مجلة جامعة الإمام عبدالرحمن بن فيصل للعلـــوم الإنســانيـــــة والتربويـــة

Journal of Imam Abdulrahman Bin Faisal University for Humanities and Educational Sciences

Peer-reviewed Journal   دورية علمية محكمة

فهرس المحتويات

Omar M. Muammar,

Professor at the College of Education,
Imam Abdulrahman Bin Faisal University

https://orcid.org/0000-0002-5774-7477

عمر محمد معمر

أستاذ بكلية التربية ،
جامعة الإمام عبد الرحمن بن فيصل

ABSTRACT

Spatial intelligence is a critical cognitive competency for success in STEM fields; however, few validated assessment methods exist to identify gifted students' spatial abilities (Andersen, 2014; Stumpf et al., 2013). This study examined the validity of the Spatial Analytical Assessment (SAA) within the Discovering Intellectual Strengths Capabilities while Observing Varied Ethnic Responses (DISCOVER) project to enhance the identification of gifted students and to support debriefing sessions about spatial analytical abilities. Participants included 3,434 students in Grades K–12 from diverse ethnic backgrounds, with Native American students representing the majority of the sample. Results indicated that different SAA activities tapped distinct aspects of spatial cognitive ability and that the internal structure of the SAA demonstrated high consistency across problem types, item complexity, and grade levels. Nonetheless, the disproportionate representation of ethnic groups limited the generalizability of the findings to the wider student population. Overall, the SAA was found to capture the multidimensional nature of spatial analytical ability, and a revised order of puzzle pages across grade categories was proposed. Future research should focus on improving the psychometric properties of the SAA through more demographically balanced sampling and refined calibration of item difficulty.

الملخص:

تُعَدّ القدرة المكانية من الكفاءات العقلية الأساسية للنجاح في مجالات العلوم والتقنية والهندسة والرياضيات (STEM)، إلا أنّ هناك ندرة في أدوات القياس المُعتمدة؛ للتحقّق من قدرات الطلبة الموهوبين المكانية، من هنا هدفت هذه الدراسة إلى فحص صدق مقياس القدرة المكانية ضمن مشروع اكتشاف القدرات الفكرية أثناء ملاحظة الاستجابات المتنوعة للطلاب من خلفيات عرقية مختلفة (DISCOVER Project)؛ وذلك بهدف تحسين عمليات الاكتشاف، والتعرّف على الطلبة الموهوبين، وتعزيز جلسات التقييم، واتخاذ القرار للمحكّمين حول القدرات المكانية التحليلية للطلبة. وقد شملت العينة (3434) طالبًا ينتمون إلى خلفيات عرقية متعددة من الصفوف الدراسي (K–12) ، وكان غالبيتهم من الطلاب الأمريكيين الأصليين. واعتمد البحث المنهج الكميّ الوصفي الارتباطي؛ للتحقّق من صدق مقياس الذكاء المكاني التحليلي. وأظهرت النتائج أنّ أنشطة مقياس القدرة المكانية تقيس أبعادًا مختلفة من القدرات المكانية، كما أظهرت البنية الداخلية للمقياس درجة عالية من الاتساق مع أنواع المشكلات، ومستويات التعقيد، والمجموعات العمرية المختلفة، إلا أنّ التمثيل غير المتوازن للمجموعات العرقية حدّ من إمكانية تعميم النتائج على مجتمع الطلاب بشكل عام. وبناء على ما توصّلت إليه الدراسة من نتائج أوصت بضرورة ترتيب صفحات الألغاز عبر الفئات العمرية والصفية المختلفة ترتيبًا مُنقحًا، إضافة لأهمية إجراء المزيد من البحوث لتحسين الخصائص السيكومترية للمقياس من خلال عينات أكثر توازنًا ديموغرافيًا، ومعايرة أدقّ لمستويات صعوبة البنود.

Journal of IAU of Humanities and Educational Sciences. Volume.4 .Issue 1. 2026

Online copy

1. Introduction

While spatial intelligence is an essential cognitive competency for success in Science, Technology, Engineering, and Mathematics (STEM) fields, very few assessment methods have been developed and validated to measure and identify gifted students' spatial analytical talents (Andersen, 2014; Aprile et al., 2020; Kell & Lubinski, 2013; del Cerro & Méndez, 2021). A review of the literature shows empirical evidence supporting the correlation between STEM performance and spatial intelligence (Buckley et al., 2018; Kell & Lubinski, 2013). Furthermore, spatial ability has demonstrated predictive validity for future performance in STEM fields (Berkowitz & Stern, 2018). STEM education is crucial for economic development in the knowledge-based era. Although many countries consider STEM fields to be a national priority, spatial intelligence—despite its role as a predictor of STEM talent—is rarely measured, if not entirely neglected, in talent search programs (Andersen, 2014; Kell & Lubinski, 2013; Stumpf et al., 2013). Spatial intelligence enables gifted students to develop the scientific and artistic thinking used by engineers in various fields such as mechanical, civil, biomedical, design, architecture, and computer engineering. This ability enables students to conceptualize and construct spatial models, analyze geometric objects, interpret diagrams, and identify functions, among other skills (del Cerro & Méndez, 2021, p. 1). Even though spatial intelligence is a critical component of analytical and creative thinking in STEM, very limited research has investigated the validity of assessments related to this cognitive construct (Maker, 2020; Yoon et al., 2023).

A high level of spatial ability has been linked not only to the arts but also to science, mathematics, and technology (Kell & Lubinski, 2013; Kozhevnikov et al., 2013). Students with high spatial ability are able to create mental representations of complex problems, then analyze those representations and manipulate them mentally for creative production (Trickett & Trafton, 2007). The capabilities of Einstein, as well as those of many scientists and engineers across different eras, cannot be fully explained without understanding spatial intelligence. Even though spatial intelligence is not a strong predictor of school success, it plays a vital role in many scientific and creative accomplishments (Maker, 2020; Das et al., 1994; Gardner, 1983; Humphreys & Lubinski, 1996; Lohman, 1996). Spatial intelligence is a crucial cognitive ability that involves the understanding and manipulation of visual and spatial information, both of which contribute to scientific reasoning and creativity (Carroll, 1993). Accurately assessing spatial intelligence is essential for identifying individuals' strengths and weaknesses in this cognitive domain. One effective and widely used tool for assessing spatial analytical intelligence is the tangram puzzle, as employed in the DISCOVER assessment project. The tangram puzzle has been used for centuries as a measure of spatial intelligence. Its historical significance and widespread application provide a strong foundation for validating its effectiveness in assessing this cognitive ability (Renavitasari & Supianto, 2018). Tangram puzzles require individuals to mentally manipulate geometric shapes to form specific spatial patterns (Ayaz et al., 2012; Kmetová & Nagyová, 2021). This directly assesses spatial reasoning skills such as mental rotation, visualization, and spatial analytical perception, which are key components of spatial intelligence. In using tangram puzzles, the assessment aligns with real-world spatial analytical tasks, enhancing its ecological validity. Moreover, tangram puzzles are culturally unbiased, as they do not rely on specific cultural knowledge or language proficiency. When applied appropriately, they serve as valuable tools for assessing spatial intelligence.

In the spatial analytical intelligence activities of the DISCOVER assessment project, a series of problems varying in complexity and required thinking processes are presented to students using tangram puzzles. Early research showed that the problems in this assessment were appropriately ordered, with puzzles arranged from the easiest to the most difficult in the booklet (Maker, 2001; Sarouphim, 2000). However, subsequent observers questioned the order of the puzzle pages based on their observations of the time required for students to progress through the booklet. This issue is important for the validity of the assessment (Maker, 2020). Another important psychometric aspect of this assessment is the way in which different problems are clustered and organized to present varying levels of

challenge to examinees. For example, to what extent do the different problems used in this assessment measure distinct cognitive competencies and abilities? Do the problems reflect varying levels of difficulty and complexity? What are the best practices for using such puzzles to assess spatial analytical intelligence effectively?

Furthermore, after several years of using tangram puzzles in the spatial analytical intelligence assessment, a substantial amount of data has accumulated in the DISCOVER database. These data can be analyzed to support the identification of spatially gifted students. This gap in the research is the focus of the current study. The author aims to validate a spatial analytical intelligence assessment using tangram puzzles to identify gifted students and support STEM education as part of the DISCOVER project.

2. Theoretical Framework

The assessment of spatial ability began in the early 1900s with the development of traditional forms of intelligence tests (Humphreys & Lubinski, 1996). From the first intelligence test devised by Binet and Simon to the present day, spatial ability items have been included in most Intelligence Quotient (IQ) tests. For example, the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV), the Stanford-Binet Intelligence Scale, the Universal Nonverbal Intelligence Test (UNIT), and Raven's Progressive Matrices all contain multiple items designed to assess spatial ability (Sattler, 2001). The Wechsler, Binet, and McCarthy scales measure cognitive processes through nonverbal spatial tasks (Das et al., 1994). Researchers studying individual differences have noted that factorial studies of spatial ability demonstrate that spatial tests are strong indicators of general intelligence, or the "g" factor (Lohman, 1979; Jensen, 1993). Most IQ tests that contain nonverbal spatial items aim to measure general intelligence by combining these items with others, rather than measuring spatial ability as a separate construct.

Historically, spatial ability as a distinct factor in human intelligence was described early in the 20th century. For much of that century, intelligence testing was dominated by the concept of a general intelligence factor, or "g," until Thurstone introduced his multidimensional theory of intelligence (Andersen, 2014). In Thurstone's theory, human intelligence consists of several primary mental abilities, one of which is spatial ability. Later, in 1938, Thurstone acknowledged the existence of the "g" factor; since then, spatial ability has been included in most intelligence theories, although it is often considered a secondary-level human ability (Chen & Siegler, 2000). Toward the end of the 20th century, spatial ability regained a more prominent role in theories of intelligence when Gardner (1983, 1993, 1999) developed the theory of multiple intelligences. Gardner identified eight types of intelligence: linguistic, spatial, musical, logical–mathematical, interpersonal, intrapersonal, bodily–kinesthetic, and naturalist. He further distinguished between two forms of spatial intelligence: spatial–artistic and spatial–analytical.

Spatial intelligence has been thoroughly studied in explicit theories of intelligence, including both differential and cognitive processing theories. From the differential theory perspective, Lohman (1996) concluded that spatial ability includes several components: visualization ability (i.e., the capacity to manipulate visual patterns successfully regardless of task speed), speeded rotation (the speed of mentally manipulating, transforming, and rotating objects), closure speed (the speed of identifying novel visual patterns), closure flexibility (the ability to find and identify novel visual patterns), and perceptual speed (the ability to quickly recognize known visual patterns or accurately compare multiple patterns). All of these cognitive processes are involved in solving tangram puzzles.

From the cognitive theorist's perspective, spatial intelligence includes three cognitive processes: encoding, transformation, and reproduction of visual images (Gardner, 1983; Lohman, 1996). Thus, individual differences in spatial ability may be related to these processes. However, Lohman (1996) noted that research in cognitive psychology does not explicitly address the sources of individual differences in spatial processing. He referred to four hypotheses that might explain such differences: (a) speed in performing analog transformations or mental rotations, (b) skill in generating mental representations that preserve configural information, (c) the amount of visual–

spatial information that can be maintained in working memory, and (d) the flexibility of strategies available for solving such tasks (as cited in Dennis & Tapsfield, 1996).

The diversity of definitions surrounding spatial intelligence has led to disagreement about how it should be defined. Carroll (1993) defined spatial ability as "an ability in manipulating visual patterns as indicated by level of difficulty and complexity in visual stimulus materials that can be handled successfully, without regard to the speed of task solution" (p. 362). Lohman (1996) described it as "the ability to generate, retain, retrieve, and transform visual images" (p. 98). Carroll's definition emphasizes outcomes and performance, irrespective of processing speed, whereas Lohman's focuses on underlying cognitive functions and processes.

In Gardner's (1993) view, both the cognitive processes and their outcomes are essential in assessing spatial intelligence. He defined it as "the capacity to perceive the visual world accurately, to perform transformations and modifications upon one's initial perceptions, and to be able to re-create aspects of one's visual experience" (p. 173). Gardner's (1983, 1993) definition captures both internal processing and external performance, as well as the domain specificity of spatial intelligence, which he divides into spatial–artistic and spatial–analytical components.

More specifically, Gardner posited that spatial abilities are not identical across individuals. For example, a person may possess a strong ability to perceive and mentally manipulate visual images but may lack the capacity to re-create those perceptions through drawing or construction. Consequently, Gardner (1983) emphasized the existence of subcomponent abilities within each type of intelligence, including both divergent and convergent thinking skills.

In the present study, the DISCOVER project defines spatial analytical ability [SAA] as "skills involving seeing how things fit together visually and in space through mental images, on paper, and in physical objects or forms" (Maker, 2020).

Assessment of Spatial Ability

Spatial ability can be measured using various item formats, including performance-based assessments, paper-and-pencil tests, and individually administered oral tests such as the Wechsler Scales and the Stanford-Binet. More recently, computerized forms of spatial tests have become available (del Cerro & Méndez, 2021). Items measuring nonverbal performance were among the earliest testing methods and are still used in contemporary intelligence assessments (Gardner, 1983; Lohman, 1996).

The advantages of performance-based assessment (PBA) include the following: (a) PBA reduces the anxiety often associated with paper-and-pencil tests; (b) it provides a calming context that minimizes test anxiety; and (c) it uses methods closely aligned with teaching and learning processes, allowing examinees to demonstrate their competencies with high levels of motivation. In addition, examinees are engaged in real-life problem-solving situations that reflect their competencies in both process and product (Miller & Miller, 2000; Sarouphim, 2001).

Consistent with new paradigms of intelligence proposed by Feldman (1991)—in which giftedness is understood to take multiple forms and to be identified through performance rather than standardized testing—spatial intelligence is assessed using a contemporary performance-based measure known as Discovering Intellectual Strengths and Capabilities While Observing Varied Ethnic Responses (DISCOVER). Spatial analytical intelligence is one of the several domains measured by DISCOVER. The DISCOVER Spatial Analytical Assessment (SAA) was developed to identify gifted students and to assess the cognitive strengths of all learners, drawing upon Gardner's (1993) theory of multiple intelligences and Sternberg's triarchic theory of intelligence (Sternberg, 1997). Students' performance on the SAA is the focus of the current study.

Spatial Analytical Assessment in DISCOVER

The DISCOVER assessment measures individuals' problem-solving abilities across various domains of intelligence (Maker, 2020). From a psychometric standpoint, the definition of superior ability depends largely on one's beliefs about the nature of giftedness.

In designing the DISCOVER assessment, Maker (1996) defined giftedness or high competence as "the ability to solve the most complex problems in the most efficient, elegant, effective, or economical ways" (p. 44). Maker drew upon a problem continuum derived from the work of Getzels and Csikszentmihalyi (1967, 1976) as a framework for developing a series of problem situations that require diverse intellectual competencies. Maker (2001) asserted that "the emphasis on problem solving was perceived as an important way to access the abilities of students from at-risk populations. When testing a student's knowledge, often we are assessing exposure, not the ability to learn the information" (p. 235). From a cognitive psychology standpoint, students who are able to learn and produce promising results are driven by unique cognitive representations, where the scope of problem solving becomes a critical gateway to a student's abilities—particularly when tasks vary in structure and difficulty (Edens and Potter, 2008; Domin & Bodner, 2012; Álvarez et al., 2020). Unlike traditional IQ tests, the DISCOVER assessment is based on the belief that the evaluation of problem-solving competencies stimulates both convergent and divergent thinking (Maker, 2004).

The assessment of spatial analytical intelligence for gifted students within the DISCOVER project has been in place for nearly four decades. Students participating in the DISCOVER assessment engage in a spatial analytical problem-solving session using tangram pieces. Following each assessment, a debriefing session is conducted during which observers complete a behavioral checklist. This checklist, developed during the initial research phase of the DISCOVER project, documents observable behaviors of high-level problem solvers and the characteristics of their constructed products. For example, within the spatial analytical component of the checklist, behaviors such as "sees which pieces will complete a puzzle without manipulating them," "makes initial shape in a visual way," and "sees how to substitute pieces for others to make same shapes or puzzles" are typical of superior problem solvers. The checklist also includes characteristics of their products, such as "makes many constructions and uses many pieces to make the initial shape" (Maker, 2020).

After completing the checklists, observers and the assessment leader discuss each student's performance, using the behavioral checklists and students' products as guides. Students are then assigned a rating for each activity based on the quality of their performance. A five-level rating scale is used to distinguish students' competencies in each domain of intelligence. In addition, a special rating—referred to as a "redo"—is assigned when a student is unable to complete a task due to external or emotional factors. For instance, a student may experience exceptional emotional distress or be called out of the room; in such cases, both the teacher and the observer may agree that the task should be rescheduled. This student is then given a "redo" for that specific portion of the assessment.

Although interest has grown in performance-based assessments such as DISCOVER—as promising alternatives to standardized IQ tests for identifying gifted students—their psychometric properties require further investigation. Since the development of the spatial analytical intelligence assessment, a number of studies have been conducted to examine several of its aspects. However, while many research paradigms have been developed to assess spatial intelligence, limited research has been conducted to validate these assessments (Aprile et al., 2020; Maker, 2020).

3. Purpose of the Study

The purpose of the current study is to provide useful information for improving decision-making regarding the assessment of students' spatial analytical intelligence using tangram-based tasks. This is accomplished by analyzing data collected over several years from the DISCOVER project. Specifically, the author aims to determine evidence-based recommendations for structuring the assessment using tangrams and to provide guidance on the psychometric properties of this type of assessment.

The following research questions guided the current study:

1. How does the number of pieces used in initial puzzle configurations (ill-defined problems) relate to students' performance on spatial analytical intelligence tasks (well-defined problems) across grade categories?

2. What are the psychometric properties (including item difficulty and performance patterns) of spatial analytical intelligence puzzles across grade categories, and how do these properties vary by task complexity and time on task?

4. Method

This study employed a quantitative, descriptive–correlational research design with an item analysis component. The research was non-experimental, as no variables were manipulated; instead, existing student performance data from the DISCOVER project were analyzed. Descriptive statistics (means and standard deviations) were used to examine the psychometric properties of the SAA items, including puzzle complexity, puzzle completion rates, and time on task across grade levels.

Correlational analyses were conducted to determine the relationships between task complexity (i.e., the number of pieces used in the initial shape—an ill-structured problem) and performance on other items (the number of puzzle pages completed—well-structured problems). In addition, item difficulty indices were calculated based on time-to-completion and performance across groups, providing evidence of construct validity and developmental trends in the spatial analytical intelligence assessment.

Participants

Male and female students from kindergarten through Grade 11 in various school districts across Arizona participated in this study. The total sample consisted of 3,434 students from 128 classrooms, including 560 kindergarten students, 2,066 elementary school students (Grades 1–5), 581 middle school students (Grades 6–8), and 277 high school students (Grades 9–11).

Participants represented diverse ethnic backgrounds, including Caucasian American, Black American, Hispanic, Native American, and Asian American students. Although the exact number of participants from each ethnic group could not be determined, the percentage distribution in the participating schools and districts reflected these backgrounds. According to the DISCOVER project, the approximate ethnic composition of the sample was as follows: 12% Caucasian American, 5% Black American, 23% Hispanic, 59% Native American, and 1% Asian American (Maker, 2020).

Instrument

The SAA is conducted as part of the DISCOVER project within regular classroom settings. Students are provided with sets of tangram pieces, including six large triangles, three medium triangles, six small triangles, three parallelograms, and three squares. Four to five students are assigned to one observer. General instructions are delivered to all students by the teacher, while basic spatial tasks are demonstrated simultaneously by each observer (Maker, 2020).

Students are then asked to complete a set of preliminary activities, using the tangram pieces to create various shapes. These tasks are designed to help students become familiar with the materials and develop the foundational skills necessary for the assessment. Following the practice session, students engage in an open-ended activity in which they are instructed to construct a given shape using as many tangram pieces as possible.

Once time is called for the open-ended task, students receive a booklet containing six structured puzzle pages. They are instructed to solve the puzzles using the tangram pieces. Observers record the time each student takes to complete each page before prompting them to proceed to the next. The puzzles are arranged in what the developers believed to be an order of increasing difficulty—from easiest to most challenging. If a student completes all six puzzles before the allotted time ends, they may choose to complete an additional "challenge" page.

Four different booklets are used for four grade-level groups: K–2, 3–5, 6–8, and 9–12. The total time allotted for completing the puzzle pages varies by developmental level: 20 minutes for Grades K–2, 25 minutes for Grades 3–5, 30 minutes for Grades 6–8, and 35 minutes for Grades 9–12.

The SAA for Grades K–12 includes five distinct booklets: one for each of Grades 3–5, 6–8, and 9–12, and two for Grades K–2. Each booklet follows the same instructions and procedures and includes two main activities: (a) the shape activity and (b) the puzzle tasks. The shape activity differs by grade level: students in Grades K–2

construct a square; those in Grades 3–5, a triangle; Grades 6–8, a parallelogram; and Grades 9–12, a rhombus. Each grade level is also assigned six puzzles, with additional challenge pages available.

The problems used in the SAA vary in structure and are aligned with the problem continuum. As shown in Table 2, the problem continuum classifies problem-solving situations based on whether the problem, method, or solution is known by either the presenter or the solver (Maker & Schiever, 2005). This framework, used in the construction of the DISCOVER assessment, distinguishes between well-structured problems, which require convergent thinking, and ill-structured problems, which require divergent thinking. In this study, the spatial analytical intelligence assessment was intentionally designed along this continuum. The initial shape activity represents a Type V problem, while the structured puzzle pages include problems ranging from Type I to Type IV, ensuring a balanced assessment of both convergent and divergent thinking skills (See Table 2).

Validity of the SAA. Among the earliest studies on the DISCOVER assessment was research conducted by Griffiths (1996) focusing on interrater reliability. Griffiths examined the level of agreement in student ratings between the researcher and observers with varying levels of experience: novice (fewer than 10 observations), experienced (10–20 observations), and expert (more than 20 observations). The findings revealed 80% agreement between novice observers and the researcher, and 100% agreement between both experienced and expert observers and the researcher on the spatial analytical component of the assessment. Griffiths also reported 95% agreement on categorical ratings across all observers experience levels and across all activities, indicating high interrater reliability and strong consensus on identifying students who demonstrated superior problem-solving skills.

In a separate study, Kassymov (2000) investigated interrater reliability among five expert observers from the DISCOVER team, each of whom had conducted more than 20 observations. He compared interrater agreement on the spatial analytical activity before and after a debriefing session and found consistent agreement levels of 80% both before and after the session.

In another phase of his research, Kassymov examined interrater reliability within a district team of 11 expert observers trained by the principal investigator of the DISCOVER project. Agreement on spatial analytical ratings within this district team was also high—80% before and 100% after the debriefing session. These findings support the reliability of the observer-based evaluation process used in the DISCOVER assessment, particularly for identifying spatially gifted students.

Sak and Maker (2003) investigated the predictive validity of the DISCOVER assessment and found that students rated as definitely superior problem solvers on certain components of the assessment scored significantly higher than their nongifted peers on standardized tests and academic performance measures. Specifically, students who demonstrated superior problem-solving skills in the spatial analytical activities scored significantly higher than those identified as nongifted (N = 82) on the Stanford 9 ($F = 7.02$, $p < .01$), the AIMS Math subtest ($F = 7.29$, $p < .01$), and on sixth-grade end-of-year science grades ($F = 4.05$, $p < .01$).

Sarouphim (2001) examined the concurrent validity of the DISCOVER assessment using the Raven's Progressive Matrices with a sample of 257 students in kindergarten and second, fourth, and fifth grades from two minority groups: Navajo Indian and Mexican American. The study found significant and positive correlations between DISCOVER assessment ratings and Raven scores. In the spatial analytical activity, moderate to high correlations were found at Grade K ($r = .351$, $p < .05$, $n = 74$), Grade 2 ($r = .398$, $p < .01$, $n = 47$), Grade 4 ($r = .495$, $p < .01$, $n = 74$), and Grade 5 ($r = .395$, $p < .01$, $n = 74$), with an overall correlation of $r = .409$ ($p < .01$, $N = 257$).

Collectively, these studies (Sak & Maker, 2003; Sarouphim, 2001) provide empirical support for the concurrent validity of the DISCOVER assessment and contribute to the understanding of the validity of the SAA. However, these findings are not yet sufficient to fully establish the psychometric soundness of the SAA. Further research is necessary to improve the identification of gifted students through spatial analytical performance, especially as evaluated during

debriefing sessions. Multiple aspects of the SAA require continued investigation to enhance its credibility and practical utility. While existing studies support its internal structure and concurrent validity, additional research is particularly needed to explore its internal dimensional structure in depth. Therefore, the current study focuses on the psychometric properties of the SAA as a tool for identifying gifted students in the domain of spatial analytical intelligence.

The results of both studies (Sak & Maker, 2003; Sarouphim, 2001) have provided evidence of the concurrent validity of the DISCOVER assessment and contributed to the understanding of the validity of the SAA, but they are still not enough. The validity of SAA in the DISCOVER project needs to be researched to provide information about improving the identification of gifted students and spatial analytical abilities during the debriefing sessions. Different aspects of the SAA need to be studied to increase its credibility. Several studies reported on the internal structure and the concurrent validity of the DISCOVER assessment and have shown support for its use; however, more studies are still needed to especially investigate the internal structure of the SAA. Therefore, this current study was focused on the psychometric aspects of the SAA as a tool for identifying gifted students in this area of intelligence.

Procedure

Data for this study were collected over a period of 11 years as part of the DISCOVER project. Information was recorded for all students in classrooms that participated in the spatial analytical intelligence assessment, in order to generate an appropriate and representative sample. These data were extracted from original assessment records and organized according to school district, grade level, teacher, and student.

After representative samples were identified for each grade level, the author entered students' performance data from the SAA into a centralized database. Student performance was evaluated across two activities: the shape construction task and the structured puzzle task. In the first activity, student performance was assessed based on their ability to construct the required shape and the number of tangram pieces used to do so,

which reflects the complexity of their spatial solution. In the second activity, which comprised six structured puzzle pages, performance was evaluated based on the number of puzzle pages completed within the allotted time and whether the student completed them independently (i.e., without clues or assistance).

5. Data Analysis

Data were analyzed using both descriptive and inferential statistical methods, employing the Statistical Package for the Social Sciences (SPSS) to address the study's two primary research questions.

For the first research question—which examined the relationship between the number of pieces used to construct the initial shape (ill-defined task) and the number of puzzle pages completed (well-defined tasks)—Pearson product-moment correlation coefficients were calculated for each grade group (K–2, 3–5, 6–8, and 9–12). This method was chosen to statistically assess the degree of association between student performance on the two types of tasks: the ill-defined shape construction (measured by the number of pieces used) and the well-defined puzzle completion (measured by the number of pages completed). Pearson's r coefficients helped indicate whether performance on the two task types was correlated or distinct. Additionally, effect sizes were computed using Cohen's d to quantify the magnitude of observed differences between performance on the two types of activities across grade levels.

For the second research question, which focused on the psychometric properties of spatial analytical intelligence assessment, descriptive statistics (means, standard deviations, percentages) were calculated across grade levels for two performance indicators: (a) the number of pieces used in the shape activity and (b) the number of puzzle pages completed. These analyses were intended to identify developmental trends, stability, and variation in performance, thus contributing to the construct validity and sensitivity of the instrument.

To examine the difficulty level of the puzzle pages, two complementary approaches were employed. First, the average time on task and corresponding standard deviations were calculated for each puzzle page to

serve as behavioral indicators of difficulty. Second, for students who reached page six, a difficulty index was derived by analyzing both the average completion time and the proportion of successful completions per grade group. These dual measures allowed difficulty to be assessed from both behavioral (time) and cognitive (accuracy) perspectives.

All statistical analyses were conducted at a significance level of α = .01. The combination of descriptive and correlational techniques provided a comprehensive understanding of the relationship between task complexity and student performance, the psychometric characteristics of the assessment, and the relative difficulty of the puzzle pages across grade levels.

6. Results

The research question in this study aimed to examine the relationship between the number of pieces used to construct the initial shape (ill-defined task) and the number of puzzle pages completed (well-defined tasks) by students in the Spatial Analytical Intelligence Assessment. Pearson correlation coefficients were calculated for each grade group, and all correlations were found to be statistically significant at an alpha level of .01. The strength of the correlations ranged from low to moderate across grade levels: (a) Grades K–2: $r(1,307) = .252$, $p < .01$; (b) Grades 3–5: $r(1,319) = .426$, $p < .01$; (c) Grades 6–8: $r(581) = .230$, $p < .01$; (d) Grades 9–12: $r(227) = .239$, $p < .01$. These results indicate a positive association between the two task types across all grade groups, with the strongest correlation observed in Grades 3–5. Effect sizes, calculated using Cohen's d, ranged from .05 to .18, suggesting small to moderate differences in performance across the two activity types. This range of effect sizes indicates that the amount of variance explained between ill-defined and well-defined tasks is limited, supporting the interpretation that the two tasks may engage distinct dimensions of spatial problem-solving ability.

The second research question aimed to investigate the psychometric properties of the puzzles used in the spatial analytical assessment. Specifically, means and standard deviations were calculated for the following two performance indicators: (a) the number of tangram pieces used in the initial shape activity, and (b)

the number of puzzle pages completed at each grade level. As shown in Table 3, both the average number of puzzle pages completed and the number of pieces used in the initial shape varied across grade levels, and generally increased with grade.

For instance, in Groups K–2 and 3–5, the average number of puzzle pages completed increased by approximately half a page per grade level. In contrast, for Group 6–8, the average number of pages completed remained relatively stable across grades. In Group 9–11, performance was more variable and did not show a consistent trend. Similarly, the average number of tangram pieces used in the initial shape activity increased steadily—by about two pieces across K–2, and by one piece per grade level in Groups 3–5 and 6–8—while performance in Group 9–11 showed inconsistency in progression.

Additionally, this research question sought to determine the difficulty level of each puzzle page by analyzing task completion time and performance patterns. In this context, longer average time on a page was interpreted as indicative of greater task difficulty, while the number of pages completed was used to infer the performance level of higher-ability students.

To assess puzzle difficulty, the author employed two complementary approaches. First, the mean and standard deviation of time spent on each puzzle page were calculated for all students within each grade category. Second, among students who reached and attempted page six, performance-based difficulty levels were calculated for each group. Table 4 and Graph 1 present data for all students who completed any puzzle page, regardless of ability level. Table 5 and Graph 2 present performance data for high-ability students within the four grade-level groups (K–2, 3–5, 6–8, and 9–12), where the assessment content was held constant across grades within each group.

Overall, the results indicated that as the number of puzzle pages increased, so did the average time required to complete them—suggesting increasing task complexity across pages. However, several exceptions to this trend were observed, reinforcing the necessity of this research to examine the ordering, structure, and difficulty of the assessment tasks (see Tables 4 and 5;

Graphs 1 and 2).

7. Discussion

This study addressed two primary research questions to inform and improve decision-making during the DISCOVER debriefing sessions and to enhance student engagement with the Spatial Analytical Intelligence Assessment. The first question examined the relationship between students' performance on the initial shape activity and the puzzle page tasks, aiming to determine whether these two components measure related or distinct aspects of spatial analytical intelligence. The second question focused on the psychometric properties of the puzzle items—specifically, the average completion time, item difficulty level, and the sequencing of the puzzle pages. The following sections present a discussion of the major findings in relation to the two research questions, their educational and psychometric implications, and the limitations of the current study.

Problem Types on SAA

The findings of this study indicate that performance on the ill-defined task (the initial shape activity) and the well-defined tasks (puzzle pages) is related but relatively independent, suggesting that these activities likely measure distinct components of spatial analytical intelligence—namely, divergent and convergent thinking, respectively. This interpretation aligns with prior research indicating that ill-defined and well-defined problem-solving are related yet separable constructs (Jaarsveld & Lachmann, 2017).

This pattern supports the theoretical framework of the problem continuum (Maker, 2004), which posits that as tasks shift from open-ended, divergent formats (e.g., Type V problems) to highly structured, convergent formats (e.g., Type I–IV problems), the degree of cognitive overlap diminishes. The observed significant but modest correlations between performance on the two task types reinforce this theory, revealing that while there is some shared variance, each task type also taps into unique aspects of cognitive functioning supporting the theoretical foundation of assessment development (Maker, 2020).

The inclusion of both task types in the assessment thus strengthens the construct validity of the Spatial Analytical Intelligence Assessment. Limiting an assessment to only one problem type may result in an incomplete profile of students' spatial problem-solving abilities, potentially overlooking individuals who excel in either divergent or convergent domains (Maker, 2020).

One interpretation of these results is rooted in the nature of the tasks themselves: the initial shape activity represents a Type V problem on the problem continuum, characterized by high openness and multiple solution paths, whereas the puzzle pages correspond to Types I–IV, which are decreasingly structured and constrained. These distinctions imply that the cognitive competencies required for each task differ meaningfully as theorized by Maker (1992).

These findings are consistent with previous research by Sak and Maker (2003), who examined the DISCOVER Math assessment and found higher correlations between adjacent problem types (e.g., Type I and II: r = .49, p < .01) compared to non-adjacent types (e.g., Type I and IV: r = .36, p < .01), supporting the idea that problem types closer on the continuum share more cognitive demands.

Further evidence for the distinctiveness of the two task types comes from the effect size analysis using Cohen's d. While all correlations were statistically significant across grade levels, effect sizes ranged from .05 to .18, reflecting small associations according to conventional benchmarks. This suggests that although performance on ill-defined and well-defined tasks is related, the degree of shared variance is limited. In practical terms, this means that student success on one type of task only partially predicts success on the other, reinforcing the conclusion that the two task types engage complementary but non-identical dimensions of spatial analytical intelligence.

Decision Making in Debriefing Sessions

The second research question aimed to determine the average number of puzzle pages completed by students at each grade level to support informed decision-making during debriefing sessions, particularly in assessing students' spatial analytical abilities. Overall, the average number of puzzle pages completed and the average number of pieces used in the shape activity

revealed a developmental progression in spatial analytical ability, especially in Groups K–2, 3–5, and 6–8 (see Table 4 and Graph 1).

In Groups K–2 and 3–5, the gradual increase in both puzzle page completion and shape complexity suggests that the assessment is sensitive to developmental differences. This pattern is consistent with theories of cognitive development, which posit that younger children demonstrate measurable improvements in problem-solving efficiency and cognitive complexity with age (Greiff & Fischer, 2013). These trends provide evidence that the assessment effectively captures growth in spatial analytical intelligence during early and middle childhood.

However, the relative stability of performance across Grades 6–8 may indicate a developmental plateau or, alternatively, a ceiling effect in the assessment tasks. That is, the puzzles administered at this stage may lack sufficient complexity to discriminate among students at different levels within the group, thereby limiting the assessment's psychometric sensitivity.

The inconsistency in results observed in Grades 9–11 raises additional concerns about the developmental calibration of the instrument. Variability in performance, without a clear upward or downward trend, may suggest that the puzzles at this level are either too easy, too difficult, or inconsistently aligned with the cognitive demands appropriate for this age group. These limitations could compromise the reliability and validity of the assessment for older students.

Taken together, these findings imply that while the assessment is a useful tool for identifying developmental trends and supporting debriefing decisions in early and middle grades, revisions are needed for higher grade levels. Enhancing the complexity and calibration of the puzzles for Grades 6–11 would strengthen the instrument's discriminatory power, ensure developmentally appropriate measurement, and promote more accurate interpretation of student abilities in debriefing contexts.

New Order of Puzzle Pages

A second aim of the second research question was to examine the difficulty level of each puzzle page across group levels, for which the assessment content was identical. Accurately inferring puzzle difficulty, however, proved challenging due to a key procedural constraint: not all students had the opportunity to attempt all six puzzle pages (Gul, 2022). In the DISCOVER Spatial Analytical Assessment, students were permitted to progress to the next puzzle only after successfully completing the previous one (Maker, 2020). As a result, the average time required to complete higher-numbered puzzles was based on a progressively smaller and more selective subset of students—typically those demonstrating higher spatial analytical ability (see Table 5).

To address this limitation, the researcher analyzed two complementary data sets: (1) all participants, regardless of how many puzzles they completed (Table 4), and (2) only those students who completed all six puzzle pages (Table 5). The second group represents high-ability performers, and their data offer a more accurate estimation of intrinsic puzzle difficulty, as all puzzles were attempted under equivalent conditions.

The researcher aimed to compare puzzle difficulty levels by evaluating discrepancies in performance between these two groups. This analysis revealed inconsistencies in the originally assumed order of puzzle difficulty. Accordingly, the determination of a precise sequence based solely on the existing data was deemed problematic. Nevertheless, the researcher proposed a revised puzzle order based on a triangulation of key indicators from Tables 4 and 5:

1. Mean completion time (Table 5) – Longer average times among high-ability students were interpreted as indicating greater intrinsic difficulty.

2. Percentage of students completing each page (%P from Table 4) – Lower completion rates suggested that puzzles might cause frustration or disengagement if placed too early in the sequence.

3. Standard deviation (SD) – High variability in completion time pointed to inconsistency or confusion, suggesting that such puzzles should be positioned later to maintain student momentum.

4. Time discrepancies between Tables 4 and 5 – Large gaps in average time between the two groups suggested that lower-ability students struggled disproportionately, further informing item placement

for accessibility.

Table 6 presents the revised puzzle order, which reflects a data-informed restructuring of the puzzle sequence to support both developmental appropriateness and student engagement. Puzzles with shorter completion times and higher success rates were moved earlier in the sequence to build confidence, while more complex or inconsistent puzzles were positioned later to maintain motivation and challenge. This approach promotes a gradual cognitive progression aligned with each group's abilities, fostering both perseverance and learning.

Although this analysis supports the need to revise the puzzle page order, the available data do not permit a definitive reordering. Consequently, the researcher recommends a follow-up validation study in which students across a range of ability levels and grade groups attempt all puzzle pages, allowing for more robust measurement of item difficulty and sequencing.

8. Limitations

Several limitations of the current study should be acknowledged. First, the sample was not demographically balanced. Native American students comprised more than half of the participants across all grade levels and were particularly overrepresented in Grades 6–8 (over 90%) and Grade 10 (83%). In contrast, students from other ethnic backgrounds—Hispanic, Caucasian, African American, and Asian American— were underrepresented. This uneven distribution may have inflated performance averages or introduced cultural factors that influenced task engagement or interpretation, thereby limiting the generalizability of the findings to more diverse student populations (Tharp, 1989; Yamauchi & Tharp, 1995).

Second, the determination of puzzle difficulty was based primarily on average completion time. However, this metric was limited by two factors: (a) not all students attempted or completed all puzzle pages, and (b) some students were provided with clues to assist with problem-solving. These conditions likely compromised the consistency and interpretability of completion times as indicators of intrinsic task difficulty, particularly for later puzzle pages attempted only by higher-performing students.

Together, these limitations highlight the need for cautious interpretation of the current findings. Future research should involve more demographically diverse samples and adopt more controlled testing procedures—such as ensuring that all students attempt all puzzle pages under uniform conditions—to improve the validity, reliability, and generalizability of results.

9. Conclusion

In conclusion, the findings of this study suggest that the current ordering of puzzle pages in the SAA requires revision. A proposed sequence based on empirical indicators of difficulty is presented in Table 6. To validate and refine this sequence, future research should involve administering all puzzle pages to students across diverse ethnic backgrounds, allowing unlimited time for completion. This would enable the calculation of more accurate difficulty indices based on the percentage of successful completions per puzzle. Additionally, analyzing average performance metrics— including the number of puzzles completed, time spent on each puzzle, and the number of pieces used in the initial shape activity—across both grade levels and ethnic groups would yield richer insights into developmental trends and cultural variability in spatial problem-solving. Such investigations would strengthen the psychometric properties of the instrument by enhancing its construct validity, reliability, and fairness. Ultimately, these efforts will support more informed and equitable decisions about students' spatial analytical abilities during DISCOVER debriefing sessions, contributing to more valid identification of talent and promoting inclusive educational practices.

Author Biography

Dr.Omar M. Muammar,

Ph.D., is a Professor of Giftedness, Creativity, and Leadership at Special Education Department - Imam Abdulrahman bin Faisal University. Former VP at IAU, consultant to Mawhiba, and Fulbright Scholar, he has led national gifted initiatives, authored research, and designed enrichment programs shaping Saudi gifted education.

Email: omuammar@iau.edu.sa.

References

Álvarez, V., Torres, T., Gangoso, Z., & Sanjosé, V. (2020). A cognitive model to analyse physics and chemistry problem-solving skills: Mental representation implied in solving actions. *Journal of Baltic Science Education*, 19(5), 730–746

. https://doi.org/10.33225/jbse/20.19.730

Andersen, L. (2014). Visual-spatial ability: Important in STEM, ignored in gifted education. *Roeper Review*, 36(2), 114121-.

https://doi.org/10.108002783193.2014.884198/

Aprile, G., Cappagli, G., Morelli, F., Gori, M., & Signorini, S. (2020). Standardized and experimental tools to assess spatial cognition in visually impaired children: A mini-review. *Frontiers in neuroscience,* 14, 562589.

https://doi.org/10.3389/fnins.2020.562589

Ayaz, H., Shewokis, P.A., Izzetoglu, M., Çakır, M.P., & Onaral, B. (2012). Tangram solved? Prefrontal cortex activation analysis during geometric problem solving. 2012 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 4724-4727.

Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? Predicting academic achievements in advanced STEM studies. *Journal of Intelligence*, 6(4), 48.

https://doi.org/10.3390/jintelligence6040048

Buckley, J., Seery, N., & Canty, D. (2018). A heuristic framework of spatial ability: A review and synthesis of spatial factor literature to support its translation into STEM *education. Educational Psychology Review*, 30(3), 947972-.

https://doi.org/10.1007/s106489432--018-z

Carroll, J. B. (1993). Human cognitive abilities*: A survey of factor analytic studies. Cambridge University Press*.

Chen, Z., & Siegler, R. S. (2000). Intellectual development in childhood. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 92–116). chapter, Cambridge: Cambridge University Press.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes*. The PASS theory of intelligence. Massachussets: Allyn and Bacon.

Del Cerro Velázquez, F., & Méndez, G. M. (2021). Systematic review of the development of spatial intelligence through augmented reality in STEM knowledge areas. *Mathematics*, 9(23), 3067.

https://doi.org/10.3390/math9233067

Dennis, I., & Tapsfield, V. (Eds.). (1996). *Human abilities: Their nature and measurement. Lawrence Erlbaum Associates.*

Domin, D.S., & Bodner, G.M. (2012). Using students' representations constructed during problem solving to infer conceptual understanding. *Journal of Chemical Education*, 89, 837843-.

Edens, K.M., & Potter, E.F. (2008). How students "unpack" the structure of a word problem: Graphic representations and problem solving. *School Science and Mathematics,* 108, 184196-.

Feldman, D. H. (1991). Has there been a paradigm shift in gifted education? In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development (*pp. 89–94). Trillium Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* Basic Books.

Gardner, H. (1999). *Intelligences reframed: Multiple intelligences for the 21st century.* Basic Books.

Getzels, J. W. & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal*,3(9), 8084-.

Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York, NY: Wiley.

Greiff, S., & Fischer, A. (2013). Measuring complex problem solving: an educational application of psychological theories. *Journal for Educational Research Online*, 5(1), 38–58.

https://doi.org/10.2565601:8019/

Griffiths, S. E. (1996). *The inter-observer reliability of the DISCOVER problem-solving assessment.* (Unpublished thesis). University of Arizona, Tucson.

Gul, N. (2022). Item difficulty in item analysis of intelligence test items. *Pakistan Social Sciences Review*, 6(II).

https://doi.org/10.35484/pssr.2022(6-ii)09

Humphreys, L. G., & Lubinski, D. (1996). Assessing spatial visualization: An underappreciated ability for many school and work settings. In C. P. Benbow & D. Lubinski (Eds.), *Psychometric and social issues: Intellectual talent* (pp. 116–140). The Johns Hopkins University Press.

Jaarsveld, S., & Lachmann, T. (2017). Intelligence and creativity in problem solving: The importance of test features in cognition research. *Frontiers in Psychology, 8*.

https://doi.org/10.3389/fpsyg.2017.00134.

Jensen, A. (1993). Spearman's g: Links between psychometrics and biology. *Annals of the New York Academy of Sciences, 702.*

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

Kell, H.J., & Lubinski, D. (2013). Spatial ability: A neglected talent in educational and occupational settings. *Roeper Review*, 35, 219 - 230.

Kmetová, M., & Nagyová Lehocká, Z. (2021). Using tangram as a manipulative tool for transition between 2D and 3D perception in geometry. *Mathematics*, 9(18), 2185.

https://doi.org/10.3390/math9182185

Kozhevnikov, M., Kozhevnikov, M., Yu, C.J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *British journal of educational psychology,* 83 Pt 2, 196209-.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measuremen*t (pp. 97–116). Lawrence Erlbaum Associates.

Maker, C. J. (1992). Intelligence and creativity in multiple intelligences: Identification and development. Educating Able Learners: *Discovering and Nurturing Talent*, XVII (4), 12- 19.

Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes, and a promising solution. *Gifted Child Quarterly*, 40, 41–50.

Maker, C. J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International,* 15, 232 - 251.

Maker, C. J. (2004). Creativity and multiple intelligences: The DISCOVER project and research. In *WORLD SCIENTIFIC eBooks* (pp. 341–392).

https://doi.org/10.11420014_9789812567192/.

Maker, C. J. (2020). Culturally responsive assessments of spatial analytical skills and abilities: Development, field testing, and implementation. J*ournal of Advanced Academics*, 31(3), 234–253. https://doi.org/10.11771932202/x20910697

Maker, C. J., & Schiever, S. W. (2005). Teaching models in education of the gifted. (3rd ed.). PRO-ED.

Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 24(4), 367–378.

Renavitasari, I. R. D., & Supianto, A. A. (2018, November). Educational game for training spatial ability using tangram puzzle. In 2018 *International Conference on Sustainable Information Engineering and Technology* (SIET) (pp. 174–179).

https://doi.org/10.1109/siet.2018.8693164

Sak, U., & Maker, C. J. (2003). Assessing and cultivating children's divergent and convergent thinking using the problem continuum. Paper presented at the 15th *World Conference for Gifted and Talented Children, Adelaide, South Australia.*

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

Sarouphim, K. M. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly*, 45(2), 130–138.

https://doi.org/10.1177001698620104500206/

Sattler, J. M. (2001). *Assessment of children: Cognitive applications*. (4th ed.). Jerome M. Sattler Publisher.

Sternberg, R. J. (1997). A Triarchic view of giftedness: Theory and practice. In Coleangelo; Davis (eds.). *Handbook of Gifted Education*. pp. 43–53.

Stumpf, H., Mills, C. J., Brody, L. E., & Baxley, P. G. (2013). Expanding talent search procedures by including measures of spatial ability: CTY's spatial test battery. *Roeper Review*, 35(4), 254–264.

https://doi.org/10.108002783193.2013.829548/

Tharp, R. G. (1989). Psychocultural variables and constants: Effects on teaching and learning in schools. *American Psychologist*, 44, 349–359.

Trickett, S. B., & Trafton, J. G. (2007). "What if…": The use of conceptual simulations in scientific reasoning. *Cognitive Science*, 31, 84375-. doi:10.108003640210701530771/

Yamauchi, L. A., & Tharp, R. G. (1995). Culturally compatible conversations in Native American classrooms. *Linguistics and Education*, 7, 349–367.

Yoon, S.Y., Hitchings, H., & Dsilva, M. (2023). A review of spatial assessments for research in science, technology, engineering, and mathematics education: A sample from northwestern SILC website. 2023 IEEE *Frontiers in Education Conference (FIE)*, 18-.

Table 1

Number of Students Across Schools, School Districts, and Groups (N = 3,434)

Schools and Districts	Groups				Total	%
	K–2	3–5	6–8	9–11		
A	40	337	132	16	525	15
B	45				45	1
C	45		1		46	1
D		1	22	7	30	1
E		58			58	2
F	272	208		65	545	16
G	91	85	95	85	356	10
H	402	320	298	9	1,029	30
I	191	221	33		445	13
J	221	89			310	9
K				27	27	1
L				18	18	1
Total	1,307	1,319	581	227	3,434	100
%	38	38	17	7	100	

Table 2

DISCOVER: Problem Type Continuum

	Problem		Method		Solution	
Type	Presenter	Solver	Presenter	Solver	Presenter	Solver
I	S	K	S	K	S	U
II	S	K	S	U	S	U
III	S	K	R	U	S	U
IV	S	K	R	U	R	U
V	S	K	U	U	U	U
VI	U	U	U	U	U	U

Note. K, known; S, specified; U, unknown; R, range. Data from Schiever and Maker (2005).

Table 3

The Average and the Standard Deviation of the Number of Pieces Used in the Initial Shape and the Number of Pages Completed by All Students at Each Grade Level (N = 3,434)

Grade	n^a	Puzzle Pages Completed		Pieces Used in Initial Page	
		M	SD	M	SD
K	560	4.6	0.9	3.3	4.2
1	356	5.1	0.9	5.1	5.4
2	391	5.6	1.0	7.0	5.5
3	305	4.5	1.1	5.1	5.1
4	435	5.1	1.1	6.3	5.9
5	579	5.6	1.1	8.6	6.4
6	239	5.2	1.4	9.8	6.7
7	202	5.2	1.6	11.4	6.8
8	140	5.1	1.3	10.2	7.8
9	149	4.2	1.7	5.4	7.1
10	53	5.2	1.4	12.4	9.0
11	25	4.6	1.1	5.8	9.1

a the ethnicity percentages in the sample contained approximately 12% Caucasian Americans, 5% Black Americans, 23% Hispanic, 59% Native Americans, and 1% Asian Americans.

Table 4

The Average Time for Completion of Puzzles by All Students Across Grade Groups (N = 3,434)

	Puzzle						
	1	2	3	4	5	6	Challenge
Group K–2							
M	2.32	2.33	4.61	3.74	8.17	6.71	5.83
n	1,307	1,302	1,287	1,185	1,033	396	87
SD	2.17	2.00	3.63	2.87	4.59	4.60	3.88
P	100	100	98	91	79	30	7
Group 3–5							
M	2.65	2.93	7.4	6.44	9.08	7.63	7.52
n	1,319	1,319	1,313	1,200	994	533	200
SD	2.35	2.73	5.63	4.77	5.59	4.80	4.80
P	100	100	99	90	75	40	15
Group 6–8							
M	5.43	5.05	9.15	6.3	6.12	6.99	7.85
n	581	580	567	496	399	267	143
SD	4.70	4.34	6.73	4.76	5.10	5.14	5.79
P	100	100	98	85	69	46	25
Group 9–11							
M	4.77	5.49	5.91	7.18	4.16	7.78	6.31
n	227	219	194	157	129	65	29
SD	4.39	4.05	5.03	4.90	3.88	6.62	7.84
P	100	96	85	69	57	29	13

Table 5

The Average Time for Completion of Puzzles by Only Students Who Completed Puzzle 6 Across Group Levels (N = 1261)

	Puzzle						
	1	2	3	4	5	6	Challenge
Group K–2 (*n* = 396)							
M	1.58	1.52	2.54	2.44	6.75	6.71	5.83
SD	0.89	0.74	1.50	1.54	3.72	4.61	3.88
Group 3–5 (*n* = 533)							
M	1.75	1.92	4.20	4.00	7.39	7.63	7.52
SD	1.32	1.22	2.65	2.45	4.25	4.80	4.81
Group 6–8 (*n* = 267)							
M	3.38	3.28	6.20	4.66	4.58	6.99	7.85
SD	2.23	2.13	4.16	2.94	3.34	5.14	5.79
Group 9–11 (*n* = 65)							
M	2.29	3.11	3.77	5.05	3.65	7.78	6.32
SD	1.89	1.95	2.79	3.39	4.29	6.62	7.98

Table 6

Summary Table of Revised Puzzle Orders

Grade Levels	Revised Order	Rationale
K-2	1 → 2 → 4 → 3 → 6 → 5 → Challenge	Start with easiest; delay Puzzle 5 (longest time & lowest completion); Challenge last to avoid early dropout.
3-5	1 → 2 → 4 → 3 → 6 → 5 → Challenge	Similar to K–2; Puzzle 4 easier than 3; 6 placed before 5 due to better handling by completers.
6-8	2 → 1 → 5 → 4 → 6 → 3 → Challenge	Group handles moderate tasks early; Puzzle 3 is inconsistent (high SD), so delayed; end with toughest.
9-11	1 → 2 → 5 → 3 → 4 → Challenge → 6	Start with familiar; Puzzle 5 is easier for older students; Puzzle 6 is most difficult and saved for last.

Figure 1

An illustration by lines of the average time completed for each page by those who completed all six puzzle pages
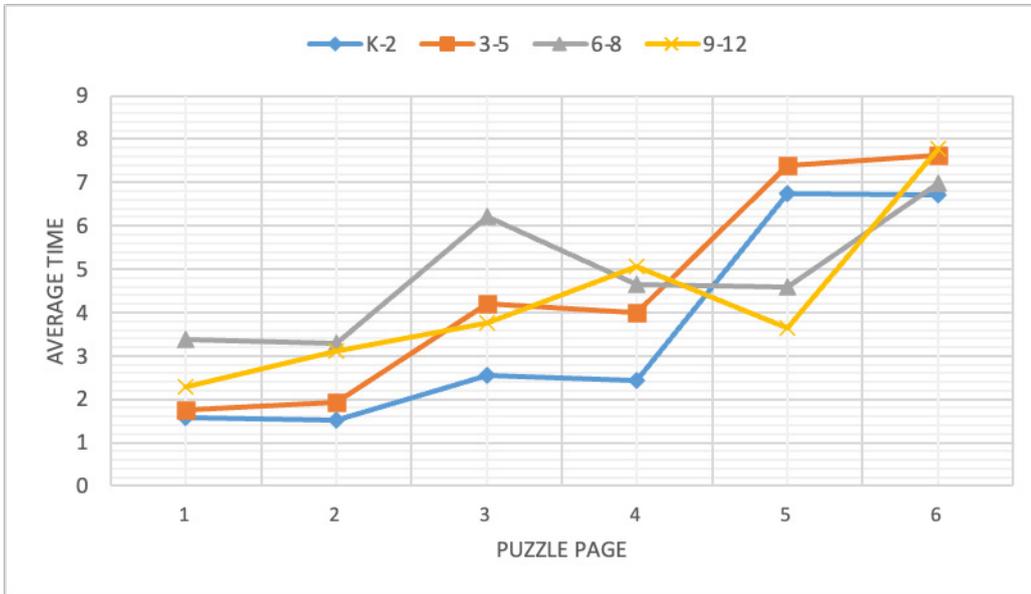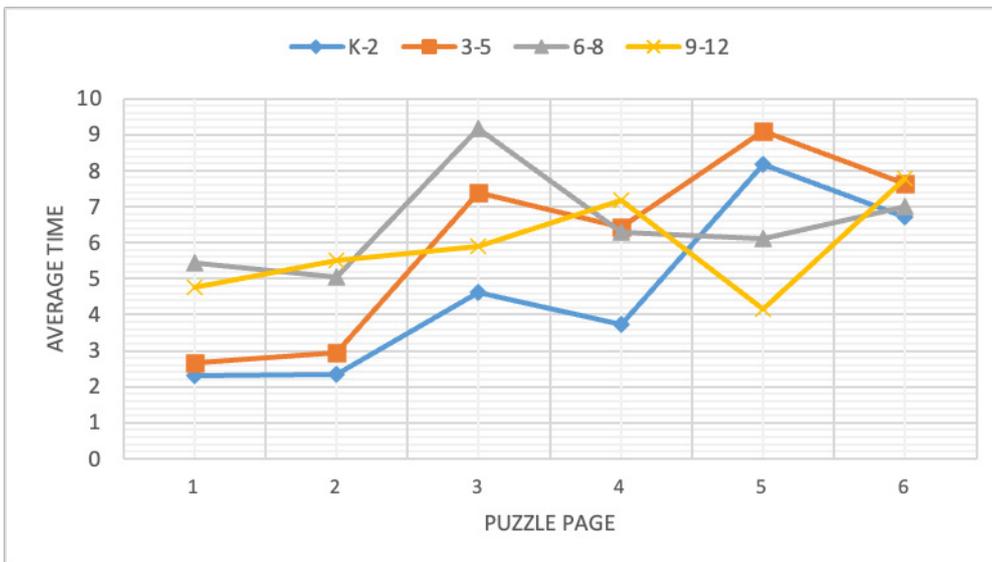

Figure 2

Average time elapsed for each page by those who completed all six puzzle pages.