DISCOVER in High School:
Identifying Gifted Hispanic and Native American Students

Ketty M. Sarouphim
Lebanese American University

Based on Gardner’s theory of multiple intelligences, the DISCOVER assessment was designed to identify gifted minority students for placement into programs for the gifted. In previous studies, the reliability and validity of the assessment in elementary grades were examined and yielded mostly positive results. In this study, similar analyses were carried out to investigate some validity aspects of DISCOVER with secondary students. The sample consisted of 303 predominantly Hispanic and Native American ninth graders. The results provided evidence for an alignment of the assessment with the theory of multiple intelligences. Also, no overall gender or ethnic differences were found in the numbers of students identified. In addition, the results suggested that the use of the DISCOVER assessment might help in reducing the problem of minority students’ under-representation in programs for the gifted, as 29.3% of the high school students who participated in this study were identified as gifted.

The field of education has witnessed lately a rise in the use of authentic assessment, also called alternative and performance-based assessment. Many educators have advocated replacing standardized tests with these new instruments which are considered to be more appealing, closer to lifelike situations, and problem-based rather than knowledge-based (Baldwin, 1985; Ford & Harmon, 2001; Maker, 1996, Sarouphim, 1999a). Also, studies have shown that minority students fare better on these measures than on traditional standardized tests (Borland & Wright, 1994; Clasen, Middleton, & Connell, 1994; Reid, Uddal, Romanoff, & Algozzine, 1999; Sarouphim, 2001), a factor that gained authentic assessment the reputation of being culturally fair and spread its use among culturally diverse groups.

In assessing giftedness, a long-standing problem has been the under-representation of minority students in programs for the gifted (Coleman & Gallagher, 1995; Gardner, 1995; Maker, 1996). Demographics on gifted education have shown that the students underrepresented are Black, Hispanic, and Native Americans, but not Asian Americans (Ford & Harmon, 2001; Hunsaker, 1994). In 1997, the National Association for Gifted Children requested educators to consider multiple criteria for identification and to practice equity in their placement decisions. However, despite sustained efforts, the under-representation of culturally diverse students still persists in today’s schools. Reid et al. (1999) stated that only about 8% of minority students qualify for placement in programs for the gifted across the country. Many educators concur that the main reason for this injustice is the use of inadequate identification procedures (Clasen et al., 1994; Cummins, 1991; Ford & Harmon). The criterion adopted for identification is, in most states, high scores (97th percentile) on standardized tests. Around 90% of school districts use the scores of intelligence or achievement tests for placement purposes, which keeps the demographics of gifted programs primarily White and middle class because, more often than not, minority and economically disadvantaged students do not score high on those standardized tests (Baker, 1996; Ford & Harmon).

Hunsaker (1994) found that testing and identification policies were the major factors hindering the access of diverse and economically disadvantaged students to gifted programs. In a study conducted on 13 districts serving mostly underprivileged and geographically isolated students, Hunsaker surveyed 39 school sites in an attempt to determine which groups
of students were underrepresented, what were the means used to increase their representation, and what were the factors that facilitated or hindered access to programs for the gifted. The results revealed that ethnic/linguistic minorities were the groups most underrepresented. To increase representation, the school sites successfully used multiple assessment criteria, such as portfolio assessment, behavioral observations, check lists and rating scales, and, most popular of all, alternative testing. Factors that facilitated the successful use of alternative procedures were supportive and understanding teachers and administrators, as well as the effectiveness of the alternate assessment procedures adopted. Hunsaker recommended that school districts develop or choose instruments and decision models for expanded assessment that examine more thoroughly the giftedness construct and its corresponding traits, aptitudes, and behaviors.

Authentic assessment carries the promise of rectifying the historic error of under-representation by providing culturally diverse students with better opportunities for accessing programs for the gifted through intelligence-fair and culturally bias-free types of measurement (Ford & Harmon, 2001; Maker, 1994; Sarouphim, 1999a). However, the use of alternative assessment is not a panacea, and the effectiveness of the instruments used should be demonstrated before educators rely on their outcome (Frechtlng, 1991; Travis, 1996; Worthen, 1993). Research on performance-based assessments using the framework of Gardner's theory of multiple intelligences is scarce (Plucker, Callahan, & Tomchin, 1996; Reid et al., 1999), and studies on the psychometric properties of performance-based assessments constitute less than 5% of published studies on the subject (Baker, O'Neil, & Linn, 1994). Nontraditional measures have been criticized for their lack of technical qualities (Worthen, 1993), excessive time administration, and subjective scoring (Travis, 1996). Plucker et al. suggested that educators who use alternative assessment, programs based on multiple intelligences, or a combination of the two must "subject the programs to rigorous evaluation" (p. 87).

Although few, some researchers have examined the psychometric properties of performance-based assessments and the effectiveness of their use with minorities. For example, Borland and Wright (1994) described an extensive method for the identification of economically disadvantaged students that included both qualitative and quantitative measures. Standardized tests, as well as classroom observations, portfolio assessment, teacher nominations, and child interviews were used for identification purposes. Validation data for two cohorts (K–2) yielded positive results. The researchers concluded that giftedness can be found in every school and that educators have no excuse for failing to identify gifted students from all backgrounds.

In a study that examined if alternative assessment might contribute to diminishing the problem of under-representation of minority students in programs for the gifted, Reid et al. (1999) compared a traditional measure, the Matrix Analogies Test–Short Form (MAT-SF) with an alternate assessment, the Problem Solving Assessment (PSA), in identifying culturally and linguistically gifted students. The PSA is a measure grounded in Gardner’s theory of multiple intelligences and Maker et al.’s (1994) assessment research. The sample consisted of 600 students of White, Black, Hispanic, American Indian, and Asian origin. The results showed that, through the use of the MAT-SF, only 22% of the students met the criteria for identification, whereas about half of the students were identified as gifted through the use of the PSA. Also, significant differences were found in the distribution of identified students. Using the MAT-SF, 11% of identified students were minority, whereas using the PSA, 39% of minority students were recommended for placement in gifted programs. Although data on the predictive validity of the PSA were not available, the researchers concluded that anecdotal evidence was favorable, indicating that most students who were placed in programs for the gifted using the PSA were successful.

In another study, Clasen et al. (1994) used a nontraditional measure for identification purposes. They tested 433 minority and nonminority students using problem solving, a drawing task, peer identification, and teacher nominations. The results showed that 24% of the students who were identified as gifted, a percentage proportionate to their representation in the schools. Also, the numbers of males and females identified were in proportion to their representation in the population.

The growth of authentic assessment has corresponded with the rise of nontraditional theories of intelligence, such as Gardner’s (1983) theory of multiple intelligences and Sternberg’s triarchic theory of intelligence (1991), as well as new conceptions of giftedness, such as Maker’s (1993), in which giftedness is defined as creative problem-solving in the "most effective, efficient, and economical ways" (p. 71). A performance-based assessment that incorporates the new conceptions of intelligence and giftedness is the DISCOVER assessment. DISCOVER is an acronym and stands for Discovering Individual Strengths and Capabilities through Observation while allowing for Varied Ethnic Responses. Grounded in Gardner’s (1983) theory of multiple intelligences and Maker’s (1993) definition of giftedness, DISCOVER is a relatively new performance-based assessment developed by June Maker and her colleagues at the University of Arizona (Maker, Nielsen, & Rogers, 1994). The assessment was specifically designed for the identification of gifted minority students and has been used with a variety of culturally diverse groups. The assessment consists of a set of activities for four aggregated grade levels: K–2, 3–5, 6–8, and 9–12. In this study, the focus is on the validity of the 9–12 version of DISCOVER.
Research on DISCOVER

A few studies have been conducted to examine the technical properties and other aspects of DISCOVER, such as gender differences. The preliminary results were promising, suggesting the effective use of DISCOVER for the identification of minority students. This research is summarized below.

In a triangulated inquiry on the inter-rater reliability of DISCOVER, Sarourhim (1999b) investigated the compatibility of ratings given to students by three independent raters: the DISCOVER observers, the classroom teacher, and the researcher. In using DISCOVER, the students’ linguistic, mathematical, and spatial intelligences are assessed through specific group activities consisting of distinguishable tasks, whereas their personal and kinesthetic intelligences are evaluated by the observers through unstructured behavior as the students perform in the group activities. For example, students who help one another and demonstrate leadership skills are given a high rating on the interpersonal intelligence; those who make pertinent remarks about themselves that correspond to their performance are given a high rating on the intrapersonal intelligence, and those who make good use of their body (i.e., incorporate their body into their construction in the spatial tasks or move gracefully throughout the administration) are given a high rating in kinesthetic intelligence. The results of this study showed that the DISCOVER observers, classroom teacher, and researcher gave similar ratings to students on the linguistic, spatial, and mathematical intelligences, but their ratings were not as similar for the personal and bodily-kinesthetic intelligences. The author concluded that the DISCOVER observers were more effective in appraising students’ intelligences when the appraisal was made through specific activities than when the appraisal depended on observing unstructured behavior. The author recommended that specific activities be developed for an accurate appraisal through DISCOVER of the whole spectrum of multiple intelligences.

Griffiths (1996) conducted two studies on the interobserver reliability of the DISCOVER assessment. In the first study, two observers separately watched videotapes of five observation sessions of the activity that assesses spatial intelligence. Participants were 25 Navajo children ranging in age from 9 to 13 years. As they viewed the tapes, the researchers sketched the children’s constructions and took notes in much the same way as the original observers on the tapes did. Then, each of the researchers independently classified the children’s problem-solving ability according to the four rating categories of “Unknown,” “Maybe,” “Probably,” and “Definitely.” A correlational analysis yielded positive and significant correlations, with the highest being 0.81, indicating a high agreement among the three observers. Percentages of agreements using Cohen’s Kappa ranged from 75 to 100%. Griffiths concluded

that the interobserver reliability of the DISCOVER assessment task for spatial activity was high. She attributed the differences in the percentages of agreements to differences in the two observers’ length of experience in the process.

In the second study, participants were observed in a live setting. Six observers with different levels of experience (novices, moderate experience, and experts) watched the students perform three of the DISCOVER assessment activities (spatial, logical-mathematical, and linguistic) and recorded separate notes. Participants were 91 students ranging in age from 5 to 11 years. Cohen’s Kappa indicated an agreement between the researcher and all six observers ranging from 80 to 100%, with the highest agreement being between the researcher and the expert observers and the lowest between the researcher and the novices. Also, the agreement among observers was 95 to 100% across all experience levels on the “Definitely” rating category. The researcher concluded that the DISCOVER assessment interobserver reliability was high and that levels of observers’ experience affected slightly, but not significantly, their rating of students’ problem-solving ability.

In another study, Sarourhim (2000) investigated the alignment of DISCOVER with the theory of multiple intelligences through a series of interobserver correlations. The sample consisted of 254 elementary students, predominantly from economically disadvantaged Native American and Hispanic groups. All participants took either the K–2 or the 3–5 version of DISCOVER, depending on their grade level. The results showed low interobserver correlations, indicating that students who were identified as gifted in one intelligence were not necessarily identified as gifted in the other intelligences. The results provided evidence that the different DISCOVER activities with distinguishable cognitive tasks may indeed measure different intelligences, a finding that might provide support to the consistency between DISCOVER and Gardner’s MI theory.

Sarourhim (2001) conducted another study to examine the concurrent validity of DISCOVER with the Raven’s Progressive Matrices. This study also examined gender differences and whether the number of minority students identified would be higher through the use of DISCOVER, than that yielded by traditional standardized tests. The results, based on a sample of Native American and Hispanic students, showed a high correlation between the students’ scores on the Raven’s and their ratings in the spatial activities of DISCOVER and low correlations between the students’ Raven scores and their ratings in the linguistic activities of DISCOVER, giving evidence about the convergent and discriminant validity of DISCOVER. The results also showed that, through the use of the DISCOVER assessment, 22.9% of the students were identified as gifted. In addition, no significant gender differences were found in identification, possibly indicating that the assessment
is mostly fair and does not discriminate against males, females, or ethnic origin.

In the present study, analyses similar to Sarouchim’s (2000, 2001) investigations were carried out to examine the validity of the DISCOVER assessment with an older population, namely secondary school students. Four questions guided the present inquiry:

1. How good is the fit between the DISCOVER assessment and the theory of multiple intelligences, which constitutes the framework of the assessment?
2. Do gender differences in identification appear through the use of the assessment with secondary students?
3. Do ethnic differences in identification appear through the use of the assessment with secondary students?
4. Could the assessment be used to reduce the problem of minority secondary students’ under-representation in programs for the gifted?

Method

Participants

The sample of this study consisted of 303 ninth graders (50.5% males and 49.5% females) taken from 4 different schools located in northern and southern Arizona. Data were collected over 4 academic years, from 1997 to 2001. The students who participated in the study were predominantly Hispanic (50%) and Native Americans from the Navajo tribe (29%). White Americans represented 21% of the sample. Participants’ socioeconomic status ranged from low to lower middle class as evidenced by the participation of the majority of the sample in the free lunch program. The ethnic and gender distribution of participants is shown in Table 1.

Instrument

The 9–12 DISCOVER assessment was designed to tap into the strengths and capabilities of secondary students through five activities: Drawing & Construction (spatial artistic), Word Play (linguistic), Writing (linguistic), Individual Tangrams (spatial analytical/logical-mathematical), and Group Tangrams (interpersonal intelligence). All five activities include a variety of problem-solving tasks that assess the underlying intelligence(s). The problems range in structure and openness from very well structured (Type I) to ill-structured or fuzzy (Type V) based on the Schieve and Maker (Maker, 1993) continuum of problem-types. (For a thorough description of the DISCOVER assessment model for grades 9–12, see Maker, 1994.)

An important feature of DISCOVER is that the assessment was designed to be culturally sensitive to diverse groups.

<table>
<thead>
<tr>
<th>Participants' Gender and Ethnic Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

The instructions are given in the native language of the students assessed. The material used was tested for its cultural appropriateness and was modified to ensure that DISCOVER is a culturally bias-free instrument. Also, the tasks were found to be engaging, the material appealing, and the problems developmentally appropriate (see section below on the development of the behavior check list).

Descriptions of activities. The rationale for focusing on spatial, linguistic, logical-mathematical, and personal intelligences, rather than on the full-spectrum of Gardner’s intelligences, is that school success depends to a great extent on high competence in these abilities. Therefore, for placement purposes, a good match must exist between the assessment and the programs for the gifted. The five activities of the assessment are:

Drawing and Construction: A box full of material is made available to each group of five to six students. The material consists of a kit for making mechanical toys, clay, pastels, drawing pencils, water colors, and a variety of construction paper. Students are asked to choose a photograph from among several found in the box and recreate it partially or fully using the material available to them. Students then use the material to make a construction of their choice. For optimal data accuracy, observers take pictures and make sketches of the constructions. More often than not, administration sessions are videotaped for preciseness and scrutiny of the processes that students use in solving problems.

Individual Tangrams: Each student is given a set of Chinese Tangrams of different geometrical shapes (triangles, rectangles, parallelograms, and squares). The activity starts by asking students to work individually on making a rhombus, then each is provided with a booklet of puzzles and is asked to complete as many puzzle sheets as possible using the Tangram pieces. When a student completes all the puzzle sheets, he or she works on additional “challenge sheets” that are of greater complexity than the puzzle booklet.

Group Tangrams: In this group activity, students put all their sets of Tangram pieces together and work jointly on making a giant triangle. They are then asked how their contribution hindered or helped the group members in achieving their goal.
Word Play: This activity consists of two tasks. In the first task, students watch a set of 13 emotion-provoking slides and write on a Word Chart all the ideas, words, or sentences that come to their mind as they watch them. The Word Chart is divided into two columns: one for words used frequently, labeled “common,” and the other for words used infrequently, labeled “uncommon.” Students are encouraged to use a thesaurus, synonym finder, or a dictionary made available to them. For the second task, students are asked to choose one photograph seen in the slides that they found interesting. Then, each one is provided with a bag that contains different items, such as coins, beads, and pieces of leather. Students are told to imagine finding one or more of the items in the scene of the photograph and to tell a story about their experience. They can either tell their story to the observer, who will record it verbatim, or they can use a tape-recorder.

Writing: Students are asked to produce a written piece about a topic of their choice. They can also choose the form of presentation (e.g., story, poem, essay, news report, play, and so forth).

The Process

Activities are carried out in the familiar classroom environment. The classroom teacher gives standardized instructions while the students work in groups of five. To avoid bias, observers rotate at the end of each activity; thus, students are typically assessed by three observers during each administration of DISCOVER. Upon completion of the tasks, observers meet to discuss the students’ performance and complete a check list for each student. Typically, students receive one rating in each activity. The ratings are Unknown, Maybe, Probably, and Definitely. Observers are guided in their evaluation by the check list items, which are established according to Gardner’s description of “core capabilities” for each intelligence. For example, superior spatial ability is demonstrated through the use of technique, color, movement, and quality of the constructions. High linguistic competence is seen through fluency and originality in the ideas and words used. Superior interpersonal skills are seen through leadership, cooperative, and instrumental behaviors.

At all times, observers must rely in their evaluation on behaviors and characteristics of products, rather than on inferences. In addition, observers focus on the processes or strategies that students use in solving problems. The “Definitely” rating is the highest and corresponds to giftedness in the intelligence measured by the respective activity. Typically, a student who receives a “Definitely” in at least two of the activities is identified as gifted. However, the criteria for identification are flexible and depend on the particular strengths of each group of students assessed, such as an outstanding quality of products or unusual speed in finishing the tasks (Sarouphim, 2000). Identification criteria also depend on the school district’s decision criteria and the focus of the program for the gifted in which students identified through DISCOVER will be placed.

Behavior checklist. The problem-solving behavior check list includes statements about superior problem-solving behaviors and characteristics of products created by the students. Each observer completes one portion of the check list that corresponds to the activity he or she observed (Sarouphim, 1999a).

Observers’ classifications range from “no strength observed” (Unknown) to “definite strength observed” (Definitely). Typically, the “Definitely” rating category is given to students who exhibit a variety of superior problem-solving behaviors as described in the check list.

Examples of check list items include: “gives descriptions easily and fluently” (linguistic), “sees which pieces will complete a puzzle without physically manipulating them” (spatial), and “persists on difficult tasks for him/her” (general). All the items in the check list indicate superior problem-solving behavior. Observers check only those they have seen the student exhibit. The greater the number of items checked, the more likely the student will receive the “Definitely” rating. Similarly, the “Unknown” rating is given to students who do not show any or most of the behaviors listed in the check list. The categories of “Maybe” and “Probably” are given to students who exhibit more or less superior problem-solving behaviors in comparison with their classmates. Of course, the more evidence observers gather on the student’s ability, the more likely that student will receive the “Definitely” rating. However, the number of items checked for each rating category is not specified and depends on each group of students assessed. The DISCOVER criteria for identification are not as well-delineated as those in standardized tests, but they are comparable to the criteria used in other performance-based assessments. The rationale is related to the major criticism of standardized tests, which is an undue emphasis on numerical data, rigid criteria, and students’ ranking.

The first DISCOVER assessment problem-solving behavior check list was developed for grades K–5 following an assessment of approximately 5,000 students using the DISCOVER process (Maker, 1996). The grades 6–8 and 9–12 check lists were adapted from the original form, but modified to match the 6–8 and 9–12 tasks. The process initially used to create the problem solving check list was as follows: After every classroom observation, observers met and decided which students in their groups were “effective, efficient, or economical” problem solvers; observers then described the behaviors they had seen students exhibit; observers also described the characteristics of the products students had created that observers considered effective, efficient, or economical. The original
observers included individuals from the same cultural and linguistic groups as the children, as well as those from other groups.

The DISCOVER assessment behavior check list was revised several times as more data were collected from observations. Some of the behaviors can be observed in only one activity, but most are relevant to two or more activities. This practice is consistent with Gardner's belief that all activities involve the use of more than one intelligence and with data from the grounded research that formed the basis for the assessment. Also, behaviors that apply to all intelligences were aggregated and included in the last section of the check list labeled “General” (Maker, 1996).

Results

To examine the fit between the DISCOVER assessment and Gardner's (1983) theory of multiple intelligences, a series of interobserver correlations were computed, resulting in a correlation matrix that shows the relationship between the rating given to students in one activity and their ratings in the other activities. Gender and ethnic differences on the DISCOVER activities were assessed through a 2 x 3 MANOVA (gender by ethnicity), and a chi square test was performed to examine gender and ethnic differences in identification (i.e., students given the “Definitely” rating in at least two activities of the assessment).

Interobserver Correlations

Although all correlations were significant (see Table 2), correlation values were generally low, with the highest being \( r = 0.567 \), between students' ratings in the activities of Word Play and Writing, both assessing linguistic intelligence. The lowest correlation (\( r = 0.193 \)) was found between students' ratings in Drawing & Construction and their ratings in Writing, activities that assess spatial and linguistic intelligences, respectively.

Gender and Ethnic Differences

Table 3 shows the mean ratings of students by gender and ethnicity. Data were coded as follows: Unknown = 1, Maybe = 2, Probably = 3, and Definitely = 4. As shown in Table 3, participants' ratings fluctuated between the Maybe and Probably rating categories, with the highest mean given to White females in Word Play (3.11) and the lowest mean given to Native American males in the Writing activity (2.62). A 2 x 3 MANOVA yielded nonsignificant gender-by-ethnicity interaction (\( F[4, 297] = 0.78, \text{ ns} \)), nonsignificant main effect for gender (\( F[4, 297] = 1.65, \text{ ns} \)), and nonsignificant ethnicity effect (\( F[3, 274] = 1.23, \text{ ns} \)), indicating that males and females, as well as participants from the different ethnic groups, received mostly similar ratings in the five activities of the DISCOVER assessment.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Observers’ Inter-rating Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Drawing &amp; Construction</td>
<td>–</td>
</tr>
<tr>
<td>Word Play</td>
<td>–</td>
</tr>
<tr>
<td>Writing</td>
<td>–</td>
</tr>
<tr>
<td>Ind. Tangrams</td>
<td>–</td>
</tr>
<tr>
<td>Group Tangrams</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: *p < 0.05; **p < 0.01

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Students’ Mean Ratings by Gender and Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mexican Americans</td>
</tr>
<tr>
<td>Drawing &amp; Construction</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>2.88</td>
</tr>
<tr>
<td>Females</td>
<td>2.83</td>
</tr>
<tr>
<td>All</td>
<td>2.85</td>
</tr>
<tr>
<td>Word Play</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>2.92</td>
</tr>
<tr>
<td>Females</td>
<td>2.95</td>
</tr>
<tr>
<td>All</td>
<td>2.93</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>2.73</td>
</tr>
<tr>
<td>Females</td>
<td>2.70</td>
</tr>
<tr>
<td>All</td>
<td>2.71</td>
</tr>
<tr>
<td>Individual Tangrams</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>2.80</td>
</tr>
<tr>
<td>Females</td>
<td>2.76</td>
</tr>
<tr>
<td>All</td>
<td>2.78</td>
</tr>
<tr>
<td>Group Tangrams</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>3.09</td>
</tr>
<tr>
<td>Females</td>
<td>3.00</td>
</tr>
<tr>
<td>All</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Note: Unknown = 1, Maybe = 2, Probably = 3, Definitely = 4.

Fall 2002, Volume XIV, Number 1 35
In terms of gender and ethnic differences in identification, no significant differences were found, $\chi^2(2, 89) = 0.066$, ns. A total of 89 students representing 29.3% of the participants were identified as gifted (see Table 4), with a slightly larger number of females. The percentages of identified participants were mostly in proportion to their ethnic distribution in the sample.

**Discussion**

The purpose of this study was to investigate some validity aspects of the grades 9–12 version of the DISCOVER assessment. Four issues were addressed: the good fit between the assessment and the theory of multiple intelligences, gender differences, ethnic differences, and whether the assessment would identify a larger pool of minority students than traditional measures. The results showed some evidence for a good fit between the assessment and Gardner's theory of multiple intelligences, as interobserver correlations of students ratings across DISCOVER tasks were mostly low, though significant. No gender or ethnic differences were found in identification. Also, 29.3% of the participants were identified as gifted, suggesting that the use of DISCOVER might help in reducing the problem of minority students' under-representation in programs for the gifted.

In this study, observers' inter-rating correlations across the different DISCOVER tasks were found to be low, indicating that individual students who received a high rating in one activity did not necessarily receive a similarly high rating in the other activities. In other words, students who were found gifted in one intelligence were not necessarily found gifted in the other intelligences assessed through DISCOVER. It is noteworthy to mention that the highest correlation ($r = 0.567$) was found between the activities of Word Play and Writing, the two activities of DISCOVER that measure linguistic intelligence, and the lowest correlation was between the two activities of Drawing & Construction and Writing ($r = 0.193$), which measure nonverbal and verbal intelligence, respectively. In other words, when the same construct was assessed, students received more similar ratings, whereas students' ratings were the least similar when the constructs assessed were different. Since some evidence for the interobserver reliability of DISCOVER was revealed in previous studies (Griffiths, 1996; Sarouphim, 1999b), this finding suggests that the DISCOVER activities with distinguishable cognitive tasks might measure different intelligences. Similar results were found in the analysis of the K–2 and 3–5 versions of DISCOVER (Sarouphim, 2000).

An interesting finding is the lack of gender and ethnic differences across activities. Scholars have long lamented the consistent finding of gender and ethnic differences in traditional tests: Boys do better than girls in the spatial and mathematical subtests, and White middle-class students receive higher overall scores than students from culturally diverse groups (Cummins, 1991; Lubinski & Benbow, 1992; Ogbu, 1994). In this study, the absence of cultural and gender bias is an important finding that might impact the practice of gifted identification. Also, this finding supports the argument advanced by proponents of performance-based assessment about the effectiveness of these instruments, especially with culturally diverse students (Borland & Wright, 1994; Clasen et al., 1994).

However, another explanation that cannot be ruled out for the lack of significant gender differences is the lack of an activity in the 9–12 DISCOVER, which assesses pure mathematical intelligence. In the elementary and middle school versions of DISCOVER, mathematical intelligence is assessed through a worksheet of open-ended problems (Sarouphim, 1999a); but, in the 9–12 version of DISCOVER, mathematical reasoning is measured through the Tangrams activity, which is not sufficient by itself to reveal faithfully students' mathematical intelligence. Therefore, a recommendation that stems from the results of this study is to develop an activity for assessing students' mathematical ability in the 9–12 DISCOVER that better fits Gardner's description of the core capabilities of logical-mathematical intelligence and is appropriate for high school students.

The use of DISCOVER to diminish the under-representation of minorities in programs for the gifted was also supported. A total of 29.3% of students were identified as gifted, a much larger percentage than the traditional 1–3% identified through the use of standardized tests (Reid et al., 1999). A significant finding is the identification of a large number of White participants, a finding that might promote the use of DISCOVER not only with culturally diverse groups, but also with the majority population.

Researchers would agree that the identification of a large number of minority students does not necessarily increase the validity of an instrument. However, the present demographics...
DISCOVER in High School

of gifted education clearly show that the potential and talent of culturally diverse students, especially those of a Black, Hispanic, or Native American ethnicity, are not properly discovered or nurtured. The debate of equity versus excellence (Ford & Harmon, 2001), about whether providing a better opportunity for minority students might lead to placing underserving students in programs for the gifted, though significant, should not constitute a barrier for opening a door to disadvantaged students that has been closed for too long. The possibility of identifying false positives exists even with the use of traditional tests.

However, this argument or any other argument of the sort does not justify the use of an assessment that lacks validity. Educators must continue to strive to find means to increase minority representation in programs for the gifted, but these means certainly should not include the use of inadequate instruments. The placement of a student, erroneously identified as gifted, in a program for those of high ability leaves that student with little hope to succeed and might be potentially harmful. Based on the information available at present on the technical properties of DISCOVER, how confident could one be selecting students on the basis of DISCOVER and placing them in programs for the gifted? The answer to this question lies within the kind of program for the gifted under consideration. A good match must exist between the program and the abilities measured in DISCOVER (i.e., focus on multiple intelligences). So far, research on DISCOVER has yielded mostly positive results in support of its effective use. However, further research is undeniably needed to investigate more thoroughly the construct validity of DISCOVER and its predictive validity before solid conclusions can be drawn.

One concern that must be raised is the criteria used in DISCOVER for identification, that is, a student who receives the “Definitely” rating in at least two activities is identified as gifted. These criteria might be troublesome and possibly confusing. For example, suppose a student receives the “Definitely” rating in the activities of Individual Tangrams and Group Tangrams, what does such performance reveal about his or her strengths? Also, are these ratings sufficient for placing that student in a program for the gifted? Even though this is a real concern, it might not constitute a serious problem because the developers of DISCOVER clearly specify that the criteria for identification are flexible and depend on the program in which students will be placed. However, the school authorities who make placement decisions based on DISCOVER must be made aware of this possible shortcoming. One recommendation that stems from this study is to revise the criteria used for identification. In particular, the developers of DISCOVER need to focus on whether the criterion of having the “Definitely” rating in any two of the DISCOVER activities is sufficient to qualify the student for placement in a program for the gifted or whether a “Definitely” rating in a specific combination of two activities would be needed for making more accurate decisions concerning placement.

Although the results of this investigation are encouraging and provide support for the use of DISCOVER, there are some limitations, as well. An important limitation is in the composition of the sample, which consisted solely of ninth graders. This limitation restricts the generalization of the results to other secondary students for whom this version of DISCOVER was designed. In further studies, participants from upper grade levels (10–12) must be included for a more solid conclusion on the use of DISCOVER with the full range of secondary students for whom the assessment was intended.

Another limitation is the lack of data on the performance of students identified through DISCOVER who were indeed placed in programs for the gifted. Such data are necessary and would provide information on the predictive validity of the assessment, which ultimately is the best indicator of the effectiveness of any instrument. Since DISCOVER is a relatively new assessment, studies on its predictive validity have not yet been published. However, work on this subject is underway. At this time, a longitudinal study is being conducted to follow the academic progress of a group of Native American students (the “Step-Up” group) who were all identified through DISCOVER and placed in different programs for the gifted. But, until the results are out, solid conclusions on the predictive validity of DISCOVER cannot yet be drawn, a limitation that school officials must take into consideration when using the DISCOVER assessment for identification and placement purposes.

**References**


---

**Author Note**

The author would like to thank Dr. C. June Maker for her contribution to this article.