al., 1999; Romanoff, et al., 2009), ancestral origin, preferred language, second language, religious preference, birthplace of parents, mobility, family income, family home, parents' occupation, parents' educational attainment, degrees earned, teaching experience, identified as gifted, and work or volunteer at a school (Nielson, 1994).

Important to consider is that other researchers have found that including measurements of creativity and creative problem solving in talent identification is beneficial: (a) both domain-general and domain-specific ability and creativity are important variables associated with outstanding achievements in productivity and performance (Feist, 2006a, 2006b; Subotnik, et al., 2011) so inclusion of both increases the predictive ability of the talent identification (Sternberg, 2010; Hedlund, et al, 2006); and (b) inclusion of measures of creativity increases the potential to identify students from groups traditionally underrepresented in programs for exceptionally talented students because they score at similar levels (Glover, 1976; Ivcevic & Kaufman, 2013; Kaufman, 2006; Kaufman, et al., 2004; Torrance, 1971) or sometimes higher (DeVries, & ShiresGolon, 2011; Ghonsooly & Showqui, 2012; Jenkins, 2005; Kharkhurin, 2012) than students from mainstream groups.

Recommendations

Assess “What they know”

Based on the results of the CDTIS project and the results of research on the DISCOVER assessments from 1992 to 2005, the most important recommendation for identification of

students with exceptional talent in STEM is to include measures in which the emphasis is on “What do the students know and what can they do?” rather than “Do the students know ____?” and “Can they reach the conclusion that ____?” By using this open-ended approach, the different experiences and different levels and types of exposure to information do not determine the levels of ability identified. Here are some examples from the different parts of the assessment:

1. In the life science performance assessment, the first task is to notice the characteristics of either flowers or insects and to make groups based on their similarities. In this example, students are not asked to remember classification schemes they have been taught, nor are they asked to identify certain characteristics from a list of those considered to be accurate. They are asked to use their ability to observe and compare characteristics, and to group the flowers or insects based on similarities. These are thinking and problem-solving skills, not memory skills.
2. In the concept mapping assessments, opportunities are provided for students to demonstrate that they know the appropriate connections among concepts, but also to make connections that might not be conventional. They also are invited to add examples from their own experience.

Another related recommendation is to include interviews. When an observer sits with a student individually and asks him or her to explain a drawing or a construction or to provide additional information about a grouping of flowers or insects, the student does not have to depend on her or his ability to write an answer. It also gives the observer an opportunity to ask a student to provide additional details or explain an answer that is not clear to the observer. Across the different assessments, interviews seemed most important in the life science performance

assessment because students often told more in their interviews than they could demonstrate when grouping flowers or insects or designing their ecosystems. An interview also was included in the mechanical-technical assessment. Students were asked to explain how their machines operated. Often students demonstrated the working of the machine with little explanation, but the explanations often helped observers understand the students' levels of knowledge and experience with mechanical-technical concepts. For students with strengths in visual and spatial abilities, making the product first and explaining it later is an important component. For those with strengths in verbal abilities, being able to explain how something is supposed to work provides important information about students' abilities.

Use Concept Maps as Assessments of Domain-Specific Ability and Developing Expertise

Conceptions of ability as “developing expertise” rather than a static trait are supported by research more frequently than previous views of ability as stable and unchangeable (Flynn, 1987, 2007; Sternberg, 1999b; Subotnik, et al., 2011); thus, assessment of the current level of expertise in students is an important way to identify those with exceptional talent. The concept mapping assessment developed by Novak and Gowin (1984) and refined during a number of projects by Ruiz-Primo and colleagues (Ruiz-Primo, 2007; Ruiz-Primo & Shavelson, 1996; Ruiz-Primo, et al., 2001) is a promising practice that needs to be extended beyond assessments in STEM to other areas in education of exceptionally talented students. Using concept maps is an excellent way to assess students' higher levels of thinking and their understanding of concepts and their interrelationships, demonstrating the degree to which their thinking is similar to the types of thinking demonstrated by experts in a domain: organized around key concepts, containing patterns of meaningful information, and showing remote connections among different hierarchies of concepts.

Another important strength of the concept mapping assessment is that it has no ceiling. Even if students have completed maps of the same concepts in the past, as their knowledge base becomes increasingly interconnected and diverse, their scores can increase, demonstrating the ways they have integrated new knowledge and experience into their understanding of the domain. Thus, concept maps can be used as pretest and posttest measures of gains from special programs and can be used to track the development of domain-specific expertise over time as recommended by Sternberg (1999). Students' maps also can be compared with the maps of experts to determine their progress in development of the types of knowledge structures of experts in each domain (Ruiz-Primo, et al., 1997).

Include Measures of both First-Order and Second-Order Knowledge in Each Domain

In the CDTIS project, six assessments in three domains were included in the identification of students with exceptional talent in STEM: math (math problem solving and spatial analytical performance assessment), life science (life science concept map and life science performance assessment), and physical science (physics concept map and mechanical-technical performance assessment). The measures of first-order knowledge gave information about students' problem solving and creative thinking, creativity-relevant skills and abilities in Amabile's (1996, 2013) theory; while the measures of second-order knowledge gave information about their development of domain-specific expertise (Amabile; Sternberg, 1999). While observers are watching students during the performance assessments, they also can evaluate students' task motivation (Amabile; Renzulli, 1978).

If selecting students for specific domains rather than STEM in general, identification procedures can be based on the two assessments in each domain rather than administering all six assessments. All students in a school or at a certain grade level should be assessed, however,

rather than relying on teacher recommendations to screen out those who participate in further assessments. Teachers often overlook those who are creative thinkers and problem solvers (McCoach & Siegle, 2003) especially if they are not high achievers on conventional measures.

Use Performance Assessments that Include Real-Life Tasks

As recommended by Maker (2005) and Hong and Milgram (2013), the tasks in performance assessments need to be as closely related as possible to the types of problems professionals solve in real life situations. In the context of identifying exceptional talent in STEM, these tasks need to be related to the problems solved by mathematicians, scientists, and engineers. In the new assessments, for instance, one of the life science tasks is to develop groups of either flowers or insects based on the similarities in their characteristics. Noticing similarities and differences in natural phenomena is an important skill of a life scientist, and one of the major tasks of life scientists is developing and validating classification schemes. In the mechanical-technical assessment, the second task is to make a vehicle that functions the way it should function and the third task is to create a new machine. Both of these tasks represent important problem-solving challenges faced by engineers: designing vehicles that function well and developing new and innovative designs that function well or perhaps better than those of the engineers in competitive companies.

Use a low Cut-Off Score (i.e., 2.0 GPA) and Administer Culturally Responsive Assessments

If the use of a GPA cut-off score is important, set it at a low level such as 2.0 and then administer concept maps, math problem solving, and the performance-based assessments in the domains of interest or in all domains of STEM. If, for instance, life science is the domain of interest, use the life science concept mapping assessment and the life science performance assessment. If physical science is the domain of interest, use the physics concept mapping

assessment and the mechanical-technical performance-based assessment. For math, both the math problem solving and the spatial analytical performance assessments can be used together. As noted above, the spatial analytical assessment can be used alone as measure of domain-general abilities across STEM areas.

Setting a low cut-off score without additional assessments is not recommended because many students would be selected who do would not demonstrate the knowledge and skills needed for exceptional performance and productivity (Subotnik, et al., 2011). However, setting a cut-off score that is too high would result in screening out many students who do have those capabilities, as demonstrated in this research.

Use the Spatial Analytical Assessment as a Domain-General Assessment for STEM

If only one performance assessment can be chosen as a way to identify exceptionally talented students in STEM, the best choice is the spatial analytical assessment. Spatial ability is (a) a key component of the abilities of adolescents who achieve advanced educational and occupational credentials in STEM careers, (b) an important component in educational and occupational settings for all individuals including those who are talented, and (c) talent searches exclude many intellectually talented students when criteria are restricted to mathematical and verbal ability measures (Wai, Lubinski, & Benbow, 2009). Wai and colleagues (2009) found that 70% of the top 1% in spatial ability were not in the top 1% in either math or verbal abilities, and would be missed in many talent searches even though their spatial abilities enabled them to be highly successful in biological sciences, physical sciences, engineering, and mathematics. A similar pattern can be found in the comparison of M1 and M2 students in the current study. Of the 8 students who would not have been accepted into the special internship program because their GPAs were lower than the published cut-off criterion, 6 had scores of definitely or wow (4

or 5 on a 5-point scale) on the spatial analytical assessment, the scores used to place students in special programs. One student scored 3 and another scored 2. The average score of these 8 students on the spatial analytical assessment was 3.75 (on a 5-point scale) and their average GPA was 2.49 (on a 4-point scale). As noted earlier, as many as 18 out of 23 probably would not have been chosen if one compares their GPAs to those of the students selected through the M1 process. Also relevant is the fact that in other studies of the numbers and percentages of students identified as exceptionally talented using DISCOVER, no ethnic, cultural, or economic biases have been found when using the spatial analytical assessment (Maker, 2020; Sarouphim, 2001, 2002, 2004).

Conduct More Research on the Validity and Reliability of the Instruments

Because of limited participation at three out of the four partner schools, the pool of participants may not have been representative of the total school population, with a greater percentage of students from homes of parents with higher educational levels, better English skills, higher incomes, and positive attitudes toward educational opportunities. Thus, the differences between those identified using the conventional methods and those identified using the new methods may have been greater than the differences found in the study. Readers are encouraged to continue this line of research, and to evaluate both the psychometric properties and the practicality of the new assessments. At this time, we have not conducted predictive validity or concurrent validity studies, but these will be essential to the future use of the new assessments.

Conclusion

Many students from cultural groups and low-income families usually underrepresented in special programs were identified as exceptionally talented using the new measures developed

during the Cultivating Diverse Talent in STEM (CDTIS) Project. These often-overlooked students were given an opportunity to participate in a special internship program at an R1 university in the southwest. Reports from teachers, graduate students, and observers confirmed that although some students from both the M1 and M2 groups needed additional academic and personal support, both groups were successful in the internship program, demonstrating that students from all cultures can be successful in high-level educational programs. Students participated in original research, created posters about their work, presented to a community of researchers and parents, and also presented to school boards and other interested groups in their local communities. Some also returned in subsequent years as interns to assist other students in the program. Many have applied and been accepted to the R1 university where the internship was held, and many have applied and been accepted to other universities in the state and region. Students identified by the new assessments (the M2 group) believed that the program had a significant impact on their goal-setting and planning for the future (Wu, Pease, & Maker, 2019).

Another important conclusion is that the use of performance-based assessments and other similar measures are important for students from low-income and culturally diverse groups, and that both the students and their parents need to be given the results in a format they can understand (Pease, et al., 2019). Students and their parents, especially parents without college degrees, need encouragement and assistance in identifying potential opportunities (Subotnik, et al., 2011; Olszewski-Kubilius, et al., 2015) to develop their abilities. Special programs that capture students' interest, enable them to engage in original research (Feist, 2006a, 2006b) and act as catalysts for their motivation are essential to their continued participation in STEM experiences and careers (Subotnik, et al., 2011). Finally, in the words of Joseph and Ford (2006,

p. 50) "...every effort must be made to ensure that belief systems, tests, policies, and procedures do not serve as gatekeepers that close the doors of opportunity for diverse students."

References

- Alfaiz, F. S., Pease, R., & Maker, C. J. (2020). Culturally responsive assessments of physical science abilities: Development, field testing, and implementation. *Journal of Advanced Academics*.
- Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997–1013.
- Amabile, M. (1996). *Creativity in context*. Boulder, Co: West-view.
- Amabile, T. M. (2013). Componential theory of creativity. In *Encyclopedia of management theory*. (pp. 134–139). Los Angeles, CA: SAGE Publications, Inc.
- An, D. & Runco, M.A. (2016). General and domain-specific contributions to creative ideation and creative performance. *Europe's Journal of Psychology*, 12(4), 523–532, doi:10.5964/ejop.v12i4.1132
- Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. *American Educational Research Journal*, 49(5), 881–908.
- Austin, L. B., & Shore, B. M. (1993). Concept mapping of high and average achieving students and experts. *European journal of ability*, 4(2), 180-195. doi: 10.1080/0937445930040207
- AZED, Arizona Department of Education statistics. (2006). *Data about gifted students*. Retrieved from <http://www.azed.gov>
- Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal*, 11, 173–177.
- Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review*, 37, 165-178, DOI: 10.1080/02783193.2015.1047480

- Bahar, K. & Maker, C. J. (2011). Exploring the relationship between mathematical creativity and mathematical achievement. *Asia-Pacific Journal of Gifted and Talented Education*, 3(1), 33 - 48.
- Bahar, K. & Maker, C.J. (2020). Culturally responsive assessment of mathematical ability: Development, field testing, and implementation. *Journal of Advanced Academics*, 1–23
DOI: 10.1177/1932202X20906130
- Barbot, B., Besancon, M., & Lubart, T. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, 52, 178-187. <http://dx.doi.org/10.1016/j.lindif.2016.06.005>
- Barbot, B., & Tinio, P. P. L. (2015). Where is the “g” in “creativity”? A specialization-differentiation hypothesis. *Frontiers in Human Neuroscience*, 8, 1041.
<http://dx.doi.org/10.3389/fnhum.2014.01041>.
- Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neuroscience and Biobehavioral Reviews*, 51, 108–117.
- Beaty, R. E., Benedek, M., Wilkins, R. W., Jauk, E., Fink, A., Silvia, P. J., Hodges, D. A., Koschutnig, K., & Neubauer, A. C. (2014). Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia*, 64, 92–98.
- Benbow, C., & Arjmand, O. (1990). Predictors of high academic achievement in mathematics and science by mathematically talented students: A longitudinal study. *Journal of Educational Psychology*, 82, 430-441. doi: 10.1037/0022-0663.82.3.430
- Brandwein, F. (1992). Science talent: The play of exemplar and paradigm in the science education of science prone young. *Science Education*, 76(6), 121-139.

- Brandwein, P. (1995). *Science talent in the young expressed within ecologies of achievement (RBDM 9510)*. Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.
- Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. (Expanded ed.). Washington, D.C.: National Academy Press.
- Brown, J., & Weiner, B. (1984). Affective consequences of ability versus effort ascriptions: Controversies, resolutions, and quandaries. *Journal of Educational Psychology*, 76, 146–158.
- Cattell, R. B. (1957). Appendix, fluid and crystallized intelligence. In *Personality and motivation structure and measurement* (pp. 871–880). New York, NY: Harcourt, Brace, and World, Inc.
- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive sciences*, 5(2), 180-195.
- Clasen, D. R., Middleton, J. A., & Connell, T. J. (1994). Assessing artistic and problem-solving performance in minority and nonminority students using a nontraditional multidimensional approach. *Gifted Child Quarterly*, 38(1), 27–32.
- Collins, K. H. (2018). Confronting color-blind STEM talent development: Toward a contextual model for Black student STEM identity. *Journal of Advanced Academics*, 29(2). 143-168. DOI: 10.1177/1932202X18757958
- Curby, T. W., Rudasill, K. M., Rimm-Kaufman, S. E., & Konold, T. R. (2008). The role of social competence in predicting gifted enrollment. *Psychology in the Schools*, 45, 729-744. doi: 10.1002/pits.20338

- Dai, D.Y. & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly*, 57(3), 151-168.
- Daley, B. J. (1999). Novice to expert: An exploration of how professionals learn. *Adult Education Quarterly*, 49, 133–147. doi: 10.1177/074171369904900401
- Deaux, K., & Emswiller, T. (1974). Explanation of successful performance on sex linked task: What is skill for male is luck for the female. *Journal of Personality and Social Psychology*, 29, 80–85.
- Deschamps, J.-C., & Beauvois, J.-L. (1994). Attributions intergroupes. In R. Y. Bouhris & J.-P. Leyens (Eds.), *Stéréotypes discrimination et relations intergroupes*. Liège: Mardaga.
- DeVries, M. & ShiresGolon, A. (2011). Making education relevant for gifted Native Americans: teaching to their learning style. In *Special Populations in Gifted Education*. J.A. Castellano & A.D. Frazier, Eds. (47–72). Waco, TX: Prufrock Press.
- Dogusoy-Taylan, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in Human Behavior*, 36, 82–93. doi:10.1016/j.chb.2014.03.036
- Dweck, C. S. (2006). *Mindsets. The psychology of success*. New York, NY: Ballantine.
- Dweck, C. S. (2012). Mindsets and malleable minds: Implications for giftedness and talent. In R. F. Subotnik, A. Robinson, C. M. Callahan, & E. J. Gubbins (Eds.), *Malleable minds: Translating insights from psychology and neuroscience to gifted education* (pp. 7–18). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut.

- 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. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.
- Erdimez, O., Tan, S., & Zimmerman, R. (2017). The use of concept maps as a tool to measure higher level thinking skills in elementary school science classes. *Journal for the Education of Gifted Young Scientists*, 5(2), 1-42. DOI: <http://dx.doi.org/10.17478/JEGYS.2016221898>
- Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Erlbaum.
- Ericsson KA. (2015). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, 90, 1471-86. <http://dx.doi.org/10.1097/ACM.0000000000000939>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406. doi:10.1037/0033-295X.100.3.363
- Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2005). Giftedness viewed from the expert-performance perspective. *Journal for the Education of the Gifted*, 28, 287-311.
- Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review*, 85, 114-121.
- Feist, G. J. (2006a). The development of scientific talent in Westinghouse finalists and members of the National Academy of Sciences. *Journal of Adult Development*, 13(1), 23-35. DOI: 10.1007/s10804-006-9002-3

- Feist, G. J. (2006b). How development and personality influence scientific thought, interest, and achievement. *Review of General Psychology, 10*(2), 163-182. DOI: 10.1037/1089-2680.10.2.163
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171.
- Flynn, J.R. (2007). *What is intelligence? Beyond the Flynn Effect*. New York: Cambridge University Press.
- Ford, D. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education, 32*(1), 4-14.
- Ford, D.Y. (2006). Desegregating gifted education: A need unmet. *Journal of Negro Education, 64*, 52-62.
- Gallagher, S.A., & Gallagher, J.J. (2013). Using problem based learning to explore unseen academic potential. *Interdisciplinary Journal of Problem Based Learning, 7*(1), 111–131.
- Gallagher, J., Harradine, C. C., & Coleman, M. R. (1997). Challenge or boredom? Gifted students' views on their schooling. *Roeper Review, 19*, 132–136.
- Gallagher, S. A., Stepien, W. J., & Rosenthal, H. (1992). The effects of problem-based learning on problem solving. *Gifted Child Quarterly, 36*(4), 195–200.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.
- Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B. Gifford & M. O'Connor (Eds.), *Changing assessment: Alternative views of aptitude, achievement, and instruction*. (pp. 77–120). Boston, MA: Kluwer.

- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic books.
- Getzels, J. & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal*, 3(9), 80-84.
- Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York, NY: Wiley.
- Getzels, J. W., & Jackson, P. W. (1958). The meaning of “giftedness”: An examination of an expanding concept. *Phi Delta Kappan*, 40, 75–77.
- Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence*. New York: Wiley.
- Gilhooly, K. J., Fioratou, E. E., Anthony, S. H., & Wynn, V. V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, 98, 611–625.
- Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, 40(1), 631–666.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. xv–xxviii). Hillsdale, NJ: Erlbaum.
- Glover, J.A. 1976. Comparative levels of creative ability among elementary school children. *Journal of Genetic Psychology*, 129, 131–135.
- Griffiths, S.E. (1996). *The inter-observer reliability of the DISCOVER problem-solving assessment*. Unpublished manuscript. The University of Arizona, Tucson.
- Ghonsooly, B. & Showqi, S. (2012). The effects of foreign language learning on creativity. *English Language Teacher*, 5, 161–167.
- Guilford, J. P., & Christensen, P. R. (1973). The one-way relation between creative potential and IQ. *Journal of Creative Behavior*, 7, 247–252.

- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, New York, NY.
- Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: A supplement to the Graduate Management Admissions Test. *Learning and Individual Differences, 16*, 101–127.
- Hong, E. & Milgram, R.M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal, 22*(3), 272–287, DOI: 10.1080/10400419.2010.503535
- İngeç, Ş. K. (2009). Analysing concept maps as an assessment tool in teaching physics and comparison with the achievement tests. *International Journal of Science Education, 31*(14), 1897-1915.
- Ivcevic, Z. & J.C. Kaufman. (2013). The can and cannot do attitude: How self estimates of ability vary across ethnic and socioeconomic groups. *Learning and Individual Differences, 27*, 144-148.
- Jellen, H. U., & Urban, K. (1989). Assessing creative potential worldwide: The first cross-cultural application of the Test for Creative Thinking–Drawing Production (TCT-DP). *Gifted Educational International, 6*, 78–86.
- Jenkins, A.H. 2005. Creativity and resilience in the African American experience. *Journal of Humanistic Psychology, 33*, 25–33.
- Jenks, G.F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography, 7*, 186-190.
- Joseph, L. M., & Ford, D. Y. (2006). Nondiscriminatory assessment: Considerations for gifted education. *Gifted Child Quarterly, 50*(1), 42–51.

- Jung, C. (2017). Arizona ELL student graduation rate lags far behind national average. *KJZZ Education Report*. Retrieved from <https://kjzz.org/content/440381/arizona-ell-student-graduation-rate-lags-far-behind-national-average>
- Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., ... Benedek, M. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence*, 57, 105–117.
- Kassymov, A. (2000). *Researchers and implementers: Comparison of inter-rater reliability*. Unpublished manuscript, University of Arizona, Tucson.
- Kaufman, J.C. (2006). Self-reported differences in creativity by gender and ethnicity. *Journal of Applied Cognitive Psychology*, 20, 1065–1082.
- Kaufman, J.C., Baer, J. & Gentile, C.A. (2004). Differences in gender and ethnicity as measured by ratings of three writing tasks. *Journal of Creative Behavior*, 38, 56–69.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken: Wiley.
- Kharkhurin, A.V. (2012). *Multilingualism and Creativity*. Bristol: Multilingual Matters.
- Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *The Journal of Secondary Gifted Education*, 16 (2/3) 57–66. <http://dx.doi.org/10.4219/jsge-2005-473>.
- Lubart, T. I., Zenasni, F., & Barbot, B. (2013). Creative potential and its measurement. *International Journal of Talent Development and Creativity*, 1(2), 41–51.
- Lubinski, D., & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math–science expertise. *Perspectives on Psychological Science*, 1, 316–345.

- Luria, S. R. O'Brien, R.L. & Kaufman, J.C. (2016). Creativity in gifted identification: increasing accuracy and diversity. *Annals of the New York Academy of Sciences*. 1377, 44-52. DOI: 10.1111/nyas.13136
- Maker, C.J. (1978). *Successful Handicapped Adults: Their Perceptions of Significant Events, Causes, and Effects*. Unpublished Doctoral Dissertation, University of Virginia, Charlottesville, VA., USA.
- Maker, C.J. (1992). Intelligence and creativity in multiple intelligences: Identification and development. *Educating Gifted Learners: Discovering and Nurturing Talent*, 12(4), 12-19.
- Maker, C. J. (1993). Creativity, intelligence, and problem-solving: A definition and design for crosscultural research and measurement related to giftedness. *Gifted Education International*, 9, 68-77.
- Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. *Gifted Child Quarterly*, 40(1), 41-50.
- Maker, C.J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International*. 15(3), 232-251.
- 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. (2020). Culturally Responsive Assessments of Spatial Analytical Skills and Abilities: Development, Field Testing, and Implementation. *Journal of Advanced Academics*.

- Maker, C. J., Alhusaini, A. A., Pease, R., Zimmerman, R., & Alamiri, F. Y. (2015). Developing creativity, talents, and interests across the lifespan: Centers for creativity and innovation. *Turkish Journal of Giftedness and Education*, 5(2), 83–109.
- Maker, C. J. & Anuruthwong, U. (2003). *The miracle of learning: The prism model*. 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.
- Maker, C. J., & Nielson, A. B. (1995). *Teaching models in education of the gifted*. Austin, TX: Pro-ed.
- Maker, C. J. & Schiever, S. W. (2010). *Curriculum development and teaching strategies for gifted learners*. (3rd Ed.). Austin, TX: Pro-Ed.
- Maker, C. J. & Zimmerman, R. (2019). *Concept Maps as Culturally Responsive Assessments of the Complexity and Interrelationships of Concepts in Science: Development, Field Testing, Implementation, and Results*. Manuscript submitted for publication.
- Margulies, J. (2002). Lead plaintiff against Michigan still hopes for a spot at its law school. *Chronicle of Higher Education*, 49(16), A23.
- McCoach, D. B., & Siegle, D. (2003a). Factors that differentiate underachieving gifted students from high-achieving gifted students. *Gifted Child Quarterly*, 47, 144–154.
- McCoach, D. B., & Siegle, D. (2003b). The structure and function of academic self-concept in gifted and general education samples. *Roeper Review*, 25, 61–65.
- Milgram, R.M, & Hong, E. (1993). Creative thinking and creative performance in adolescents as predictors of creative attainments in adults: A follow-up study after 18 years. *Roeper Review*, 15(3), 135-139.

- Milgram, R. M. & Hong, E. (1999). Creative out-of-school activities in intellectually gifted adolescents as predictors of their life accomplishment in young adults: A longitudinal study. *Creativity Research Journal*, 12, 77-87. doi:10.1207/s15326934crj1202
- Miller, L. S. (2004). *Promoting sustained growth in the representation of African Americans, Latinos, and Native Americans among top students in the United States at all levels of the education system*. Storrs, CT: The National Research Center on the Gifted and Talented.
- Miller, J. D., & Kimmel, L. G. (2012). Pathways to a STEMM profession. *Peabody Journal of Education*, 87(1), 26–45.
- National Science Board [NSB]. (2010). *Preparing the Next Generation of STEM Innovators: Identifying and Developing our Nation's Human Capital*. Arlington, VA: National Science Foundation.
- Neisser, U. E. (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, D.C.: American Psychological Association.
- Newman, T. & Sternberg, R.J. (Eds.) (2004). *Students with both gifts and learning disabilities*. Boston: Kluwer Academic Publishers.
- Nielson, A. B. (1994). Traditional identification: Elitist, racist, sexist? New evidence. *CAG Communicator: The Journal of the California Association for the Gifted*, 3, 18-19, 26-31.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York, NY: Cambridge University Press.
- Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, 39, 36–45.

- Olivas, M. (1999). Higher education admissions and the search for one important thing. *University of Arkansas at Little Rock Law Review*, 12, 993-1024.
- Olszewski-Kubilius, P. Subotnik, R., & Worrell, F.C. (2015). Antecedent and concurrent psychosocial skills that support high levels of achievement within talent domains. *High Ability Studies*, 26(2), 195-210. <http://dx.doi.org/10.1080/13598139.2015.1095077>
- Ortiz, S. O. (2002). Best practices in nondiscriminatory assessment. In A. Thomas & J. Grimes (Eds.). *Best Practices in School Psychology IV* (pp. 1321-1336). Bethesda, MD: National Association of School Psychologists.
- Pease, R. Vuke, M.H., Maker, C. J., & Muammar, O.M. (2019). A practical guide for implementing the STEM assessment results in classrooms: Using strength-based reports and real engagement in active problem solving (REAPS). *Journal of Advanced Academics*.
- Plucker, J. A. (1999). Reanalysis of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior*, 33, 136-137.
- Plucker, J. A., Burroughs, N., & Song, R. (2010). *Mind the (other) gap!* Bloomington: Center for Evaluation and Education Policy, Indiana University.
- Plucker, J.A. & Zabelina, D. (2009). Creativity and interdisciplinarity: one creativity or many creativities? *ZDM Mathematics Education*, 41, 5-11. DOI 10.1007/s11858-008-0155-3
- Reid, C., Udall, A., Romanoff, B. & Algozzine, B. (1999). Comparison of traditional and problem solving assessment criteria. *Gifted child Quarterly*, 43, 244-251.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60, 180-184.

- Renzulli, J. S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J.E. Davidson (Eds.), *Conceptions of Giftedness*, (pp. 53-92). New York, NY: Cambridge University Press.
- Rimm, S. B., Siegle, D., & Davis, G. A. (2018). *Education of the gifted and talented* (7th ed.). Boston, MA: Pearson.
- Rogers, T. (2009). Native American poverty: A challenge too often ignored. *Spotlight on Poverty and Opportunity*. Retrieved from <http://www.spotlightonpoverty.org/ExclusiveCommentary.aspx>
- Rojas, L. F. (2015). Factors affecting academic resilience in middle school students: A case study. *Gist: Education and Learning Research Journal*, 11, 63-78.
- 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.
- Ruiz-Primo, M. (2003). *A framework to examine cognitive validity*. Paper presented at the meeting of the American Education Research Association, Chicago.
- Ruiz-Primo, M. (2007). Assessment in science and mathematics: Lessons learned. In M. Hoepfl & M. Lindstrom (Eds.), *Assessment of Technology Education*, CTTE 56th Yearbook (pp. 203–232). Woodland Hills, CA: Glencoe-McGraw Hill.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of research in science teaching*, 33(6), 569-600. DOI: CCC 0022-4308/96/060569-32

- Ruiz-Primo, M., Shavelson, R., Li, M., & Schultz, S. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, 7(2), 99-141.
- Ruiz-Primo, M., Shavelson, R., & Schultz, S., (1997). *On the validity of concept map based assessment interpretations: An experiment testing the assumption of hierarchical concept-maps in science*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Runco, M. A. (2007). *Creativity*. Amsterdam: Elsevier.
- Runco, M. A., & Albert, R. S. (1986). The threshold theory regarding creativity and intelligence: An empirical test with gifted and nongifted children. *Creative Child and Adult Quarterly*, 11, 212–218.
- Sak, U., & Maker, C. J. (2003). The long-term predictive validity of a performance-based assessment used to identify gifted CLD students. In *Proceedings of the 15th Biennial World Conference of the World Council for Gifted and Talented Students*.
- Sak, U. & Maker, C. J. (2006). Developmental variation in children's creative mathematical thinking as a function of schooling, age, and knowledge. *Creativity Research Journal*, 18(3), 279-291.
- Sanchez, C. (2017). English Language Learners: How your state is doing. *National Public Radio Education*. Retrieved from <https://www.npr.org/sections/ed/2017/02/23/512451228/5-million-english-language-learners-a-vast-pool-of-talent-at-risk>
- 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. *Prufrock Journal*, 15(2), 61–69.
- Sarouphim, K. M. & Maker, C. J. (2010). Ethnic and gender differences in identifying gifted students: A multi-cultural analysis. *International Education*, 39(2), 42–56.
- Sawyer, R. K. (2006). *Explaining creativity: the science of human innovation*. New York: Oxford University Press.
- Schiever, S. W., & Maker, C. J. (1997). *Enrichment & acceleration: An overview & new direction*. In N. Colangelo & G. A. Davis (Eds.), *Handbook of Gifted Education*, (2nd Ed. (pp. 113–125). Needham Heights, MA: Allyn & Bacon.
- Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, 63(3), 225.
- Shavelson, R. J., Carey, N. B., & Webb, N. M. (1990). Indicators of science achievement: Options for a powerful policy instrument. *Phi Delta Kappan*, 71(9), 692–97.
- Shavelson, R., Ruiz-Primo, M., Li, M., & Ayala, C. (2002). *Evaluating new approaches to assessing learning*. CSE: Technical Report 604. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing, University of California.
- Siegle, D. (2013). *The underachieving gifted child: Recognizing, understanding, and reversing underachievement*. Waco, TX: Prufrock Press.
- Siegle, D., & McCoach, D. B. (2005). *Motivating gifted students*. Waco, TX: Prufrock Press.
- Siegle, D., McCoach, D.B., & Roberts, A. (2017). Why I believe I achieve determines whether I achieve. *High Ability Studies*, 28(1), 59–72.

- Siegle, D., Rubenstein, L. D., Pollard, E., & Romey, E. (2010). Exploring the relationship of college freshman honors students' effort and ability attribution, interest, and implicit theory of intelligence with perceived ability. *Gifted Child Quarterly*, 54, 92–101.
- Silvia, P. J. (2008). Another look at creativity and intelligence: exploring higher-order models and probable confounds. *Personality and Individual Differences*, 44, 1012–1021.
- Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational and Psychological Review*, 27, 599–606. DOI 10.1007/s10648-015-9299-1
- Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, 4, 79–85.
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68–85.
- Simonton, D. K. (1994). *Greatness: Who makes history and why*. New York: Guilford Press.
- Smith, J. L., Cech, E., Metz, A., Huntoon, M., & Moyer, C. (2014). Giving back or giving up: Native American student experiences in science and engineering. *Cultural Diversity and Ethnic Minority Psychology*, 20(3), 413–429. DOI: 10.1037/a0036945
- Smith, G., & Garrison, G. (2005). The impending loss of talent: An exploratory study challenging assumptions about testing and merit. *Teacher's College Research*, 107(4), 629-653.

- Snyder, T., & Dillow, S. (2012). *Digest of education statistics 2011*. National Center for Education Statistics by the U.S. Department of Education. Washington, DC: Institute of Education Sciences.
- Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a basis for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology*, 31(2), 344–376.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.
- Sternberg, R. J. (1997). *Successful intelligence*. New York, NY: Plume.
- Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology*, 24, 359-375.
- Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge, United Kingdom: Cambridge University Press.
- Sternberg, R. J. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology*, 39(2), 189–202.
- Sternberg, R.J. (2006). The nature of creativity. *Creativity Research Journal*, 18, 87–98.
- Sternberg, R.J. (2010). Assessment of gifted students for identification purposes: New techniques for a new millennium. *Learning and Individual Differences*, 20, 327-336.
- Sternberg, R.J. & Coffin, L.A. (2010). Admitting and developing ‘‘new leaders for a changing world.’’ *New England Journal of Higher Education*, 24, 12–13.
- <http://dx.doi.org/10.1016/j.lindif.2009.08.003>

- Subotnik, R. F., Olszewski-Kubilius, P. & Worrell, F. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, 12(1), 3-54. DOI: 10.1177/1529100611418056
- Subotnik, R. F., Tai, R. H., Rickoff, R., & Alamarode, J. (2009). Specialized public high schools of science, mathematics, and technology and the STEM pipeline: What do we know now and what will we know in 5 years? *Roeper Review*, 32, 7-16. DOI: 10.1080/02783190903386553
- Tan, S., Erdimez, O., & Zimmerman, R.H. (2017). Concept mapping as a tool to develop and measure students' understanding in science. *Acta Didactica Napocensia*, 10(2), 109-122. ISSN 2065-1430
- Tan, S. & Maker, C. J. (2015). *The predictive validity of the Raven Progressive Matrices for identifying academic achievement of the Diné children*. Unpublished Manuscript. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.
- Teasdale, T.W. & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, 36, 121-126. <http://dx.doi.org/10.1016/j.intell.2007.01.007>
- Torrance, E.P. (1971). Are the Torrance tests of creative thinking biased against or in favor of disadvantaged groups? *Gifted Child Quarterly*, 15, 75–80.
- Torrance, E.P. (1973). Non-test indicators of creative talent among disadvantaged children. *Gifted Child Quarterly*, 17, 3–9.
- Torrance, E. P. (1977). *Creativity in the classroom*. Washington, DC: National Education Association.

- Torrance, E. P. (2008). *Torrance Tests of Creative Thinking: Norms-Technical manual, verbal forms A and B*. Bensenville: Scholastic Testing Service.
- Troiano, A.B. & Bracken, B.A. (1983). Creative thinking and movement styles of three culturally homogeneous kindergarten groups. *Journal of Psychoeducational Assessment*, 1, 35–46.
- Urban, K. K., & Jellen, H. G. (1996). *Test for Creative Thinking–Drawing Production (TCT-DP) Manual*. Frankfurt: Swets Test Services.
- Vanides, J., Yin, Y., Tomita, M., & Ruiz-Primo, M. A. (2005). Concept maps. *Science Scope*, 28(8), 27-31.
- Van Tassel-Baska, J. (2002). Planning effective curriculum experiences for gifted learners. *Understanding Our Gifted*, 15(1), 6–8.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wai, J., Lubinski, D., & Benbow, C. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817-835 DOI: 10.1037/a0016127
- Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J.H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, 102(4), 860–871.
- Wallach, M. A. (1971). *The intelligence-creativity distinction*. Morristown, NJ: General Learning Press.
- Wallach, M.A. (1976). Tests tell us little about talent. *American Scientist*, 64(1), 57–63.
- Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic

- engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, 28, 12–23.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
- Weisberg, R. W. (2006). *Creativity: Understanding innovation in problem solving, science, invention, and the arts*. Hoboken: Wiley.
- Weisberg, R. W. (2015). Expertise, nonobvious creativity, and ordinary thinking in Edison and others: Integrating blindness and sightedness. *Psychology of Aesthetics, Creativity, and the Arts*, 9(1), 15-19. <http://dx.doi.org/10.1037/aca0000001>
- Westby, E.L. & Dawson, V.L. (1995). Creativity: Asset or burden in the classroom. *Creativity Research Journal*, 8(1), 1-10.
- Whitmore, J.R., & Maker, C.J. (1985). *Intellectual giftedness in disabled persons*. Austin, TX: Pro-Ed.
- Winner, E. (2000a). The origins and ends of giftedness. *American Psychologist*, 55(1), 159–169.
- Winner, E. (2000b). Giftedness: Current theory and research. *Current Directions in Psychological Science*, 9(5), 153–156.
- Wu, I-C., Pease, R., & Maker, C. J. (2019). Students' perceptions of a special program for developing exceptional talent in STEM. *Journal of Advanced Academics*, 30(4) 474–499. DOI: 10.1177/1932202X19864690
- Xie, Y., Fang, M., & Kimberlee, S. (2015). STEM education. *Annual Review of Sociology*, 41, 331-357. DOI:10.1146/annurev-soc-071312-145659
- Zimmerman, R., Maker, C. J., Gomez- Arizaga, M.P., Pease, R. (2011). The use of concept maps in facilitating problem solving in earth science. *Gifted Education International*, 27(3),

274-287.

Zimmerman, R., Maker, C. J., & Alfaiz, F.A. (2019). *Culturally responsive assessments of life science abilities: Development, field testing, and implementation*. Manuscript submitted for publication.

Tables and Figures for Version 3

Table 1

Maker and Schiever Problem continuum modified from Getzels and Csikszentmihalyi's research

Problem	Type	Problem		Method		Solution	
		Presenter	Solver	Presenter	Solver	Presenter	Solver
Closed	I*	Specified	Known	Known	Known	Known	Unknown
	II*	Specified	Known	Known	Unknown	Known	Unknown
Semi-open	III	Specified	Known	Range	Unknown	Known	Unknown
	IV	Specified	Known	Range	Unknown	Range	Unknown
Open-ended	V	Specified	Known	Unknown	Unknown	Unknown	Unknown
	VI*	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Note: Presented in “Problem Continuum” by Maker, J., & Schiever, W. (2010). *Curriculum*

development and teaching strategies for gifted learners (3rd ed.). Austin, TX: Pro-Ed.

*Types I, II, and VI were Types I, II, and III in Getzels and Csikszentmihalyi's continuum.

Table 2*Core competencies, problem types, and examples of tasks in M2 STEM assessments*

Assessment	Core Competencies	Problem Type/Task
Concept Maps	Understanding of the complexity of concepts and their interrelationships	Using the concepts given, create a hierarchical map showing how the concepts are related. You can add other concepts and examples. Scoring: Propositions (Closed), Hierarchy (Semi-Open), Crosslinks (Open), Examples (Closed, Semi-Open, Open)
Math Problem Solving	The use of abstract models, numbers, mathematical figures, and objects that symbolize abstract ideas; the ability to handle long chains of reasoning, to discover a promising idea and draw out its implications, and to recognize significant problems and solve them	Closed: Solve this equation: $12x + 3 = 5$; $x = \underline{\hspace{1cm}}$ Semi-Open: Create as many fractals as possible. Open: Write as many problems as possible that would fit a graph, and then solve the problems created.
Life Science Problem Solving	Observation through the five senses; classifying, identifying, and collecting data in both natural and man-made environments; keen observation; understanding of, or interest in, the natural world; accurate perception of shapes, forms, and texture; and comparison of the different characteristics of different species	Semi-Open: Make groups of flowers or insects based on similarities, give each group a name, and explain the reason(s) for the name. Open: Make an ecosystem that shows interrelationships and interdependencies. Label and describe the elements of the ecosystem. Closed: Scoring includes accuracy of scientific ideas in both the semi-open and open tasks.
Mechanical Technical Problem Solving	To understand, create, and repair machines or other devices that perform or help perform human tasks	Closed: Make a gear box. Semi-Open: Make one of the vehicles. Open: Make a machine that moves.
Spatial Analytical Problem Solving	To see things accurately and clearly through one's eyes and seeing images clearly in one's mind; then producing, describing, or modifying them in a visual or other form; to make transformations and modifications on one's initial perceptions and to re-create aspects of a visual experience, even when the physical stimuli are not present	Closed: Solve puzzles that have only one correct way to be solved. Semi-Open: Solve puzzles that have several correct ways to be solved. Open: Make a rhombus with as many pieces as you can.

Table 3*Ethnicity of students selected by existing methods (M1) and new assessments (M2)*

	Ethnicity					Row Totals
	White	Hispanic	Navajo	African American	Asian American	
M1 Students	8	6	1	2	3	20
M2 Students	2	5	13	1	2	23
Column Totals	10	11	14	3	5	43

Table 4*Results of Chi-square analysis of Gender by Group*

	Gender		Row Totals
	Female	Male	
M1 Students	13	7	20
M2 Students	13	10	23
Column Totals	26	17	43

Note. 0 cells (.0%) have an expected count less than 5. The minimum expected count is 7.91.

Table 5*Results of Chi-square analysis of Ethnicity by Group*

	Ethnicity					Row Totals
	White	Hispanic	Native American	African American	Asian American	
M1 Students	8	6	1	2	3	20
M2 Students	2	5	13	1	2	23
Column Totals	9	11	15	3	6	43

Note. 5 cells (50.0%) have an expected count less than 5. The minimum expected count is 1.40.

Table 6*Results of Chi-square analysis of Primary Language by Group*

	Language				Row Totals
	English	Spanish	Navajo	Other (Tagalog, Vietnamese, Mai Mai)	
M1 Students	18	1	0	1	20
M2 Students	17	3	1	2	23
Column Totals	35	4	1	3	43

Note. 6 cells (75.0%) have an expected count less than 5. The minimum expected count is .47.

Table 7*Results of Chi-square analysis of Highest Level of Parent Education by Group*

Parent Education					
	Middle School	High School	BA Degree	Master's or PhD Degree	Row Totals
M1 Students	1	8	5	6	20
M2 Students	0	15	5	3	23
Column Totals	1	23	10	9	43

Note: 4 cells (50.0%) have an expected count less than 5. The minimum expected count is .47.

Figure 1. Problem solving as a unifying concept for assessment of general abilities

(intelligence), domain specific knowledge and abilities, and creativity relevant processes and abilities; with personal and environmental factors as influences on all abilities, skills, and knowledge

Figure 2. Comparison of ratings of M1 and M2 students on measures of domain specific and domain general knowledge and skills (GPA, concept maps, and mathematical problem solving)

Note: UW GPA-Unweighted Grade Point Average; LSCM-Life Science Concept Map; PCM-Physics Concept Map; Math-Mathematical Problem-Solving Assessment; M1-Students selected by conventional methods (GPA, teacher recommendation, and self-statement); M2-Students selected by new methods (concept maps, mathematical problem solving, and performance-based assessments)

Figure 3. *Comparison of ratings of M1 and M2 students on performance-based assessments of creative problem solving*

Note: PBA-Performance-Based Assessment; M1-Students selected by conventional methods (GPA, teacher recommendation, and self-statement); M2- selected by new methods (concept maps, mathematical problem solving, and performance-based assessments)