Gifted Child Quarterly
http://gcq.sagepub.com/

The Gifted Rating Scales-School Form : An Analysis of the Standardization Sample Based on Age, Gender, Race, and Diagnostic Efficiency

Steven I. Pfeiffer and Tania Jarosewich
Gifted Child Quarterly 2007 51: 39
DOI: 10.1177/0016986206296658

Gifted Child Quarterly
Volume 51 Number 1
Winter 2007 39-50

10.1177/0016986206296658
http://gcq.sagepub.com
hosted at
http://online.sagepub.com

The Gifted Rating Scales-School Form

An Analysis of the Standardization Sample Based on Age, Gender, Race, and Diagnostic Efficiency

Steven I. Pfeiffer
Florida State University

Tania Jarosewich
Censeo Group LLC

Abstract: This study analyzes the standardization sample of a new teacher rating scale designed to assist in the identification of gifted students. The Gifted Rating Scales–School Form (GRS-S) is based on a multidimensional model of giftedness. Results indicate no age or race/ethnicity differences on any of the scales and small but significant differences in favor of females on three of the six scales: Artistic Ability, Motivation, and Leadership Ability. Diagnostic efficiency statistics and receiver operating curve analysis support the validity of the Intellectual Ability scale in identifying intellectually gifted students. The Intellectual Ability scale was successful in both correctly identifying students with high IQ scores (test sensitivity) and correctly identifying students without high IQ scores (test specificity). The present findings extend the analysis of the standardization sample reported in the test manual and provide additional support for the psychometric qualities of the GRS-S as a valid gifted screening tool.

Putting the Research to Use: A number of issues compromise the ability to identify gifted students. Definitions of *gifted* and *talented* are inconsistent across states. Experts disagree on how to conceptualize and define giftedness. In addition, there are few technically adequate screening instruments to assist in the identification of gifted students. The present study reports on an analysis of the standardization sample of a new teacher rating scale designed to assist in the identification of gifted students. The Gifted Rating Scale–School Form (GRS-S) is a 72-item teacher rating scale based on a multidimensional model of giftedness. In this study, we have shown that the GRS-S is a technically sound screening instrument that works equally well with African American, Asian American, Hispanic, and White children across the entire age range: 6.0 to 13.11 years. Small but significant differences in favor of females were found on three of the six GRS-S scales: Artistic Ability, Motivation, and Leadership Ability. However, gender differences were not found on the Intellectual Ability, Academic Ability, and Creativity scales of the GRS-S. The GRS-S holds promise for practitioners in helping to identify gifted students. The GRS-S also holds promise for researchers in providing a technically sound instrument to measure multiple manifestations of giftedness. Policy makers will welcome the GRS-S as an easy-to-use, score, and interpret test with strong face validity that provides a level playing field and strong face validity for students from diverse ethnic and racial backgrounds.

Keywords: *gifted rating scales; gifted identification; gifted screening*

Many argue that American society today does not place a high priority on educating its most talented young citizens, even though they will be tomorrow's leaders. The No Child Left Behind Act (2002) focuses attention and resources on how to reach our most poorly educated and those students who are lagging behind. However, there is considerably less interest in America's brightest and most able (Borland, 1996; Gallagher, 2003; Pfeiffer, 2002). For example,

Authors' Note: Please address correspondence to Steven I. Pfeiffer, Department of Educational Psychology & Learning Systems, Florida State University, 306 Stone Building, Tallahassee, FL 32306; e-mail: pfeiffer@coe.fsu.edu.

99.9% of special education funding goes to the lower end of the ability continuum (Sternberg, 1996), and only two cents out of every $100 allocated for education is directed to gifted students (Winner, 1997).

There are a growing number of leaders in American society who are beginning to speak out in support of addressing the unmet needs of the gifted. The positive psychology movement has helped to bring attention to the unique needs of America's most talented (Pfeiffer, 2001; Seligman, 1998; Seligman & Czikszentmihalyi, 2000).

At the same time, many public schools—both in the United States and abroad—remain ill equipped to meet the needs of gifted and talented students. Too few educators have the knowledge, experience, time, or resources to design effective programs that meet the needs of the gifted. One important first step in serving the gifted is accurately and efficiently identifying gifted students. A recent survey of gifted experts reported in this journal highlighted the identification process as the second most frequently cited issue facing the field. Forty-one percent of the 64 authorities agreed that identification of the gifted remains problematic (Pfeiffer, 2003).

One of the problems is that the gifted field has too few technically sound screening instruments. The ubiquitous IQ test is almost routinely used—irrespective of the particular cut score that a school district or state adopts for inclusion/exclusion—to determine whether a student qualifies for gifted placement. There are few screening tools available to complement the IQ test in providing a more comprehensive picture of the student's abilities. A recently published article reviewed three of the more popular teacher rating scales designed to identify gifted students (Jarosewich, Pfeiffer, & Morris, 2002). The investigators initially identified 31 rating scales and then narrowed down their list to the 3 most widely used instruments that are currently available and employ the teacher as informant. The three scales reviewed were the Scales for Rating the Behavioral Characteristics of Superior Students (Renzulli et al., 1997), the Gifted and Talented Evaluation Scales (Gilliam, Carpenter, & Christensen, 1996), and the Gifted Evaluation Scale (2nd ed.; McCarney & Anderson, 1989). Although the review noted that each of the three rating scales had unique qualities, the article cautioned that all three scales had serious technical shortcomings that limited their diagnostic usefulness. Specific concerns included nonrepresentative standardization normative samples, low interrater reliability, and lack of evidence for diagnostic accuracy (Jarosewich et al., 2002).

Development of a New Gifted Rating Scale (GRS)

Recognizing that hundreds of thousands of school-age children in the United States and worldwide are tested annually for gifted consideration and that the gifted field did not have a technically adequate screening tool to assist in the identification of gifted students, we undertook to develop a new gifted screening instrument, the GRS (Pfeiffer & Jarosewich, 2003). Six principles guided the development of the GRS. The screening instrument was designed to be user friendly, requiring minimal training to administer, score, and interpret. It was developed to be scientifically sound, reliable, and valid. It included a standardization sample that matched the latest U.S. census in terms of race/ethnicity, parent education level, and regional representation. It was based on a multi-abilities conceptualization of giftedness and a straightforward interpretive model that simplified the screening of gifted children. It was intended to be a clinically flexible tool that could complement an IQ test and other procedures (e.g., auditions, portfolio samples, nonverbal tests) as part of a comprehensive test battery. It was linked to the new Wechsler Intelligence Scale for Children– Fourth Edition (WISC-IV) and Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III), which was accomplished by colinking the standardization of the GRS with the standardization of the new WISC-IV and WPPSI-III (Pfeiffer & Jarosewich, 2003).

The present study investigated the effect of age, gender, and race with the GRS-S standardization sample.[1] We explored whether possible differences exist on each of the GRS-S scales for gender, race, and age. The study also investigated the diagnostic validity of the GRS for the total sample of the standardization sample for which both GRS and WISC-IV scores were collected. Diagnostic efficiency statistics and a receiver operating curve (ROC) analysis were undertaken for all students in the standardization normative sample for which both GRS and WISC-IV data had been collected.

The analyses described in this study have not been reported elsewhere. The present analyses of the standardization sample are intended to extend the information reported in the user manual.

Method

Participants

Data used in the present study were obtained from the GRS standardization sample. The full standardization sample was used for the MANOVA analyses. This sample consisted of 291 boys (49%) and 301 girls (51%). The age group of the sample was stratified within eight 12-month age bands from 6.00 to 13.11, with each age composing 13% of the standardization population.

During standardization sampling, the test publisher, Harcourt Assessment, intentionally stratified the standardization sample to closely approximate the U.S. population on important demographic characteristics, such as race/ethnicity, parent education level, and regional representation (U.S. Bureau of Census, 2000). For, example, 64% of the sample was Caucasian (*n* = 379), 16% African American (*n* = 96), 16% Hispanic (*n* = 94), and 4% (*n* = 23) Asian American. Tables 4.4 to 4.6 in the GRS test manual report data on the race/ethnicity, parent education level, and regional representation of the sample stratified within the eight age bands (Pfeiffer & Jarosewich, 2003, pp. 26-27).

We used a subsample of the standardization data that included all subjects with both GRS and WISC-IV data for the analysis of efficiency statistics and ROC analysis. This subsample consisted of 196 boys (51%) and 185 girls (49%). Sixty-seven percent of the sample was Caucasian (*n* = 254), 17% African American (*n* = 63), 10% Hispanic (*n* = 39), 5% Asian (*n* = 20), less than 1% Native American (*n* = 3), and less than 1% Other (*n* = 3). The age group of the sample ranged from 6.00 to 13.11, with each age composing approximately 13% of the standardization population; this subsample was similar to the total standardization group and similarly closely matched the U.S. population on race/ethnicity, parent education level, and geographic representation (U.S. Bureau of Census, 2000).

Instrument

The GRS (Pfeiffer & Jarosewich, 2003) includes a Preschool/Kindergarten Form (GRS-P) for ages 4.0 to 6.11 and a School Form (GRS-S) for ages 6.0 to 13.11. The GRS-P consists of five scales with 12 items each, for a total of 60 items; the GRS-S consists of six scales with 12 items each, for a total of 72 items. The items of the GRS-P represent skills and behaviors developmentally appropriate for preschool and kindergarten students, whereas the items of the GRS-S reflect more developmentally advanced skills or behaviors. The GRS-S includes a sixth scale, Leadership, which is not included in the GRS-P. Both forms yield raw score totals on all scales, which are converted to age-based *T* scores and associated cumulative percentages. This article focuses exclusively on the GRS-S.

The GRS is based on a multidimensional model of giftedness that incorporates the Munich Model of Giftedness and Talent (Zigler & Heller, 2000) and the typology that appears in the U.S. Department of Education Report, *National Excellence: A Case for Developing America's Talent* (Ross, 1993). Below is a brief description of each of the six GRS-S scales.

- *Intellectual Ability*. This scale measures the student's verbal and/or nonverbal mental skills, capabilities, or intellectual competence. Items on this scale rate a student's abstract reasoning, problem solving, mental speed, and memory.
- *Academic Ability*. This scale measures the student's skill in dealing with factual and/or school-related material. Items rate advanced competence and high levels of proficiency in reading, math, and other aspects of the school curriculum.
- *Creativity*. This scale measures the student's ability to think, act, and/or produce unique, original, novel, or innovative thoughts or products. Items rate how a student solves problems, experiments with new ideas, formulates a solution to a group project, and/or uses imagination.
- *Artistic Talent*. This scale measures the student's potential for or evidence of ability in drama, music, dance, drawing, painting, sculpture, singing, playing a musical instrument, and/or acting. Items rate how a student approaches activities, completes assignments, and/or uses art supplies or artistic media.
- *Leadership Ability*. This scale measures the student's ability to motivate others toward a common or shared goal. Items rate a student's conflict resolution skills, initiative in group situations, and understanding of social dynamics and interpersonal communication.

- *Motivation*. This scale refers to the student's drive or persistence, desire to succeed, tendency to enjoy challenging tasks, and ability to work well without encouragement or reinforcement. The motivation scale is not viewed as a type of giftedness but rather as the dynamic energy that drives or impels a student to achieve.

Each item is rated by a teacher on a 9-point scale divided into three ranges: 1 to 3 = *below average*, 4 to 6 = *average*, and 7 to 9 = *above average*. The GRS-S manual provides a classification system that indicates not whether a student is gifted but rather the likelihood that a student is gifted, based on their *T* score. The higher the student's *T* score on one or more of the gifted scales, the higher the probability that they are, in fact, gifted compared to their same-age peers. The *T* scores were computed based on each age group and thus were age adjusted so that the classificatory ranges may be applied across age bands. A *T* score of less than 55 (less than 69%) indicates a low probability of gifted, a score between 55 and 59 (69% to 83%) indicates a moderate probability, a score between 60 and 69 (84% to 97%) indicates a high probability, and a score higher than 70 (98+%) indicates a very high probability.

Test development followed a carefully prescribed set of steps, including a survey of gifted experts, focus groups, and pilot and field testing. As mentioned earlier, standardization was colinked with standardization of the new WISC-IV (and WPPSI-III in the case of the GRS-P). Final item selection was guided by factor structure, item mean scores, bias (parent education level, gender, and ethnicity), and interrater and test-retest reliability. The GRS test manual reports evidence of high reliability and validity. Based on the standardization sample, coefficient alpha reliabilities ranged from .97 to .99 and standard error of measurements ranged from 1.0 to 1.41 across the six scales and eight age ranges. Test-retest reliability coefficients, based on a sample of 160 students ages 6.0 to 13.11 and a median retest interval of seven days, ranged from .83 on the Artistic Talent scale to .97 on the Academic Ability scale. Interrater reliability, based on 152 students, ranged from .70 to .79 for students ages 6.0 to 9.11 and .64 to .75 for students ages 10.0 to 13.11. The test manual also provides evidence in support of internal structure and convergent and divergent validity (Pfeiffer & Jarosewich, 2003). The reader interested in more detailed information on the reliability, validity, and normative data is directed to a recent article that reviewed the GRS (Margulies & Floyd, 2004).

Procedure

The test publisher, Harcourt Assessment, provided the authors with a data file that included the data for the entire standardization sample. The subsample consisted of the entire standardization sample for which both GRS-S and WISC-IV data were collected.

Three sets of statistical analyses were conducted. First, using the full standardization sample data, a separate multivariate analysis of variance (MANOVA) was conducted to test for differences on each of the six GRS-S scales for gender, race, and age. Given the small sample sizes in some of the cells, a single MANOVA to test for the three independent variables could not be conducted. The standardization sample included three Native American students and five students whose race was identified as Other. Because of the small *n*, data for these students were not included in these analyses. The follow-up of significant effects was conducted using univariate ANOVAs (Newton & Rudestam, 1999).

The second set of analyses, using the subsample of students who received both the GRS-S and WISC-IV, calculated diagnostic efficiency statistics for GRS-S Intellectual Ability *T* cut scores of 55 and 60 and WISC-IV IQ Full Scale (FS) scores of 115 and 130. GRS-S *T* scores of 55 and 60 were used because the GRS classification system developed by the authors proposes that a *T* score between 55 and 59 (69% and 83%) indicates a moderate probability of giftedness and a *T* score between 60 and 69 (84% and 97%) indicates a high probability of gifted (Pfeiffer & Jarosewich, 2003). We set the WISC-IV FS IQ at 115 and 130 because most school systems presently use the 130 IQ (2 *SD* above the mean) for operationally defining intellectually gifted, and some authorities (e.g., Renzulli, 1978) suggest using a less stringent IQ score of 1 *SD* above the mean for identifying gifted students. We computed five diagnostic efficiency statistics. The sensitivity (SE) of a test, also called the true positive rate, is the proportion of people who have the attribute (in this case, intellectual giftedness operationally defined as IQ $\geq$ 130) who are correctly detected by the test (in this instance, the Intellectual Ability scale of the GRS). The specificity (SP) of a test is the proportion of people without the attribute (i.e. not intellectually gifted) who are correctly labeled by the Intellectual Ability scale of the *GRS* as not gifted. One can combine the SE and SP into a single index called the likelihood ratio (LR+), which is defined as follows:

Table 1
Correlation Coefficients Among GRS-S Scale Scores

	Intellectual	Academic	Creativity	Artistic	Leadership	Motivation
Intellectual	1.00	0.936**	0.829**	0.620**	0.673**	0.795**
Academic		1.00	0.815**	0.651**	0.710**	0.842**
Creativity			1.00	0.722**	0.680**	0.720**
Artistic				1.00	0.580**	0.672**
Leadership					1.00	0.800**
Motivation						1.00

Note: GRS-S = Gifted Rating Scales–School Form.
**$p \leq .01$ (two-tailed)

$$\text{LR+} = \text{SE} = \text{True Positive Rate} \\ = 1 - \text{SP} = \text{False Positive Rate}$$

The LR+ is an index of the accuracy of a test and depicts what the odds are that a positive test result comes from a person who has the attribute. When the LR+ is 1, the test is useless and does not contribute to making an accurate classification (Streiner, 2003). In addition to the LR+, there is an equivalent formula for a negative test result:

$$\text{LR–} = \text{SP} = \text{True Negative Rate} \\ = 1 - \text{SE} = \text{False Negative Rate}$$

The overall correct classification (OCC) is (the True Positive + the True Negative)/*N*. The OCC incorporates the SE and SP of a test into an overall index of diagnostic accuracy. The five diagnostic statistics—SE, SP, LR+, LR–, and OCC—are not affected by the prevalence or base rate of the condition in the population. Diagnostic efficiency statistics such as positive predictive power, negative predictive power, and kappa, on the other hand, are affected by prevalence (Meehl & Rosen, 1955; Streiner, 2003). Because giftedness is by definition a low prevalence phenomenon and because only 2% of our subsample obtained WISC-IV IQ ≥ 130, the four diagnostic efficiency statistics were excellent choices for evaluating the diagnostic efficiency of the GRS (D. L. Streiner, personal communication, May 24, 2005).[2]

We also conducted a ROC analysis, which graphically depicts the diagnostic efficiency across the entire range of standard scores. The ROC analysis graphs the SE and 1-SP associated with each possible cut score on the test. The area under curve (AUC) value describes the overall fit of classification, with a score of 1.0 representing a perfect fit, a score of 0.5 chance-level prediction, and a score of less than 0.5 below-chance prediction. The AUC, like SE, SP, and the LR+, is independent of the prevalence or base rate. Two additional advantages of the ROC analysis is that it is independent of test cut scores used in making classifications and an excellent way to communicate effect size to audiences (Hsu, 2002).

Results

Preliminary diagnostics indicated that multivariate normality and homogeneity of variance assumptions were met. For all significant findings, η^2 was calculated to determine effect size. An effect size of $\eta^2 = .01$ was defined as a small effect, $\eta^2 = .06$ as a medium effect, and $\eta^2 = .14$ as a large effect (Sprinthall, 2000).

Correlations between the GRS-S scales are presented in Table 1. The highest obtained correlation coefficient among GRS-S scale scores was .936, between Intellectual Ability and Academic Ability. The three lowest obtained correlation coefficients all included the Artistic Ability scale (.580 with Leadership, .620 with Intellectual Ability, and .651 with Academic Ability). All correlation coefficients were significant at the $p \leq .01$ level.

Analysis of GRS-S Standardization Sample by Gender

The MANOVA corresponding to gender yielded a significant result, $F(6, 585) = 12.26$, $p \leq .001$ (Wilks' Lambda = 0.89), with an effect size of $\eta^2 = .11$. Descriptive statistics for each GRS-S scale for gender are presented in Table 2. The scale scores for girls were significantly higher on the Artistic Talent scale, $F(1, 590) = 24.53$, Mse = 2,425.47, $p \leq .001$, $\eta^2 = .04$, Motivation scale, $F(1, 590) = 17.85$, Mse = 1,818.13, $p \leq .001$, $\eta^2 = .03$, and Leadership scale,

Table 2
Mean Scores and Standard Deviations for GRS-S Scale Scores by Gender

	Girls (n = 301)		Boys (n = 291)	
	M	*SD*	*M*	*SD*
Intellectual	50.13	10.24	49.99	10.43
Academic	51.23	9.88	50.65	10.01
Creativity	50.70	10.00	49.98	9.54
Artistic	52.28	10.16	48.23	9.72
Leadership	51.67	9.68	49.15	9.81
Motivation	52.59	9.99	49.09	10.20

Note: GRS-S = Gifted Rating Scales–School Form.

Table 3
Mean Scores and Standard Deviations for GRS-S Scale Scores by Race

	Asian American (n = 23)		African American (n = 96)		Caucasian (n = 379)		Hispanic (n = 94)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Intellectual	51.39	9.21	48.06	10.21	50.99	10.47	48.01	9.65
Academic	52.87	9.41	48.86	10.62	51.82	9.93	49.09	8.88
Creativity	50.04	7.61	48.55	10.54	51.28	9.74	48.47	9.15
Artistic	50.70	9.52	48.33	10.87	51.02	10.15	49.22	9.26
Leadership	52.91	10.40	49.16	10.52	50.96	9.99	49.03	7.85
Motivation	53.17	10.07	48.93	10.36	51.55	10.41	49.55	9.12

Note: GRS-S = Gifted Rating Scales–School Form.

$F(1, 590) = 9.87$, Mse = 936.95, $p \leq .01$, $\eta^2 = .02$. The mean score for girls on the Artistic Talent scale was $M = 52.28$ and for boys was $M = 48.23$, yielding a 4-point difference in favor of females. The girls' mean scores for the Motivation scale was $M = 52.59$ and boys' $M = 49.09$ (a 3.5 point difference), whereas the girls' mean score for the Leadership scale was $M = 51.67$ and the boys' $M = 49.15$ (a 2.5 point difference).

Analysis of GRS-S Standardization Sample by Race

The MANOVA comparing GRS-S scales based on race did not yield significant results at the .01 level, $F(18, 1{,}649) = 1.01$, $p = .443$ (Wilks' Lambda = 0.97), with an effect size of $\eta^2 = .01$. Table 3 presents means and standard deviations based on race. Although not statistically significant at the .01 level, the trend was consistently in favor of Asian American and White students obtaining slightly higher GRS ratings than African American and Hispanic children. However, the differences by race/ethnicity were, at most, modest. For example, the mean scale score for Asian American students was 3.3 points higher than the mean scale score for African American and Hispanic students on the Intellectual Ability scale score (approximately 1/3 *SD* higher).

Analysis of GRS-S Standardization Sample by Age

The MANOVA corresponding to age group did not yield significant results, $F(42, 2{,}719) = 1.11$, $p = .28$ (Wilks' Lambda = .92). Mean scores and standard deviations for GRS-S scale scores by age are presented in Table 4. GRS means scores did not vary across the almost 8-year age span of 6 years, 0 months, through 13 years, 11 months.

Analysis of Diagnostic Validity

Diagnostic efficiency values. Table 5 provides a summary of the five diagnostic efficiency values (i.e., SE, SP, OCC, LR+, and LR–) for GRS-S Intellectual Ability *T* scores of 55 and 60. WISC-IV FS IQ scores

Table 4
Mean Scores and Standard Deviations for GRS-S Scale Scores by Age Group

	6.00 to 6.11 (n = 75)		7.00 to 7.11 (n = 74)		8.00 to 8.11 (n = 74)		9.00 to 9.11 (n = 74)		10.00 to 10.11 (n = 74)		11.00 to 11.11 (n = 73)		12.00 to 12.11 (n = 74)		13.00 to 13.11 (n = 74)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Intellectual	49.55	10.19	49.82	11.42	49.92	10.45	50.15	9.36	50.15	9.67	50.34	10.35	50.12	10.30	50.31	11.18
Academic	50.39	10.56	50.86	11.57	50.70	10.06	51.55	9.02	51.46	9.47	51.53	9.42	50.88	9.18	50.19	10.36
Creativity	50.05	10.38	49.96	10.97	51.03	9.73	50.95	8.48	49.57	9.68	50.79	10.48	50.27	8.69	50.15	9.91
Artistic	50.15	10.25	49.34	10.73	49.93	11.52	50.99	8.52	50.26	8.59	50.60	10.75	50.16	9.24	50.89	11.42
Leadership	50.92	10.02	50.86	11.63	50.45	8.63	51.47	8.56	51.92	9.34	50.23	11.45	48.50	9.43	49.11	8.98
Motivation	50.41	9.54	50.80	11.76	50.15	9.85	50.80	9.64	53.04	9.60	50.32	10.73	51.00	9.61	50.43	11.12

Note: GRS-S = Gifted Rating Scales–School Form.

Table 5
Results of Diagnostic Efficiency Statistics

Category	Sensitivity	Specificity	Overall Correct Classification	LR+	LR–
WISC-IV FS IQ ≥ 130					
GRS Intelligence *T* score 60	0.86	0.82	0.82	4.78	5.86
GRS Intelligence *T* score 55	1.00	0.66	0.66	5.56	≥ 1.00
WISC-IV FS IQ ≥ 115					
GRS Intelligence *T* score 60	0.50	0.86	0.80	3.58	0.16
GRS Intelligence *T* score 55	1.00	0.66	0.72	2.94	≥ 1.00

Note: WISC-IV = Wechsler Intelligence Scale for Children–Fourth Edition; FS = full scale; GRS = Gifted Rating Scale; LR = likelihood ratio.

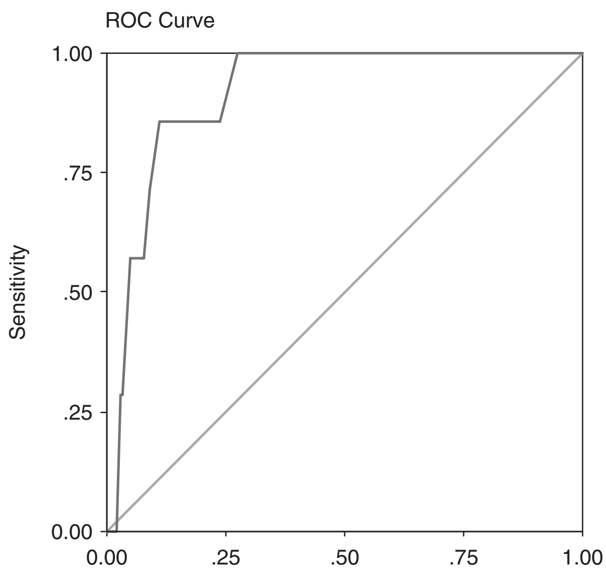
were set at 115 and 130, as mentioned earlier, because the great majority of school systems presently use the 130 IQ score (2 SD above the mean) for operationally defining intellectually gifted, whereas some authorities (e.g., Renzulli, 1978) suggest using the less conservative IQ score of 115. The AUC, which will be reported shortly, provides additional information on diagnostic validity independent of cut score.

Irrespective of whether the WISC-IV FS IQ score was set at 115 or 130, the GRS-S Intellectual Ability *T* score of 60 was more efficient than a *T* score of 55. The OCC rate was .82 versus .66, in favor of *T* = 60 when the WISC-IV FS IQ was set at 130, and .80 versus .72, in favor of *T* = 60 when the WISC-IV FS IQ was set at 115.

Recall that the SE of a test, also called the true positive rate, is in this instance the proportion of people who are intellectually gifted based on a WISC-IV IQ ≥ 130 and who are correctly identified by the GRS Intellectual Ability scale. The SP of a test is the proportion of people who are not intellectually gifted (again, based on IQ < 130) and who are correctly identified by the GRS Intellectual Ability scale as not gifted. Using a *T* score of 60 and with WISC-IV IQ ≥ 130, the SE was .86 and the SP was .82 for the GRS Intellectual Ability scale. Lowering the *T* score to 55 (with the WISC-IV IQ ≥ 130), the test SE was 1.0, but the test SP dropped to 0.66. Similar test SE and test SP resulted with GRS cut scores of 60 and 55 and WISC-IV IQ ≥ 115 (see Table 5).

The LR+ is another index of the accuracy of the test unaffected by prevalence or base rates. When the LR+ = 1, the test is no better than chance in helping to make a diagnostic or classification decision. Using a *T* score of 60 and with the WISC-IV IQ ≥ 130, the LR+ = 4.78. This means that a obtaining a *T* score of 60 or above on

Figure 1
ROC Curve



Note: ROC = receiver operating curve.

the GRS Intellectual Ability scale is almost 5 times as likely for those students who have IQ ≥ 130 as for those who do not. With a *T* score of 60 and WISC-IV IQ ≥ 130, the LR– was 5.86. In other words, obtaining a *T* score below 60 on the GRS Intellectual Ability scale is almost 6 times as likely to have come from a student who does not have IQ ≥ 130.

The LR+ was well above chance but not as strong with the WISC-IV IQ set at 115. The LR+ was 3.58 and 2.94 for cut scores of 60 and 55, respectfully. When the test SE is too high, however, relative to the SP, as was the case for IQs at both 115 and 130 and a

T score of 55, the LR– did not exceed chance. The LR– was also nonsignificant for a *T* score of 60 and IQ ≥115.

ROC values. The ROC analysis for the GRS-S Intellectual Ability scale performed significantly above chance, with an estimated AUC of 0.92 (Standard Error of Measurement = 0.03, $p < .01$) and a 95% confidence interval of .86 to 0.98. An AUC of 1.0 would represent a perfect fit and an AUC of 0.5 chance level prediction. Recall that the AUC is a type of effect size (Hsu, 2002). As Figure 1 depicts, the ROC falls far above the diagonal line, which represents chance level prediction, indicating that the GRS-S Intellectual Ability scale works exceptionally well as a diagnostic screening index for intellectual giftedness across all cut scores.

Discussion

The present study investigated the possible effect of age, gender, and race with the GRS-S standardization sample. The study also explored the diagnostic validity of the GRS-S for the portion of the standardization sample for which both GRS and WISC-IV data were collected during the norming of both instruments. The findings reported in this article have not appeared elsewhere and are intended to extend the information available in the GRS test manual.

GRS-S mean scores differed significantly by gender for three of the six scales in the standardization sample. Although the gender differences for the three scales were modest (i.e., a mean point difference of 2.5 points for Leadership Ability, 3.5 points for Motivation, and 4 points for Artistic Talent), they are nonetheless noteworthy, particularly because girls' mean scores were somewhat higher than boys' mean scores for all six GRS-S scales. Nationwide standardization sampling of the GRS-S and WISC-IV followed a carefully prescribed and rigorous set of norming procedures. It is unlikely that the GRS-S standardization sample is unrepresentative or biased in a way that might explain the small yet consistent gender differences. A more likely explanation is that teachers, who serve as raters with the GRS, tend to perceive female students in their classes as slightly more motivated and talented artistically and as displaying somewhat stronger leadership ability, when compared to their male counterparts. These findings of modest gender differences on three of the scales in favor of females apply to students in Grades 1 through 8, who constituted the standardization sample. The GRS-S was not designed or standardized for older students in Grades 9 to 12. It is important to reiterate that although the gender differences were statistically significant, they were quite small. Research indicates that gifted girls outperform gifted boys in classroom achievement throughout the school years, maintaining higher grades in all subjects (Kerr, 1997). However, adolescence appears to present subtle yet pernicious sociocultural influences that moderate gifted female achievement (Kerr & Nicpon, 2003). Interesting enough, girls did not obtain higher mean scores than boys on either the Intellectual Ability or the Academic Ability scales of the GRS-S. Future research may want to follow up on the present findings by examining the possible interaction of gender-by-age and whether teacher gender differentially influences the ratings of male and female students.

Whereas we identified small but significant gender differences on three of the six GRS-S scales, we did not find any significant race/ethnicity or age differences on any of the GRS-S scales for the standardization sample. Asian American and White students in the standardization sample obtained slightly higher mean scale scores when compared to African American and Hispanic students. However, these differences were in all instances quite small and did not reach significance. This is an important and favorable finding, especially as the gifted field has been concerned with the underrepresentation of African American, Hispanic, and Native American students in gifted education programs (Ford, 1998, in press; Pfeiffer, 2002). Of course, every test is culturally loaded to some extent (Barona & Pfeiffer, 1992; Flanagan, McGrew, & Ortiz, 2000; Jensen, 1974). For example, picture vocabulary tests and portions of the Verbal scale of the WISC-IV and Stanford Binet are highly culturally loaded, whereas nonverbal matrix tests and digit span memory tests are less highly culturally loaded (Jensen, 1974, 2004; Naglieri & Ford, 2003; Sattler, 2001). The fact that the GRS-S appears to be less highly culturally loaded and works equally well across different racial/ethnic groups makes it an attractive gifted screening tool.

Mean scores did not differ by age group on the GRS-S standardization sample, another encouraging finding. Teachers, gifted educators, school psychologists, and other school personnel who help to identify gifted students can be reassured that the GRS-S works equally well across the age span of 6 years, 0 months, to 13 years, 11 months.

We were not surprised that the highest correlation between GRS-S scales was for Intellectual Ability and Academic Ability, long believed to represent a similar if not identical underlying factor. With a correlation of .936, these two scales have 87% shared variance. Also not surprising, in all instances the three lowest correlations between GRS-S scales included the Artistic Ability scale. For example, the Artistic Ability scale correlated .62 with the Intellectual Ability scale; these two scales have 38% shared variance. We look to future research that will incorporate large enough samples to further explicate the relationship between the GRS-S scales. Future GRS-S research may be able to help answer the related questions of whether we can reliably measure the multiple manifestations of giftedness and if one underlying *g* factor explains most of the reliable variance accounted for in a student's ratings.

The second purpose of the study was to examine the diagnostic validity of the GRS-S for that portion of the standardization sample for which both GRS-S and WISC-IV scores were collected during standardization sampling. We examined how successfully the GRS-S Intellectual Ability scale classified individuals as high IQ (as a proxy for intellectually gifted) or not high IQ (as a proxy for not intellectually gifted). Of course, one daunting challenge in undertaking any test validation study is the need to identify an established "gold standard" that meets the criteria for defining the construct of interest, in this instance, intellectual giftedness. We used the WISC-IV FS IQ as the criterion measure because it is almost universally accepted as the putative measure of intelligence within the gifted field.

Although infrequently reported, the validity of any instrument or screening procedure should include reporting the diagnostic performance of the test (Cohen, 1990; Kessel & Zimmerman, 1993; Robins, Schoff, Glutting, & Abelkop, 2003; Streiner, 2003). An analysis of the SE, SP, OCC, and LR+ support the validity of the GRS-S in identifying intellectually gifted students. The GRS-S was successful in both correctly identifying individuals who are intellectually gifted (test SE) and correctly identifying individuals who are not intellectually gifted (test SP), especially when intellectually gifted is operationally defined as intellectual ability in the top 5%. Using a *T* score of 60, as suggested in the test manual as demarcating a high probability of gifted, the OCC rate for intellectually gifted was .82 with a WISC-IV IQ of 130 and .80 with a WISC-IV IQ of 115.

Where exactly to set the cut score for any test is never a simple decision. It should take into account the purpose of the test (e.g., screening, classification, diagnosis) and the relative risk that the user is willing to accept in making Type I versus Type II errors. The GRS-S is designed as a screening tool to assist in the identification of gifted students. It was neither designed nor purported to be a stand-alone gifted instrument. As a screening tool, the fact that the GRS-S will miss correctly identifying very few truly intellectually gifted students is very good news. The fact that the GRS-S may overidentify as high probability gifted a substantial number of students who, on more in-depth assessment, are found not to be intellectually gifted—at least based on high IQ scores—is not troubling in the least, given the stated purpose of the GRS-S as a screening tool.

It will be important for subsequent GRS-S research with independent samples to cross-validate the present findings. New studies will want to extend the present investigation by validating the other GRS-S scales. This will not be an easy task, as establishing "gold standards" for the GRS-S scales Creativity, Leadership Ability, and Artistic Ability will require thoughtful ingenuity. The interested reader is directed to review the test manual, where preliminary validity evidence is provided in support of the GRS-S scales and external criteria (Margulies & Floyd, 2004). However, much more scale validation work is warranted. The present study does underscore, however, that the GRS-S holds potential as a new screening test that can assist in the identification of gifted students. It is particularly noteworthy that the GRS-S works equally well across racial/ethnic groups, holding out the promise that the GRS-S can play an important role in helping to identify typically underrepresented gifted minority children.

Notes

1. Appreciation is extended to PsychCorp for permission to use the GRS-S standardization sample data.

2. Thanks to David L. Steiner, PhD, C. Psych, Professor, Department of Psychiatry, University of Toronto, for his guidance regarding selection of diagnostic statistics.

References

Barona, A., & Pfeiffer, S. I. (1992). Effects of test administration procedures and acculturation level on achievement test scores. *Journal of Psychoeducational Assessment*, *10*, 124-132.

Borland, J. H. (1996). Gifted education and the threat of irrelevance. *Journal for the Education of the Gifted*, *19*, 129-147.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45,* 1304-1312.
Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.
Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education*, *32*, 4-14.
Ford, D. Y. (in press). Serving gifted minority group children. In S. I. Pfeiffer (Ed.), *Handbook of serving the gifted and talented: A psychological approach*. New York: Kluwer Academic/ Plenum Publishers.
Gallagher, J. J. (2003). Issues and challenges in the education of gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 11-23). Boston: Allyn & Bacon.
Gilliam, J. E., Carpenter, B. O., & Christensen, J. R. (1996). *Gifted and Talented Evaluation Scales*. Waco, TX: Prufrock.
Hsu, L. M. (2002). Diagnostic validity statistics and the MCMI-III. *Psychological Assessment*, *14*, 410-422.
Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment*, *20*, 322-336.
Jensen, A. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs*, *90*, 185-244.
Jensen, A. (2004). The mental chronometry of giftedness. In D. Boothe & J. L. Stanley (Eds.), *In the eyes of the beholder: Critical issues for diversity in gifted education* (pp. 157-166). Waco, TX: Prufrock.
Kerr, B. (1997). Developing talents in girls and young women. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 483-497). Boston: Allyn & Bacon.
Kerr, B., & Nicpon, M. F. (2003). Gender and giftedness. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 493-505). Boston: Allyn & Bacon.
Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problems, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment*, *5*, 395-399.
Margulies, A. S., & Floyd, R. G. (2004). Test review: The Gifted Rating Scales. *Journal of Psychoeducational Assessment*, *22*, 275-282.
McCarney, S. B., & Anderson, P. D. (1989). *Gifted Evaluation Scale, second edition. Technical manual*. Columbia, MO: Hawthorne Educational Services.
Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*, 194-216.
Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*, 155-160.
Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant: Answers to your data analysis questions*. Thousand Oaks, CA: Sagens.
No Child Left Behind Act. (2002). P.L. 107-110. Available at http://www.nochildleftbehind.gov/
Pfeiffer, S. I. (2001). Professional psychology and the gifted: Emerging practice opportunities. *Professional Psychology: Research & Practice*, *32*, 175-181.
Pfeiffer, S. I. (2002). Identifying gifted and talented students: Recurring issues and promising solutions. *Journal of Applied School Psychology*, *1*, 31-50.
Pfeiffer, S. I. (2003). Challenges and opportunities for students who are gifted: What the experts say. *Gifted Child Quarterly*, *47*, 161-169.
Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales*. San Antonio, TX: Harcourt Assessment.
Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, *60*, 180-184, 261.
Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. I. (1997). *Scales for rating the behavioral characteristics of superior students*. Mansfield Center, CT: Creative Learning Press.
Robins, P. M., Schoff, K. M., Glutting, J. J., & Abelkop, A. S. (2003). Discriminative validity of the Behavior Assessment System for Children-Parent Rating Scales in children with recurrent abdominal pain and matched controls. *Psychology in the Schools*, *40*, 145-154.
Ross, P. O. (1993). *National excellence: A case for developing America's talent*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Jerome M. Sattler Publisher.
Seligman, M. E. P. (1998). The gifted and the extraordinary. *APA Monitor*, *29*, 2.
Seligman, M. E. P., & Czikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5-14.
Sprinthall, R. C. (2000). *Basic statistical analysis*. Boston: Allyn & Bacon.
Sternberg, R. J. (1996). The sound of silence: A nation responds to its gifted. *Roeper Review*, *18*, 168-173.
Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, *81*, 209-219.
U.S. Bureau of the Census. (2000). *Current population survey, October 2000: School enrollment supplement file* [CD-ROM]. Washington, DC: Author.
Winner, E. (1997). Exceptionally high intelligence and schooling. *American Psychologist*, *52*, 1070-1081.
Zigler, A., & Heller, K. A. (2000). Conceptions of giftedness from a meta-theoretical perspective. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 3-21). Amsterdam: Elsevier.

Steven I. Pfeiffer, PhD, is a professor at Florida State University, where he serves as director of the doctoral program in combined counseling/school psychology. Prior to his tenure at Florida State, he was executive director of Duke University's Talent Identification Program. He is on the board of SENG and president-elect of the APA Consortium of Combined-Integrated Doctoral Programs in Professional Psychology. He frequently lectures on the psychological needs

of the gifted. He is editor of *Handbook of the Gifted & Talented: A Psychological Approach*, published by Kluwer/Plenum.

Tania Jarosewich, PhD, is president of Censeo Group LLC, with headquarters in Cleveland, Ohio. Prior to her present position, she held a position in program evaluation at Indiana University. She received her doctorate from Kent State University and completed a research postdoctorate at the Duke University Talent Identification Program. She is coauthor of the Gifted Rating Scales.