Turkish Journal of Giftedness and Education
2018, Vol 8, Issue 2, 114-142
Türk Üstün Zeka ve Eğitim Dergisi
2018, Cilt 8, Sayı 2, 114-142

Alhusaini, A. A., & Maker, C. J. (2018). Who is gifted? The stability of scores on the DISCOVER assessment and the Raven's progressive matrices in Diné gifted children. *Turkish Journal of Giftedness and Education, 8*, 2, 114-142
Received: 17 August 2018; Accepted: 15 Nov 2018

Who is Gifted? The Stability of Scores on the DISCOVER Assessment and the Raven's Progressive Matrices in Diné Gifted Children

Kimler Üstün Yeteneklidir? Navaholu Üstün Yetenekli Çocukların DISCOVER ve Raven'in İlerlemeli Matrisi'ne Göre Değerlendirilmesi

Abdulnasser A. Alhusaini[1] & C. June Maker [2]

Abstract

The purpose of this study was to investigate the stability of the Discovering Intellectual Strengths and Capabilities through Observation while allowing for Varied Ethic Responses (DISCOVER) assessment, when used across time, to identify 74 Diné gifted children, and the Raven's Progressive Matrices (RPM), when used across time, to identify 52 Diné gifted children. Students were tested when they were in the second, third, and fourth grades. An analysis of the data using three methods (viz., calculating the change differences on each student's scores, obtaining correlation coefficients, and applying a linear single regression analysis across the three testing periods) provides evidence of the stability of three of the five DISCOVER activities—Spatial Analytical, Spatial Artistic, and Oral Linguistic, as well as the overall score. The overall DISCOVER score provides greater evidence of stability than do students' scores from the RPM, which varied from year to year. Based on the results of this study, the authors concluded that the DISCOVER assessment is a culturally fair instrument, and is more appropriate when used to identify Diné gifted children than traditional tests. Future researchers may consider conducting a long and large-scale longitudinal investigation into the same research problem, as well as designing a mixed-method study to investigate how Diné children understand the RPM problems to highlight any potential cultural components.

Key Words: DISCOVER, RPM, Identifying Gifted Students, Diné Children

Öz

Bu çalışmanın amacı, Navaholu (Amerika'da yaşayan bir yerli grup) öğrencilerin tanılanmasında kullanılan DISCOVER (Entelektüel Yeteneğin ve Potansiyelin Gözlem yoluyla Keşfedilmesi) değerlendirmesinden ve Raven'in İlerlemeli Matrisi'nden elde edilen verilerin farklı zaman aralıklarındaki tutarlılığını araştırmaktır. Çalışma kapsamında 2. 3. ve 4. sınıfta öğrenim gören 74 Navaholu üstün yetenekli öğrenciye DISCOVER, 52 Navaholu üstün yetenekli öğrenciye ise Raven uygulanmıştır. Veri analizinde fark analizleri, korelasyon analizi ve basit doğrusal regresyon analizi olmak üzere üç yöntemden yararlanılmıştır. Verilerin analizi sonucunda DISCOVER aktivitelerinden üçünün (Uzamsal Analitik, Uzamsal Sanat ve Dil) ve toplam puanın tutarlı olduğu bulunmuştur. Ayrıca DISCOVER'ın, Raven'in İlerlemeli Matrisi'ne göre farklı yıllardaki ölçümlerinde daha tutarlı olduğu sonucuna ulaşılmıştır. Çalışmanın bulguları DISCOVER değerlendirmesinin farklı kültürlerde kullanılabilecek bir ölçek olduğunu ve Navaholu üstün yetenekli öğrencileri tanılamada geleneksel testlere kıyasla daha uygun olduğunu ortaya koymuştur. İleri araştırmalarda, daha uzun yılları ve daha çok örneklemi kapsayan boylamsal araştırmalar gerçekleştirilebilir. Ayrıca potansiyel kültürel bileşeni vurgulamak için Navaholu çocukların Raven'in İlerlemeli Matrisi problemlerini nasıl anladıklarına yönelik karma araştırmalar da yapılabilir.

Anahtar Sözcükler: DISCOVER, Raven'in İlerlemeli Matrisi, üstün yetenekli öğrencilerin tanılanması, Navaholu çocuklar

[1] *Corresponding author, PhD, Assistant Prof, Department of Special Education, University of Jeddah; Jeddah, Saudi Arabia Kingdom; aalhusaini@msn.com*
[2] *PhD, Prof., Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, AZ, USA*

ISSN 2146-3832, http://www.tuzed.org

Introduction

Although the term *gifted* has been used frequently in today's schools, answering the fundamental question, "*Who is Gifted*?" has not been easy (cf. Borland, 2008; Reis & McCoach, 2000). To provide a conservative operational definition, a gifted student would be one who has met the determined criteria set by policy-makers in a state or school district. Therefore, identifying gifted students has been completely dependent on the definition(s) of giftedness adopted by state, local, or national policy-makers. Consequently, in the field of education of gifted students, definitions of giftedness vary from one place to another and also across programs. Alhusaini (2006) summarized the four major factors that have influenced the criteria for defining giftedness, especially when establishing a new program for gifted students: (a) the philosophy of the program, including vision, mission, theoretical framework, and objectives; (b) the country's needs, community's needs, or cultural values of the population; (c) the amount of available money or funding allocated to establish the program; and (d) the instruments or criteria that were available and could be adapted to the program. Those four major factors have made a student be considered gifted in one country, but not in another, or in one state, but not in another, or even in one program in a state, but not in another.

Programs for Gifted Students

A well-developed program for gifted students should include several related and consistent components. Each component should be linked to the previous one and foundational to the next step. The authors of the current study believe that a well-developed program for gifted students has five major components: (a) philosophy, (b) definition, (c) identification, (d) teaching, and (e) final evaluation (See Figure 1). While we agreed that identifying gifted students has been a very important step, educators should view identification as one of many steps that have contributed to the program. Identification, therefore, has been used in a well-developed program for gifted students to reflect the adopted operational definition of giftedness and to tailor teaching practices.

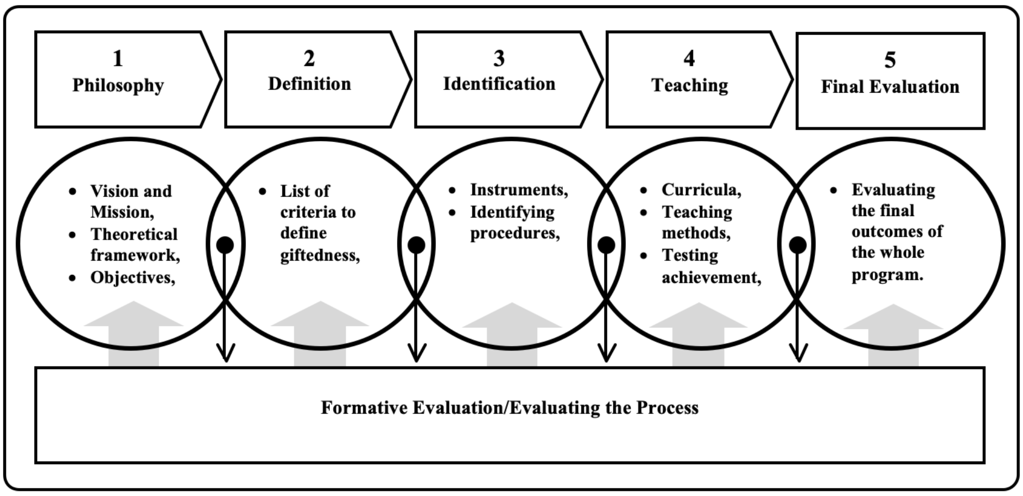

Figure 1. The Major and Essential Components in Programs for Gifted Students

In the field of educational testing and measurement, a large body of research was found on test reliability. Miller (n. d.) defined reliability as "…the extent to which a questionnaire, test, observation or any measurement procedure produces the same results on repeated trials" (p. 1). Kubiszyn and Borich (2007) said further, "a reliable test will yield stable scores over repeated administrations, assuming the trait being measured has not changed" (p. 326). In the current study, we used the term *stability* to refer to the consistency with which the instruments were able to identify the same children as gifted throughout multiple years and then continue serving them in a program for gifted students over time. In the field of educating gifted students, educators not only wanted to reflect the adopted operational definition of giftedness, but also they wanted to use stable instruments. In fact, the education of gifted students has been viewed as an investment by countries or societies (cf. Clinkenbeard, 2007; Shah 2011); consequently, educators of gifted students wanted to invest their efforts in the right children over time. From this perspective, investigating the stability of the instruments used in education of the gifted is highly important.

Procedures for Identifying Gifted Students

With respect to the official adopted definition of giftedness, many different practices could be used as procedures for identification. Yamin (2006), for example, described one of the most popular procedures that could be used in public schools that might contain a large number of students, which he called "multiple steps of identification." The main idea of the multiple steps of identification process was that educators divided the selection procedure into two or more stages (cf. Castellano, 2003; Coleman, 2001, 2003; Yamin, 2006). For instance, the first stage was the screening process, by which gifted students were nominated by their teachers or selected based on their achievements. The second stage was the identification process, in which students who passed the screening process were carefully examined using individual and standardized instruments to test their IQ and creativity. According to Yamin (2006), the multiple steps of identification process might be useful in saving time and money by focusing on only the recognized group of students, rather than all students, in the public school. Students who passed the second stage would be labeled as "gifted" and served in a program for gifted students. Similarly, Castellano (2003) proposed an extended form of the multiple steps of identification process with three stages: (a) general screening or student search, (b) review of students for eligibility, and (c) services options match. Using the same idea, Renzulli (1990) presented six stages of the multiple steps of identification process in his model.

Although in many programs for gifted students, the multiple steps of identification process has been used widely due to its practicality and use of multiple criteria to select gifted students (Castellano, 2003), educators continue to question and criticize the use of this process. For instance, in the field of special education, measurement errors yielded two probable outcomes: (a) positive, which meant including a student to be served in the special program even though he or she was not truly qualified; and (b) negative, which meant excluding a student from the services even though he or she was qualified. These measurement errors could occur in the field of giftedness for many reasons. For example, the probability of misidentifying gifted students who were underachievers (cf. Clemons, 2008; Wellisch & Brown, 2011; Ziegler, & Stoeger, 2012); twice exceptional (cf. Beckley, 1998; Brody & Mills, 1997; Nowak, 2001); and those who were from Culturally and

Linguistically Diverse (CLD) groups (cf. Ford, 1998; Ford, Grantham, & Whiting, 2008; Ramos, 2010) has been common. Problems exist at each step of the process. At the referral stage, for example, only one teacher's opinion could prevent a child from participating in screening or identification. At both the screening and identification stages, the instruments—if biased or inappropriate—have the potential to prevent CLD students from being identified as gifted.

Identifying Gifted Students from CLD Groups

Researchers have not agreed on a specific selection procedure to use in screening for gifted students from CLD groups; however, they have recommended using instruments that could be culturally fair (cf. Cole & Zieky, 2001; Gregory, 2004; Sattler, 1988). Often, non-verbal tests are recommended (e.g., Raven's Progressive Matrices, Naglieri Nonverbal Abilities Test, and CogAt Nonverbal). Although most educators understand that using certain instruments has resulted in overrepresentation of the dominant cultural group and underrepresentation of minority groups, Erdimez and Maker (2012) synthesized data from 2006 from both the Office for Civil Rights and the Arizona Department of Education. They found that Caucasian students made up 47.04 % of the overall student population and 64.65 % of the students served in programs for gifted students. African-American students made up 5.13% of the overall student population and 2.88% of the students in programs for gifted students. Mexican-American students made up 39.67% of the overall student population and 23.21% of the students in programs for gifted. American Indian students made up 5.64% of the overall student population and 3.23% of the students in programs for gifted students. Nationally, the picture was better, but certain groups remained underrepresented: Caucasian students made up 57.7% of the school-aged population and 67.7% of programs for the gifted; African-American students made up 14.8% of the school population and 9.1% of students in programs for the gifted; Mexican-American students, 19.7% and 12.8%; American Indian/Alaskan Native students, 0.9% and 0.9% (Snyder & Dillow, 2012). These statistics showed that, despite efforts to develop better ways to identify gifted CLD students, the disparity in their proportional representation in programs for the gifted remained.

Giftedness as Domain General Versus Domain Specific

In the recent literature, arguments have been made for both types of giftedness (i.e., domain-general and domain-specific). The idea of domain-general giftedness could be traced back to Spearman's *g* theory, in which cognitive ability was viewed as one domain and measured by traditional IQ tests (Spearman, 1923). A more recent conceptualization of the domain-general idea of giftedness in educational practice has been the idea that a student who scored high on an IQ test would perform high in any specific area (cf. Clark, 1997; Colangelo & Davis, 1991; Coleman & Cross, 2001). On the other hand, the idea of domain-specific giftedness is that cognitive ability occurs in many different domains (cf. Feldhusen, Hoover, & Sayler, 1990; Gilliam, Carpenter, & Christensen, 1996). Therefore, a student who scored high in one domain as measured by an instrument that included multiple domains would perform high in that specific domain. A simple example of a theory that was developed from a domain-specific standpoint was Gardner's (1983; 1999) theory of Multiple Intelligences (MI), in which intelligence was no longer seen as restricted to a single domain, but different across many domains (i.e., Spatial, Linguistic, Logical Mathematical, Musical, Naturalist, Existential, Interpersonal, Intrapersonal, and Bodily-Kinesthetic).

Non-Verbal Group Administered Tests Versus Performance-Based Assessments

Non-Verbal Group Administered Tests. Tests have been categorized as non-verbal if items contained visual stimuli such as concrete objects or line drawings and required a non-verbal response, such as putting pieces together, completing a visual pattern, pointing to an answer, or filling in a circle under a picture (Lohman, 2005). Non-verbal tests have been used to measure students' skills independently of their language proficiency. They have been used frequently in today's schools to measure students' intelligence and then to identify gifted students. Due to the fact that non-verbal tests did not contain items with verbal stimuli, they were accessible for children from various cultural backgrounds in different programs for gifted students (Bittker, 1991).

The most popular example of non-verbal group administered tests has been the Raven's Progressive Matrices (RPM). The RPM was published in 1936, and was developed based on Spearman's *g* theory, After 76 years, during which all educational fields have developed extensively in both theoretical and practical aspects, the RPM still is being used as a screening or identification tool, particularly in the field of education of gifted students. Educators who have supported the use of RPM have discussed many advantages such as (a) not having a verbal component, which helped to minimize cultural bias (Saldaña, 2001); (b) assisting educators in labeling or ranking students among their peers (Baska, 1986); (c) having reasonable psychometric properties (Raven, 2000); and (d) having easy administration and scoring (Neisser, 1997). Despite these advantages, other researchers have criticized the RPM in different areas, such as for (a) not measuring all students' strengths (Glasser, 1993), (b) not providing specific information about children's academic strengths (Johnson, 2006), and (c) including questions with only one correct answer.

Recently, Tan and Maker (2012) investigated the predictive validity of the RPM for identifying academic achievement of Diné children. They compared second grade students' RPM scores with the same students' fourth grade scores on the Iowa Tests of Basic Skills (ITBS) and the Comprehensive Tests of Basic Skills (CTBS). They found that the RPM significantly predicted overall achievement and students' performance in reading and math. However, they did not find that the RPM was a significant predictor of language assessment scores. In an extended review, Raven (2000) presented evidence of the RPM's reliability and validity from different studies (Raven, 1941; Raven & Walshaw, 1944; Byrt & Gill, 1973; Kratzmeier & Horn, 1979).

The reliability of the RPM was examined, especially in its use with different ethnic groups. For example, in South Africa, Owen (1992) investigated the reliability of the RPM with different ethnicities (Caucasian, Indian, and African students). Overall, he found that the RPM was a reliable instrument. However, when conducting a discriminate analysis, the researcher found that the largest proportion of the African and Caucasian students had a similar pattern with their groups. He also found large mean differences, especially between Caucasian and African students. Similarly, Jensen (1973) examined the reliability of the RPM on a sample of Caucasian, African-American, and Mexican-American students from kindergarten to eighth grade. Large mean differences were found among the three groups of students. He also concluded that the RPM was significantly culturally biased against African-American students.

Performance-Based Assessments. For the last 20 years, performance-based assessments have frequently been investigated by researchers and used in schools. Pinchok and Ploeg (2009) stated that in the second half of the 20th century researchers considered the necessity of getting quality feedback from learners, involving the judgment of teachers and experts, and applying modern research findings in education. Specifically, the first official attempts to use performance-based assessments took place in the measurement of the levels of Bloom's Taxonomy. Educators formulated instructional objectives for each level of Bloom's Taxonomy and developed a flexible method to measure exactly how well students achieved instructional objectives. Later, researchers extended and improved the uses of performance-based assessments to include measurement of both the students' cognitive abilities and achievement. Van Tassel-Baska, Johnson, and Avery (2002) argued that performance-based assessments were domain-specific, provided an alternative perspective for assessing students' performance, were designed to have a parallel structure with curricula, and included open-ended problem solving. Notwithstanding these advantages, many criticisms have been presented of performance-based assessments, such as that the assessments are not structured, and are subjective, time-consuming, and expensive.

The Discovering Intellectual Strengths and Capabilities through Observation while allowing for Varied Ethic Responses (DISCOVER) assessment has been developed as a performance-based assessment to address the needs of students from diverse backgrounds who have been underrepresented in programs for gifted students (Maker, Nielson, & Rogers, 1994) by combining Gardner's (1983, 1999), Sternberg's (1985, 1997), and Getzels and Csikszentmihalyi's (1967, 1976) theories. The assessment was designed to tap into the strengths of students from kindergarten to high school levels. Maker (1996) argued that using a performance-based assessment was more appropriate than using traditional measures, because the former was closer to lifelike situations, and was problem-based rather than knowledge-based. She also argued that using the traditional three-stage referral, scoring, and selection process was not appropriate for identifying CLD children who are gifted, because teachers may not recognize the abilities of students whose cultures and languages were different from their own (McBee, 2006; 2010). Results from research about the DISCOVER assessment have indicated that it was a useful and successful instrument for identifying students' strengths, and that it has reasonable psychometric evidence of reliability and validity (Griffiths, 1997; Lori, 1997; Nielson, 1994; Sak & Maker, 2003; Sarouphim, 2000).

Conducting the current study has been highly important for several reasons. As a research implication, in his extended review, Raven (2000) presented evidence of the RPM's stability across time with different groups of young people based on the following studies: Raven (1941); Raven and Walshaw (1944); Byrt and Gill (1973); and Kratzmeier and Horn (1979). However, he stated that a study of the RPM's stability across time with the same group of students needed to be conducted. The current study would also be the first study in which the DISCOVER assessment's stability across time has been investigated. As a theoretical implication, the results of the current study would provide evidence of giftedness—whether as domain-general or as domain-specific—due to the fact that two instruments, one designed based on each theoretical perspective, were used with the same group of students. Therefore, the theoretical validity of both Spearman's (1923) *g* and Gardner's (1983; 1999) MI would be examined equally. As a practical implication, investigating the stability of both instruments across time would assist educators in choosing instruments that have been found to be consistent and stable across time, especially when identifying students from CLD groups.

Purpose

The purpose of this study was to investigate the stability of both the DISCOVER assessment and the RPM when used across time to identify Diné gifted children. Students were tested using both instruments when they were in the second, third, and fourth grades. In this study, we examined the stability of the scores on the two instruments across the three testing periods. The questions that guided the study were as follows:

1. What was the consistency of scores on the DISCOVER assessment across the three years?
   (a) Scores on each of the five activities (viz., Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic)
   (b) Overall scores
2. Which of the five DISCOVER activities were the most consistent across the three years?
3. What was the consistency of the RPM scores across the three years?
4. Which of the two instruments yielded the most consistent scores across the three years?
   (a) The DISCOVER assessment
   (b) The RPM
5. Which of the overall scores was the most consistent in predicting giftedness in students across the three years?
   (a) The DISCOVER assessment
   (b) The RPM

Method

Research Design

The authors investigated the stability of both the DISCOVER assessment and the RPM when used to assess giftedness in Diné children across three years. From a holistic perspective, the current study met all the criteria for longitudinal research: (a) the data were collected for two or more distinct periods, (b) the same participants were used in each testing period, and (c) analysis involved comparison of data between or among periods (Menard, 1991; Miller & Brewer, 2003). However, from a methodological perspective, the current study would be considered quantitative research in which a correlational design was used to answer the research questions (Fraenkel & Wallen, 2010; Kubiszyn & Borich, 2007).

Settings

During the Systematic Training of Educational Programs for Underserved Pupils (STEP-UP) project that was implemented from 1990 to 1993, four methods (viz., the DISCOVER assessment, Cognitive Abilities Test, Raven, and teacher recommendations) were used to identify gifted Diné children and place them into the STEP-UP classes. The second author of the current study was the coordinator of the STEP-UP project in the Southwestern region of the United States. She continued following the same children and has retested these students each year from the second through tenth grades (i.e., for 9 years) using almost the same instruments. The Diné children who were participants in this study were placed in self-contained classrooms in four different schools (A, B, C, and D) in grades 2, 3, and 4. In school A (grades K to 8), the total population was approximately 630 students, and 98.99% of students were Diné. In school B (grades K to 6), the total population was approximately 390 students, with 99% of students identifying themselves as Diné. In school C (grades K to 12), the total population was 430 students, and 98% of students were identified as Diné. In school D (grades K to 6), the total population was 574 students, of which 99% were identified as Diné. All schools were located in the Diné Nation, and all were in rural, low-income areas. At least 94% of the students at each of these schools came from low-income families. More details about the start of the longitudinal data collection are available in the STEP-UP project report (Maker, 1993).

Participants

When the STEP-UP project was initiated, the total number of Diné students who were enrolled in the self-contained classrooms was approximately 76 male and female children. (See Table 1 for an explanation of the selection procedure and population of children who participated in each school.) During the data collection periods, some students moved from schools or were absent during the administration of tests. Therefore, in the current study, the authors considered only the data of students who were present for all testing periods. For the DISCOVER assessment, the total number of students who met this condition and were involved in the current study was 74 Diné children, approximately 97.36% of the original sample; and for the RPM, the total number of student who met this condition and were involved in the current study was 52 Diné children, approximately 68.42% of the original sample.

Table 1. Methods of Placement and Numbers of Children Placed in the STEP-UP Classes

School	PS & S	PS only	PS & T	T only	T & S	S only	Total
A	6	9	0	1	1	2	19
B	0	11	1	1	2	2	17
C	10	0	0	0	0	8	18
D	3	5	1	1	2	6	22
Totals	19	25	2	3	5	18	76

Notes: Adapted from the Report on project STEP-UP (Maker, 1993); PS = the DISCOVER assessment; S = Standardized (i.e., Cognitive Abilities Test and Raven); T = teacher recommendation.

Instruments

The DISCOVER assessment. In the past two-and-a-half decades, Maker and her colleagues have been developing and testing a measure to assess students' abilities to solve a variety of problem types in five different domains of intelligence: Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic. The assessment was designed based on a theoretical framework in which three theories were combined—Gardner's (1983; 1999), Sternberg's (1985; 1997), and Getzels and Csikszentmihalyi's (1967; 1976). The assessment was performance-based and criterion-referenced to address the needs of students from diverse backgrounds who have been underrepresented in programs for gifted students (Maker, Nielson, & Rogers, 1994).

Assessment procedure. During the administration of the assessment, students were engaged using hands-on problem-solving activities in their familiar environments. The classroom teacher gave directions to the students and teams of three to five certified observers monitored the problem-solving behaviors of up to five students in each group. When the activities used for the assessment were over, the observers met to complete a behavior checklist for each student and came to consensus on the rating of each student for each of the five domains based on a scale of four categories (i.e., Definitely, Probably, Maybe, and Unknown).

Reliability evidence. The assessment has been found to be a reliable instrument in several studies. For instance, Griffiths (1997) investigated the inter-rater reliability and found that expert DISCOVER observers, those who had conducted 30 or more assessments, agreed between 92% and 100% of the time, as $r = .92$ to 1.00—more than novice DISCOVER observers, those who had conducted less than 10 assessments, who agreed between 47% and 92% of the time, $r = .47$ to .92. However, across all experience levels, observers agreed 95% ($r = .95$) of the time on the highest rating of problem solving. In another study, Kassymov (2000) investigated the inter-rater reliability among the DISCOVER team members and found it was 81% overall and 100% on the highest rating.

Validity evidence. The assessment also has been found to be valid in several studies of different types of validity: construct, concurrent, and predictive. For instance, Sarouphim (1999a) evaluated the construct validity of the DISCOVER assessment, and found that the percentage of students who received the highest ratings was similar across different cultural backgrounds. Sarouphim (1999b) also studied the concurrent validity in particular by investigating the correlation between the RPM and the five activities of the DISCOVER assessment. Sarouphim found that the highest correlations between the RPM and DISCOVER were with the spatial artistic ($r = .58$, $p < .01$), spatial analytical ($r = .39$, $p < .01$), and math ($r = .35$, $p < .01$) activities of the DISCOVER assessment. The lowest correlations, as expected, were with the oral ($r = .20$) and written linguistic ($r = .093$) activities, so she interpreted these results as associations between two non-verbal logical reasoning tasks that demonstrated the concurrent validity of the DISCOVER assessment. Additionally, Lori (1997) investigated the relationship between Oral Linguistic abilities and the personal traits of Bahraini students. He found that significant relationships existed between students' Oral Linguistic abilities and their interpersonal and intrapersonal traits.

Sak and Maker (2003) examined the predictive validity of the DISCOVER assessment in two studies. They found significant differences between students who were identified as gifted and those not identified as gifted across two instruments (viz., the Arizona Instrument to Measure Standards [AIMS] and the Stanford 9 Achievement Test). They also concluded that the results supported the use of the DISCOVER assessment as an instrument to identify gifted students, and that kindergarten results could predict achievement as much as 6 years later. Erdimez and Maker (2012) also investigated the predictive validity of the DISCOVER assessment for students' achievement as measured by the ITBS and CTBS. The DISCOVER assessment was administered at the beginning of the third grade, and the achievement tests were administered at the end of fourth grade. Erdimez and Maker found that as a model, students' scores on all components of the DISCOVER assessment accounted for 43.9% of the variance in total achievement scores of the students. Therefore, they concluded that strong evidence for the predictive validity of the assessment was obtained.

Raven's Progressive Matrices. This test has been referred to as Raven's Matrices, Raven, or RPM. The RPM was published in 1936 and used as a nonverbal group-administered test to assess general cognitive ability from five to 80 years of age, based on Spearman's *g* theory (1923). The RPM was developed as a norm-referenced, multiple-choice test, and administered solely in pencil-paper format. In the current study, the Colored Progressive Matrices (CPM) was used with Diné children in second grade because they were within its target group of children aged five through 11 years; also, the Standard Progressive Matrices (SPM) was used with Diné children in third and fourth grade because they were within its target group of children ages five and over.

Assessment procedure. The CPM was administered to children by giving them two sets (A and B) of 36 mostly colored background items, listed in order of difficulty. The students' task was to identify the missing element that completed a pattern. The SPM was administered to children by giving them five sets (A to E) of 12 items each, a total of 60 items. All items were presented against a black and white background and listed in order of difficulty. Similarly, the students' task was to identify the missing element that most appropriately completed a pattern.

Reliability and validity evidence. The RPM has been used widely in both research and identification of gifted students. Sattler (1988) stated that test-retest reliabilities ranged from .71 to .92 and concurrent validity estimates ranged from .55 to .86. Also, in the most recent version of the CPM manual, the publisher stated that the split-half reliability of the CPM was found to be .97. To determine concurrent validity, the correlation between the CPM and the CPM-Parallel was examined and found to be .87, with a standard error of measurement for the CPM standardized score of 2.62. See Table 2 for a summary of the similarities and differences between the DISCOVER assessment and RPM.

Data Collection

The data were collected from the archives of the DISCOVER projects at the University of Arizona. More specifically, the author selected students who were enrolled in classrooms that were part of the STEP-UP project. In the DISCOVER project archives, longitudinal data about students who were enrolled in the STEP-UP project's classrooms were available for researchers, including results that were collected using a variety of tests (See the report from the STEP-UP project [Maker, 1993] for more details).

Table 2. The DISCOVER Assessment Versus Raven's Progressive Matrices

Items	The DISCOVER Assessment	Raven's Progressive Matrices
Author	C. June Maker	John C. Raven
Date of publication	1987	1936
Theoretical framework	Gardner (1983; 1999), Sternberg (1985; 1997), and Getzels and Csikszentmihalyi (1967; 1976)	Spearman (1923)
Target group	Students from preschool to the twelfth grade	People from age five to the elderly
Measurement	Students' abilities to solve a variety of problem types in different domains: Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic	People's cognitive abilities (i.e., educative and reproductive) based on Spearman's *g* theory
Psychometric properties	Found to be reliable and valid	Found to be reliable and valid
Standardization	Standardized in the administration, scoring, and interpretation	Standardized in the administration, scoring, and interpretation
Administration	Mainly group-administered (maximum of 5 students in each group)	Mainly group-administered (no maximum numbers)
Approximate time	Two and a half hours	From 15 to 30 minutes
Kind of referenced	Criterion-referenced	Norm-referenced
Type of assessment	Performance-based assessment	Solely pencil-paper test
Task	Hands-on problem-solving activities	Multiple-choice
Versions	(a) Preschool (b) Kindergarten to the twelfth grade (i.e., K to 2nd, 3rd to 5th, 6th to 8th, and 9th to 12th)	(a) Colored Progressive Matrices (b) Standard Progressive Matrices (c) Advanced Progressive Matrices
Uses	With all students, especially to identify CLD gifted students	With all students, especially CLD students
Results	Ordinal ratings: Definitely, Probably, Maybe, and Unknown	Numeric value and percentile scores
Interpretation	'Definitely' score in two activities indicated giftedness	A cut-point (e.g., 95th percentile) indicated high ability

Procedure

The DISCOVER assessment was designed to assess students' performance in five separate domains: Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic. Thus, a numerical rating of each student's overall performance has not been employed in assessments, but students' performance in each domain have been reported as "Definitely" a superior problem solver in the activity = 4, "Probably" a superior problem solver = 3, "Maybe" a superior problem solver = 2, and "Unknown" if the student was a superior problem solver = 1. On the other hand, the RPM was designed to assess students' general cognitive abilities by transferring a raw score of correct answers to standardized scores such as percentile, which allowed researchers or teachers to compare a student to the norm group and then make a decision about his or her cognitive ability. In the DISCOVER assessment, the authors of the current study obtained the numeric values for each domain separately to indicate students' ability in that domain, and they used the average of all numeric values of all domains to indicate students' overall ability.

Data Analysis

For questions one and two, the frequencies and percentages of students' performance in each of the five DISCOVER activities were analyzed to examine how they were distributed at each grade level. The differences among students' scores across the three testing periods (between grades 2 and 3, 3 and 4, and 2 and 4) were then calculated for each of the five DISCOVER activities (Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic). To determine the stability of scores on each activity, the authors formulated two critical criteria: (a) the change in each student's score was considered low if it was categorized between 0 to ±1 and high if it fell between >1 to 4, and (b) the cut-off point for making decisions about the stability of each activity was that ≥ 70% of the students' scores must not change, categorized between 0 to ±1. If an activity met these two criteria, it would be considered as stable for use with Diné students.

For question three, students' RPM raw scores from all three years were converted into a Z-score (M = 0.00, SD = 1.00). The average of the highest converted Z-score values was 2.5117/4 = 0.6279. Therefore, the authors decided to categorize the RPM scores into four ordinal levels, each with a value of 0.6279, to be equivalent to the rating level of the DISCOVER assessment. Then, the differences among students' scores across the three testing periods (between grades 2 and 3, 3 and 4, and 2 and 4) were calculated in the same manner the authors used to calculate these differences in the data from the DISCOVER assessment. We also conducted a further analysis to examine the correlation between the RPM scores in grades 2 and 3, between grades 3 and 4, and between grades 2 and 4.

For questions four and five, the authors used regression analysis. Gravetter and Wallnau (2009) emphasized that basic correlation techniques would be appropriate to use when evaluating or testing the significance of associations between two variables. However, if researchers were interested in investigating how well a variable(s) predicts the other(s), the appropriate analysis would be regression, which they defined as a "statistical technique for finding the best-fitting straight line for a set of data […] and the resulting straight line [has been] called the regression line" (p. 566). Therefore, the authors used the SPSS software (i.e., Premium Grad Pack 21 for Mac) to apply regression analyses and answer the last two research questions.

Results

What Was the Consistency of Scores on the DISCOVER Assessment across Three Years?

Spatial Artistic. The activities of the Spatial Artistic subsection of the DISCOVER assessment were designed to assess which students exhibited strength in spatial abilities. In the second grade, 47.29% of students were rated "Definitely" superior problem solvers; in the third grade, the proportion of "Definitely" superior problem solvers was 35.13%; and in the fourth grade, 55.40% of students were rated as "Definitely" superior in this category (See Table 3 for more details). When analyzing the difference between students' scores in grades 2 and 3, the percentage of low change scores (0 and ±1) was 71.62; between grades 3 and 4, 74.32; and between grades 2 and 4, 77.02. The average percentage of students' scores categorized between 0 and ±1 (low change) across all three years was 74.32. Therefore, the Spatial Artistic activities of the DISCOVER assessment met the two criteria and was stable when used with Diné students (See Table 4).

Table 3. Ratings of Students' Performance in the Spatial Artistic Activities across the Three Years

Student	Calculation	The DISCOVER Assessment Rating Levels				
		Mis.=0	**U=1**	**M=2**	**P=3**	**D=4**
In Second Grade	*f*	0	5	23	11	35
	%	0%	6.75%	31.08%	14.86%	47.29%
In Third Grade	*f*	0	4	26	18	26
	%	0%	5.40%	35.13%	24.32%	35.13%
In Fourth Grade	*f*	1	2	13	17	41
	%	1.35%	2.70%	17.56%	22.97%	55.40%

Note. The sample size was 74 students; Mis. = missing value or unrated student; U = unknown; M = maybe; P = probably; D = definitely; *f* = frequency; % = percentage.

Table 4. ***The Stability of Students' Scores in the Spatial Artistic Activities***

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	**High Change >1 to 4**
***X* of 2nd–*X* of 3rd**	**Number of Students (%)**	53 (71.62%)*	21 (28.37%)
***X* of 3rd–*X* of 4th**	**Number of Students (%)**	55 (74.32%)*	19 (25.76%)
***X* of 2nd–*X* of 4th**	**Number of Students (%)**	57 (77.02%)*	17 (22.97%)
Average Number of Students		55	19
Average Percentages of Overall Stability		74.32%*	25.68%

Note. The sample size was 74 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage; * = cut-off point used by the authors to determine stability of test was that ≥ 70% of the students' scores must not change more than ±1, thus * indicates the test has been fairly stable.

Spatial Analytical. The activities of the Spatial Analytical subsection of the assessment were designed to identify students who were superior analytical thinkers. In the second grade, 43.24% of students were rated as "Definitely" superior in this category; in the third grade, 44.59%; and in the fourth grade, 50.00% (See Table 5 for more details). When analyzing the difference in students' scores between grades 2 and 3, the percentage of low change scores (0 and ±1) was 75.67; between grades 3 and 4, 81.08; and between grades 2 and 4, 74.33. The average percentage of the students' scores categorized between 0 to ±1 (low change) across all three years was 77.02. Therefore, the Spatial Analytical activities of the DISCOVER assessment met the two criteria and were considered stable when used with Diné students (See Table 6).

Table 5. Ratings of Students' Performance in the Spatial Analytical Activities across the Three Years

Student	Calculation	The DISCOVER Assessment Rating Levels				
		Mis.=0	U=1	M=2	P=3	D=4
In Second Grade	*f*	2	7	20	13	32
	%	2.70%	9.45%	27.02%	17.56%	43.24%
In Third Grade	*f*	2	5	12	22	33
	%	2.70%	6.75%	16.21%	29.72%	44.59%
In Fourth Grade	*f*	0	6	13	18	37
	%	0%	8.10%	17.56%	24.32%	50.00%

Note. The sample size was 74 students; Mis. = missing value or unrated student; U = unknown; M = maybe; P = probably; D = definitely; *f* = frequency; % = percentage.

Table 6. *The Stability of Students' Scores in the Spatial Analytical Activities*

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	High Change >1 to 4
X of 2nd–*X* of 3rd	Number of Students (%)	56 (75.67%)*	18 (24.33%)
X of 3rd–*X* of 4th	Number of Students (%)	60 (81.08%)*	14 (18.92%)
X of 2nd–*X* of 4th	Number of Students (%)	55 (74.33%)*	19 (25.67%)
Average Number of Students		57	17
Average Percentages of Overall Stability		77.02%*	22.97%

Note. The sample size was 74 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage; * = cut-off point used by the authors to determine stability of test was that ≥ 70% of the students' scores must not change more than ±1, thus * indicates the test has been fairly stable.

Logical Mathematical. The activities of this subsection were designed to identify those students whose strengths were in mathematics. In the second grade, 40.54% were rated as "Definitely" superior problem solvers in this category; in the third grade, 37.83% were rated as such; and in the fourth grade, the percentage was 12.16 (See Table 7 for more details). When analyzing the difference in students' scores between grades 2 and 3, the percentage of low change scores (0 and ±1) was 64.86; between grades 3 and 4, 62.16; and between grades 2 and 4, 66.22. The average percentage of students' scores categorized between 0 to ±1 (low change) across all three years was ≈ 64.40. Based on this analysis, the Logical Mathematical activities of the DISCOVER assessment did not meet the two criteria and this subsection was not found to be stable when used with Diné students (See Table 8).

Table 7. Ratings of Students' Performance in the Logical Mathematical Activities across the Three Years

Student	Calculation	The DISCOVER Assessment Rating Levels				
		Mis.=0	U=1	M=2	P=3	D=4
In Second Grade	*f*	3	6	30	5	30
	%	4.05%	8.10%	40.54%	6.75%	40.54%
In Third Grade	*f*	1	15	20	10	28
	%	1.35%	20.27%	27.02%	13.51%	37.83%
In Fourth Grade	*f*	11	10	14	30	9
	%	14.86%	13.51%	18.91%	40.54%	12.16%

Note. The sample size was 74 students; Mis. = missing value or unrated student; U = unknown; M = maybe; P = probably; D = definitely; *f* = frequency; % = percentage.

Table 8. ***The Stability of Students' Scores in the Logical Mathematical Activities***

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	**High Change >1 to 4**
***X* of 2nd–*X* of 3rd**	**Number of Students (%)**	48 (64.86%)	26 (35.14%)
***X* of 3rd–*X* of 4th**	**Number of Students (%)**	46 (62.16%)	28 (37.84%)
***X* of 2nd–*X* of 4th**	**Number of Students (%)**	49 (66.22%)	25 (33.78%)
Average Number of Students		≈ 47.66	≈ 26.34
Average Percentages of Overall Stability		≈ 64.40%	≈ 35.60%

Note. The sample size was 74 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage.

Oral Linguistic. The activities of this subsection of the assessment were designed to identify which students had strong verbal communication abilities. In the second grade, 36.48% of students were rated as "Definitely" superior in this category; in the third grade, 33.78%; and in the fourth grade, 43.24% (See Table 9 for more details). When analyzing the difference in students' scores between grades 2 and 3, the percentage of low change scores (0 and ±1) was 66.22; between grades 3 and 4, 75.68; and between grades 2 and 4, 75.68. The average percentage of the students' scores categorized between 0 to ±1 (low change) across all three years in this category was ≈ 72.51. Therefore, the Oral Linguistic activities of the DISCOVER assessment met the two criteria and this part of the assessment was found to be stable when used with Diné students (See Table 10).

Table 9. Ratings of Students' Performance in the Oral Linguistic Activities across the Three Years

Student	Calculation	The DISCOVER Assessment Rating Levels				
		Mis.=0	**U=1**	**M=2**	**P=3**	**D=4**
In Second Grade	*f*	2	10	20	15	27
	%	2.70%	13.51%	27.02%	20.27%	36.48
In Third Grade	*f*	3	3	15	28	25
	%	4.05%	4.05%	20.27%	37.83%	33.78%
In Fourth Grade	*f*	2	11	9	20	32
	%	2.70%	14.86%	12.16%	27.02%	43.24%

Note. The sample size was 74 students; Mis. = missing value or unrated student; U = unknown; M = maybe; P = probably; D = definitely; *f* = frequency; % = percentage.

Table 10. ***The Stability of Students' Scores in the Oral Linguistic Activities***

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	**High Change >1 to 4**
***X* of 2nd–*X* of 3rd**	**Number of Students (%)**	49 (66.22%)	25 (33.78%)
***X* of 3rd–*X* of 4th**	**Number of Students (%)**	56 (75.68%)*	18 (24.32%)
***X* of 2nd–*X* of 4th**	**Number of Students (%)**	56 (75.68%)*	18 (24.32%)
Average Number of Students		≈ 53.66	≈ 20.34
Average Percentages of Overall Stability		≈ 72.51% *	≈ 27.49%

Note. The sample size was 74 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage; * = the cut-off point used by the authors to determine stability of test was that ≥ 70% of students' scores must not change more than ±1, thus * indicates the test has been fairly stable.

Written Linguistic. The activities in the Written Linguistic subsection of the DISCOVER assessment were designed to assess those students possessing strong writing skills. In the second grade, 36.48% of students were rated to be "Definitely" superior problem solvers in this category; in the third grade, 36.48%; and in the fourth grade, 18.91% (See Table 11 for more details). When analyzing the difference in students' scores between grades 2 and 3, the percentage of students with low change scores was 68.92; between grades 3 and 4, 71.62; and between grades 2 to 4, 67.57. The average percentage of students' scores categorized between 0 to ±1 (low change) across all three years was ≈ 69.37. Therefore, the Written Linguistic activities of the DISCOVER assessment did not meet the two criteria and this subsection was not found to be stable when used with Diné students (See Table 12).

Table 11. Ratings of Students' Performance in the Written Linguistic Activities across the Three Years

Student	Calculation	The DISCOVER Assessment Rating Levels				
		Mis.=0	U=1	M=2	P=3	D=4
In Second Grade	*f*	4	9	19	15	27
	%	5.40%	12.16%	25.67%	20.27%	36.48%
In Third Grade	*f*	7	5	16	19	27
	%	9.45%	6.75%	21.62%	25.67%	36.48%
In Fourth Grade	*f*	6	8	20	26	14
	%	8.10%	10.81%	27.02%	35.13%	18.91%

Note. The sample size was 74 students; Mis. = missing value or unrated student; U = unknown; M = maybe; P = probably; D = definitely; *f* = frequency; % = percentage.

Table 12. ***The Stability of Students' Scores in the Written Linguistic Activities***

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	**High Change >1 to 4**
***X* of 2nd–*X* of 3rd**	**Number of Students (%)**	51 (68.92%)	23 (31.08%)
***X* of 3rd–*X* of 4th**	**Number of Students (%)**	53 (71.62%)*	21 (28.38%)
***X* of 2nd–*X* of 4th**	**Number of Students (%)**	50 (67.57%)	24 (32.43%)
Average Number of Students		≈ 51.33	≈ 22.67
Average Percentages of Overall Stability		≈ 69.37%	≈ 30.63%

Note. The sample size was 74 students; X = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage; * = the cut-off point used by the authors to determine stability of test was that ≥ 70% of students' scores must not change more than ±1, thus * indicates the test has been fairly stable.

The overall DISCOVER assessment. Each student's scores on the five activities of the DISCOVER assessment were averaged to indicate his or her overall score on the assessment. Next, the differences in students' scores across the three testing periods (between grades 2 and 3, 3 and 4, and 2 and 4) were calculated. When analyzing the difference between students' scores in grades 2 and 3, the percentage of students with low change scores was 85.14; between grades 3 and 4, 89.19; and between grades 2 and 4, 86.49. The average percentage of students' scores categorized between 0 to ±1 (low change) across all three years was ≈ 86.93. Therefore, the overall (i.e., averaged) score of the DISCOVER assessment met the two criteria, and was found to be stable when used with Diné students (See Table 13 for more details).

Table 13. *The Stability of Students' Average Scores across All Activities of the DISCOVER Assessment*

Comparison	Calculation	Categories of Changes between each Student's Score across the 3 Years	
		Low Change 0 and ±1	High Change >1 to 4
X **of 2nd–*X* of 3rd**	**Number of Students (%)**	63 (85.14%)*	11 (14.86%)
X **of 3rd–*X* of 4th**	**Number of Students (%)**	66 (89.19%)*	8 (10.81%)
X **of 2nd–*X* of 4th**	**Number of Students (%)**	64 (86.49%)*	10 (13.51%)
Average Number of Students		≈ 64.33	≈ 9.64
Average Percentages of Overall Stability		≈ 86.93% *	≈ 13.02%

Note. The sample size was 74 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage; * = the cut-off point used by the authors to determine stability of test was that ≥ 70% of students' scores must not change more than ±1, thus * indicates the test has been fairly stable.

Which of the Five DISCOVER Activities Were the Most Consistent across the Three Years?

Based on the answer to the first research question, the DISCOVER assessment subsections that were most consistent among the others have been listed from high to low: (a) Spatial Analytical activities, with an average stability percentage of 77.02 (See Tables 5 and 6); (b) Spatial Artistic activities, with an average stability percentage of 74.32 (See Tables 3 and 4); (c) Oral Linguistic activities, with an average stability percentage of ≈ 72.51 (See Tables 9 and 10); (d) Written Linguistic activities, with an average stability percentage of ≈ 69.37 (See Tables 11 and 12); and (e) Logical Mathematical activities, with an average stability percentage of ≈ 64.40 (See Tables 7 and 8).

What Was the Consistency of the RPM Scores across the Three Years?

The Colored Progressive Matrices (CPM) test was used to measure students' general intelligence when they were in the second grade. The students' raw scores in the second grade ranged from 16 to 34 ($M = 27.32$, $SD = 3.96$). When they were in the third and fourth grades, students were measured using the Standard Progressive Matrices (SPM) test. The students' raw scores in the third grade ranged from 16 to 47 ($M = 32.46$, $SD = 7.74$). The students' raw scores in the fourth grade ranged from 21 to 46 ($M = 35.82$, $SD = 5.80$). To achieve an accurate result and complete the analysis, students' raw scores from all three years were converted into a *Z*-score ($M = 0.00$, $SD = 1.00$). Based on this sample data, the average of the highest converted *Z*-score values was 2.5117/4 = 0.6279. Therefore, the authors decided to categorize the RPM scores into four ordinal levels, each with a value of 0.6279, to be equivalent to the rating level of the DISCOVER assessment. Then, the differences among students' scores across the three testing periods (between grades 2 and 3, 3 and 4, and 2 and 4) were calculated in the same manner the authors used to calculate these differences in the data from the DISCOVER assessment.

The percentage of students with low change scores between grades 2 and 3 was 28.85; between grades 3 and 4, 17.31; and between grades 2 and 4, 48.08. The average percentage of students' scores categorized between 0 to ±0.6279 (low change) across all three years was ≈ 31.40. Therefore, the RPM test did not meet the two criteria and it has not been found to be stable when used with Diné students (See Table 14 for more details). Further analysis was conducted to examine our conclusion. The correlation between scores in scores in grades 2 and 3 was $r\ (50) = -.159$, $p = .259$; between grades 3 and 4 was *r* (50) = –.126, *p* = .375; and between grades 2 and 4 was *r* (50) = –.652, *p* = .000. Based on these correlations, the authors confirmed that the RPM was not found to be a stable instrument when used with Diné students.

Table 14. ***The Stability of Students' Scores on the RPM***

Comparison	**Calculation**	**Categories of Changes between each Student's Score across the 3 Years**	
		Low Change 0 and ±0.6279	**High Change >0.6280**
***X* of 2nd–*X* of 3rd**	**Number of Students (%)**	15 (28.85%)	37 (71.15%)
***X* of 3rd–*X* of 4th**	**Number of Students (%)**	9 (17.31%)	43 (82.69%)
***X* of 2nd–*X* of 4th**	**Number of Students (%)**	25 (48.08%)	27 (51.92%)
Average Number of Students		≈ 16.33	≈ 35.67
Average Percentages of Overall Stability		≈ 31.40%	≈ 68.60%

Note. The sample size was 52 students; *X* = each individual student's score; 2nd = second grade; 3rd = third grade; 4th = fourth grade; % = percentage.

Which of the Two Instruments Yielded the Most Consistent Scores across the Three Years?

First. A linear single regression analysis was used to test how well students' ratings on each of the five DISCOVER activities taken independently in the fourth grade were predicted by their scores in the second and third grades. On the basis of the sample data, (a) Spatial Artistic scores in the fourth grade were predicted by students' Spatial Artistic performance both in the second grade, accounting for 5.6% of the variance, $F(1, 71) = 4.290$, $p = 0.041$, and in the third grade, accounting for 9.6% of the variance, $F(1, 71) = 7.565$, $p = 0.007$; (b) Spatial Analytical scores in the fourth grade were predicted by students' Spatial Analytical performance both in the second grade, accounting for 6.5% of the variance, $F(1, 70) = 4.933$, $p = 0.029$, and in the third grade, accounting for 11% of the variance, $F(1, 70) = 8.727$, $p = 0.004$; (c) Logical Mathematical scores in the fourth grade were not predicted by students' Logical Mathematical performance both in the second and third grades; (d) Oral Linguistic scores in the fourth grade were predicted by the students' Oral Linguistic performance both in the second grade, accounting for 14.7% of the variance, $F(1, 68) = 11.775$, $p = 0.001$, and in the third grade, accounting for 11% of the variance, $F(1, 67) = 7.162$, $p = 0.009$; (e) Written Linguistic scores in the fourth grade were not predicted by the second grade scores at significant level, but the third grade scores predicted students' Written Linguistic performance in the fourth grade, and accounted for 13.5% of the variance, $F(1, 60) = 9.392$, $p = 0.003$ (see Table 15 for more details).

Second. A linear single regression analysis also was used to test how well students' overall scores on both the DISCOVER assessment and the RPM in the fourth grade were predicted by their scores in the second and third grades. On the basis of the sample data, the overall DISCOVER scores in the fourth grade were predicted at a significant level by students' overall DISCOVER scores both in the second grade, accounting for 14.6% of the variance, $F(1, 72) = 12.348$, $p = 0.000$, and in the third grade, accounting for 18.3% of the variance, $F(1, 72) = 16.169$, $p = 0.000$. For the RPM, the converted *Z*-scores were analyzed using linear single regression analyses. Students' scores on the RPM in the fourth grade were predicted at a significant level by their scores in the second grade, accounting for 42.5% of the variance, $F(1, 50) = 36.981$, $p = 0.000$; however, RPM scores in the fourth grade were not predicted at a significant level by students' scores in the third grade (see Table 16 for more details).

Table 15. ***Prediction of Students' Performance in each of the DISCOVER Assessment Activities***

	The DISCOVER Assessment	Prediction	R^2	*B*	*SE*	*β*	*p*
1.	**Spatial Artistic**	**2nd to 4th**	0.056 *	0.215	0.847	0.238	0.041
		3rd to 4th	0.096 **	0.260	0.829	0.310	0.007
2.	**Spatial Analytical**	**2nd to 4th**	0.065 *	0.241	0.970	0.256	0.029
		3rd to 4th	0.110**	0.350	0.942	0.332	0.004
3.	**Logical Mathematical**	**2nd to 4th**	0.015	0.107	0.934	0.123	0.344
		3rd to 4th	0.047	0.165	0.916	0.219	0.087
4.	**Oral Linguistic**	**2nd to 4th**	0.147**	0.387	1.026	0.384	0.001
		3rd to 4th	0.096**	0.402	1.057	0.310	0.009
5.	**Written Linguistic**	**2nd to 4th**	0.052	0.193	0.906	0.229	0.068
		3rd to 4th	0.135**	0.356	0.878	0.367	0.003

Note. The sample size was 74 students; R^2 = R Squared; *B* = Unstandardized Coefficient; *SE* = Std. Error; *β* = Standardized Coefficient; *p* = Significance level; * = Significant at $p > .05$; ** = Significant at $p > .01$; predictions with respect to grade level.

Table 16. ***Prediction of the overall DISCOVER scores and the converted Z-scores of the RPM***

	Test	Prediction	R^2	*B*	*SE*	*β*	*p*
1.	**DISCOVER**	**2nd to 4th**	0.146**	0.308	0.514	0.382	0.000
	(*N* = 74)	**3rd to 4th**	0.183**	0.386	0.503	0.428	0.000
2.	**RPM**	**2nd to 4th**	0.425**	0.652	0.765	0.652	0.000
	(*N* = 52)	**3rd to 4th**	0.015	-0.125	1.001	-0.125	0.374

Note. *N* = Sample size; R^2 = R Squared; *B* = Unstandardized Coefficient; *SE* = Std. Error; *β* = Standardized Coefficient; *p* = Significance level; * = Significant at $p > .05$; ** = Significant at $p > .01$; predictions with respect to grade level.

Which of the Overall Scores Was Most Consistent in Predicting Giftedness in Students across the Three Years?

The DISCOVER assessment. Based on the procedure used in the DISCOVER assessment to determine giftedness, a student would be considered gifted if he or she achieved two "Definitely" rankings. In the second grade, 44 students were identified as gifted. A linear single regression analysis was used to test how well the DISCOVER assessment was able to identify those gifted children when they were in the third and fourth grades. On the basis of the sample data, the overall DISCOVER scores in the third grade were predicted significantly by the scores of those who had been identified as gifted when they were in the second grade, accounting for 7.7% of the variance, $F(1, 72) = 6.046$, $p = 0.016$; and in the fourth grade, accounting for 7.1% of the variance, $F(1, 72) = 5.576$, $p = 0.020$.

The RPM. Based on the RPM manuals (CPM and SPM), students' percentiles rankings were obtained. In the State of Arizona, ≥ 97th percentile has been used as the criterion for identifying students as gifted, so this criterion was used to test the stability of the RPM to identify gifted students. In the second grade, 4 students were considered gifted. A linear single regression analysis was used to test how well the RPM was able to identify those 4 students in the third and fourth grades. On the basis of the sample data, the converted *Z*-scores of the RPM in the third grade were not predicted at a significant level by the scores of those students who were identified as gifted when they were in the second grade, but the scores from the second grade did predict giftedness in the fourth grade at a significant level, accounting for 9.6% of the variance, $F(1, 50) = 5.318$, $p = 0.025$. (See Table 17 for more details).

Table 17. ***The Ability of the DISCOVER Assessment and the RPM to Identify the Same Gifted Students across the Three Years***

	Test	Gifted Student in the 2nd Grade		Prediction	R²	B	SE	β	p
		f	%						
1.	**DISCOVER**	**44 (*N* = 74)**	**59.45%**	**3rd**	0.077*	-0.224	0.478	-0.278	0.016
				4th	0.071*	-0.239	0.479	-0.268	0.020
2.	**RPM**	**4 (*N* = 52)**	**7.69%**	**3rd**	0.038	0.729	0.990	0.196	0.163
				4th	0.096*	-6.687	5.571	-0.310	0.025

Note. R^2 = R Squared; *B* = Unstandardized Coefficient; *SE* = Std. Error; β = Standardized Coefficient; *p* = Significance level; * = Significant at $p > .05$; ** = Significant at $p > .01$; % = Percentage; *f* = Frequency; predictions with respect to grade level; *N* = sample size.

Discussion

At the beginning of this article, the authors raise the fundamental question, "*Who Is Gifted*?" and highlight the major factors that influence the criteria for defining and then identifying gifted students. We agree that some of those factors are external; they are not always under the control of educators at the school level, and are often determined by a country's needs, a community's needs, a population's cultural values, and the amount of available money or funding allocated to establish the program. However, we strongly believe that educators still have the power to control the most important internal factors, such as the theoretical framework and the instruments or criteria that are adopted for use in the program. When educators control internal factors in programs for gifted students, the fundamental question, "*Who Is Gifted*?" will be restated in a focused and researchable way. For instance, will the theoretical framework continue to be valid with different populations? Will different instruments developed based on different theoretical frameworks be stable enough to identify the same group of students across grade levels and administrations?

In the current longitudinal investigation, the authors were able to deal with the last question by examining the stability of two different instruments to identify the same group of students across three years. The RPM was developed based on Spearman's (1923) *g* theory (domain general) as a non-verbal, solely pencil-paper, multiple-choice, and group-administered test. The DISCOVER assessment, on the other hand, was developed based on Gardner's (domain specific) MI theory (1983; 1999) as an "intelligence-fair" (Gardner, 1992) assessment, administered in the children's home or dominate language, performance-based, and using hands-on activities with a small group of peers. Therefore, with respect to the existing limitations, the authors conducted this longitudinal investigation not only to decide which of the two different theoretical frameworks was more valid, or which of the two different instruments was more stable when used to identify Native American students (i.e., Diné), but also to address some potential reasons why Native American groups are underrepresented in programs for gifted students in the State of Arizona and nationally (Erdimez & Maker, 2012).

The Stability of Scores on the DISCOVER Assessment

When examining the stability of each of the five DISCOVER activities (Spatial Artistic, Spatial Analytical, Logical Mathematical, Oral Linguistic, and Written Linguistic), we find that Spatial Analytical, Spatial Artistic, and Oral Linguistic activities are stable when used to identify Diné gifted students across the three years. On the other hand, the Logical Mathematical and Written Linguistic activities are found not to be stable when used to identify Diné gifted students across the three years. The Logical Mathematical and Written Linguistic activities were designed to measure what Gardner (1992) calls "second-order knowledge," or what is learned in school, while the other activities were designed to measure "first-order knowledge," or what is learned from experience with concrete objects and personal interactions. All students come to school possessing first-order knowledge. As the students moved from one grade to the next, they achieved formal educational objectives, increased their understanding about content, and shaped their skills in both math (Beecher & Sweeny, 2008; Draper, 2002) and writing (Raimes, 1987). Their performance on these two tasks changed as they progressed in school. In addition, because the Spatial Artistic, Spatial Analytical, and Oral Linguistic components of the DISCOVER assessment are hands-on activities that use concrete materials, they might have resulted in a more stable finding than the Logical Mathematical and Written Linguistic activities did, as the latter two are solely pencil-paper activities. Despite the fact that two of the five DISCOVER assessment activities were found to be unstable, the average score of all five activities was found to be stable when used to identify Diné gifted students across the three years. The authors also use regression analyses for each of the five DISCOVER activities and of the overall score to provide further evidence of its stability. These stability scores are similar to the findings of researchers who examined the reliability of the assessment (Griffiths, 1997; Kassymov, 2000).

The Stability of Scores on the RPM

When evaluating the stability of the RPM, the authors find that the RPM is not stable when used to identify Diné gifted students across the three years. The authors also conducted further analyses by using Pearson's correlation, and find that the RPM is not a stable instrument when examined in this way. The regression analyses also indicate that the RPM is not a stable instrument. To explain this finding, Raven (2000) described the processes of normalizing the RPM in both the United States and the United Kingdom. One of the major criticisms was that in the United States Native American students were not included in the norm group (Raven, 2000). Because the RPM is considered a language-free and culturally fair instrument (Lohman, 2005), many educators assume that it will result in stable scores across years. However, in the current study, students' scores in the RPM varied from year to year. Perhaps a hidden or subtle cultural component is present. For instance, Jensen (1973) examined the reliability of the RPM on a sample of three ethnic groups, and found that it was a culturally biased instrument against African-American students. Additionally, in South Africa, Owen (1992) investigated the reliability of the RPM with different ethnicities, and found that the largest proportion of the African and Caucasian students had a pattern similar to their own cultural/ethnic group.

Predicting Giftedness

The DISCOVER assessment was useful to identify 59.45% of the 74 Diné gifted children who had been place in a special program for children with high potential when they were in the second grade; on the other hand, the RPM was able to identify only 7.69% of the 52 Diné children from the same group when they were in the second grade. This finding was consistent with Sarouphim's two studies, in which she found that the DISCOVER assessment was a culturally fair assessment that increased the representation of minority students in programs for gifted learners. Sarouphim (2001) investigated concurrent validity of the DISCOVER assessment, gender differences, and identification of minority students with a sample of 257 students that was mostly comprised of students of Diné and Mexican-American ethnicities. She concluded that DISCOVER identified 22.9% of all participants, a much higher percentage than traditional IQ tests. Sarouphim (2002) studied a sample of 303 high school students who were mostly Mexican-American and Diné, and concluded that 29.3% of the participants were identified as gifted. Therefore, throughout the literature on the DISCOVER assessment, researchers frequently provide evidence that the DISCOVER assessment is able to identify more CLD gifted students than traditional instruments do. However, in the current study, the authors find that the DISCOVER assessment is able to identify the same gifted Diné children across three years at a higher rate than the RPM does. This is likely because the DISCOVER assessment is a performance-based assessment, was developed based on multiple intelligence theories (MI), has five domains, assesses students in their familiar environments, contains open-ended tasks, and requires higher levels of thinking, Researchers have found that performance-based assessments are more powerful in identifying and increasing the representation of CLD students in programs for the gifted than are traditional instruments (Clasen, Middleton, & Connell, 1994; Sarouphim, 2001; 2002; Van Tassel-Baska et al., 2002).

Limitations

We conducted this study with 74 Diné children who participated in the DISCOVER assessment and with 52 Diné children who participated in the RPM; therefore, the results should not be generalized to other ethnic groups. Another limitation of the current study was the fact that the authors were interested in investigating the two instruments' stability by focusing on individual students' scores across the three years rather than calculating groups means. Therefore, the problem addressed in the current study was not one of simple test re-test reliability, which can be analyzed using the correlation coefficients. Because of the 1-year gap between each testing period, correlation coefficient was not useful to answer the research questions. Additionally, we could not use any growth modeling methods because the sample size was small, the data for the DISCOVER assessment were ordinal, and growth modeling methods use groups means. Additionally, we could use neither the growth modeling methods nor repeated measures ANOVA because their assumptions are not useful to answer the research questions. For instance, these analyses measure the increase in students' scores (e.g., achievement); however, we wanted to measure the stability of students' scores across three years. Therefore, we calculated the change differences and then the obtain percentages of students based on the two criteria we formulated. Even though the data analysis was a major limitation, using multiple analyses to answer all of our research questions (viz., the change differences, correlation coefficient, and regression) provided overall consistent results regarding the two instruments' stability.

Implications

Theoretical Implications

Alhusaini and Maker (in press) studied the predictive validity of the DISCOVER assessment to identify general creativity in Diné children as measured by the TCT-DP. Students were tested using the DISCOVER assessment when they were in the third grade, and then tested by the TCT-DP three years later when they were in sixth grade. The researchers found that only Spatial Analytical and Oral Linguistic activities predicted students' general creativity. In recent studies, Tan and Maker (2012) examined the predictive validity of the RPM to identify students' overall achievement, and achievement in various content areas. Students were tested using the RPM when they were in the second grade, and then tested by the ITBS and CTBS when they were in fourth grade. The researchers found that the RPM scores accounted for 19.5% of the variance in overall achievement scores, 13.3 % of the variance in reading, 4.6% of the variance in language, and 26.6 % of the variance in math. In contrast, Erdimez and Maker (2012) also examined the predictive validity of the DISCOVER assessment. Participants were tested using the DISCOVER assessment when they were in the third grade, and then tested by the ITBS and CTBS when they were in fourth grade. The researchers found that as a model, students' scores on all activities of the DISCOVER assessment accounted for 43.9% of the variance in total achievement scores of the students. When synthesizing the results across all three studies, all of which were conducted with Diné children, we interpret that an instrument developed from a domain-specific perspective is a better predictor of students performance then an assessment that was developed from a domain-general perspective.

In the current study, we approached the problem of the underrepresentation of Native American students in programs for gifted students by examining the stability of two different instruments and then eventually evaluating two different theories. All of our analyses resulted in further evidence for the stability of the DISCOVER assessment, which can be interpreted as establishing theoretical validity for the MI theory. Therefore, the results of this study might be seen as evidence of giftedness as domain-specific as measured by the DISCOVER assessment rather than of giftedness as domain-general as measured by the RPM. However, we believe that giftedness is a combination of both strength in a specific domain (viz., domain-specific) and overall skill (viz., domain-general) that can be integral across all domains. Amabile (1996) developed a componential theory of creativity, in which she argued that all general creative skills, specific skills, and task motivation are important for creative performance. Even though the final result of the DISCOVER assessment provides information about students' strengths across five domains, Maker and her colleagues have considered in the assessment's behavioral checklist both general problem solving behaviors and domain-specific problem solving behaviors.

Implications for Practice

The results of the current study should be read as an important indicator for educators to avoid using a single test to identify gifted students, especially those who are from CLD groups. In this context, we believe that the DISCOVER assessment with its five domains will be an appropriate instrument to be used with Diné students, because it has been found to identify giftedness across a wide range of strengths and to be stable when used to identify Diné children. Using performance-based assessments developed from a domain-specific perspective is a potential solution to the problem of underrepresentation of Native American groups in programs for gifted students in Arizona and nationally, as these kinds of instruments were originally designed to overcome the limitations of traditional instruments.

Implications for Research

The authors were able to conduct a short-scale longitudinal investigation based on 3 years' worth of data for 74 gifted students who participated in the DISCOVER assessment, and for 52 students who participated on the RPM. Future researchers might consider conducting a long and large-scale longitudinal investigation into the same research problem. However, we believe that what is most needed is a mixed-methods study to investigate how did Diné children understand the RPM test problems by using the following steps: (a) separating the RPM into sections with similar problems; (b) analyzing students' scores on those sections; (c) conducting item analysis to identify patterns of which items were missed consistently; and (d) interviewing the children about why they answered the way they did, especially on missed problems. This type of study would provide valuable information about the logical thought processes of Diné children.

Conclusion

Within the limits of available data, the authors have examined the stability of both the DISCOVER assessment and RPM when used to identify Diné gifted children. Students were tested using both instruments when they were in the second, third, and fourth grades. In the current study, consistent evidence was found that three (viz., Spatial Analytical, Spatial Artistic, and Oral Linguistic) of the five DISCOVER activities were stable, as were the overall scores of the DISCOVER assessment. Although the scores from grade two were significant predictors of scores in grade four, they were not consistent with scores in grade three, nor did third grade scores predict fourth scores. The RPM was not considered to be a stable instrument when used to identify Diné gifted children. Finally, the DISCOVER assessment was able to identify a greater number of Diné gifted children when they were in the second grade and predict their giftedness across the three years more than the RPM was found to identify and predict. Gardner's statement (n. d.) is important to consider as readers think about these results: "I believe that the brain has evolved over millions of years to be responsive to different kinds of content in the world. Language content, musical content, spatial content, numerical content, etc."

References

Alhusaini, A. (2006). *Factors that have influenced the criteria for defining giftedness in Saudi Arabia*. Unpublished manuscript. King Abdul-Aziz University.

Alhusaini, A. & Maker, J. (in press). The predictive validity of the DISCOVER assessment to identify general creativity in Diné children. Manuscript submitted for publication. *Journal of Creative Behavior*.

Amabile, M. (1996). *Creativity in context*. Boulder, CO: Westview.

Baska, L. (1986). The use of the Raven Advanced Progressive Matrices for the selection of magnet junior high school students. *Roeper Review, 8*(3), 181-184. doi:10.1080/02783198609552969

Beckley, D. (1998). *Gifted and learning disabled: Twice-exceptional students*. Storrs, CT: Neag Center for Gifted Education and Talent Development.

Beecher, M., & Sweeny, M. (2008). Closing the gap with curriculum enrichment and differentiation: One school's story. *Journal of Advanced Academics*, *19*(3), 502-530.

Bittker, M. (1991). Patterns of academic achievement in students who qualified for a gifted program on the basis of nonverbal tests. *Roeper review*, *14*(2), 65-67. doi:10.1080/02783199109553389

Borland, H. (2008). Identification. In J. Plucker & C. Callahan (Eds.), *Critical issues and practices in gifted education* (pp. 261-280). Waco, Texas: Prufrock Press Inc.

Brody, E., & Mills, J. (1997). Gifted children with learning disabilities: A review of the issues. *Journal of Learning Disabilities, 30*(3), 282-296. doi:10.1177/002221949703000304

Byrt, E., & Gill, E. (1973). *Standardization of raven's standard progressive matrices and mill hill vocabulary for the irish population: Ages 6–12*. Unpublished master's thesis. University College Cork.

Castellano, A. (2003). *Special populations in gifted education: Working with diverse gifted learners.* Boston, MA: Pearson Education, Inc.

Clark, B. (1997). *Growing up gifted.* Upper Saddle River, NJ: Merrill.

Clasen, R., Middleton, A., & Connell, T. (1994). Assessing artistic and problem-solving performance in minority and nonminority students using a nontraditional multidimensional approach. *Gifted Child Quarterly*, *38*(1), 27-32. doi:10.1177/001698629403800104

Clemons, L. (2008). *Underachieving gifted students: A social cognitive model*. Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Clinkenbeard, R. (2007). Economic arguments for gifted education. *Gifted Children*, *2*(1), 5-9.

Colangelo, N., & Davis, A. (1991). Introduction and historical overview. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 3-13). Needham Heights, MA: Allyn & Bacon.

Cole, S., & Zieky, J. (2001). The new faces of fairness. *Journal of Educational Measurement, 38*(4), 369-382. doi:10.1111/j.1745-3984.2001.tb01132.x

Coleman, L. J., & Cross, L. (2001). *Being gifted in school: An introduction to development, guidance, and teaching*. Waco, TX: Prufrock Press.

Coleman, M. R. (2001). Surviving or thriving? 21 gifted boys with learning disabilities share their school stories. *Gifted Child Today*, *24*(3), 56-63.

Coleman, M. R. (2003). The identification of students who are gifted. *The ERIC Clearinghouse on Disabilities and Gifted Education*. Retrieved from: http://ericec.org/digests/e644.html

Draper, J. (2002). School mathematics reform, constructivism, and literacy: A case for literacy instruction in the reform-oriented math classroom. *Journal of Adolescent and Adult Literacy*, *45*(6), 520-529.

Erdimez, O., & Maker, J. (2012). *The predictive validity of the DISCOVER performance-based assessment to identify the academic achievement of Diné students.* Unpublished manuscript. University of Arizona.

Feldhusen, F., Hoover, M., & Sayler, F. (1990). *Identification of gifted students at the secondary level.* Monroe, NY: Trillium.

Ford, Y. (1998). The under-representation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education, 32*(1), 4-14. doi:10.1177/002246699803200102

Ford, Y., Grantham, C., & Whiting, W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children, 74*, 289-306.

Fraenkel, J. & Wallen, N. (2010) *How to design and evaluate research in education*. New York, NY: McGraw-Hill.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: BasicBooks.

Gardner, H. (1992). *Multiple intelligences: The theory in practice*. New York: Basic Books.

Gardner, H. (1999). *Intelligence reframed*. New York: Basic Books.

Gardner, H. (n. d.). BrainyQuote.com. Retrieved April 24, 2013, from BrainyQuote.com Web site: http://www.brainyquote.com/quotes/quotes/h/howardgard194106.html

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: Wiley.

Gilliam, E., Carpenter, O., & Christensen, R. (1996). *Gifted and talented evaluation scales: A norm-referenced procedure for identifying gifted and talented students*. Austin, TX: PRO-ED.

Glasser, W. (1993). *The quality school teacher.* New York: Harper Collins.

Gravetter, J., & Wallnau, B. (2009). *Statistics for the behavioral sciences*. Belmont, CA: Cengage Learning.

Gregory, J. (2004). *Psychological testing: History, principles, and applications*. Boston: Allyn & Bacon.

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

Jensen, A. (1973). *How biased are culture-loaded tests?* [Document Resume]. Retrieved from: http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED080644

Johnson, M. (2006). Internet use and cognitive development: A theoretical framework. *E-Learning, (3)*4, 565-573.

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

Kratzmeier, H., & Horn, R. (1979). *Manual: Raven-matrizen-test, standard progressive matrices*. Weinheim, Germany: Beltz Test.

Kubiszyn, T., & Borich, G. (2007). *Educational testing and measurement*. Hoboken, NJ: John Wiley and Sons.

Lohman, F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted child quarterly*, *49*(2), 111-138. doi:10.1177/001698620504900203

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

Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. *Gifted Child Quarterly, 40*, 41-50. doi:10.1177/001698629604000106.

Maker, C. J. (1993). *Report on project STEP-UP in Arizona*. Unpublished report. University of Arizona.

Maker, C. J., Nielson, B., & Rogers, A. (1994). Giftedness, diversity, and problem solving: Multiple intelligences and diversity in educational settings. *Teaching Exceptional Children, 27*(1), 4-19.

McBee, M. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education, 17*(2), 103-111.

McBee, M. (2010). Examining the probability of identification for gifted programs for students in Georgia elementary schools: A multilevel path analysis study. *Gifted Child Quarterly, 54*(4), 283-297. doi:10.1177/0016986210377927

Menard, S. (1991). *Longitudinal research*. Newbury Park: Sage Publications.

Miller, L., & Brewer, D. (2003). *The a-z of social research*. London: Sage.

Miller, M. (n. d.) *Graduate research methods: Reliability and validity*. Western International University, Retrieved from: http://michaeljmillerphd.com/res500_lecturenotes/reliability_and_validity.pdf

Neisser, U. (1997). Rising scores on intelligence tests. *American Scientist, 85*(5), 440–447.

Nielson, B. (1994). Traditional identification: Elitist, racist, sexist? New evidence. CAG Communicator: *The Journal of the California Association for the Gifted, 24*(3), 18-19, 26-31.

Nowak, M. (2001, November). Twice-exceptional (gifted + learning disabled) students, the equality ideal, and the reward structure of the educational system. Paper presented at the annual meeting of the American Anthropological Association, Washington, DC.

Owen, K. (1992). The suitability of raven's standard progressive matrices for various groups in South Africa. *Personality and Individual Differences, 13*, 149-159. doi:10.1016/0191-8869(92)90037-P

Pinchok, N., & Ploeg, A. (2009). *A brief on performance-based assessment technical considerations from an international perspective*. The Great Lakes East Comprehensive Center. Retrieved from: http://www.learningpt.org/greatlakeseast/newsletters/OH_DOE_Perf_Assess_pilot_project.pdf

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning, (37)*3, 439-468. doi:10.1111/j.1467-1770.1987.tb00579.x

Ramos, E. (2010). Let us in: Latino underrepresentation in gifted and talented programs. *Journal of Cultural Diversity 17*(4), 151-3.

Raven, J. (1941). Standardisation of progressive matrices. *British Journal of Medical Psychology, 19*(1), 137-150. doi:10.1111/j.2044-8341.1941.tb00316.x

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1-48.

Raven, J., & Walshaw, B. (1944). Vocabulary tests. *British Journal of Medical Psychology, 20*, 185-194.

Reis, S., & McCoach, B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly, 44*(3), 152-170.

Renzulli, J. (1990). A practical system for identifying gifted and talented students. *Early Childhood Development, 63*, 9-18.

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.

Saldaña, D. (2001). *Cultural competency: A practical guide for mental health service providers.* Austin, TX: Hogg Foundation for Mental Health.

Sarouphim, K. (1999a). DISCOVER: A promising alternative assessment for the identification of gifted minorities. *Gifted Child Quarterly, 43*(4), 244-251. doi:10.1177/001698629904300403

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

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

Sarouphim, K. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly, 45*, 130-138. doi:10.1177/001698620104500206

Sarouphim, K. (2002). DISCOVER in high school: Identifying gifted Hispanic and Native American students. *Journal of Advanced Academics, 14*(1), 30-38. doi:10.4219/jsge-2002-385

Sattler, M. (1988). *Assessment of children*. (3rd ed.). San Diego: Jerome M. Sattler.

Shah, N. (2011). *Investing in gifted education could cost little, report finds.* Retrieved from: http://blogs.edweek.org/edweek/speced/2011/11/_gifted_education_has_been.html

Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition.* London: Macmillan.

Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. (1997). *Thinking styles*. New York: Cambridge University Press.

Tan, S., & Maker, C. J. (2012). *The predictive validity of the raven progressive matrices for identifying academic achievement of the Diné children*. Unpublished manuscript. University of Arizona.

Snyder, T. D., & Dillow, S. A. (2012). Digest of Education Statistics, 2011. NCES 2012-001. *National Center for Education Statistics.*

Van Tassel-Baska, J., Johnson, D., & Avery, L. (2002). Using performance tasks in the identification of economically disadvantaged and minority gifted learners: Findings from project STAR. *Gifted Child Quarterly, 46(2)*, 110-123. doi:10.1177/001698620204600204

Wellisch, M. & Brown, J. (2011). Where are the underachievers in the DMTG's academic talent development? *Talent Development & Excellence, 3*(1) 115-117.

Yamin, T. (2006). *Practical procedures to select gifted students in public schools.* Unpublished Article, Arabian Gulf University.

Ziegler, A., & Stoeger, H. (2012). Shortcomings of the IQ-based construct of underachievement. *Roeper Review, 34*(2), 123-132. doi:10.1080/02783193.2012.660726