DOCUMENT RESUME

ED 368 096 EC 302 841

AUTHOR Saccuzzo, Dennis P.; And Others
TITLE Use of the Raven Progressive Matrices Test in an Ethnically Diverse Gifted Population.
PUB DATE 94
CONTRACT R206A00569
NOTE 15p.; In: Saccuzzo, Dennis P.; And Others. Identifying Underrepresented Disadvantaged Gifted and Talented Children: A Multifaceted Approach. (Volumes 1 and 2); see EC 302 840.
PUB TYPE Reports - Research/Technical (143) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability Identification; Aptitude Tests; Comparative Analysis; Cultural Differences; *Culture Fair Tests; *Disadvantaged Youth; Educational Diagnosis; Elementary Education; Elementary School Students; *Gifted; Intelligence Tests; *Minority Group Children; Screening Tests; Sex Bias; Test Bias
IDENTIFIERS *Raven Progressive Matrices; San Diego Unified School District CA; *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

The efficacy of use of the Raven Progressive Matrices Test (RPM) in the selection of gifted children from traditionally underrepresented groups was investigated in a large-scale study with a diverse population. A total of 16,985 subjects were given the Raven Progressive Matrices Test. These included 22.7 percent Latinos, 37 percent Whites, 14 percent African-Americans, 2.8 percent Asians, 8.4 percent Filipinos, and 5.6 percent Indochinese, each of whom had been identified as potentially gifted on the basis of a case study analysis by a school psychologist. The sample of children certified as gifted based on Raven performance was compared to a group certified as gifted based on individual administrations of the Wechsler Intelligence Scale for Children - Revised (WISC-R), and to actual enrollment ratios of cultural subgroups for the school district. Findings support the position that the RPM is a more equitable test than the WISC-R for evaluating students from ethnically diverse populations, as well as for ensuring greater gender equity. (PB)

CHAPTER 1

Use of the Raven Progressive Matrices Test in an Ethnically Diverse Gifted Population

Dennis P. Saccuzzo*

Nancy E. Johnson

Tracey L. Guertin

San Diego State University

This research was funded by Grant R206A00569, U.S. Department of Education, Jacob Javits Gifted and Talented Discretionary Grant Program.

The authors express their appreciation to the San Diego Unified City Schools, to Gifted and Talented Education (GATE) Administrator David P. Hermanson, and to the following school psychologists: Will Boggess, Marcia Dijiosia, Eva Jarosz, Dimaris Michalek, Lorraine Rouse, Ben Sy, and Daniel Williams.

Correspondence should be addressed to Dennis P. Saccuzzo, Joint San Diego State/University of California, San Diego Doctoral Training Program, 6363 Alvarado Court, Suite 103, San Diego, California 92120-4913 (Telephone: 619-594-2844 / FAX: 619-594-6780 / e-mail: dsaccuzz@sunstroke.sdsu.edu).

Abstract

The efficacy of use of the Standard Raven Progressive Matrices Test (RPM) in the selection of gifted children from traditionally underrepresented groups was investigated in a large-scale study with a diverse population. A total of 16,985 Raven subjects included 22.7% Latinos, 37% Whites, 14% African-American, 2.8% Asian, 8.4% Filipinos, and 5.6% Indochinese, each of whom had been identified potentially gifted based on a case study analysis by a school psychologist. The resultant sample of children certified gifted based on Raven performance was compared with a group certified gifted based on individual administrations of the Wechsler Intelligence Scale for Children - Revised (WISC-R). The Raven gifted sample was also compared to actual enrollment ratios for the school district to ascertain if equity in representation could be achieved for gifted programs using the Raven. Although chi square results indicated that use of the RPM as opposed to the WISC-R led to increased equity for all groups, full equity was not achieved. When compared to the district population, the Raven-selected group showed underrepresentation for Latinos and African-Americans and overselection for Whites, Asians, and Filipinos. Chi square comparisons based on the expectation of equal gender representation revealed that the WISC-R disproportionately overselected boys, while the RPM showed no gender differences. The findings support the position that the RPM is a more equitable test than the WISC-R for ethnically diverse gifted populations and for girls.

Use of the Raven Progressive Matrices Test in an Ethnically Diverse Gifted Population

Professionals in the field of Gifted Education continue to express considerable interest in developing alternatives to traditionally used standardized tests in identifying gifted students. Such alternatives are particularly needed for students from traditionally underrepresented groups such as African-Americans and Latino/Hispanics. One such alternative, repeatedly mentioned in the literature, is the Raven Progressive Matrices (RPM) Test (Raven, 1958, 1960; Raven et al., 1986). A number of investigators have pointed to the RPM as a culturally reduced measure of cognitive ability with considerable promise as a measure of giftedness in traditionally underrepresented and culturally diverse students (Baska, 1986; James, 1984; Powers & Barkan, 1986; Powers, Barkan, & Jones, 1986).

The standard form of the RPM consists of 60 matrix problems, which are separated into five sets of 12 designs each. Within each set of 12, the problems become increasingly difficult, and each of the five sets is progressively more difficult. Each individual design has a missing piece. The student's task is to select the correct piece to complete the design from among six to eight alternatives. Correct responses are based on various organizing principles such as increasing size, reduced or increased complexity, and number of elements.

Because RPM stimuli are visually presented, it is easy to mistake the test as one of visual perception or spatial reasoning. It is neither. As Cherkes-Julkowski, Stolzenberg, and Segal (1990) have noted, "The Raven is as close to a study of pure thinking processes in the absence of the influence of specific content acquisition as is available" (p.7). As Snow and Colleagues have shown using radex and hierarchical models, the RPM is the best available measure of general intelligence (Marshalek, Lohman, & Snow, 1983; Snow, Kyllonen, & Marshalek, 1984). As a measure of general intelligence the RPM correlates highly with verbal measures of ability, even though the stimuli themselves are completely nonverbal (Carpenter, Just, & Shell, 1990). In fact, Positron Emission Tomography (PET) Scans, which produce computer generated images of the brain, have shown that the entire brain is involved in solving RPM problems, with the three most used areas being the right cerebral hemisphere, the left temporal lobes, and the left frontal lobes (Haier, Siegel, Nuechterlein, Hazlett, Wu, Paek, Browning, & Buchsbaum, 1988). The left temporal lobe involvement is most likely due to the use of verbal codes in solving RPM problems.

Because its stimuli are nonverbal, the RPM can be administered fairly in American schools to individuals who speak a language other than English. Because stimuli are visually presented, rather than spoken, they are not transitory. Thus, the stimulus remains in front of the student, which reduces the role of memory and even attentional factors in performance (Cherkes-Julkowski et al., 1990). Solving RPM problems does not depend heavily, as do all language based tests, on acquired knowledge, specific cultural experiences, or reading ability. As Carpenter et al. (1990) have noted, "The Raven measures the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge derived either from schooling or previous experience." In sum, the Raven measures general intelligence and correlates with measures of linguistic ability. It uses nonverbal stimuli and does not require a specific knowledge base.

Previous investigations have found that the Raven has not only been effective in identifying traditionally underrepresented children for gifted programs, but also correlates with their success (Baska, 1986). In one study, Powers, Barkan, and Jones (1986) found no significant differences between Hispanic and Anglo-American children's mean scores, score variability, and test reliability for the RPM. Other studies have supported the validity of the RPM for Hispanic (Powers & Barkan, 1986) and Navajo students (Sidles & MacAvoy, 1987).

As Raven (1989) has noted, the Raven Progressive Matrices Tests have been used in over 1,600 published psychological studies (see Court, 1988; Court & Raven, 1977, 1982), making the RPM among the most researched psychological tests. Until recently, however, the RPM has not been used widely in applied clinical and educational settings. A major problem had been the lack of adequate U. S. norms (Kaplan & Saccuzzo, 1989, 1993). An extensive and relatively current set of norms, which include U. S. as well as worldwide norms, is now available (Raven, Summers et al., 1986). More than 30,000 students

aged 5 to 18 were chosen to be representative of school districts across the United States in approximately 30 norming studies. Ethnicity and socioeconomic factors made independent contributions to the variance. Ethnic differences, which were attributed to differences in birth weight, infant mortality, and the incidence of serious childhood illness, showed a decline compared to earlier reports (Burciaga, 1973; Hoffman, 1983; Jensen, 1980). There were, for example, no major Hispanic/White differences in the Ontario-Montclair School District of California. Moreover, the RPM was found to have equal predictive validity within each group (Hoffman 1983, 1986).

It should be noted that while ethnic differences may be declining, there remain differences in the mean scores as a function of ethnicity as well as socioeconomic background. Nevertheless, the question remains as to whether the RPM can be used to achieve a more equitable selection of students than can be obtained with more widely used traditional tests, such as the Wechsler Intelligence Scales.

In the present study we present the results of a Jacob Javits grant whose purpose was to evaluate the efficacy of the RPM in the selection of traditionally underrepresented groups for gifted programs. The data were collected over a three year period and include an archival data pool of over 5,000 administrations of the Wechsler Intelligence Scale to children from 8 major ethnic backgrounds.

Method

Subjects:

The subjects were 16,985 children who were referred and evaluated for the gifted program at San Diego City schools between the Fall of 1991 and the Spring of 1993. Students were classified into one of eight ethnic backgrounds based on self-report of the parent as follows: 3,864 Latino/Hispanic, 6,286 White, 2,389 African-American, 483 Asian, 75 Native-American, 104 Pacific Islander, 1,419 Filipino, and 958 Indochinese. There were 1407 classified as "Other". Of the 16,985 subjects, 51.5% (8,740) were female, 48.5% (8,245) male. The distribution by grade level was as follows: 24 first-, 7,664 second-, 1,467 third-, 819 fourth-, 3,737 fifth-, 748 sixth-, 2,122 seventh-, 263 eighth-, and 90 ninth-graders. There were 51 cases where data on grade was missing.

As a comparison sample, the files of all children evaluated for giftedness in the San Diego City School System between 1984 and 1990 were examined. During this time period, the district had used the WISC-R as the primary tool for identifying giftedness. A total of 9315 students had been given the WISC-R during this time period.

A second point of comparison was based on actual enrollment figures by ethnic background during the course of the study. The average enrollment ratio in the district as a whole by ethnic background between 1991 and 1993 was as follows: 27.2% Latino/Hispanic, 37.2% White, 16.2% African-American, 2.3% Asian, 0.6% Native-American, 0.7% Pacific Islander, 8% Filipino, and 7.7% Indochinese.

Procedure:

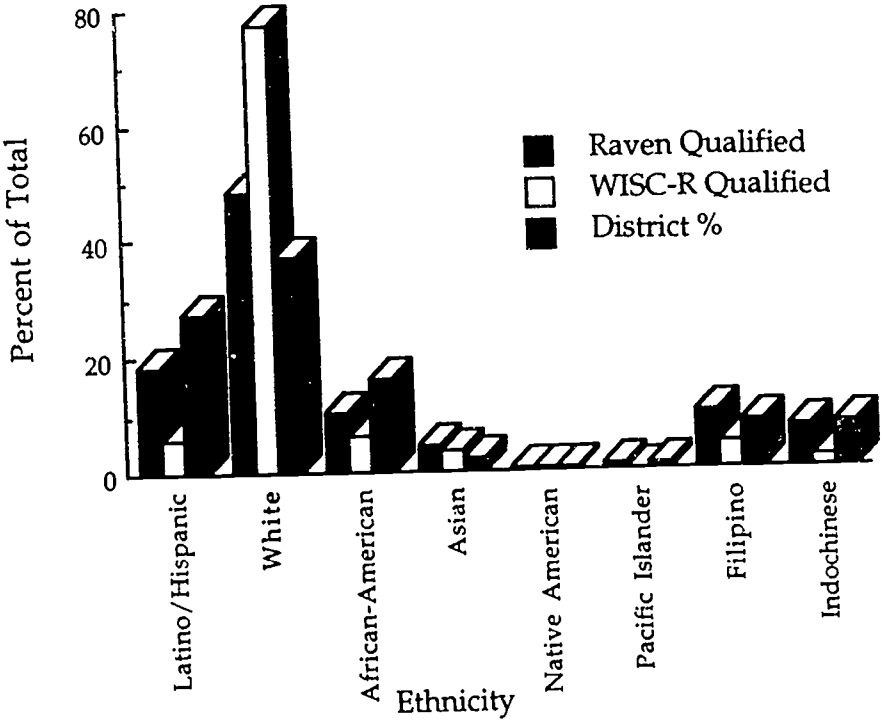
Children were given either a WISC-R (entire 1984-1990 sample) or a Standard Raven (entire 1990-1993 sample) by a district school psychologist. The WISC-Rs were individually administered; the Raven was group administered. As a part of the evaluation process the school psychologists conducted a case study analysis of each child to evaluate for the presence of risk factors and level of achievement (see Saccuzzo, Johnson, & Russell, 1992). Five risk factors were considered. These were: cultural/language, economic, emotional, environmental, and health. Achievement was evaluated in terms of standard scores on the California Test of Basic Skills (CTBS) or the Abbreviated Stanford Achievement Test (ASAT).

Children were certified as gifted if they obtained a Full Scale IQ of 130 on the WISC-R or an IQ equivalent of 130 on the Raven (i.e., achieved a score in at least the 98th percentile) based on the Smoothed U.S. Norms reported by Raven et al. (1986). In addition, children who had two or more of the risk factors were certified as gifted if they had a Full Scale WISC-R IQ of 120 or Raven IQ equivalent of 120. Finally, children who obtained a WISC-R or Raven score of 3 standard deviations above the mean were placed in a special "Seminar" program.

Results

Figure 1 illustrates the ethnic composition of the district for all eight ethnic backgrounds compared to all children who were certified as gifted with the WISC-R (entire 1984-1990 sample) and to all children who were certified as gifted with the Raven (entire 1990-1993 sample). Inspection of Figure 1 reveals that the RPM led to increased equity for all ethnic groups. For example, only about 20% of the expected number of Latino/Hispanic were selected with the WISC-R, while about 80% of the expected were selected using the RPM. For the Whites, about 200% of the expected were selected with the WISC-R, while only about 120% of the expected were selected with the RPM. Pacific Islanders, Native-Americans, and Indochinese were all greatly underselected by the WISC-R, but selected according to the expectation based on their numbers with the Raven. Filipinos went from about 60% of expectation with the WISC-R to about 120% with the Raven. African-Americans went from about 40% of the expectation to over 60%. Thus, the RPM provided a more equitable distribution for all ethnic backgrounds.

Figure 1. Ethnic composition of those certified gifted on the basis of WISC-R scores (1984-1990) versus Raven's Progressive Matrices scores (1990-1993).



To demonstrate statistically the superiority of the RPM over the WISC-R in terms of equity, we used the following procedure. Chi Squares were computed for each ethnic group using the number of children certified by use of the WISC-R in each ethnic background as the expected and the number of children in that ethnic background selected by use of the Raven as the observed. Table 1 provides a summary of the findings. As Table 1 shows, there were significant differences in the direction of increased equity for all groups but Native-Americans (where the *n* was small) and Asians.

Table 1. *Chi Square results using children certified with the WISC-R as the expected and children certified with the Raven as the observed*

Ethnicity	*df*	*n* (Expected)	*n* (Observed)	Chi Square
Latino/Hispanic	1	413	1304	1922.23 ***
White	1	5535	3441	792.20***
African-American	1	454	714	148.90***
Asian	1	260	306	8.14
Native American	1	20	36	12.80
Pacific Islander	1	10	50	160.00***
Filipino	1	321	713	478.70***
Indochinese	1	112	513	1435.72***

* $p < .05$
** $p < .01$
*** $p < .001$

Despite the substantial improvement in equity when the RPM was used, inspection of Figure 1 indicates that full equity was not achieved. To verify this observation statistically, Chi Squares were again computed using the proportionate number of children in each ethnic background as the expected compared to the number actually certified with the Raven as the observed.

Table 2 presents a summary of these findings and reveals significant discrepancies for all groups except Native-Americans, Pacific Islanders, and Indochinese. Latino/Hispanics and African-Americans were significantly underselected compared to the expectation while Whites, Asians, and Filipinos were significantly overselected.

Table 2. *Chi Square results comparing the actual proportion of children in each ethnic background with the number actually certified with the Raven*

Ethnicity	*df*	*n* (Expected)	*n* (Observed)	Chi Square
Latino/Hispanic	1	2009	1351	215.51***
White	1	2743	3568	244.69***
African-American	1	1200	766	156.96***
Asian	1	170	319	130.59***
Native American	1	41	39	0.10
Pacific Islander	1	55	48	0.89
Filipino	1	595	739	34.85***
Indochinese	1	569	532	2.41

* $p < .05$
** $p < .01$
*** $p < .001$

To examine possible ethnic differences on the Raven independently of risk factors, we used the following system. First, we considered only those children who scored in the 98th percentile or better on the Raven. Next, we determined the number of children in the school district in each ethnic background that would represent 2 percent of that group. For example, given the total number of Latino/Hispanics in the district, 905 would be in the top two percent. The actual number of children in the top two percent for each ethnic background was used as the expected, while the number of children in each background who actually scored in the upper 2% (i.e., 98th percentile or better) was used as the observed in a series of Chi Square analyses. Table 3 summarizes these analyses and shows significant discrepancies for four groups: Latino/Hispanics and African-Americans, who were underselected; and Whites and Asians who were overselected compared to the expectation.

Table 3. *Chi Square results comparing the actual number of children in the top 2% of each ethnic background with the number who actually scored in the upper 2%*

Ethnicity	*df*	*n* (Expected)	*n* (Observed)	Chi Square
Latino/Hispanics	1	905	379	305.72***
Whites	1	1238	2009	480.16***
African-Americans	1	541	209	203.74***
Asians	1	77	173	119.69***
Native Americans	1	18	21	0.50
Pacific Islanders	1	25	12	6.76
Filipinos	1	268	316	8.59
Indochinese	1	256	199	12.69

* $p < .05$
** $p < .01$
*** $p < .001$

To investigate the possibility of gender differences in WISC-R and Raven performance, the number of boys and girls who actually scored in the 98th percentile and above on each test in each ethnic group was compared to the expectation that there would be equal numbers of boys and girls. As can be seen in Table 4, there were no significant differences in performance for boys and girls of any ethnic background on the Raven.

Table 4. *Chi Square results comparing the number of boys and girls of each ethnic background who actually scored in the top two percent on the Raven*

Ethnicity	*df*	Total *n*	Boys (Observed)	Chi Square
Latino/Hispanics	1	374	181	0.19
Whites	1	2002	1057	3.13
African-Americans	1	206	116	1.64
Asians	1	172	79	0.57
Native Americans	1	21	12	0.21
Pacific Islanders	1	12	6	0
Filipinos	1	315	170	0.99
Indochinese	1	198	100	0.01

* $p < .05$
** $p < .01$
*** $p < .001$

For the WISC-R, by contrast, significantly more White boys (2632) than girls (2272) achieved an IQ score in the top two percent (Table 5). Moreover, the trend was in favor of boys on the WISC-R for every ethnic group except Pacific Islanders.

Table 5. *Chi Square results comparing the number of boys and girls of each ethnic background who actually scored in the top two percent on the WISC-R*

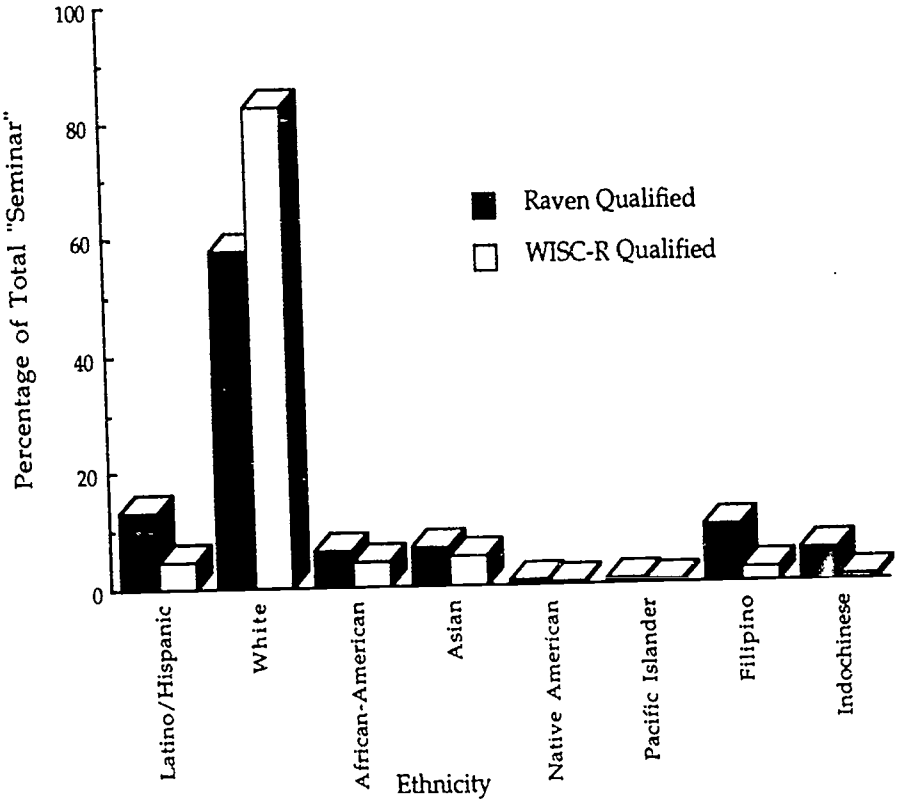
Ethnicity	*df*	*n*	Boys (Observed)	Chi Square
Latino/Hispanic	1	265	152	2.87
White	1	4904	2630	13.21***
African-American	1	252	129	0.07
Asian	1	202	109	0.63
Native American	1	13	9	0.96
Pacific Islander	1	9	3	0.50
Filipino	1	182	107	2.81
Indochinese	1	61	34	0.40

* $p < .05$
** $p < .01$
*** $p < .001$

In order to examine performance of individuals at the highest level of ability (i.e., three standard deviations above the mean), it was necessary to generate local norms, since the norms provided in the manual only go up to the 99th percentile. In constructing local norms for the 99.1 through 99.9 percentile, we used the following procedure. Based on the table of smoothed North American norms provided by Raven, et al. (1986), we selected all those children in our study who had obtained a raw score in the 99th percentile for each age range in the table. We then examined the frequency distribution of raw scores for each age range and attempted to break the scores down into ten groups occurring with equal frequency, representing the 99.0 through 99.9 percentile.

In the analyses that followed, the local norms (Guertin, Johnson, Saccuzzo, & Lopez, 1992) were used to identify students who scored three standard deviations above the mean (i.e., above the 99.87 percentile). Figure 2 illustrates the proportion of students from the entire 1990-1993 sample, by ethnic background, who obtained scores at the 99.9 percentile on the local Raven norms versus the proportion, by ethnic background, who obtained WISC-R Full Scale IQ's of 145 or above. As inspection of Figure 2 shows, there was an increase in the representation of children selected for the schools' very gifted "Seminar" program for all ethnic backgrounds except for the Whites, who represent about 35 percent of the district. With the WISC-R, Whites represented more than 80% of the children selected for the "Seminar" program. With the Raven, Whites represented less than 60%.

Figure 2. Ethnic composition of the 99.9th percentile: WISC-R versus Raven's Standard Progressive Matrices



To demonstrate statistically the superiority of the RPM over the WISC-R for selection at the highest level of ability, Chi Square analyses were computed for each ethnic background using the proportionate number of children selected by the WISC-R as the expected and the number selected by the Raven as the observed, as was done in Table 1. Results were similar to those found for the children who scored above 130. In fact, there were significant increases in the direction of increased equity for all ethnic backgrounds except Asians and Native Americans (See Table 6).

Table 6. *Chi Square results using children certified "Seminar" with the WISC-R as the expected and children certified "Seminar" with the Raven as the observed*

Ethnicity	*df*	*n* (Expected)	*n* (observed)	Chi Square
Latino/Hispanics	1	24	69	84.38***
Whites	1	430	300	38.79***
African-Americans	1	22	33	5.50*
Asians	1	26	33	1.88
Native Americans	1	1.4	3	1.83
Pacific Islanders	1	1	4	9.00**
Filipinos	1	13	49	99.69***
Indochinese	1	4	29	156.25***

* $p < .05$
** $p < .01$
*** $p < .001$

As with the children who scored two standard deviations above the mean, there were inequities for children who scored three standard deviations above the mean. A total of 520 children in our sample scored in the 99.9th percentile on the Raven, as determined by local norms. We compared the number who actually scored at that level with the number expected based on district proportion for each ethnic group. Results, summarized in Table 7, indicate that Latino/Hispanics and African-Americans were underrepresented while Whites, Asians, and Indochinese were overrepresented.

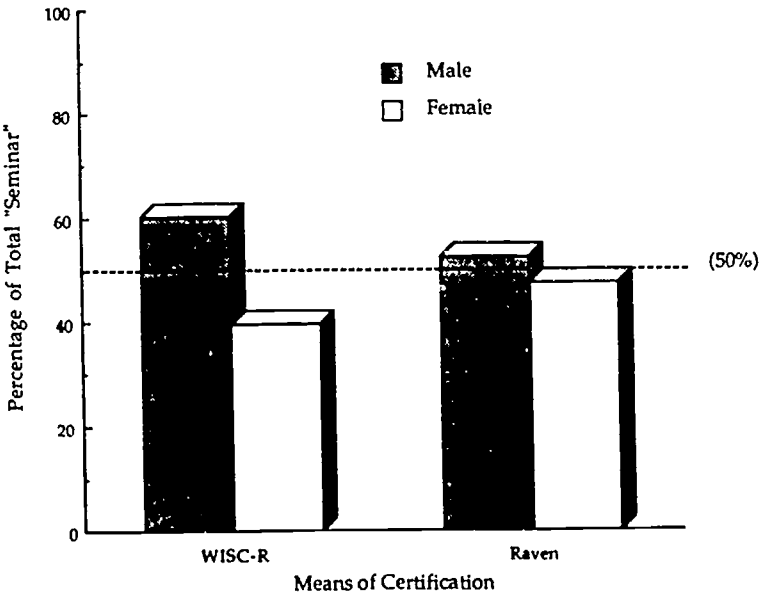
Table 7. *Chi Square results comparing children certified "Seminar" using the Raven to their proportionate numbers in the district*

Ethnicity	*df*	*n* (Expected)	*n* (Observed)	Chi Square
Latino/Hispanics	1	141	69	36.77***
Whites	1	193	300	59.32***
African-Americans	1	84	33	30.36***
Asians	1	12	33	36.75***
Native Americans	1	3	3	0.00
Pacific Islanders	1	4	4	0.00
Filipinos	1	42	49	1.17
Indochinese	1	41	29	3.51

* $p < .05$
** $p < .01$
*** $p < .001$

Figure 3 illustrates the gender composition for children who scored either at or above the 99.9 percentile on the Raven or had a Full Scale WISC-R IQ of 145 or greater.

Figure 3. Gender composition of the 99.9th percentile: WISC-R versus Raven's Standard Progressive Matrices



As Figure 3 shows, the substantial gender imbalance in favor of males with the WISC-R was essentially eliminated with the RPM. Chi square analyses statistically demonstrated the reduction in gender bias at the highest levels of ability. Results are summarized in Tables 8 and 9, where actual numbers of boys and girls who scored three standard deviations above the mean on the WISC-R and the Raven, respectively, were compared with the expectation of equal numbers of boys and girls.

Table 8. *Chi Square results comparing boys and girls certified "Seminar" using the WISC-R with the expectation of equal gender numbers*

Ethnicity	*df*	Total *n*	Boys (Observed)	Chi Square
Latino/Hispanic	1	52	29	0.35
White	1	925	570	22.47***
African-American	1	47	26	0.27
Asian	1	56	30	0.14
Native American	1	3	2	0.17
Pacific Islander	1	2	2	1.00
Filipino	1	29	21	2.91
Indochinese	1	8	5	0.25

* $p < .05$
** $p < .01$
*** $p < .001$

Table 9. *Chi Square results comparing boys and girls certified "Seminar" using the Raven with the expectation of equal gender numbers*

Ethnicity	*df*	*n* (total)	Boys (Observed)	Chi Square
Latino/Hispanic	1	69	34	0.007
White	1	300	156	0.24
African-American	1	33	18	0.14
Asian	1	33	15	0.14
Native American	1	3	1	1.16
Pacific slander	1	4	3	0.5
Filipino	1	32	32	2.30
Indochinese	1	29	15	0.02

* $p < .05$
** $p < .01$
*** $p < .001$

The efficacy of the Raven was compared to that of the WISC-R for predicting achievement in a series of correlations between intelligence tests and achievement tests used by the school district. Each child received a test of intelligence and an achievement test during the same school year. Table 10 demonstrates correlations between performance on the Raven, as expressed in z-scores, and on the subtests of the Comprehensive Test of Basic Skills (CTBS) for 1707 children, as well as correlations between CTBS and WISC-R for another 1925. As can be seen, the Raven correlated more highly with CTBS Total Language subscores than did the WISC-R for African-American and White children, but not for Latinos. In fact, for African-Americans no significant correlations were found between WISC-R scores and CTBS Total Language. Raven performance correlated more highly with CTBS Total Math scores than did WISC-R only for White children. For all three ethnic backgrounds, WISC-R VIQ scores were more highly correlated with CTBS Total Reading scores than were Raven scores.

Table 10. *Correlation(r) of Raven or WISC-R performance with achievement test (CTBS) performance for children of three ethnic backgrounds*

		Subtest		
		CTBSTL	CTBSTR	CTBSTM
Whites				
RAVEN	(n = 901)	.1907**	.1727**	.2187*
VIQ	(n = 1566)	.1638**	.2479**	.1836**
PIQ	(n = 1566)	.1081**	.1274**	.1648**
FSIQ	(n = 1566)	.1623**	.2292**	.2069**
African-Americans				
RAVEN	(n = 276)	.2372**	.1916**	.1257*
VIQ	(n = 221)	.0691	.2740**	.1440
PIQ	(n = 221)	.0184	.1126	.1771**
FSIQ	(n = 221)	.0541	.2406**	.2002**
Latinos/Hispanics				
RAVEN	(n = 530)	.1563**	.0797*	.1327**
VIQ	(n = 138)	.1899*	.1365	.3150**
PIQ	(n = 138)	- .0476	- .0196	.1708*
FSIQ	(n = 138)	.0761	.0725	.2718**

* $p < .05$
** $p < .01$

Correlations between the Raven and the Abbreviated Stanford Achievement Test (ASAT) for children of six ethnic backgrounds are summarized in Table 11. Correlation coefficients ranged from .2235 to .3109, with a median $r = .25$. Clearly, the Raven correlates more highly with the ASAT than with the CTBS in all three domains measured (language, reading, and math). In addition, Raven—ASAT correlations were higher than WISC-R—CTBS correlations.

Table 11. *Correlation (r) of Raven performance with achievement as measured by the Abbreviated Stanford Achievement Test for children of six ethnic backgrounds (Raven performance was expressed in Z scores.)*

Ethnicity	*n*	Total Language	Total Reading	Total Math
African-American	1581	.2478**	.2639**	.2695**
Asian	305	.2386**	.2654**	.2634**
White	4020	.2245**	.2252**	.2494**
Filipino	1002	.2235**	.2438**	.2429**
Indochinese	587	.2582**	.2484**	.2584**
Latino	2528	.2468**	.2981**	.3109**

** $p < .01$

Discussion

Our results lead to two clear conclusions. First, considering only children referred for giftedness testing, the RPM produces far better equity for all ethnic backgrounds when compared to the WISC-R. While the RPM overselected Whites, it did so at a substantially reduced rate (i.e., 120 percent over expectation with the Raven vs. 200 percent over expectation with the WISC-R). Moreover, while it did not produce complete equity, even when considering such risk factors as low socioeconomic or cultural differences, the RPM led to substantially increased selection ratios for traditionally underrepresented groups such as Latino/Hispanics and African-Americans. Moreover, the RPM did lead to an equitable selection for Native-Americans, Pacific Islanders, and Indochinese, all of whom had been underrepresented with the WISC-R.

The success of the RPM with Indochinese is of interest in terms of evaluating non-English speaking children. In the past the district had difficulties evaluating giftedness in Indochinese children who spoke little or no English. Since use of the RPM enabled evaluation of ability independently of language, it was possible to assess these children with the least bias heretofore achieved. With the RPM Indochinese were selected almost exactly in proportion to their numbers in the district as a whole.

The advantage of the RPM was not only evident in terms of producing a more equitable distribution, but it was more effective than the WISC-R in predicting language achievement for African-American and White children, as well as in predicting math achievement for Whites. Although our data cannot be used to directly compare the WISC-R with the Raven for predicting scores on the ASAT, RPM scores were more highly correlated with ASAT scores than with CTBS scores for all groups. However, the WISC-R correlated more highly than the RPM with CTBS Total Reading for all groups and with Total Math for African-Americans and Latinos. It would be instructive to directly compare WISC-R — ASAT and Raven — ASAT correlations in the same sample to more completely compare efficacy for predicting achievement.

The Raven's ability to predict achievement, even language achievement, is due to its high correlation with Spearman's (1904, 1927a, 1927b) *g* factor (see Carpenter et al., 1990; Marshalek et al., 1983; Snow et al., 1984). Since tests of language achievement are highly correlated with *g*, *g*-saturated tests such as the RPM share common variance with them. Thus, the RPM can have clear advantages for measuring abilities for individuals who speak a language other than English or are from a different

culture (Court, 1991). Moreover, since it does not depend on an explicit knowledge base, as does the WISC-R and other verbally weighted standardized tests, the RPM is better suited to traditionally underrepresented children. It must be emphasized, however, that the RPM is not simply for underrepresented children, as it led to a more equitable distribution across ethnicity, even for Whites.

Previous studies that have compared the WISC-R and the Raven in the selection of gifted children (e.g., James, 1984; Pearce, 1983; Tulkin & Newbrough, 1968; Meeker, 1973; Kier, 1949) have been primarily concerned with the correlation between the two measures. Such studies have reported correlations in the .70's and have been generally supportive of the Raven as an alternative. Our findings suggest that where equity is a concern, the Raven is a far better measure.

It should be noted, however, that the population of gifted children selected by the Raven is not identical to that selected by the WISC-R. The differences are important. First, there is the group of children who would be selected by either test. These children tend to be verbally advanced, highly intelligent, and high in achievement. Second, there is the group of children who would be qualified by a WISC-R, but not by a Raven. In our experience, such children tend to be verbally advanced and highly motivated. Teachers of academic gifted programs readily accept these children as gifted, especially in the early years. Due to the unreliability of IQ scores at the upper IQ levels, especially for the younger age levels, the IQ's of many of these verbally advanced children show regression to the mean as they mature.

A third group of children are those who would not qualify with the WISC-R, but would with the Raven. Such children are of extremely high potential, but may be only average (or even below average) in achievement. Others may be of very high ability, but poorly motivated. Our experience has revealed that this type of student is not always accepted by teachers as gifted. Yet, it is this very type of child — the one with raw, undeveloped potential — for whom the present investigation was aimed. For programs using the Raven on a widescale basis, teacher training is often needed to help integrate these students into the gifted classroom.

One of the reasons tests like the WISC-R continue to be used, in spite of their obvious selection bias, is that they are reliable and objective. To date, no other approach matches standardized tests in terms of reliability, predictive validity, and objectivity. Yet, given the huge selection bias inherent in the WISC-R, it is difficult to imagine how it, and others like it, can continue to be used in today's litigious environment and heightened awareness of civil rights. A simple examination of Figure 1 would find the WISC-R completely unacceptable as a tool to select for giftedness.

The Raven clearly fared better in terms of equity without sacrificing objectivity, reliability, and predictive validity. Nevertheless, it is also clear that the Raven, even when used in conjunction with a consideration of risk factors, still falls short of producing equity across all ethnic groups.

Based on our present findings, and on previous reviews of psychometric tests (Kaplan & Saccuzzo, 1993), we are convinced that as traditional tests are presently used, there exists not a single one that would produce an equitable selection for gifted programs. If the goal is to select children for an academic program using an objective, reliable measure with high predictive validity, then traditional tests, as they are presently used, fall short in terms of equity. The search for multiple intelligences, as suggested by Gardner (1983), is, of course, one viable option. The use of portfolios and other subjective approaches, however, while promising on a small scale, faces numerous obstacles in terms of objectivity, reliability, and predictive validity.

The question remains, are traditional tests beyond redemption? We believe that there remain potentially promising options. One such option, suggested by Raven (1989), is the use of local ethnic norms for selection purposes in gifted programs. A second, suggested by Carlson (1989) and Colleagues (Carlson & Dillon, 1978; Carlson & Wiedl, 1978), is to use traditional tests in creative ways. For example, using a dynamic testing approach with the Raven, Carlson and Wiedl (1979) were able to eliminate Hispanic/White and Black/White differences in IQ. Before we abandon what remains the most objective, reliable, and valid approach to selection, it behooves us to determine whether innovative uses can rescue traditional assessment procedures, or if they should be abandoned in favor of less psychometrically sound, but perhaps more equitable, approaches.