**Culturally Responsive Assessments of Spatial Analytical Skills and Abilities:
Development, Field Testing, and Implementation**

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

The University of Arizona

Maker, C. J. (2020). Culturally responsive assessments of spatial analytical skills and abilities: Development, field testing, and implementation. *Journal of Advanced Academics*. 31(3) 234–253. DOI:10.1177/1932202X20910697

Author Note

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

Correspondence concerning this article 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. 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.

Abstract

A persistent problem in education is underrepresentation of certain cultural and linguistic groups such as American Indian, African American, and Hispanic, in special programs for exceptionally talented students, especially in science, engineering, technology, and mathematics (STEM). The spatial analytical task, a performance-based assessment with demonstrated reliability and validity as an instrument to identify exceptionally talented students, was included with new instruments created in the Cultivating Diverse Talent in STEM (CDTIS) project. A continuum of problems, including closed, semi-open, and open-ended, was an important component of the assessment, enabling the assessment of creative problem solving as well as assessment of skills such as seeing how things fit together visually and in space; through mental images, on paper, and in physical objects or forms. The spatial analytical assessment was implemented as part of a battery of instruments to identify students to participate in a special internship program. The assessment has multiple purposes in addition to its use as an identification tool: evaluation of special programs, planning differentiated instruction, and as a pre- and post-measure of student gains. The test-retest reliability, and construct, concurrent, and predictive validity for secondary students need continued investigation in future studies.

Keywords: identification of gifted students, exceptional talent, STEM, creative problem solving, spatial intelligence

Culturally Responsive Assessments of Mathematical and Spatial Analytical Skills and Abilities: Development, Field Testing, and Implementation

Understanding of most fields in STEM is dependent on the logic, reasoning, and ability to discern patterns characteristic of mathematical and spatial analytical domains of intelligence (Gardner, 1983; NSB, 2010; Sternberg, 1999; Super & Bachrach, 1957; Wai, Lubinski, & Benbow, 2009; Webb, Lubinski, & Benbow, 2007). Mathematical abilities have long been considered necessary in all areas of science, technology, and engineering. However, spatial abilities have not always been viewed as essential to STEM careers, and have not been included in some standardized tests used to select students for special opportunities, including the long-standing talent search at Johns Hopkins University (NSB, 2010; Stanley, 1996; Wai, et. al., 2009; Webb, et. al., 2007). Combining Wai and colleagues' detailed longitudinal analysis of the contributions of spatial ability to success in STEM careers with the previous research of Super and Bachrach (1957) and analysis of data from the massive Project TALENT database (Wise, McLaughlin, & Steel, 1979), Wai and colleagues (2009) reached three main conclusions: (a) spatial ability is a key component of the abilities of adolescents who achieve advanced educational and occupational credentials in STEM careers, (b) spatial ability is an important component in educational and occupational settings for both talented individuals and the general population, and (c) talent searches exclude many intellectually talented students when criteria are restricted to mathematical and verbal ability measures. One of their most notable findings was that 70% of the top 1% in spatial ability were not in the top 1% in either math or verbal abilities. These individuals would be missed in exclusive talent searches, yet their spatial abilities enabled them to be highly successful in biological sciences, physical sciences, engineering, and mathematics.

Defining Spatial Ability

Spatial ability has been defined either broadly or narrowly by different theorists and researchers. For example, one of the earliest researchers, L. L. Thurstone (1935) identified spatial ability as one of his seven primary factors of intellect, and divided it into the ability to (a) recognize an object when it is seen from several angles, (b) imagine movement or internal “displacement” among the parts of a configuration, and (c) think about spatial relations when the observer’s body position is part of the problem. Later, Gardner (1983) defined it by listing essential capacities: (a) perceiving the visual world accurately, (b) performing transformations and modifications on initial perceptions, and (c) re-creating aspects of one’s visual experience even when the physical stimulus is absent. His definition included both perceiving and manipulating existing forms and creating new forms. Lohman (1994) defined it more narrowly as “the ability to generate, retain, retrieve, and transform well-structured visual images.” (p. 1000) Comparison of these definitions with Sternberg’s (1985, 1997, 2005) separation of intelligence into three types: analytical, creative, and practical. Using this lens, one can see that Gardner included both creative and analytical forms of ability while Lohman’s definition is focused mainly on the analytical form of spatial ability. For the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) projects research and development projects, spatial ability was divided into two aspects: creative/artistic and analytical. Experiences leading to this division are described in the following section.

Assessing Spatial Ability

Although achievement and ability tests used to select middle and high school students for special programs such as talent searches and entrance into higher education programs usually do not include spatial ability subtests or subsections, most intelligence tests do include them, and are

divided into three sections: verbal, quantitative, and non-verbal/spatial. Some theorists suggest that these abilities correspond to three types of symbol systems used for representation (i.e., verbal, numerical, and visual/imagistic), and that each individual's experience with or proclivity for a particular symbol system is an important aspect of her or his level of ability to do certain types of tasks (Gardner, 1983; Sternberg, 1999; Wai, et. al, 2009). In other instances, when developing tests of intelligence with the goal of assessing individuals in ways that are less culturally and economically biased than verbal tests, researchers rely on items that are intended to measure spatial reasoning rather than verbal reasoning, based on the underlying belief that such items have similar meaning across cultures or can be designed to minimize the effects of cultural referents (Naglieri, 2003; Lohman; 1996). Well-known examples are the Raven Progressive Matrices (Raven, Court, & Raven, 1998), the Test of Nonverbal Intelligence (TONI) (Brown, 2003), and the Naglieri Nonverbal Abilities Test (NNAT) (Naglieri, 2003). The results of Wai and colleagues' analysis (Wai, et. al., 2009) seem to be supportive of this approach in that students otherwise missed would be identified. These methods, however, are more frequently used at elementary levels than at higher levels of education.

Maker (2005) and colleagues, during their 20 years of research on the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) assessments, found several problems with current assessments of spatial ability: (a) the figures or pictures included in the tests are not culturally appropriate or are not recognizable aspects of the students' environments, (b) figures and pictures often are combined with verbal explanations that do not match the pictures or figures, and (c) the figures or pictures are not complete, so they convey a message different from the text of the item. For example, in one often-used test of cognitive ability, a pictorial item is presented as a representation of a city block with a girl

supposedly walking around the block and back to her home. Test takers are supposed to tell how many corners she turned on her walk. However, the picture is incomplete, and shows only two corners, with the picture ending on two sides. If a student is from a city and is familiar with city blocks, he or she can easily get the item correct simply from reading and remembering personal experiences, so the item is a test of both linguistic ability and spatial memory. If, however, a student has never been in a city or is not from a city with square blocks, the picture is misleading. Based on their spatial analysis of the picture, they might conclude the girl turned two corners, and get the item wrong. Using this experience, the DISCOVER research team designed performance assessments involving little or no language, and drawing upon students' proficiency in perceiving, manipulating, analyzing, and creating products with forms and shapes rather than with words and numbers.

The assessment described in this article was used as part of a battery of assessments to select students to participate in the Cultivating Diverse Talent in STEM (CDTIS) project, which was a collaboration among four departments (education, pharmacy, biology, and biomedical engineering) at an R1 University in the Southwest and four public schools in the same state. The spatial analytical assessment was initially developed as part of the DISCOVER assessment battery for students in grades K to 12. It was field tested and implemented with diverse groups of students in the USA and other countries (Maker, 2005; 1997; Sak & Maker, 2003; 2006; Sarouphim & Maker, 2010). Describing these field tests is beyond the scope of this paper; however, the main aspects of its development, revision, and implementation are summarized. The DISCOVER spatial analytical assessment was considered to be an excellent measure of the skills underlying the ability to become STEM innovators, so it was not revised or field tested during the CDTIS project.

Conceptual Framework

The theories that guided development of the assessment of spatial abilities all had problem solving as a key component, and all provided a different and complementary perspective. Within the general framework of Gardner's (1983, 1999) and Sternberg's (1997, 2005) definitions of intelligence as the ability to solve a variety of types of problems, Maker (1993) offered a definition of giftedness or exceptional talent that provided the conceptual base for a number of alternative assessments: "...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) To make this definition operational, the continuum developed by Maker and Schiever based on the early work of Getzels and Csikszentmihalyi (1967, 1976) was used to create problems ranging from *closed* (well-defined, having a clear method for developing a solution, having a correct answer) to *semi-open* (well-defined, having one or more methods, having more than one appropriate answer) to *open-ended* (ill-defined, having an unlimited number of methods, and having an unlimited number of appropriate answers) to assess students' ability to solve a variety of types of problems, but to focus on solving the most complex problems. Within this framework, two types of knowledge were important (Gardner, 1992): first-order, the knowledge developed through experience and interaction with the environment (fluid intelligence, Cattell [1957]), and second-order, the knowledge learned through formal schooling (crystallized intelligence, Cattell). The Spatial Analytical assessment is a measure of first-order knowledge and fluid intelligence.

All assessments included components of domain-specific knowledge and skills, creativity-relevant processes, and task motivation both in the tasks themselves and the behaviors

observed or scored that indicate superior abilities in a domain (Amabile, 1996, 2013). For example, domain-specific knowledge and skills are found mainly in the assessments of second-order knowledge, such as in the math assessment, students worked problems involving mathematical operations and concepts, while in the life science and physics Concept Map assessments, students mapped the relationships and hierarchies of concepts from science disciplines. Creativity-relevant processes included behaviors in the Spatial Analytical assessment such as makes puzzles in more than one way, takes apart puzzles when necessary to solve all the puzzles on a page; in the Concept Map assessments, makes many connections among concepts; in the Life Science assessment, makes many different and original groupings of flowers or insects; and in the Mechanical-Technical assessment, makes a unique machine different from others and different from his or her previous construction. Task motivation was included mainly in the behaviors to be observed: continuously working, does not want to quit even when others have stopped, persists in solving difficult problems, and verbalizes enjoyment of the task. An in-depth description of the components of the conceptual framework is included in another publication (Maker, 2019).

During the initial development of the DISCOVER assessments (Maker, 2005; Sarouphim, 2000), which included two spatial tasks, observers and researchers found that some students performed very well on the tasks requiring artistic/creative forms of spatial ability, but not on the tasks requiring analytical forms. In the task requiring artistic/creative ability, students created visual representations of forms found in nature, items in photographs, and images they saw in their minds using various types of art materials. In the task requiring analytical forms of spatial ability, students solved puzzles and created shapes such as a square, triangle, and rhombus using as many Tangram pieces as possible. Although some students performed well on

both, most did not. As a result, the assessments were renamed “Spatial Artistic” and “Spatial Analytical” to make this distinction clear.

The spatial artistic task, then, seemed to be an assessment of artistic/creative spatial intelligence while solving puzzles was a spatial analytical task. To solve the puzzles, individuals must analyze the space available and find the Tangram pieces that fit into the puzzle outline. Making a large shape requires a holistic view of a space, but still includes the ability to analyze the pieces available and decide which pieces are needed and how they need to be placed to make the shape. However, this task also included elements of practical intelligence (Sternberg, 2005) because of the need to apply knowledge and skills in a hands-on situation. For the spatial analytical assessment, a closed problem was a puzzle page that had only one solution; a semi-open problem was a puzzle page that had more than one solution, but included constraints; and the open-ended problem was to create a certain shape with as many puzzle pieces as possible.

Early in the process of development, the DISCOVER research team identified Tangrams as a math manipulative that could be used in a flexible way to design assessments. One of the most important considerations was the possibility of using the pieces in a variety of ways; another was that the shapes were geometric, and thus could be used to distinguish those who could analyze the properties of shapes as they worked puzzles and made constructions. However, the team also recognized that the possibilities were limited if only the usual 7-piece set of Tangrams was used; therefore, three sets were combined to form a large set of 21 pieces. The set contained six small triangles, six large triangles, three medium triangles, three parallelograms, and three squares.

Initially, following the problem continuum, the tasks included first solving puzzles with either one or a limited number of correct answers (closed problem), then solving puzzles with

more possible methods and solutions (semi-open), then making a shape with as many pieces as possible and making anything the participants wanted to make (open-ended). However, observers soon learned that the most engaging part of the assessment was solving puzzles, and that some students were unable to solve the puzzles in a reasonable amount of time, so they had to either extend the time for the assessment, making it boring for those who had completed all activities, or change the order of the tasks. Interestingly, however, most students were so engaged in solving the puzzles they did not want to stop until they had completed all of them. These observations resulted in changing the order of the tasks and making the final task (make anything you want to make) optional. The first task, before giving students the puzzle books, then, was an open-ended problem: make a large shape (rhombus) with as many of the Tangram pieces as possible.

In addition, the team also learned that a time limit was necessary. Otherwise, the assessment would be impractical for a number of reasons, including that to enable all to complete the puzzles would take too much time, and that while the slower students were completing the puzzles, the faster ones would have nothing to do. Thus, time limits were developed to enable the most able students to complete the shape and all the puzzles, and for some exceptional students to complete the optional tasks, which eventually included both a “challenge page” (a very difficult puzzle) and the open-ended task of making anything with the pieces.

Another addition resulting from field testing was the incorporation of clues. Some students have had experiences with Tangrams and others have not. To minimize the effects of prior experiences and to assist in identifying students who were able to learn from a minimum of new information, a set of clues was designed. Clues also served as a way to assist students in achieving some level of success on the task so they would not become frustrated and stop

working. Clues were designed to progress from general reminders of the parameters of the task to more specific methods for solving the puzzles.

During the field testing of the DISCOVER tasks (Maker, 1997), after each assessment, observers were asked to identify which students they believed were superior problem solvers in the domain being assessed, and to tell what behaviors they observed that led them to believe these students were superior problem solvers. If an observer noted a general observation such as “motivated”, the observer was asked, “What did the student do or say that led you to believe he or she was motivated?” The purpose of this question was to identify observable behaviors rather than inferences about behavior. Answers to these questions were compiled during assessments of over 5000 students, and when no new behaviors were being identified, the DISCOVER research team grouped them into categories based on the ways they were alike. Two categories were identified for each domain: problem solving processes and characteristics of products. These behaviors were then included on observer note forms and observation checklists. Other behaviors identified across domains were grouped into the categories of general problem-solving processes and characteristics of products. Some examples of effective spatial analytical problem-solving processes included (a) completes a puzzle without physically manipulating pieces (e.g., chooses shapes without turning); (b) sees how to substitute pieces for others to make the same shape or puzzle; and (c) makes more than one initial shape with many pieces. Characteristics of products included (a) complexity, (b) use of many pieces to make the initial shape, and (c) originality (e.g., making a 3-D shape rather than a flat one). General problem-solving processes included behaviors such as inventing new strategies, showing involvement in the task (e.g., not being distracted, focusing on own problem solving rather than that of others), persisting on difficult tasks, showing non-verbal enjoyment of task, and verbalizing enjoyment of the task. General

characteristics of products included (a) new, original, or different; and (b) constructed from a unique perspective. See Rogers (1998) for a description of the checklist, its development, and its use.

Field testing of the DISCOVER Spatial Analytical assessment took place in a number of schools: two K-12 schools in an American Indian Nation with 95 to 99% American Indian students, one high school with 98 to 99% American Indian students, two schools in an urban area in a Southwestern state with a variety of ethnic groups (African American, Caucasian, and Hispanic, with some Asian Americans). One school in the urban area was in a high socioeconomic region and the other in a low socioeconomic region of the city. All schools in the American Indian Nation were in low-income areas.

Implementation

During the Cultivating Diverse Talent in STEM (CDTIS) project, the Spatial Analytical Assessment was part of a battery of instruments used in partner schools to select students completing their junior year in high school who had exceptional talent in STEM to participate in a summer internship program on the campus of an R1 University in the Southwest. After the summer program, Keep Engaging Youth in Science (KEYS), students were invited to participate in specially designed programs in their schools, to mentor peers and younger students in their schools, to become student counselors in the KEYS program during the years after they participated, and to continue their research at a university or in another appropriate setting (Wu, Pease, & Maker, 2019).

The first step in implementation was to secure parent permission and student assent for assessment. Assisted by teachers and administrators in the schools, all juniors were given the opportunity to participate; follow-up by teachers and project liaisons at the schools was also

necessary as a means of encouraging as many students as possible to participate in the assessment and avoid missing talented students. In Table 1, demographic data about the students who were assessed, their schools, and the ethnic make-up of the schools are presented.

The assessment team consisted of those who had designed the spatial analytical assessment as well as some doctoral students in education of the gifted who had experience at the secondary level. If observers for the performance-based assessment had not been involved in the field testing, they received training from one of the research team members and then participated as co-observers before being on their own as observers.

Science and math teachers as well as administrators and department chairs at the schools were provided with an individual profile of the strengths of each student assessed (Pease, Vuke, Maker, & Muammar, 2019). Each individual profile included a line indicating the average of all the students assessed at that school, or in the case of the schools with American Indian students, students assessed at two schools with similar populations. The report also included descriptions of the assessments, the kinds of occupations or activities in which the abilities were important, and ways to develop these strengths in the students in their classes. Teachers were encouraged to focus on each student's strengths regardless of whether or not the student was considered to be exceptionally talented in some area. Students and parents also received a shorter report. Their report included an individual profile and a one-page description of the abilities measured by the assessment. In this description, emphasis was placed on the ways the information could be used. More information about the use of the results and the programs designed for the CDTIS Project can be found in another publication (Pease, et.al., 2019).

Schools

The assessments were conducted in the project's partner schools: four high schools located in the Southwestern United States (Table 1). The number of students in these schools ranged from 94 to 2245, and more than 71% received free or reduced-price lunches. In three schools, 97 to 100% were American Indian, and 83.4% of the students in one school were Hispanic. The schools were located in poverty areas with an unemployment rate that ranged from 7.7% to 45.8%. Many different academic programs were offered in the larger schools, such as Honors and Advanced Placement (AP) classes and programs for gifted students. However, the smaller schools had a limited number of programs. A more detailed description of the schools in the CDTIS project is included in the first article in this issue (Maker, 2019).

Description of the Spatial Analytical Assessment

The Spatial Analytical assessment consisted of two separate tasks: making a large rhombus-like figure (making a rhombus with exactly equal sides is not possible with all 21 Tangram pieces) with as many Tangram pieces as possible and solving puzzles of increasing difficulty. To minimize the effects of prior experiences with Tangrams, students were provided instruction before the first task and guided in constructing basic shapes. They were shown how to put two large triangles together to make a larger triangle, a square, and a parallelogram. Then, they were shown how to use different smaller shapes (a parallelogram and two small triangles, a square and two small triangles, and two medium sized triangles) to make a large triangle. The principle of substituting or trading was introduced at this time. Observers asked the students in their groups to make each of the shapes on the table in front of them so the observer could make certain each student understood the instructions. Then, the first task was introduced.

To minimize the effects of language and knowledge of geometric shapes, a drawing of a rhombus was shown. Students were told that a rhombus has approximately equal sides, and that

it does not need to have right angles like a square does. They also were told that their rhombus might not have exactly equal sides, but they could be approximately equal and still qualify as a rhombus. Students were given 10 minutes to complete the shape. The observer counted the number of pieces in the shape, noted whether it was complete, and noted whether it had approximately equal sides.

All students in the group were then given a book of puzzles organized from easy to difficult, and were given some simple instructions that included telling the observer when they had completed a puzzle page so the observer could check their work. The first page (Figure 1) was one that could be worked by all students at this age; the middle puzzles (Figure 2) could be worked by many of the students; the last puzzles were completed by very few students (approximately 5%) (Figure 3). If a student had worked on a particular puzzle for at least 5 minutes, he or she was asked if he or she wanted a clue. If the answer was “yes”, one of the specific clues was given. The clues were given in a defined order based on the amount of specific information contained in the clue, and the same words were used for clues for all students. For instance, the first clue in the series was “You have enough pieces to do the page.” The sixth clue (out of seven) was “Make this shape in a different way.” The observer then noted how the student used the clue and recorded when specific clues, such as five, six, and seven, were given. The first three were considered general hints, often were given to the entire group, and were reminders of instructions that had already been given, so they were not recorded.

If a student completed the puzzles within the allotted time, he or she was given the opportunity to work a “challenge page.” If the student did not want to work the challenge page, an option was to make something with the Tangram pieces. However, most students who

completed the puzzles chose to work the challenge puzzle. If the first challenge was completed, and time remained, another, more difficult challenge was available.

To assign ratings of student performance, observers first ranked the students based on the number of puzzles they completed, the number of pieces they used in the large shape, the number and specificity of the clues they were given, the order (in their group) in which they completed the most difficult puzzles, and the effective problem solving behaviors they exhibited. At the highest level were the students who completed the shape with many pieces and finished all the puzzles and the challenge page without specific clues. Next were those who completed everything, but had a small number of specific clues, and then those who completed the large shape with many pieces and the six pages of the puzzle book, again without specific clues; proceeding from the highest number of puzzle pages completed without specific clues, then with specific clues, until all students were listed. The last step was to assign overall ratings of unknown, maybe, probably, definitely, and wow (See Maker, 2019, this issue, for an explanation of the ratings) 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. If no natural breaks could be easily identified, problem solving behaviors were reviewed to decide which rating to give to students close to the end of one rating and the beginning of another.

Results

Results of Prior Research on the Spatial Analytical Assessment

Several studies have been conducted on the final version of the spatial analytical assessment. All have shown it to be a reliable and valid measure of ability across cultures and

economic groups (Maker, 2005; Sak & Maker, 2003; Sarouphim, 2000; 2001; 2002; Sarouphim & Maker, 2010). It has remained unchanged since these studies were completed.

Reliability. Reliability data are available from previous research, and will be summarized here. In one of the first studies (Griffiths, 1996), inter-observer agreement was calculated by comparing the decisions made by the researcher (an expert observer) with decisions made by a novice (fewer than 10 assessments), experienced (more than 10 but less than 20) and expert (more than 20 assessments) observers. Griffiths found 80% agreement between the researcher and novice observers and 100% agreement between the researcher and expert observers. Perhaps more importantly for identification of exceptional talent, she found 95% agreement across all tasks in the original assessment and all levels of observer experience, showing that observers can recognize the highest levels of performance regardless of the experience they have had with the assessment. Kassymov (2000) also found a high level of agreement between expert members of a school district assessment team before (80%) and after (100%) debriefing sessions. This study and other experiences demonstrated the importance of the debriefing sessions in which student performance was discussed among all observers.

Validity. Sak and Maker (2003) studied the validity of the spatial analytical assessment administered in kindergarten to predict the performance of students in grade 6. They found that those who were given the rating of definitely on the spatial analytical assessment performed significantly higher than those not given this high rating on the Stanford 9 ($F = 7.02$, $p < .01$) and Arizona Instrument to Measure Standards (AIMS) Math subtests ($F = 7.29$, $p < .01$) and end-of-year grades in Science ($F = 4.05$, $p < .01$). In studies of concurrent validity, Sarouphim (2001) found significant correlations between scores on the spatial analytical assessment and the Raven Progressive Matrices at kindergarten, second, fourth, and fifth grades and across the total group

of participants ($r = .409$, $p < .01$, $N = 257$), which consisted of students from two cultural groups: Hispanic and American Indian. In a study of the assessment used with three ethnic groups (American Indian, 29% of the sample; Hispanic American, 50% of the sample, and Caucasian, 21% of the sample) at the secondary level, Sarouphim (2002) found no significant differences in the ratings of students across ethnic and gender groups, indicating that the assessment was not biased in favor of or against any of the three groups, nor was it biased against or for male and female students.

Results of Current Research

Students were selected for the special CDTIS programs using two different methods, conventional (M1) and the new performance assessments (M2). The conventional method, which had previously been used as the only method of selection, consisted of Grade Point Average (GPA), teacher recommendation, and self-statement. Overall results are provided in another publication (Maker, 2019). However, in this section the results of the Spatial Analytical and Math assessments in each partner school are compared to demonstrate the utility of the assessment to identify exceptional talent in students from diverse cultural groups and low-income families (Table 2). Math was chosen as the comparison assessment because it is the most closely related to Spatial Analytical ability.

Discussion

Ratings of all students assessed at the partner schools were higher on the Spatial Analytical assessment (a measure of first order knowledge) than on the math assessment (a measure of second order knowledge). In fact, the Spatial Analytical ratings of all students assessed at the partner schools were the same as the students selected using conventional methods at two out of the three schools; and the Spatial Analytical ratings of students selected

for the internship program using the new assessments (M2) was much higher than the ratings of students selected using the conventional methods (M1). These results demonstrate that the Spatial Analytical assessment can be a useful addition to a battery of assessments to select students for special programs.

A closer look at these results shows that the spatial analytical assessment can be a way to determine which students demonstrate potential for success in STEM areas even though their grades are too low for them to be considered for special programs for exceptionally talented students. One of the criteria for identification of students for the special internship program in STEM was a GPA of 3.2. Of the 11 students who would not have been accepted into the special internship program based on their GPA scores, 9 scored at the levels considered to be exceptional (*definitely* or *wow*; 4 or 5 on a 5-point scale). One student scored 3 and another scored 2, but were placed because of extremely high scores on other assessments. The average score of these 11 students on the spatial analytical assessment was 3.90 (on a 5-point scale) and their average GPA was 2.68 (on a 4-point scale).

An important comparison is with the results of other research on assessment of spatial ability. Wai and colleagues (Wai et al., 2009) provide a comparison of composite means and standard deviations of 9th, 10th, 11th, and 12th grade students from four socioeconomic (SES) quartiles on spatial ability. Across all grade levels, both males and females from lower quartiles scored lower than males and females from the higher quartiles. For example, at grade 11, the mean for males in the first quartile (lowest) was 69.95 and the mean for females was 57.57 while in the highest quartile, the mean for males was 87.59 and the mean for females was 74.98; a difference of 17.64 for males and 17.41. The pattern was the same across all grade levels.

Students from low income and minority groups score either higher or at similar levels on the DISCOVER spatial analytical assessment as demonstrated in a number of studies.

The results of this study are similar to other research on the DISCOVER assessments in which patterns of ability were included. Sarouphim (2002), in a study of high school students (grade 9) found that Spatial Analytical ratings for Hispanic, American Indian, and White students were not significantly different, while the ratings for written and oral language assessments were lower for Hispanic and American Indian students than for White students, but the differences were not statistically significant. These findings are different from results of many studies showing that scores of students from diverse ethnic backgrounds are significantly lower than scores of students from mainstream backgrounds on many assessments, which has led to the common practice of assigning weighted scores for students from different backgrounds.

Further support for the claim that this assessment can be useful in selecting students from underserved groups can be found in the results of comparisons of the ratings on the Spatial Analytical assessment and the GPAs of students selected by conventional methods. Considering the fact that GPAs are on a 4-point scale, the difference between an average of 3.8 and an average of 3.1 is educationally and statistically significant. Maker (2019) found significant differences between students selected by the DISCOVER problem solving assessments and those selected by conventional methods in three areas: GPA, ethnicity, and parent level of education. She concluded that the differences in ethnicity could not be attributed to the selection methods because of the makeup of the partner schools (high percentages of Hispanic and American Indian students); however, GPA and parent level of education could be attributed to the differences in methods. Maker (2019) recommends that if only one of the performance assessments is used to

identify students as exceptionally talented in STEM, the spatial analytical assessment would be the best choice.

Suggested Uses of the Assessment

In this study, the CDTIS research team used the entire assessment battery of six separate DISCOVER assessments to identify students with exceptional talent in STEM. However, all of the assessments described in this special issue can be used as stand-alone measures of problem-solving ability in aspects of STEM.

The Spatial Analytical assessment perhaps is the most defensible if only one assessment of exceptional talent in STEM is desired. Researchers have identified it as a general ability needed for most STEM occupations and for success in advanced academic studies in STEM areas (Super & Bachrach, 1957; Wai, et al., 2009). In addition, studies have shown that it has little cultural or gender bias (Sarouphim, 2002; Sarouphim & Maker, 2010), so it would be useful in “casting a wide net” (NSB, 2010) to find students from underrepresented groups who are exceptionally talented. Studies similar to those conducted by Erdimez and Maker (2015) and Tan and Maker (2015) would be useful. They found that the combined DISCOVER assessments (Spatial Artistic, Spatial Analytical, Oral Linguistic, Math, and Written Linguistic) explained more of the variance (43.9%, $F [5, 55] = 7.818, p = .000$) in total achievement scores of American Indian students (Erdimez & Maker) than did a well-known and often-used test of non-verbal intelligence (19.5%, $F [1, 49] = 11.84, p = .001$) (Tan & Maker). These studies provide evidence that the Spatial Analytical assessment can be combined with other assessments of specialized scientific ability such as physics or biology to identify students with exceptional talent. Use of this assessment could add an important dimension to an assessment battery

consisting of multiple choice or other assessments already being administered as part of an existing program.

In classrooms, the results of both assessments can be used to inform instruction, and to differentiate instruction for individuals and small groups. A simple way to use the assessment results is to form small groups of students who perform at similar levels in overall scores (e.g., needs improvement, meets expectations, exceeds expectations), and to pace instruction or increase the complexity of content based on the ability levels of the students in the groups. Another way to use the results is to examine performance on the three types of problems. Students who can solve the closed problems correctly, but perform less well on the semi-open, and have difficulties with the open-ended problems can benefit both from solving more challenging closed problems (building on strengths) and from practicing the solving of simple semi-open and open-ended problems (remediating weaknesses). This kind of teaching will help them begin to develop creative thinking processes. Other students may be proficient in solving the open-ended problems, but lack correct information and knowledge of methods for solving the closed and semi-open problems. They may need additional practice with the closed problems (remediating weaknesses) and additional more complex opportunities to solve open-ended problems (building on strengths). If performance of all students is analyzed across problem types, the same process can be used to plan instruction to develop strengths and remediate weaknesses for the class as a whole.

Many students enjoy creating problems for other students to solve, and using this interest can help both the teacher and the students. Students can draw upon their experiences with the open-ended problems and create tasks for others. For instance, they can create puzzles using the tangram pieces, perhaps problems of all types, ranging from closed to open, to gain experience

with and understanding of the different processes needed to develop and solve problems with different degrees of open-endedness. Additional ideas for using the assessment to inform instruction can be found in the final article in this issue (Pease, et al., 2019).

Research Needed

Some observers and researchers (Muammar, Maker, & Kuang, 2015) have noted that some puzzle pages may not be placed in the appropriate order of difficulty, and have recommended testing all puzzles with different groups to determine the average length of time needed to complete each puzzle page independently rather than as part of a puzzle book. A change such as this can enable more students to complete more puzzles before needing a clue, and can possibly increase the validity of the assessment. However, a study such as this is difficult to do, and the usability of the results is questionable. For example, puzzles would need to be worked by each individual in a different order, and the average time calculated across participants. Then, researchers would need to test the new placement of puzzle pages. The reason for additional studies of any new arrangement is that observers have found that students learn from solving puzzles early in the sequence, and they apply this learning to their solving of puzzles later in the sequence. Therefore, changing the order of puzzles can change the time needed for each puzzle.

In addition, more studies of the use of the spatial analytical assessment with different groups would be helpful. Most of the research conducted at the high school level to date has been with Hispanic, American Indian, and Caucasian students. Its use with high school students from Asian and African American populations has not been studied; additional research is needed to determine the appropriateness of the high school tasks for these groups.

Another area of research that would be important is to investigate the relationship between the DISCOVER Spatial Analytical assessment and assessments of three-dimensional spatial analytical tasks. This research would be even more powerful if the studies included both performance-based tasks and paper-and-pencil tasks because most currently used assessments of spatial ability are paper-and-pencil tasks.

Finally, although concurrent, predictive, and construct validity studies have been conducted, more are needed, especially those validating the use of the spatial analytical assessment in predicting performance in STEM careers and advanced educational programs. The first in such a series could be a long-term follow-up of the students assessed for the CDTIS project.

References

- Alfaiz, F. S., Pease, R., & Maker, C. J. (2019). *Culturally responsive assessments of physical science skills and abilities: Development, field testing, and implementation*. Manuscript submitted to the *Journal of Advanced Academics*. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, AZ.
- 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.
- Bahar, A. K. (2019). Culturally responsive assessments of mathematical skills and abilities: Development, field Testing, and implementation. Manuscript submitted to the *Journal of Advanced Academics*. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, AZ.
- Brown, L. (2003). Test of nonverbal intelligence. In *Handbook of nonverbal assessment* (pp. 191–221). New York, NY: Springer.
- 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.
- 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.
- 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: Kluver.
- 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. & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York, NY: Wiley & Sons.
- Griffiths, S.E. (1996). *The inter-observer reliability of the DISCOVER problem-solving assessment*. Unpublished manuscript. Department of Disability and Psychoeducational Studies, The University of Arizona, Tucson, AZ.
- Jenks, G.F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186-190.
- Kassymov, A. (2000). *Researchers and implementers: Comparison of inter-rater reliability*. Unpublished manuscript. Department of Disability and Psychoeducational Studies, The University of Arizona, Tucson, AZ.
- Lohman, D. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 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. (1994). Authentic assessment of problem solving and giftedness in secondary school students. *Journal of Secondary Gifted Education*, 2(1), 19–29.

- Maker, C. J. (1997). DISCOVER Problem-Solving Assessment. *Quest, The Journal of the Division of Research and Evaluation, National Association for Gifted Children*, 8(1), 3, 5, 7, 9.
- 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. (2019). *Identifying exceptional talent in Science, Technology, Engineering, and Mathematics (STEM): Increasing diversity and assessing creative problem solving*. Manuscript submitted to the *Journal of Advanced Academics*. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, AZ.
- 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. & Schiever, S. W. (2010). *Curriculum development and teaching strategies for gifted learners*. (3rd Ed.). Austin, TX: Pro-Ed.
- Muammar, O., Maker, C. J., & Kuang, C. C. (2015). *Spatial intelligence: The internal structure of the DISCOVER spatial analytical assessment*. Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, AZ.

Naglieri, J. A., (2003). Naglieri nonverbal ability tests: NNAT and MAT-EF. In R.S. McCallum (Ed). *Handbook of Nonverbal Assessment* (pp. 175–189). Springer Book Archive, DOI 10.1007/978-1-4615-0153-4_9

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.

Pease, R., Vuke, M., Maker, C. J., & Muammar, A. A. (2019). *A practical guide for implementing the STEM assessment results in classrooms: Using strength-based reports and Real Engagement in Active Problem Solving (REAPS)*. Manuscript submitted to the *Journal of Advanced Academics*. Department of Disability and Psychoeducational Studies, The University of Arizona, Tucson, AZ.

Raven, J. C., Court, J. H., & Raven, J. (1998). *Manual for the Standard Progressive Matrices*. London: H. K. Lewis.

Renzulli, J. S. (1978). What makes giftedness? Re-examining a definition. *Phi Delta Kappan*, 60, 180–184, 261.

Sak, U. & Maker, C. J. (2003). The long-term predictive validity of a performance-based assessment used to identify gifted CLD Students. *Proceedings of the 15th Biennial World Conference of the World Council for Gifted and Talented Students*. Adelaide, Australia: World Council for Gifted and Talented Students.

Sarouphim, K. M. (2000). Internal structure of DISCOVER: A performance-based assessment. *Journal for the Education of the Gifted*, 23(3), 314–327.

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. & Maker, C. J. (2010). Ethnic and gender differences in identifying gifted students: A multi-cultural analysis. *International Education*, 39(2), 42–56.
- Schiever, S. W., & Maker, C. J. (1997). *Enrichment & acceleration: An overview & new directions*. In N. Colangelo & G. A. Davis (Eds.), *Handbook of Gifted Education*, (2nd Ed. (pp. 113–125). Needham Heights, MA: Allyn & Bacon.
- Stanley, J. C. (1996). In the beginning: The study of mathematically precocious youth. In C. P. Benbow & D. Lubinski (Eds.), *Intellectual talent* (pp. 225–235). Baltimore: Johns Hopkins University Press.
- 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. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology*, 39(2), 189–202.
- Super, D. E., & Bachrach, P. B. (1957). *Scientific careers and vocational development theory*. New York, NY: Bureau of Publications, Teachers College, Columbia University.
- Tan, S. & Maker, C. J. (2012). *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.

- Thurstone, L. (1935). *The vectors of mind; multiple-factor analysis for the isolation of primary traits* (University of Chicago science series). Chicago, Ill.: The University of Chicago Press.
- Wai, J., Lubinski, D., & Benbow, C. P. (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.
- Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, 99, 397–420.
- Wise, L. L., McLaughlin, D. H., & Steel, L. (1979). *The Project TALENT data bank*. Palo Alto, CA: American Institutes for Research.
- Wu, I., Pease, R., & Maker, C. J. (2017). Students' perceptions of a special program for developing exceptional talent in STEM. *Journal of Advanced Academics*, 30(4), 474-499.
DOI: 10.1177/1932202X19864690

Table 1

Number of high school juniors who participated in the spatial analytical assessment, by school, gender, and ethnicity

<u>School</u>	<u>Year</u>	<u>Total</u>	<u>Gender</u>		<u>Ethnicity</u>				
			<u>Female</u>	<u>Male</u>	<u>Native American</u>	<u>White not Hispanic</u>	<u>African American</u>	<u>Hispanic</u>	<u>Asian-Pacific Islander</u>
School A	1	75	37	38	3	4	2	64	2
	2	100	53	47	3	11	6	77	3
School B	1	21	11	10	21	0	0	0	0
	2	25	14	11	24	0	0	0	1
School C	1	12	6	6	12	0	0	0	0
	2	18	13	5	17	0	0	0	1
School E	1	63	37	26	60	2	0	1	0
Total		314	171	143	140	17	8	142	7

Table 2

Average Ratings for Spatial Analytical, Math, and GPA Across Schools and Methods of Selection

Ability Assessed	Average Scores of All Students Assessed at Partner Schools				Averages of Selected Students	
	A (Tuba)	B (GHHS)	C (Shonto)	D (DV)	M1	M2
Spatial Analytical	2.7	3.1	3.1	2.7	3.1	3.6
Math	2.0	2.9	2.0	2.2	3.0	3.0
GPA	NA	NA	NA	NA	3.8	3.1

Note 1: NA means that GPAs were not available for all students assessed using the new methods. Researchers were allowed to collect GPAs only for those students who applied for the internship program or were selected to participate in it.

Note 2: All ratings for Math and Spatial analytical are on a 5-point scale. GPA is on a 4-point scale.

Figure 1. Puzzle 1

Figure 2. Puzzle 3

Figure 3. Puzzle 6