Culturally Responsive Assessment of Physical Science Skills and Abilities:

Development, Field Testing, Implementation, and Results

Fahad S. Alfaiz, Ph.D., King Saud University

Randy Pease, M.A., and C. June Maker, Ph.D., Litt.D., The University of Arizona

Author Note

Fahad S. Alfaiz, Self-Development Skills Department, King Saud University, Riyadh, Saudi Arabia; Randy Pease and C. June Maker, Department of Disability and Psychoeducational Studies, The University of Arizona, Tucson, AZ, USA

Correspondence concerning this article should be addressed to Fahad S. Alfaiz, Self-Development Skills Department, King Saud University, Riyadh, Saudi Arabia. E-mail:

falfaiz@ksu.edu.sa

This research was supported by the National Science Foundation, Grant #1321190, Cultivating Diverse Talent in STEM, PI Uwe Hilgert and Co-PIs C. June Maker, Frans Tax, and Martha Lindsey, University of Arizona, and Harold Begay, Tuba City Public Schools. It was approved by a Native American Research Review Board, #NNR-13.166. We acknowledge that there is no conflict of interest related to this research.

Publication Information

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*, 31(3), 298-328. <https://doi.org/10.1177/1932202X20920572>

Abstract

During the Cultivating Diverse Talent in STEM (CDTIS) project, a team of scientists, teachers, and a researcher developed a performance-based assessment of high school students' creative problem solving skills and ability to apply physical science principles in practical ways. It was one of six measures to identify exceptionally talented students. Students identified using conventional methods (M1), with an average GPA of 3.93, had an average rating of 2.95 on a 5-point scale on the mechanical-technical assessment. The M2 students, who were from schools with high percentages of Hispanic, American Indian, and low SES students, had an average GPA of 3.07 and an average rating of 3.27, demonstrating that this assessment can be an important way to change the cultural and economic balance of students identified as exceptionally talented in STEM. Other researchers are encouraged to examine the validity of the mechanical-technical assessment to identify exceptionally talented students in different groups.

Keywords: STEM, mechanical-technical, performance-based assessment, exceptional talent, creative problem solving, physical sciences

Culturally Responsive Assessment of Physical Science Skills and Abilities:**Development, Field Testing, Implementation, and Results**

The applications of Science, Technology, Engineering, and Mathematics (STEM) have become part of our daily lives. In the labor market, the demands for STEM occupations have increased over non-STEM jobs (U.S. Department of Commerce [USDC], 2011, 2017), and the average wages of specialists in STEM are much higher than those in most other disciplines (Sawchuk, 2018; USDC, 2017). The National Science Board (NSB, 2010) has called for a sustained effort to stop importing STEM talent and develop STEM innovators: “creators of significant breakthroughs or advances in scientific and technological understanding” (p. vii). One of the main recommendations of the NSB was for researchers to “cast a wide net”, identifying and nurturing many types of talent in all demographic groups of students. Clearly, finding students from all groups and levels of society, developing their problem solving skills, and preparing them to be professionals in STEM are important to the nation’s sustainable development and the students’ future lives.

Creative Problem Solving in Physical Science

Creativity is defined as novel and appropriate ways to produce a product, respond to a question, or solve an open-ended problem (Amabile, 2013). It has four components; three of which are related to the individual: domain relevant skills, creativity-relevant processes, and task motivation. The fourth component is related to the individual’s social environment (Amabile, 1983, 1996, 2013). Domain relevant skills are foundations of creativity and include “knowledge, expertise, technical skills, intelligence, and talent in the particular domain” (Amabile, 2013, p. 135). Creativity-relevant processes are cognitive and personal characteristics. An example of creativity-relevant processes is the ability to synthesize information in flexible categories. Task

motivation includes both internal and external factors that play important roles in doing a task. The social environment plays a vital role in enhancing or inhibiting creativity (Amabile, 2013).

The current vision of science education is that students should learn how to solve complex and ill-defined problems in creative ways rather than solve them in conventional ways (Wang et al., 2008). Employing creative problem solving in education helps students to improve their self-confidence and academic performance (Elvan et al., 2010) and is an important component in the development of STEM innovators (Dino, 2017). For all these reasons, applying higher level thinking skills and creative problem solving strategies are the aims of science educators (DeHaan, 2009).

Physical Science Majors and Ethnicities

A degree is required for physical science careers. Among physical science majors, the percentage of individuals from different ethnic groups is not proportional to their numbers in the general population. The imbalance exists in both universities and the workforce. For example, in 2012 approximately 65.3% of bachelor's degree students in physical science majors were White, 12 % were Asian, 5.9% African-American, 9.6% Hispanic, 0.4% American Indian or Alaskan Native, and 6.8% were other races or unknown (National Science Foundation, 2015). In the workforce, Whites represented 68% of all physical science careers, while 16% were Asians, 8% African-Americans, 6% Hispanics, and 2% other cultural and ethnic groups including American Indians (Funk & Parker, 2018). One of the main reasons for the disparities between these groups in physical science jobs is the academic achievement gaps leading to different percentages being served in programs at higher levels of education (Carter & Darling-Hammond, 2016; Duncan & Murnane, 2014; Gonzalez & Kuenzi, 2012).

In K-12 settings, disparities have existed between the number of White, Hispanic, African American, and American Indian students in general education programs and the number in special programs for exceptionally talented students and students with disabilities (Donovan & Cross, 2002; U.S. Department of Education, Office for Civil Rights, 2016). In 2012 in the state where the project was located, for example, the population of Asian and Pacific Islanders in K-12 schools was 2.9% and 7.2% of students identified as exceptionally talented, Whites consisted of 43% of the total students and 56% of those identified as gifted. In contrast, Hispanics consisted of 44% of the total students in schools and only about 27% of those who were identified as gifted, 5.4% of African-Americans were in schools but only 2.8% of students identified as exceptionally talented, and American Indians consisted of 5.1% of the population in schools, but only 1.3% of those identified as exceptionally talented (Arizona Minority Education Policy Analysis Center [AMEPAC], 2013). In contrast, the percentage of students other than White was higher in programs specifically provided for students with disabilities (AMEPAC). In the state where the project was located, for instance, 15% of American Indian students were enrolled in programs for students with disabilities compared with 11.4% of White students. Test bias and inequity in general education were some factors that played important roles in the disparity between the percentage of students of minority groups in various educational programs (Skiba et al., 2008).

The Need for Developing a new Assessment

Conventional methods of identification for special programs for exceptionally talented students have not been considered to be appropriate to measure students' abilities and skills because of the heavy emphasis on standardized achievement and aptitude tests (Shute et al., 2016). Often, only one method of assessment is used, and the measures include only a few

cognitive abilities (Flanagan & Ortiz, 2001; Gardner, 1983, 1992). Important factors such as motivation, effort, and creativity are not considered (Maker, 1996, 2005; Renzulli, 1978).

Use of conventional methods to assess students from different cultures and socioeconomic levels may negatively influence their future academic lives because the assessments do not include their strengths in creativity and problem solving (Glover, 1976; Ivcevic & Kaufman, 2013; Kaufman, 2006; Kaufman et al., 2004; Torrance, 1971), and their scores on widely used achievement tests are lower than the scores of students from White and Asian students and those from high socioeconomic levels. For example, on the state science test, approximately 76.2% of Asian and Pacific Islander and 73.9% of White students passed the science test whereas only 45.8% of African American, 45% of Hispanic, and 33.1% of American Indian students passed the test (AMEPAC, 2013). In the United States, the percentage of minority groups has increased over the years; the percentage of students of color is projected to reach 50% of the student population by 2050 (Carter & Darling-Hammond, 2016). Thus, the number of exceptionally talented students who are overlooked will increase if changes are not made in methods for identifying them.

The disparity among socioeconomic (SES) levels has been another factor in students' performance in school (Duncan & Murnane, 2014). The time spent on extracurricular and other educational programs among students from high and middle SES levels was higher than that of students from low SES levels (Oakes et al., 2013). An important factor that prevented families from low SES levels from registering their children in extracurricular or other educational activities was the decrease in yearly income since the 1970s when compared with the income of families from high and middle SES levels (Duncan & Murnane). For all these reasons, the Cultivating Diverse Talent in STEM (CDTIS) Project was funded by the National Science

Foundation. During the CDTIS project, the education research team developed an assessment with the potential to identify the strengths of students from low SES groups; an assessment that would be a “fair” measure, appropriate for students from all cultures and economic levels. Both poverty and cultural differences were factors that affected the disparity between the percentages of students from different demographic groups in various educational programs, particularly those designed for exceptionally talented (Skiba et al., 2008).

Because developing an assessment to measure competence in physical science is important, the process of developing valid measures requires considering different dimensions of knowledge and abilities (Shen et al., 2017). Until recently, science assessments did not include the following three dimensions of science learning: (a) the core ideas of the science disciplines such as physical sciences, engineering, and technology; (b) “the practices through which scientists and engineers do their work”; and (c) the “key crosscutting concepts” that connect science majors (National Research Council, 2014, p. Sum 1).

Conceptual Framework for Developing the Physical Science Assessment

During the implementation of the CDTIS project, the research team used a new definition of exceptional talent to guide the development of instruments and the identification of students to be served:

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.* (Maker, forthcoming-a, p. 21)

In addition to the mechanical-technical performance-based assessment described in this article, three new assessments were developed, one existing assessment was modified, and one of the original DISCOVER assessments was included in the battery of instruments used to assess and identify students for a special internship program on the campus of an R1 University in the southwestern USA. Assessment of physical science ability included two parts: (a) concept maps (Maker & Zimmerman, 2019) to assess the first component of exceptional talent and (b) the performance-based assessment of creative problem solving described in this article (mechanical-technical) to assess the second component of the definition of exceptional talent.

People need to develop problem solving skills to face a variety of challenges in their lives and make decisions about ways to solve them (Maker, forthcoming-a; Zimmerman et al., 2019), and STEM innovators need to create advances in technology and engineering (Dino, 2017). To solve these problems and create advances, people need creativity and the motivation to create solutions to pressing problems in effective ways (Maker). In school, students use problem solving skills to make sense of knowledge they acquire rather than simply remembering information (Elvan et al., 2010). For these reasons, the research team used the idea of problem solving and three major theories to develop the physical science assessments: The Theory of Multiple Intelligences (Gardner, 1983), the Theory of Successful Intelligence (Sternberg, 1997, 1999), and the Prism of Learning (Maker & Anuruthwong, 2003; Maker et al., 2015). In addition, the developers of the assessment combined the ideas of first and second-order knowledge (Gardner, 1992; Maker, 2005) and a continuum of problem types (Getzels & Csikszentmihalyi, 1967, 1976; Maker & Schiever, 2005, 2010) when they created problem solving tasks for the physical science assessments.

The mechanical-technical performance-based assessment of the second component of the definition of exceptional talent is a measure of first-order knowledge (Gardner, 1992). Students acquire knowledge and skills from different sources other than formal education, especially from interacting with their environments and the contexts in which they live (Gardner; Maker, 2005). The concept map assessment of the first component of the definition of exceptional talent is a measure of second-order knowledge (Gardner), which is related closely to what students learn in school. Rather than assessing only whether students have acquired particular knowledge, it is an assessment of the ways they think about the concepts they have learned (Austin & Shore, 1993; Erdimez et al., 2017; İngeç, 2009; Lopez et al., 2011; Novak & Canas, 2006; Ruiz-Primo et al., 2001).

The continuum of problems adapted from the work of Getzels & Csikszentmihalyi (1967, 1976) and modified by Maker and Schiever (2005, 2010) was the framework for developing problem solving tasks. In this framework, tasks are on a continuum that ranges from *closed*, with a well-defined problem that has a known or specified method and a right answer; to *semi-open*, with a well-defined problem, more than one method, one or a defined range of solutions; and *open-ended*, with either a well-defined problem or a problem that is structured or defined by the problem solver, methods determined by the problem solver, and an infinite number of solutions, which are created and selected by the problem solver. The closed problems require mainly knowledge and convergent thinking while the open-ended problems require creative and flexible thinking. Semi-open problems require both types of thinking. In the mechanical-technical assessment, the closed problem was to make a gear box using the materials given, the semi-open problem was to choose one of two vehicles and make it without being given information about how to put the pieces together. The open-ended problem was to make a machine, and no

restrictions were given except that it needed to be different from the vehicle they made and of their own design. Additional information about the conceptual framework used to develop all assessments in the CDTIS project can be found in the article by Maker (forthcoming-a).

Method

The mechanical-technical assessment was developed by the education research team as part of the CDTIS project. The team consisted of an education researcher with extensive experience developing assessments to identify exceptional talent in diverse groups; a scientist with experience in education; a public-school teacher with extensive experience in teaching gifted students, teacher education, and development of assessments; a post-doctoral scholar with experience in math and science education for gifted students at several levels; and graduate students in education of gifted students. The CDTIS Project was designed to identify and nurture exceptionally talented students in STEM. In this project, the education research team developed four new assessments: life science problem solving (Zimmerman et al., 2019), life science and physical science concept maps (Maker & Zimmerman, 2019), and the mechanical-technical performance-based assessment. One assessment (mathematical problem solving) was modified to include a greater variety of problems (Bahar & Maker, 2020), and one performance-based assessment from the existing DISCOVER high school battery, spatial analytical problem solving (Maker, forthcoming-b), was included in the set of instruments used to select students for a special internship program on the campus of an R1 university in the southwestern USA.

Development

Belief in the need for a separate assessment of the mechanical-technical aspects of spatial ability was the result of observations during the original DISCOVER assessments (Maker, 2005). Because the original version of the high school DISCOVER assessment included a wide variety

of spatial tasks based on Gardner's definition, observers had many opportunities to watch the different ways students demonstrated their spatial ability. The original spatial assessment included many art supplies, such as paints, brushes, pencils, charcoal, clay, and a mechanical construction option using a material called Capsela. It was a construction kit consisting primarily of gears and motors in spherical plastic capsules that could be connected to form various static or dynamic toys suitable for land or water. The capsules also included an electric motor, gears, switches, and other items; materials included wheels, propellers, lights, wires, a battery holder, and other supporting parts. In the spatial assessment, students created visual depictions of nature, various items in pictures, and other products using various materials. Photographs included items that could be made with the Capsela mechanical materials.

Observers noticed that students often were divided into two groups, those who enjoyed visual arts, which included painting, drawing, sketching, molding, and sculpting, and a smaller group of students who enjoyed the mechanical-technical Capsela materials. The students who created constructions using the Capsela materials described interests that were more aligned with mechanical-technical areas of visual spatial abilities. This formed the basic idea for the creation of the mechanical-technical assessment.

The CDTIS research team found that Capsela was no longer available and materials needed updating to better reflect the current level of technology. After reviewing many robotic materials, such as the LEGO Mindstorms EV3 systems, the team settled on Thames and Kosmos Remote-Control Machines Science Experiment Kits (Figure 1). The materials included a variety of gears, wheels, chain links, plastic construction pieces with connectors, infrared remote controls, and motors to construct vehicles and machines using the remote controls. The parts could be used to construct complex vehicles and numerous configurations of mechanical items.

Each student was given a container of materials that included gears, remote-control devices, wheels, connectors, and a variety of lengths of pieces that could be used for construction.

Students were then given tasks that were assessments of the mechanical-technical abilities in the Prism model, different types of problem solving, and first order knowledge.

Assessment Square

In the development process, the team used the assessment square proposed by Ruiz-Primo and colleagues (Ruiz-Primo, 2003, 2007, 2009; Shavelson et al., 2002). The assessment square was comprised of four components that included construct, observation, assessment, and interpretation (Ruiz-Primo, 2009).

Construct. The first step in developing the assessment was to define what would be measured. The construct definition “drives the tasks or situations, response demands, and scoring system that comprise a learning assessment” (Shavelson et al., 2002, p.5). The purposes of each task were to (a) measure specific problem solving skills, (b) measure core competencies of the ability being assessed, (c) be developmentally appropriate for high school students, (d) measure first-order knowledge, and (e) elicit behaviors that differentiate among different levels of ability.

The overall definition of the construct came from the Prism of Learning Model (Maker & Anuruthwong, 2003, in Maker et al., 2015, p. 92): Mechanical-Technical ability is the ability “...to understand, create, and repair machines or other devices that perform or help perform human tasks.” Some core competencies from Gardner’s (1983) theory of Multiple Intelligences and Sternberg’s Theory of Successful Intelligence (1997, 1999) were guides in the development of tasks for the assessment. Core competencies in logical-mathematical ability were recognizing patterns, recognizing significant problems, and then solving the problems; and core competencies Gardner identified for spatial ability were perceiving the visual world accurately

and making transformations and modifications of visual elements. From Sternberg's theory, aspects of analytical, creative, and practical intelligences were included. In addition, the research team considered mainly first-order knowledge (Gardner, 1992) as the type of knowledge being assessed.

Observation. The second step in the assessment square was to collect initial data about behaviors that could be observed. The research team first tried the tasks of the mechanical-technical assessment on themselves, friends, their children, and students in classrooms to ensure these tasks were engaging and elicited observable behaviors that could distinguish levels of ability. Some tasks were modified based on these observations. Modifications of the mechanical technical assessment involved time frames given to some of the tasks. For example, in the gear construction, Task 1, too much time was allotted initially, as students were able to complete the closed task of gear construction in ten minutes or less, and not the 25 minutes originally given. Another time modification was made with the vehicle construction, Task 2. The researchers found that 50 minutes was not required for most students to complete this task and changed the time limit to 25 minutes. If a student took longer than 25 minutes, he or she was allowed to complete the construction before moving on to Task 3, the open-ended construction; however, observers were asked to encourage the students to complete their vehicles so they would have time to complete their constructions (Task 3). A final modification to time was made in the open-ended construction. Students were allowed the time required in Task 3, deconstruction, to be added to their final construction. If this happened, the observer deconstructed the students' products, allowing students the extra five minutes for Task 3.

Modifications were made to the scoring criteria included in the rubrics. After many student observations, the researchers included specific movement, side to side, forward, backward, and

circular motion to the criteria for a rating of Definitely. Although the use of two motors was originally part of the Definitely criteria, observers agreed to add one motor with a unique design to the scoring criteria. In the Wow category, examples were added, such as machines that were assigned this rating were those made by students who (a) used chains that worked to power gears or wheels, (b) replicated the motor using the gears in the kit so that the machine worked exactly like the inside of the motor, and (c) used controlled steering.

Assessment. The third step was to implement the assessment tasks with a larger population. The research team field-tested the assessment over a two-year period with 334 students in partner high schools.

Interpretation. The last step was to develop a scoring system to rate students' products and to create a systematic method to interpret the results. The team returned to earlier steps as needed during the field-testing and implementation. During the development and field testing, the research team made changes to the assessment in response to the comments and suggestions of observers who were experts, scientists, and science teachers. In addition, science teachers and the science department chairs in partner schools observed the assessments and made suggestions for improvement or comments about the appropriateness of the assessment. A scientist, a teacher of physical science, and experts in education of gifted students were members of the assessment teams. They developed and applied the rubrics and reached consensus on all the ratings to be assigned. Thus, the construct validity of the assessment was established for the purpose for which it was used in the CDTIS project with the populations included.

Assessment Tasks

Gear Construction (Task 1). This was a closed problem type. Each student had a bag that contained a picture card (Figure 2) and various sizes of gears, beams, and drive axles.

Vehicle Construction (Task 2). This was a semi-open problem type. Students were shown pictures of two vehicles, and each student was asked to make one of the two models (Figure 3). They were shown the final product and the parts needed to construct it, but not the methods to use to put the parts together. Model 2 was more complex and more difficult to make than Model 1.

Machine Construction (Task 3). This was an open-ended problem type. Students were asked to make a machine that moved by its own power. They had the flexibility to make any machine they wanted to make using the materials available; thus, the problem was not well-defined (Maker & Schiever, 2010). The construction needed to be different from the vehicle they had made in the previous task and of their own design, not a copy.

Observer Instructions

Observer instructions were developed as a guide to make certain the assessment was administered consistently. Observation note forms were developed to help observers notice and record students' behaviors while they were working on the tasks. Observation note forms had spaces for the observer to write notes for up to five students and included checklists of behaviors in two domains: (a) mechanical-technical problem solving behaviors, such as invents new strategies, construction driven by belts and gears, attention to design, use of balance, understanding of rotation of gears, use of force, and understating of power and force; and (b) motivational characteristics, such as stays on task, asks questions, continuously working, and focused on task.

Field Testing and Revision

The field tests of the mechanical-technical assessment were conducted four times over a period of one month at four high schools in the southwestern United States. An average of four

observers participated in each field test. Students sat around tables in groups of 3 to 5 and one observer was assigned to each table.

Participants

Fifty-eight high school students participated in field tests of the assessment at four high schools (Table 1). The number of students in these schools ranged from 94 to 2245, and most of the students (71%, 90%, 92%, 99%) received free or reduced-price lunches. Almost all of the students in three schools were American Indian (97%, 98.4%, and 98.6%), and the majority of students in one school was Hispanic (83%). These schools were located in poverty areas with an unemployment rate ranging from 7.7% to 45.8%. Different academic programs were offered in the larger schools, such as Honors and Advanced Placement (AP) classes and programs for exceptionally talented students. Information about the students is included in Table 2. A signed assent form from students and a consent form from parents were collected prior to the field tests.

Field Test 1

The mechanical-technical assessment was tested at School A with 19 students and administered by certified observers who were part of the research team or graduate students in special education for gifted students. When students sat at their tables, each observer distributed a container to each student that included a variety of materials (e.g., remote control, geared motor, receiver, and separator tool). Examples of mechanical-technical materials are in Figure 1. The observer gave directions for the tasks and took photographs of students' products during the vehicle and machine construction. The observer also interviewed each student about his/her machine and asked the student to demonstrate its movement after the construction task was completed. The mechanical-technical assessment was designed to last for approximately two hours. However, based on this field test, the research team decided to modify the time allowed

for each component of the assessment: (a) demonstration and exploration, 5 minutes; (b) gear construction, 5 minutes; (c) vehicle construction, 25 minutes; (d) machine construction, 50 minutes. The amount of time allowed for the entire assessment was 90 minutes, so the final 5 minutes could be used for machine construction for students who needed additional time, and could be used for deconstruction of machines. If students did not have time for deconstruction, the observer took the machines apart for the students.

Field Tests 2, 3, and 4

The mechanical-technical assessment was piloted at School B with 15 seniors, at School C with 15 seniors, and at School D with 9 seniors. All procedures and scoring were the same as in Field Test 1. No revision was made during field tests 2, 3, and 4.

Implementation

The mechanical-technical assessment was part of the Cultivating Diverse Talent in STEM (CDTIS) project. The purpose of the CDTIS project was to increase the number of students from diverse ethnic and economic groups identified as exceptionally talented, and to provide opportunities for them to further develop their knowledge and abilities in STEM areas, eventually leading to a more diverse work force and more equitable representation from different groups.

Participants

The mechanical-technical assessment was conducted with students in grade 11 in two groups (M1 and M2) in a state located in the southwestern USA (Table 2). The M1 group included 20 students from different cultural groups and socioeconomic (SES) levels in urban and rural areas. They were chosen by a committee based on three conventional measures of achievement: (a) overall grade point average (GPA), (b) teacher recommendations, and (c) student self-statements. Students who who were identified by the selection committee were

invited to participate in the research. The research team attempted to choose those who were from areas of the state similar to those of the partner schools; however, this was not possible in many cases because of the pool of students. The M1 students who agreed to participate in the M2 assessments before, during, and after the program were included in the research. Although they participated in the new assessments, their scores on the M2 assessments were not considered in the selection process. M1 students came from a variety of schools and types of schools: most were from public schools in urban areas, some 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. They and other students selected using the conventional methods participated in a special internship program in the laboratories of scientists on the campus of an R1 university along with the M2 students.

The second group (M2) included 23 students who were selected from the 334 students from partner high schools who participated in the assessments (Table 3): four schools in Year 1 (A, B, C, and D) and three schools in Year 2 (A, B, and C). School D did not participate in the study in Year 2 due to the time-consuming task of obtaining parental permission. As noted in an earlier section, all partner schools were located in poverty areas with an unemployment rate that ranged from 7.7% to 45.8% and most of the students (71%, 90%, 92%, 99%) received free or reduced-price lunches. To select M2 students to participate in the summer internship program, their results on the mechanical-technical assessment were combined with the results of the other

five assessments: (a) math problem solving assessment (Bahar & Maker, 2020), (b) spatial analytical performance assessment (Maker, forthcoming-b), (c) physics concept maps (Maker & Zimmerman, 2019), (d) life science concept maps (Maker & Zimmerman, 2019), and (e) the life science performance assessment (Zimmerman et al., 2019). Student performance on all assessments was considered for final nomination. For example, any students who scored Definitely or Wow on all assessments were selected first, then students who scored Definitely or Wow on five of the assessments were chosen second and so on. If other placements were available, students with the highest scores in specific areas or overall were included.

Description of Forms and Tasks

After the field testing was completed, the research team finalized forms to be used in the implementation phase. The final forms included one for recording notes during observations and another that included observer instructions. The form for recording notes during observations enabled each observer to record observations for up to five students. At the top, space was provided to write information such as the name of the school, the observer, the date the assessment was conducted, and the time the assessment began and ended. Codes were provided for observers to note descriptions of students' products easily and quickly while observing students (e.g., F = flexible use of materials, CX = very complex, M = movement). Space was also provided to write comments about each student's performance while the student worked on the assessment tasks. At the bottom of the form was a checklist to record each student's behaviors in two domains: (a) mechanical-technical problem solving behaviors, such as invents new strategies, attention to design, use of levels and balance, use of rotation of gears, use of power supply, and use of inclined planes and (b) motivational characteristics, such as stays on task, asks questions, continuously working, and focused on task.

Specific observer instructions were developed so that observers implemented the assessment in a consistent way across groups. After students sat at their tables, the observers first introduced themselves to the students and recorded their names on the observation notes form and used the space provided to make a diagram showing where students were sitting in relation to the observer. The observer recorded the beginning and ending time at the top of the form. At the end of each task, the observer asked the students to take apart their constructions and return the materials to the bags in their kits. When the observer talked to the students about what they made, the observer did not give it a name, but referred to it as “your construction”.

Before the students began the assessment, observers demonstrated the basic functions of the remote control, geared motor, and receivers. They encouraged students to explore how these items might work together. Students were then allowed to explore all construction materials and the ways they could be used. Both of these activities took approximately ten minutes to complete.

Gear Construction (Task 1)

The observer distributed plastic bags with gear materials to each student, showed the picture of a four-gear construction (Figure 2), gave the picture card to each student and asked students to make the gear box. When students built their gear boxes, an observer checked each student’s work to ensure that the student understood how the gear box worked. Approximately five minutes were allotted for this task.

Vehicle Construction (Task 2)

The observer showed students two pictures of vehicles and asked them to choose one vehicle to make. When the students chose one of the models, they were not allowed to change to the other vehicle. After the students made their vehicles, the observer at the students’ table took a

photograph of each vehicle, making certain a nametag or other identifying information was with it and could be seen in the photograph, then the student demonstrated the vehicle's movement. Students took their constructions apart at the end of this task. The vehicle construction task lasted approximately 25 minutes.

Machine Construction (Task 3)

After students finished the vehicle and took it apart, observers asked students to make a machine that moves by its own power. They encouraged students to make an original construction using as many materials as needed. When each student completed a construction, the observer took a photograph, making certain a nametag or other identifying information was with it and could be seen in the photograph. The observer asked the student to demonstrate how the machine moved, then interviewed the student and asked him or her to tell about the construction. The time allocated for the machine construction was approximately 50 minutes. At the end of the assessment, the observer asked students to take apart their constructions carefully and to return all materials to the plastic bags in their kits. If students did not have time for deconstruction, the observer took apart the machines.

Ratings

Five levels were used to rate the vehicle construction (Task 2) and machine construction (Task 3): *Unknown*, *Maybe*, *Probably*, *Definitely*, and *Wow*. These ratings are explained in depth in another publication (Maker, forthcoming-a). After each observer listened to and transcribed the audio recordings of interviews of students in his or her group, the observer gave a preliminary rating for each student's vehicle and machine based on a rubric that included characteristics of both the machine and vehicle (Table 4). When each observer completed this process, all observers met together in one room and did the following: (a) each observer placed

his or her photographs of both constructions (vehicles and machines) on a table in rating categories based on his or her preliminary rating, (b) all observers in the group reviewed and discussed their observations and interviews related to each photograph of a product (vehicles and machines) and then reached consensus on the ratings to assign to all students, and (c) a final rating was written on each photograph, then a final score was recorded in the database.

A student's construction might not have all the characteristics listed for each level, but it should have many of these characteristics. Each level of rating needed to include the positive characteristics of the levels below it. For example, stability was essential for the rating of Probably, so an unstable construction was rated Definitely only if the unique characteristics far outweighed its instability. Criteria for scoring the mechanical-technical assessment are presented in Table 4. Examples of students' constructions (Task 3) rated Unknown, Probably, and Wow are presented in Figures 4, 5, and 6. In figure 4, a student was rated "Unknown" because he did not complete the machine. The observers did not know what the student was trying to build and it lacked stability. In the interview, the student did not explain the idea of his construction. In figure 5, "Probably" was given because the student had a design that was different from the vehicle made in Task 2 and three motors were included in his construction. The construction had five gears that moved forward and backward. In figure 6, "Wow" was given because four wheels and two gears were used in the construction and it included a hand that rotated 360 degrees. In addition, the construction had three motors, was complex, had a number of functions, and was unique. The gears of the construction were powered by the three motors and it moved right, left, and in circles. The student demonstrated an understanding of gear ratios.

Results

The students from partner schools who participated in the new assessments for the selection process had ratings in all categories. Across the schools, most were given a final rating of maybe or probably, which is similar to what would be found with a normal curve. The percentages and ratings follow: unknown, 5.1%; maybe, 42.6%; probably, 19.3%; definitely, 19.3%; and wow, 1.6%.

The average of ratings of M1 students was 3.00 while the average of the ratings for M2 students was 3.27. A comparison of their scores in all rating categories is shown in Figure 7. No student in either group had a rating of unknown. Across the rating categories, more M1 students were given ratings of maybe and probably than were the M2 students; more than twice as many M2 students had ratings of definitely, while one M1 student and no M2 students had a rating of wow.

Discussion

The mechanical-technical assessment was developed to achieve a fair assessment of creative problem-solving abilities of different cultural, ethnic, and economic groups. One of the important reasons for the underrepresentation of certain groups in special programs for exceptionally talented students is that their scores tend to be significantly lower than those of students from White and Asian students from middle and high income groups. Many of the M1 students came from schools in high-income areas with advanced placement and other programs for exceptionally talented students, special charter schools, and other similar schools. Mainly as a result of the selection methods, the M1 students' Grade Point Averages (GPAs), on a 4-point scale, ranged from 3.71 to 4.0 with an average of 3.93. Although the students in the M2 group came from schools in high poverty areas with significant percentages of underrepresented minority groups, their scores on the mechanical-technical performance assessment were similar.

In fact, the average scores of the M2 group were slightly higher. Their GPAs were significantly lower, ranging from 2.11 to 4.00 with an average of 3.07. Another factor to consider is that American Indian, Hispanic, and low-income students from partner schools had ratings in all categories, especially definitely (the rating normally used to identify those who are gifted), with some even achieving a rating of wow, the rating showing performance seldom seen in students. In several studies, researchers have found that on almost all measures of achievement (grades, class rank, and standardized test scores, state achievement tests, and National Assessment of Educational Progress [NAEP]), students from these groups are severely underrepresented in the top 1%, 5%, and 10% (Miller, 2004; Plucker et al., 2010). This pattern was clearly not evident when the mechanical-technical performance assessment was used as a method for selection.

Another comparison can be made by combining M1 and M2 students in one group. If this is done, one would find that 25.6% of exceptionally talented students in mechanical-technical ability were from the M2 group (ratings of definitely), while 11.6% were from the M1 group (ratings of definitely and wow). One way of looking at this pattern of results is that the emphasis on first-order knowledge (Gardner, 1992) enables students who have not had opportunities such as advanced educational experiences to demonstrate knowledge they have gained through their life experiences. Another way of looking at the results is that they include creative problem solving and practical application of knowledge rather than accumulation of knowledge, which is an important component of this assessment. Other researchers also have found that inclusion of measures of creativity in assessments enables identification of higher percentages of students from traditionally underserved groups (Sternberg, 2010), and that students from these groups score at similar and sometimes higher levels than those from mainstream groups (DeVries & Shires-Golon, 2011; Ivcevic & Kaufman, 2013; Jenkins, 2005; Kaufman, 2006; Kaufman et al.,

2004; Torrance, 1971). Use of the continuum of problem types (Getzels & Csikszentmihalyi, 1967, 1976) modified by Maker and Schiever (2005, 2010) was the primary method for including both domain-specific knowledge and skills and creativity-relevant processes needed for creative problem solving in the mechanical-technical domain (Amabile, 2013). Other comparisons of the students in the M1 and M2 groups can be found in another publication (Maker, forthcoming-a).

Using the idea of the assessment square proposed by Ruiz-Primo and colleagues (Ruiz-Primo, 2003, 2007, 2009; Shavelson et al., 2002) was an important element in the research, as they defined the construct, developed tasks, tried the tasks on themselves and others, field tested the assessments with populations similar to those the assessments were designed to identify, and developed scoring rubrics based on these field tests. During the field testing of the assessment, the research team made many changes to increase its validity and practicality.

The mechanical-technical assessment is important for identification and placement of students in programs to develop exceptional talents in physical sciences because it is appropriate for students from mainstream cultures as well as students from underrepresented groups. This assessment can be used to evaluate the quality of physical science programs; it gives teachers and administrators important information about student outcomes. The performance based assessment can give feedback on how well students have developed creative problem solving and critical thinking in the mechanical-technical domain.

Suggested Uses

The assessment can be used in different settings and for different purposes. First, to identify students with creative problem solving ability to be served in special programs, the performance-based assessment can be used by itself or with the physical science concept maps

(Maker & Zimmerman, 2019) to select students who are exceptionally talented in physical science. To select students for such programs, attention should be focused on their performance on all problem types ranging from closed to open-ended tasks. Students need to be assessed on both first and second order knowledge (Gardner, 1992; Maker, 2005) to give a program administrator a comprehensive picture of students' abilities. The concept map assessment is a measure of mainly second-order knowledge. Students have the opportunity to demonstrate knowledge they learned in school, the ways they organize concepts, and the ways they think about what they learned (Maker & Zimmerman). The mechanical-technical assessment is mainly an assessment of first order knowledge (Gardner) gained from experience. Being able to think of a machine that is different from the vehicle and the ability to make it with the remote control and other available parts are excellent indicators of second-order knowledge, creative problem solving, and the creative application of physical science principles. The mechanical-technical assessment is an appropriate tool to measure the abilities of students from cultural groups often underrepresented in programs for exceptionally talented students (e.g., American Indian and Hispanic) or low SES students because it is not affected by their exposure to information or their school achievement. The two assessments (physical science concept maps and mechanical-technical performance assessments), when combined, are measures of analytical, creative, and practical abilities as Sternberg (1997, 2005) recommends. They measure domain-specific knowledge, domain-general ability, and creativity-relevant processes (Amabile, 1983, 1996, 2013).

Second, in science classrooms, activities similar to the mechanical-technical assessment can be used to observe comprehension and application of the laws of physics. Tasks like the gear and vehicle construction can help students demonstrate an understanding of basic principles

involved in transfer of energy to power vehicles. Constructing a machine is a task that can be used to assess students' higher level application, analysis, and innovative use of physics principles. Students can develop diagrams, illustrations, and models of various mechanical and technical constructions, and then make them with materials that are similar to the assessment materials. If the actual tasks and exact materials in this assessment are used in instructional settings, the assessment will not be a valid measurement of mechanical-technical creative problem solving when used to identify students who are exceptionally talented in physical science because some of those being assessed might have experience with the materials and tasks and others might not. Observers will not know the difference. Therefore, if teachers implement these ideas, they need to use tasks and materials that are similar but not the same. For example, use kits that have similar functions and can be used to make many, varied constructions, not just one or two, such as the following: Space Explorers Science Plastic Experiment Kit, Remote-Control Machines: Construction Vehicles, and Remote-Control Machines DLX.

Third, this new assessment can be combined with other types of measures to form a battery of instruments that is potentially more fair to students from underrepresented groups than conventional assessments. For example, in the Cultivating Diverse Talent in STEM (CDTIS) project, six assessments were used to select students: mathematical problem solving (Bahar & Maker, 2020), life science problem solving (Zimmerman et al., 2019), life science and physical science concept maps (Maker & Zimmerman, 2019), spatial analytical problem solving (Maker, forthcoming-b), and the mechanical-technical assessment. Using these assessments allows those who implement assessments and program managers to have a complete picture of the abilities of exceptionally talented students in different domains.

Fourth, the mechanical-technical assessment can be used alone or with other assessments to determine the quality of teaching in physical science classrooms. Tasks such as those in the mechanical-technical assessment can be demonstrated in staff development programs to assist teachers in designing learning experiences to meet the differing needs of students (Hodges, 2005; Sandri, 2013; Tracy, 2015). Students can demonstrate their strengths in understanding gears, for example, when they construct a gear box. To build on this strength, the teacher can then suggest that they build a unique gear box and a machine that uses one or more gear boxes. If students have difficulty with tasks such as these, the teacher can review the ways gears operate and provide different models for the students to construct until they can demonstrate an understanding of the underlying principles. Similar learning experiences can be designed as a result of observing students' construction of vehicles and machines.

Finally, a related classroom use is to observe students' strengths in solving closed, semi-open, and open-ended problems. For example, if a student can easily make a construction identified by the teacher, but cannot think of a different one to make, this student needs more practice with the semi-open problems in which he or she is given choices among several constructions. After practice with semi-open problems, students can more easily move to open-ended problems in which they must define the problem themselves, create their own methods, and develop unique solutions (Maker, forthcoming-a; Maker & Schiever, 2005).

Research Needed

The four partner schools in the CDTIS research project were in areas with populations of Hispanic and American Indian students; therefore, studies of usefulness and effectiveness of the assessments in more schools and schools with other cultural groups is essential. Small numbers of White, Asian, and African American students attended the partner schools, so schools with

larger percentages of students from these groups need to be included in future investigations. The usefulness of the assessment with students from high and middle SES levels also needs to be studied to show its practicality in schools with multiple ethnic and cultural groups and a variety of socioeconomic levels.

Within all these populations, the inter-rater reliability and the construct, concurrent, and predictive validity of the assessment need to be studied. Some examples follow:

- **Inter-Rater Reliability.** To test the reliability of the assessment of products, two teams of qualified observers can rate all products separately and compare the scores. This can be done at the time the assessment is completed. To assess the reliability of the entire process, a study similar to that conducted by Griffiths (1996) can be designed. In the Griffiths study, two observers watched the same group of students without discussing their performance. They also conducted the final discussions of performance and rated students' products separately. In a study like this, the observations of student processes as well as the ratings of products can be compared.
- **Construct Validity.** Using Amabile's (1996) Consensual Assessment Technique (CAT), physical science teachers and experts can be asked to rate the photographs of constructions and transcripts of interviews about student constructions on creativity and technical quality. Using this method, experts use a 7-point scale from low to high. After these two ratings, if experts are then asked to tell what qualities distinguished between the highest- and lowest-rated products, their lists of qualities can be compared with the constructs underlying the assessments and the rubrics for assigning ratings. Because the experts provide two different ratings (creativity and

technical quality), these two aspects of mechanical-technical ability can be distinguished.

- **Concurrent Validity.** Even though teachers often are not proficient in identifying talented students, a study in which physical science teachers are asked to identify students they believe are talented can be conducted (Sarouphim, 1999, 2009). These ratings can be compared with the ratings made by observers during the assessment.
- **Predictive Validity.** Studies of the validity of the assessment in predicting the performance of students when placed in special programs for exceptionally talented students are important. A number of studies of the predictive validity of the original DISCOVER assessments can provide examples of research needed on the new assessments. For instance, the performance of exceptionally talented students identified using the mechanical-technical assessment can be evaluated using methods such as achievement tests, grades, and teacher observations in physical science classes and programs. The correlations of assessment scores with indicators of performance during the program are indicators of the predictive validity of the assessment as is the amount of variance explained by the physical science assessment. Further investigation of the effectiveness of the mechanical-technical assessment to predict the academic performance of undergraduate students is another area of research.
- Other areas of research include teachers' use of the individualized assessment profiles to inform instruction, differentiate curriculum, and provide programming for exceptional students in STEM (Pease et al., forthcoming).

Conclusion

A performance-based assessment of mechanical-technical ability was developed to measure high school students' creative problem solving and ability to apply physical science principles in practical ways. The assessment was field tested and then conducted during two years to identify exceptionally talented students in STEM among students in minority and low-income groups. The research team developed rubrics containing criteria such as uniqueness and complexity that were used as guides in the assignment of ratings in five categories (unknown, maybe, probably, definitely, and wow) using the procedures developed during initial research on the DISCOVER assessments (Maker, 2005). Students were rated as unknown because their final constructions showed a minimal use of materials, resulting in an incomplete and non-functioning machine, whereas the students who scored definitely or wow demonstrated elaborate machine construction with multi-functioning components.

When the ratings of the performance of students identified by conventional methods (GPA, teacher recommendation, and self-statements) were compared with the ratings of performance of students identified by the new methods, the average ratings were very similar and slightly higher for the students from partner schools, which had high minority and low-income populations. The research team concluded that using the mechanical-technical assessment helped to identify exceptionally talented students in physical science among students from minority (e.g., American Indian and Hispanic) and low-income groups that have been missed when conventional assessments are used as selection tools.

References

- Amabile, T. M. (1983). *The social psychology of creativity*. Springer.
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Westview Press Inc.
- Amabile, T. M. (2013). Componential theory of creativity. In E. H. Kessler (Ed.), *Encyclopedia of management theory*. (pp. 134-139). SAGE.
- Arizona Minority Education Policy Analysis Center (2013).
<https://higherred.az.gov/sites/default/files/AMEPAC%20MSPR%202013%20-%20Final.pdf>
- 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
- Bahar, A. K., & Maker, C. J. (2020). Culturally responsive assessments of mathematical skills and abilities: Development, field testing, and implementation. *Journal of Advanced Academics*, 1-23. doi:1932202X20906130.
- Carter, L., & Darling-Hammond, L. (2016). Teaching diverse learners (5th ed.). In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching on teaching* (pp. 593-637). American Educational Research Association.
- DeHaan, R. L. (2009). Teaching creativity and inventive problem solving in science. *Life Sciences Education*, 8(3), 172-181. doi: 10.1187/cbe.08-12-0081
- DeVries, M. & Shires-Golon, A. (2011). Making education relevant for gifted Native Americans: Teaching to their learning style. In J.A. Castellano & A.D. Frazier (Eds.), *Special Populations in Gifted Education* (pp. 47-72). Prufrock Press.

- Dino, R. (2017). Connected but different: Comparing and contrasting creativity, innovation and entrepreneurship. In J. A. Plucker (Ed.), *Creativity and innovation: Theory, research and practice* (pp. 23-34).
- Donovan, S., & Cross, C. T. (2002). *Minority students in special and gifted education*. National Academy Press.
- Duncan, G. J., & Murnane, R. J. (2014). *Restoring opportunity: The crisis of inequality and the challenge for American education*. Harvard Education Press.
- Elvan, Ý. N. C. E., Güven, E., & Aydogdu, M. (2010). Effect of problem solving method on science process skills and academic achievement. *Journal of Turkish Science Education*, 7(4), 13-25.
- 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), 1-20.
- Flanagan, D., & Ortiz, S. (2001). *Essentials of cross-battery assessment*. Wiley.
- Funk, C. & Parker, K. (2018, January 9). Diversity in the STEM workforce varies widely across jobs. Pew Research Center. <https://www.pewsocialtrends.org/2018/01/09/stem-appendix-detailed-tables-and-charts/>
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B. Gifford & M. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 77-120). Kluwer.
- Getzels, J., & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal*, 3(9), 80-84.
<http://psycnet.apa.org/psycinfo/1967-16699-001>

- Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. Wiley.
- Glover, J.A. (1976). Comparative levels of creative ability among elementary school children. *Journal of Genetic Psychology*, 129(1), 131–135. <https://doi-org.sdl.idm.oclc.org/10.1080/00221325.1976.10534020>
- Gonzalez, H. B., & Kuenzi, J. J. (2012). Science, technology, engineering, and mathematics (STEM) education: A primer. Congressional Research Service, Library of Congress.
- Griffiths, S.E. (1996). *The inter-observer reliability of the DISCOVER problem-solving assessment*. [Unpublished manuscript]. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.
- Hodges, G. C. (2005). Creativity in education. *English in Education*, 39(3), 47-61. doi:10.1111/j.1754-8845.2005.tb00624.x
- 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. <https://doi.org/10.1080/09500690802275820>
- 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. <https://doi.org/10.1016/j.lindif.2013.07.011>
- Jenkins, A. H. (2005). Creativity and resilience in the African American experience. *Journal of Humanistic Psychology*, 33(1), 25–33. https://doi.org/10.1207/s15473333thp3301_3
- Kaufman, J.C. (2006). Self-reported differences in creativity by gender and ethnicity. *Journal of Applied Cognitive Psychology*, 20, 1065–1082. <https://doi.org/10.1002/acp.1255>

- 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(1), 56–69.
<https://doi.org/10.1002/j.2162-6057.2004.tb01231.x>
- Lopez, E., Kim, J., Nandagopal, K., Cardin, N., Shavelson, R. J., & Penn, J. H. (2011). Validating the use of concept-mapping as a diagnostic assessment tool in organic chemistry: Implications for teaching. *Chemistry Education Research and Practice*, 12(2), 133-141. doi: 10.1039/C1RP90018H
- 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.
<https://doi.org/10.1177/001698629604000106>
- Maker, C. J. (2005). *The DISCOVER Project: Improving assessment and curriculum for diverse gifted learners*. Senior Scholars Series Monograph. National Research Center on the Gifted and Talented.
- Maker, C. J. (forthcoming-a). Identifying exceptional talent in STEM: Increasing diversity and assessing creative problem solving. *Journal of Advanced Academics*.
- Maker, C. J. (forthcoming-b). 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. <https://doaj-org.sdl.idm.oclc.org/article/53758c04b3fc485bbe99fd8b883f9bea>

- 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., & Schiever, S. W. (2005). *Teaching models in education of the gifted* (3rd ed.). Pro-ed.
- Maker, C. J., & Schiever, S. W. (2010). *Curriculum development and teaching strategies for gifted learners* (3rd ed.). 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.
- 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.
- National Research Council (2014). *Developing assessments for the next generation science standards*. The National Academies Press.
- National Science Board (2010). *Preparing the next generation of STEM innovators: identifying and developing our nation's human capital*. National Science Foundation.
- National Science Foundation (2015, May). *Science and engineering degrees, by race/ethnicity of recipients: 2002–12*. <https://www.nsf.gov/statistics/2015/nsf15321/pdf/tab4.pdf>
- Novak, J. D., Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization*, 5(3), 175-184.
doi:10.1057/palgrave.ivs.9500126

Oakes, J., Lipton, M., Anderson, L., & Stillman, J. (2013). *Teaching to change the world*.

Paradigm Publishers.

Pease, R., Vuke, M., Maker, C. J., & Muammar, O.M. (forthcoming). 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., Burroughs, N., & Song, R. (2010). *Mind the (other) gap!* Bloomington: Center for Evaluation and Education Policy, Indiana University.

Renzulli, J. S. (1978). What makes giftedness? Re-examining a definition. *Phi Delta Kappan*, 60(3), 180-184

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). Glencoe-McGraw Hill.

Ruiz-Primo, M. (2009). *Towards a framework for assessing 21st century science skills*. Paper prepared for the Workshop on Exploring the Intersection of Science Education and the Development of 21st Century Skills, National Research Council.

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. doi: 10.1207/S15326977EA0702_2

Sandri, O. J. (2013). Exploring the role and value of creativity in education for sustainability.

Environmental Education Research, 19(6), 765–778.

<http://dx.doi.org/10.1080/13504622.2012.749978>

Sarouphim, K. M. (1999). Discovering multiple intelligences through a performance-based

assessment: Consistency with independent ratings. *Exceptional Children*, 65(2), 151-161.

doi:10.1177/00144029990650020

Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted lebanese

students: Is DISCOVER effective? *Journal for the Education of the Gifted*, 33(2), 275-

295. doi:10.1177/016235320903300206

Sawchuk, S. (2018, May 22). Is STEM oversold as a path to better jobs? Which STEM jobs are in demand and pay well? It's complicated. *Education Week*.

<https://www.edweek.org/ew/articles/2018/05/23/is-stem-oversold-as-a-path-to.html>

Shavelson, R., Ruiz-Primo, M., Li, M., & Ayala, C. (2002). *Evaluating new approaches to*

assessing learning. CSE: Technical Report 604. Center for Research on Evaluation,

Standards, and Student Testing/ University of California, Los Angeles.

Shen, J., Liu, O. L., & Chang, H. (2017). Assessing students' deep conceptual understanding in

physical sciences: An example on sinking and floating. *International Journal of Science*

and Mathematics Education, 15(1), 57-70. doi:10.1007/s10763-015-9680-z

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. (2016). Advances in the science of

assessment. *Educational Assessment*, 21(1), 34-59. doi:10.1080/10627197.2015.1127752

Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C.

(2008). Achieving equity in special education: History, status, and current challenges.

Exceptional Children, 74(3), 264–288. doi: 10.1177/001440290807400301

- Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life*. First Agency Publishing.
- Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology*, 24(4), 359-375. doi:10.1006/ceps.1998.0998
- Sternberg, R. J. (2005). The theory of successful intelligence. *Revista Interamericana De Psicología: Interamerican Journal of Psychology*, 39(2), 189-202.
doi:10.1177/026142940001500103.
- Sternberg, R.J. (2010). Assessment of gifted students for identification purposes: New techniques for a new millennium. *Learning and Individual Differences*, 20(4), 327-336.
<https://doi.org/10.1016/j.lindif.2009.08.003>
- 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.
<https://doi.org/10.1177/001698627101500201>
- Tracy, B. (2015). *Creativity and problem solving*. AMACOM.
- U.S. Department of Commerce (2011). <https://www.usa.gov/>
- U.S. Department of Commerce (2017, March 30).
<https://www.commerce.gov/news/reports/2017/03/stem-jobs-2017-update>
- U.S. Department of Education, Office for Civil Rights (2016). *Civil rights data collection*.
<https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>
- Wang, H., Chang, C., & Li, T. (2008). Assessing creative problem-solving with automated text grading. *Computers & Education*, 51(4), 1450-1466. doi:10.1016/j.compedu.2008.01.006
- Zimmerman, R., Maker, C. J, & Alfaiz, F. S. (2019). *Culturally responsive assessments of life science skills and abilities: Development, field testing, implementation, and results*.

Manuscript submitted for publication.

Table 1

Number of high school seniors who participated in the field testing of mechanical-technical assessment

School	Field Test	Mechanical-Technical
School A	Field test 1	19
School B	Field test 2	15
School C	Field test 3	15
School D	Field test 4	9
Total		58

Table 2

Number of high school juniors in partner schools who participated in the mechanical-technical assessment, by school, gender, and ethnicity

<u>School</u>	<u>Year</u>	<u>Gender</u>		<u>Ethnicity</u>				
		<u>Female</u>	<u>Male</u>	<u>American Indian</u>	<u>White not Hispanic</u>	<u>African American</u>	<u>Hispanic</u>	<u>Asian and Pacific Islander</u>
School A	1	40	41	4	4	3	68	2
	2	54	47	3	11	6	78	3
School B	1	11	10	21	0	0	0	0
	2	14	10	23	0	0	0	1
School C	1	7	8	15	0	0	0	0
	2	13	5	18	0	0	0	0
School E	1	44	30	74	0	0	0	0
Total		183	151	158	15	9	146	6

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*Criteria for scoring the mechanical-technical assessment*

Unknown	For vehicle only <ul style="list-style-type: none"> Students completed the vehicle with lots of assistance, or did not complete the vehicle For machine only <ul style="list-style-type: none"> Students did not complete a machine
Maybe	Vehicle and/or machine <ul style="list-style-type: none"> Has 1 motor Goes backward and forward For machine only <ul style="list-style-type: none"> Is similar to the vehicle the student made in function and design May have a unique design, but is not stable
Probably	Vehicle and/or machine <ul style="list-style-type: none"> May demonstrate an attempt to use gears Is stable (doesn't fall apart when it moves) Goes backward and forward Has 1 or more motors that function For machine only <ul style="list-style-type: none"> May have a design that is similar to the vehicle the student made, but has other unique or functional characteristics
Definitely	Vehicle and/or machine <ul style="list-style-type: none"> Has 2 or more motors that function For machine only <ul style="list-style-type: none"> May use only 1 motor, but has a unique design Has a design that is different from the vehicle the student made Has gears powered by a motor or motors Has chains to connect gears or wheels Demonstrates understanding of gear ratio Moves right and left or goes in circles Has gears that have a function
Wow	For vehicle only <ul style="list-style-type: none"> The vehicle has same characteristics as model 1 or 2 For machine only <ul style="list-style-type: none"> Has gears with a definite or unique function Has chains Is complex Has a number of functions Is unique

Figure 1

Examples of the materials used in the mechanical-technical assessment

Figure 2

Example of a gear construction

Figure 3

Model 1 and 2 choices for vehicle construction

Figure 4

An example of the mechanical-technical product with a final rating of unknown

Figure 5

An example of the mechanical-technical product with a final rating of probably

Figure 6

An example of the mechanical-technical product with a final rating of wow

Figure 7

Comparison of ratings on the mechanical-technical performance-based assessment for M1 and M2 students