HHS Public Access
Author manuscript
Gift Child Q. Author manuscript; available in PMC 2015 September 02.

Published in final edited form as:
Gift Child Q. 2008 ; 52(1): 19–29. doi:10.1177/0016986207311055.

Identifying Young Gifted Children Using the Gifted Rating Scales–Preschool/Kindergarten Form

Steven I. Pfeiffer and **Yaacov Petscher**
Florida State University

Abstract

This article reports on an analysis of the diagnostic accuracy of a new teacher rating scale designed to assist in the identification of gifted preschool and kindergarten students. The Gifted Rating Scales–Preschool/Kindergarten Form (GRS-P) is based on a multidimensional model of giftedness. An examination of the standardization sample using diagnostic efficiency statistics provides support for the diagnostic accuracy of the GRS-P Intellectual Ability and Academic Ability scales identifying intellectual giftedness, irrespective of the IQ cut score used to demarcate giftedness. The present findings extend the analysis of the standardization sample reported in the test manual and provide additional support for the GRS-P as a gifted screening tool.

Keywords

gifted identification; gifted rating scales; GRS; preschool gifted

There is universal support for the early identification of, and intervention with, developmentally delayed children at an early age (Guralnick & Bennett, 1987; Meisels, 1989; Pfeiffer & Reddy, 1998). Increasingly, educators are recognizing the benefits of early identification and intervention for another special needs group, preschool and kindergarten students who show exceptionally high potential or ability (Borland & Wright, 1994; Jackson, 2003; Morelock & Feldman, 1992; Pfeiffer, 2002).

Early recognition and appropriate educational intervention for gifted preschool and kindergarten students increases the probability of future extraordinary achievement and reduces the risk for later social, behavioral, emotional, and/or educational problems (Harrison, 2004; Hodge & Kemp, 2000; Morelock & Feldman, 1992; Pfeiffer & Stocking, 2000; SankarDeLeeuw, 2002). However, many early childhood programs are not equipped to meet the needs of young students with precocious intellectual and academic abilities and/or special talents. Too few preschool and kindergarten teachers are trained or have the resources to identify or design effective programs that meet the psychoeducational needs of the young gifted child (Bloom, 1985; Hotulainen & Schofield, 2003; Jackson, 2003).

Accurate identification is one of the first, important steps when planning services for gifted preschool and kindergarten students. A recent survey of gifted experts highlighted the identification process as a critical issue facing the field (Pfeiffer, 2003). One reason for the difficulty in identifying young gifted students, in addition to no consensus on a definition of giftedness, is that the field has lacked technically sound screening instruments to complement the IQ test, especially screening tools designed for the young gifted child.

The IQ test is almost routinely used to determine whether a student qualifies for early gifted placement (Pfeiffer, 2002; Sparrow, Pfeiffer, & Newman, 2005). Unfortunately, there are few scientifically sound, standardized screening instruments available to complement the IQ test in providing a more comprehensive picture of a young student's potential or actual performance. Existing teacher rating scales suffer from serious technical shortcomings that limit their diagnostic utility (Jarosewich, Pfeiffer, & Morris, 2002). Some even suggest that teachers may be ill equipped to make valid judgments about a young child's intellectual abilities (Ehrlich, 1986; Fatouros, 1986; Hadaway & Marek-Schroer, 1992).

Recognizing the need for a technically sound screening instrument to assist in the identification of preschool and kindergarten students, Pfeiffer and Jarosewich (2003) developed a new gifted rating scale, the Gifted Rating Scales (GRS). The GRS 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. This article focuses exclusively on the GRS-P.

The present study reports on the diagnostic accuracy of two scales of the GRS-P with the standardization sample, which consists of preschool and kindergarten students, ages 4:0–6:11 years. The analyses described in this study have not been reported elsewhere and are intended to extend the information reported in the user manual and in a recent article by Pfeiffer, Petscher, and Jarosewich (2005). A similar analysis of the diagnostic accuracy of the GRS-S standardization sample for students in Grades 1–8 appears in *Gifted Child Quarterly* (Pfeiffer & Jarosewich, 2007).

METHOD

Participants

The test publisher, the Psychological Corporation, provided the authors with the GRS-P standardization sample data file. The full standardization sample consisted of 188 boys and 187 girls. The age-group of the sample was stratified within four 6-month age bands: 4:00–4:5, 4:6–4:11, 5:0–5:5, 5:6–5:11, and one 12-month age band (6:0–6:11), with each age band comprising 20% of the standardization population. The GRS-P standardization sample was stratified to approximate the U.S. population on essential demographic characteristics including race/ethnicity, parent education level, and regional representation (U.S. Bureau of the Census, 2000). For, example, 62.67% of the sample was Caucasian, 16% African American, 16% Hispanic, and 2.67% Asian American. Tables 4.1–4.3 in the test manual provide data on the race/ethnicity, parent education level, and regional representation of the GRSP sample stratified within the five age bands (Pfeiffer & Jarosewich, 2003, p. 25).

A subsample from the total standardization sample was used for the analysis of diagnostic accuracy of the GRS-P. The subsample consisted of 126 children from the standardization sample, which constituted the entire standardization sample for which *both* GRS-P and Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III) scores were available (33.6% of the total standardization sample). Fifty-four percent of the sample (*n* = 68) were male and 46% (*n* = 59) were female. Seventy-seven percent of the participants reported they were Caucasian (*n* = 98), followed by 13% African American (*n* = 16), 6% Hispanic (*n* = 8), 3% Asian (*n* = 4), and 1% Native American (*n* = 1). Table 1 provides a summary of the study sample by race/ethnicity and age band.

Instruments

Gifted Rating Scales–Preschool/Kindergarten Form—The GRS-P is designed for ages 4:0 to 6:11 and consists of five scales with 12 items each for a total of 60 items (Pfeiffer & Jarosewich, 2003). To facilitate compatibility, the GRS-P was co-linked during standardization with the new WPPSI-III. The items of the GRS-P represent skills and behaviors developmentally appropriate for preschool and kindergarten students. The GRS-P yields raw score totals on all five scales, which are converted to age-based *T* scores and associated cumulative percentages. The GRS-P 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 GRS-P scales:

- *Intellectual Ability*. This scale measures the young child's verbal and nonverbal mental skills and intellectual competence (e.g., memory, problem solving, and mental speed).
- *Academic Ability*. This scale measures the young child's skill in dealing with factual and/or school-related early childhood curriculum material.
- *Creativity*. This scale measures the young child's ability to think and/or produce unique, novel, or innovative thoughts or products (e.g., imaginative play, original thinking).
- *Artistic Talent*. This scale measures the young child's potential for, or evidence of, ability in drama, dance, drawing, singing, playing a musical instrument, and/or acting.
- *Motivation*. This scale refers to the young child's tendency to enjoy challenging tasks and ability to work well without encouragement. The motivation scale is *not* viewed as a type of giftedness. The scale is designed to measure the student's level of drive, persistence, and desire to succeed.

Each item is rated by the teacher on a 9-point scale divided into three ranges: 1–3, below average; 4–6, average; and 7–9, above average. The GRS-P manual provides a classification system that indicates, based on their *T* score, the *likelihood* that the preschooler or kindergartener is gifted. The higher the child's *T* score on one or more of the gifted scales, the greater the likelihood that he or she is gifted compared with other preschool or kindergarten peers. A *T* score below 55 (below 69%) indicates a low probability of gifted, a score between 55 and 59 (69–83%) indicates moderate probability, a score between 60 and 69 (84–97%) indicates high probability, and a score above 70 (98+%) indicates a very high probability.

Information on scale development is available in the test manual (Pfeiffer & Jarosewich, 2003). Final item selection was guided by factor structure, item mean scores, item bias (parent education level, gender, and ethnicity), and interrater and test-retest reliability. The test manual reports coefficient alpha reliabilities that range from .97 to .99 and standard errors of measurement that range from 1.0 to 1.73 across the five scales and five age ranges. Test-retest reliability coefficients range from .84 on the Creativity scale at age 5:0–5:11 to .97 on the Intellectual Ability, Academic Ability, and Creativity scales at age 6:0–6:11. Interrater reliability ranged from .70 for Artistic Talent to .80 for Academic Ability and .84 for Intellectual Ability. The test manual also provides evidence in support of internal structure and convergent and divergent validity (Pfeiffer & Jarosewich, 2003). More detailed information on the psychometric qualities of the GRS-P appears in Margulies and Floyd (2004) and Pfeiffer et al. (2005).

Wechsler Preschool and Primary Scale of Intelligence– Third Edition—The WPPSI-III (Wechsler, 2002) is a widely used measure of intellectual ability for preschool and kindergarten-age students (Sattler, 2001). Considerable research supports the concurrent and predictive validity of earlier versions of the WPPSI scale with children as young as 4 1/2 years of age (Kaplan, 1996; Kaplan, Fox, & Paxton, 1991; Schneider & Gervais, 1991). The WPPSI-III consists of 14 subtests, with seven verbal, five performance, and two processing speed subtests. The overall IQ represents an index of general mental ability. The manual reports evidence of high subtest reliability (.83-.95), internal consistency, test-retest stability, and validity (Wechsler, 2002).

Procedure

The diagnostic accuracy of the GRS-P was calculated applying diagnostic efficiency statistics for *both* the GRS-P Intellectual Ability and Academic Ability scales, using *T* cut scores set at 55, 60, and 70, and WPPSI-III Full Scale (FS) IQ scores of 115, 120, 125, and 130. GRS-P *T* scores of 55, 60, and 70 were selected because the GRS classification system proposes that a *T* score between 55 and 59 (69–83%) indicates a moderate probability of giftedness, a *T* score between 60 and 69 (84–97%) indicates a high probability of giftedness, and a *T* score of 70 and above (98+%) indicates a very high probability of giftedness (Pfeiffer & Jarosewich, 2003). WPPSI-III FS IQ cut scores were set at 115, 120, 125, and 130 because authorities in the gifted field, state regulations for gifted education, and state and local gifted officials have not reached consensus on where to operationally “draw the line” in demarcating intellectually gifted. This is especially true among the very young, where developmental issues and a greater concern for inclusiveness and not missing potentially gifted young students are prevailing considerations (Morelock & Feldman, 1992; Pfeiffer, 2002; Stephens & Karnes, 2000). Four different IQ cut scores, ranging from 115 (1 *SD* above the mean) to 130 (2 *SD* above the mean), provide four different proxies for intellectual giftedness. Including four different IQ cut scores provides a more comprehensive diagnostic picture of how well the GRS-P works when compared with the WPPSI-III.

We computed five diagnostic efficiency statistics to provide multiple perspectives on the diagnostic accuracy of the GRS-P. The *sensitivity* (SE) of a test, also known as the true positive rate, is the proportion of people who have the attribute (in this particular case, intellectual giftedness operationally defined as IQ ≥ 115, 120, 125, and/or 130) who are correctly detected by the test, applying *T* scores of 55, 60, and 70. The *specificity* (SP) of a test is the proportion of people who do *not* have the attribute (i.e., *not* intellectually gifted) who are correctly identified by the Intellectual Ability scale and/or Academic Ability scale of the GRS-P as *not* gifted. One can combine the sensitivity and specificity into a single index called the *likelihood ratio* (LR+), which is mathematically defined as the sensitivity divided by (1 – specificity) (also known as the false positive rate). The likelihood ratio is an index of the accuracy of a test and depicts what the odds are that a positive test result (e.g., GRS-P *T* score above 60) comes from a young student who is, in fact, gifted. When the LR+ is 1.0 or approximates 1.0, the test is not diagnostically useful because it 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, the LR−. It is mathematically defined as the specificity divided by (1 – sensitivity) (also known as the false negative rate).

The *overall correct classification* (OCC) is the True Positive + True Negative divided by *N*. The OCC incorporates both sensitivity and specificity into an overall index of diagnostic accuracy. The five diagnostic statistics SE, SP, LR+, LR−, and OCC were selected because *none* are affected by base rate or prevalence (Meehl & Kline, 1988; Streiner, 2003), which must be considered when measuring the diagnostic accuracy of a test with a low prevalence phenomenon such as giftedness.

We also ran receiver operating curve (ROC) analyses, which graphically depict the diagnostic accuracy across the entire range of standard scores on the GRS-P. The ROC analysis graphs the sensitivity and 1–specificity associated with each possible cut score on the test. The area under the curve (AUC) value describes the classification fit with a score of 1.0 representing a perfect fit, a score of 0.5 chance-level diagnostic prediction, and a score of less than 0.5 a below-chance prediction. The AUC, like sensitivity, specificity, and the likelihood ratio, is not influenced by the prevalence or base rate. It is considered a sufficient means to communicate effect size to audiences (Hsu, 2002; Swets, 1988).

RESULTS

Diagnostic Efficiency Values

Table 2 provides a summary of the five diagnostic efficiency values (i.e., SE, SP, OCC, LR +, and LR−) for GRS-P Intellectual and Academic Ability *T* scores of 55, 60, and 70, and WISC-IV FS IQ scores ≥115, ≥120, ≥125 and ≥130. WISC-IV FS IQ scores were set at 115, 120, 125, and 130, as mentioned earlier, to provide the reader with multiple perspectives from which to evaluate the diagnostic accuracy (and usefulness) of the GRS-P.

Diagnostic accuracy when FSIQ ≥ 115—The first set of GRS-P diagnostic efficiency values was calculated based on intellectual giftedness represented as an FSIQ ≥ 115. As mentioned earlier, test sensitivity is the proportion of children who are intellectually gifted based on a given IQ score and who are correctly identified by the GRS-P Intellectual Ability and/or Academic Ability scale. A *T* score of 55 on both the Intellectual or Academic scales yielded the highest sensitivity (.83 and .81, respectively). In other words, the majority of the preschool and kindergarten sample who obtained FSIQ scores ≥ 115 also obtained *T* scores ≥ 55. In contrast, a *T* score of 70 on both the Intellectual and Academic scales yielded low sensitivity (.21 for both scales).

Test specificity is the proportion of children who are *not* intellectually gifted (again, in this instance based on FSIQ ≥ 115) and who *are* correctly identified by the GRS-P Intellectual Ability and/or Academic Ability scale as *not* gifted. A *T* score of 70 on both the Intellectual and Academic scales yielded the highest specificity (.96 for both scales). In other words, almost *none* of the sample who obtained FSIQ scores ≤115 also obtained *T* scores of ≥70 (e.g., only 3 of the 126 participants obtained an FSIQ ≤ 115 and a *T* score ≥70 on the GRS-P Academic Ability scale). *T* scores of 60 and 55 both generated moderately high test specificity (.87 and .81, and .61 and .63, respectively).

OCC is a composite index of the diagnostic accuracy of a test. The OCC is calculated as the True Positive + True Negative divided by *N*. For all six *T* scores, with an FSIQ ≥ 115, the OCC indexes were similar and moderately high (.68-.73); the four highest OCCs were Intellectual Ability and Academic Ability *T* scores of 60 or 70.

The likelihood ratio is another index of the test's accuracy and depicts what the odds are that a positive test result comes from a student who is gifted (LR+) and a negative test result comes from a student who is *not* gifted (LR−). The LR+ was substantial for all six *T* scores, with three yielding scores in the 5+ range (i.e., it is more than five times as likely for students with *T* scores at this level or above to have IQ ≥ 115). The LR− was significant and substantial for three of the six *T* scores; Intellectual Ability and Academic Ability *T* scores of 55 yielded scores in the 3+ range.

Diagnostic accuracy when FSIQ ≥120—The second set of GRS-P diagnostic efficiency values were calculated based on intellectual giftedness represented as FSIQ ≥ 120, 5 scale score points higher than the above IQ cut score. Recall that we are examining multiple IQ cut scores ranging from 115 to 130 to provide a more comprehensive test of the diagnostic accuracy of the GRS-P. A *T* score of 55 on Intellectual Ability (.93) and Academic Ability (.89) yielded the highest sensitivity, similar to the findings when the IQ was set ≥115. A *T* score of 70 using either the Intellectual or Academic Ability scale yielded the lowest sensitivity (.21 and .29, respectively), again consistent with the IQ set ≥115.

A *T* score of 70 on Intellectual Ability and Academic Ability provided the highest specificity (.96 for both), and a *T* score of 55 on Intellectual Ability and Academic Ability yielded the lowest specificity (.67 and .59, respectively). These findings for test specificity at IQ ≥120 are consistent with when the IQ was set ≥115.

The OCCs for the six *T* scores ranged from .62 (Academic Ability *T* score of 55) to .81 (Academic Ability *T* score of 70). Four of the *T* scores yielded high OCC scores in the .7–.8 range. The LR+ was substantial for all six *T* scores, with the Academic Ability *T* score of 70 yielding an LR+ of 7.25 and the Intellectual Ability *T* score of 70 yielding an LR+ of 5.25. The LR− was substantial for four of the six *T* scores; the Intellectual Ability *T* score of 55 yielded an LR− of 9.57, and the Academic Ability *T* score of 70 yielded an LR− of 5.36.

The overall pattern of results for the diagnostic accuracy of the GRS-P is similar for IQ ≥ 115 and IQ ≥ 120.

Diagnostic accuracy when using FSIQ ≥ 125—The third set of GRS-P diagnostic efficiency values were calculated on the basis of intellectual giftedness represented as FSIQ ≥ 125. A *T* score of 55 using either the Intellectual Ability (.94) or Academic Ability (.55) scale yielded the highest sensitivity. A *T* score of 70 for both the Intellectual and Academic Ability scales yielded the lowest sensitivity (.29 for both). These test sensitivity findings are similar to those reported above for FSIQ ≥ 115 and FSIQ ≥ 120. A *T* score of 70 for both the Intellectual and Academic Ability scales yielded the highest test specificity (.93 for both), whereas a *T* score of 55 yielded the lowest specificity (.52 for Intellectual Ability, .54 for Academic Ability, respectively). *T* scores of 60 for both the Intellectual and Academic Ability scales fell within the midrange for both test sensitivity and specificity (.65-.77).

The OCC index was high with a *T* score of 70 (.84 for Intellectual Ability; .83 for Academic Ability) and moderately high with a *T* score of 60 (.68 for both). The OCC was low, however, with a *T* score of 55 (.57 and .55). The LR+ was substantial for a *T* score of 70 (more than 4 for both the Intellectual and Academic Ability scales); the other four *T* scores yielded significant although not as substantial an LR+ index (1.91–2.41). The LR− or false negative rate was substantial for an Intellectual and Academic Ability *T* score of 55 (8.67 and 4.50, respectively). A *T* score of 60 generated a significant but not substantial false negative rate (2.96 and 2.09 for Intellectual and Academic Ability, respectively). However, a *T* score of 70 for both Intellectual and Academic Ability produced a false negative rate not much better than chance.

Diagnostic accuracy when using FSIQ ≥ 130—The final set of GRS-P diagnostic efficiency values was calculated on the basis of intellectual giftedness represented at FSIQ ≥ 130, which is a cut score two *SD* above the mean (top 5%). A *T* score of 55 for both the Intellectual Ability and Academic Ability scales yielded the highest test sensitivity (1.0 and .91, respectively). In other words, *every* preschool and kindergarten student who obtained an FSIQ score ≥130 obtained a *T* score of 55 or higher on the GRS-P Intellectual Ability scale.

In contrast, test sensitivity was low for a *T* score of 70 (.27 for both scales) and intermediate for a *T* score of 60 (.73 for Intellectual Ability; .55 for Academic Ability).

A *T* score of 70 yielded the highest test specificity (.92 for *both* the Intellectual Ability and Academic Ability scales). A *T* score of 60 yielded test specificity values of .65 for Intellectual Ability and .70 for Academic Ability, whereas a *T* score of 55 generated the lowest test specificity values (.50 for Intellectual Ability, .52 for Academic Ability). The OCC was highest for a *T* score of 70 (.85 for Intellectual Ability; .85 for Academic Ability) and moderately high for a *T* score of 60 (.65 for both scales). The OCC was low with a *T* score of 55 (.54 and .52), consistent with OCC results when FSIQ ≥ 125.

The LR+ or false positive rate was substantial for a *T* score of 70, with the LR+ index 3.38 for both scales. In other words, it is almost four times as likely for students with *T* scores ≥70 on either the Intellectual Ability or Academic Ability scale to have IQ ≥ 130. *T* scores of 55 and 60 also yielded substantial indexes (1.83–2.09). The LR− was substantial for a *T* score of 55 on the Academic Ability scale (5.78) but not for Intellectual Ability (0.50). LR− was not significant for four of the six *T* scores when FSIQ ≥ 130.

Receiver Operating Curve Values

The ROC analysis for the GRS-P Intellectual Ability and Academic Ability scales performed significantly above chance at each of the four FSIQ scores, 115, 120, 125, and 130 (see Figure 1). For FSIQ ≥ 115, the estimated AUC for both Intellectual Ability and Academic Ability was 0.79 (*SEM* = 0.04, $p < .001$) with a 95% confidence interval of 0.71 to 0.87 for FSIQ ≥ 115. In other words, 79% of students in a population who are intellectually gifted would be identified as gifted based on the screening test. For FSIQ ≥ 120, the estimated AUC for Intellectual Ability was 0.82 (*SEM* = 0.04, $p < .001$) with a 95% confidence interval of 0.74 to 0.90; the estimated AUC for Academic Ability at the same IQ cut score was 0.81 (*SEM* = 0.04, $p < .001$) with a 95% confidence interval of 0.74 to 0.89. For FSIQ ≥ 125, the estimated AUC for Intellectual Ability was 0.78 (*SEM* = 0.05, $p < .001$) with a 95% confidence interval of 0.72 to 0.90; the estimated AUC for Academic Ability was 0.81 (*SEM* = 0.05, $p < .001$) with a 95% confidence interval of 0.72 to 0.90. Finally, for FSIQ ≥ 130, the estimated AUC for Intellectual Ability was 0.76 (*SEM* = 0.07, $p < .01$) with a 95% confidence interval of 0.63 to 0.89; the estimated AUC for Academic Ability was 0.80 (*SEM* = 0.06, $p < .01$) with a 95% confidence interval of 0.69 to 0.91.

Recall that the AUC is a type of effect size (Hsu, 2002; Swets, 1988). As Figure 1 depicts, irrespective of the FSIQ cut score, the ROCs for Intellectual Ability and Academic Ability fall far above the diagonal line, which represents chance-level prediction. In other words, both the GRS-P Intellectual Ability and Academic Ability scales work exceptionally well in predicting intellectual giftedness.

DISCUSSION

The present study investigated the diagnostic accuracy of the Intellectual Ability and Academic Ability scales of the GRS-P with a subset of the standardization sample. The findings reported in this article are intended to extend information available in the test manual and published elsewhere (Margulies & Floyd, 2004; Pfeiffer & Jarosewich, 2007).

The results of this investigation confirm that the GRS-P works quite well in identifying intellectual giftedness among preschool and kindergarten students. Designed as a screening instrument, the GRS-P demonstrates compelling evidence of *criterion-referenced validity.* The test was able to do what it is intended to do, specifically, identify young children who are intellectually gifted as “test-positive” and those young children who are not intellectually gifted as “test-negative.” The results of five diagnostic efficiency tests and ROC analyses confirm that the GRS-P is accurate in identifying intellectual giftedness. This proved to be true irrespective of whether intellectual giftedness was operationally defined by an FSIQ cut score ≥115, 120, 125, or 130, using *T* cut scores of 55, 60, and 70.

All analyses of the AUC for both the Intellectual Ability scale and Academic Ability scale yielded highly significant results; the GRS-P performed significantly above chance at all four IQ cut scores. Furthermore, the OCC values were significant at all four IQ cut scores (OCC in the high .70s for IQ ≥ 115; in the .80s for IQ ≥ 120 through 130).

The sensitivity and specificity of a test are not absolute values. Test sensitivity and test specificity will vary with the samples on which they are based and with the critical cut score values chosen (Verhulst & Koot, 1992). The sample used for the present study was part of the standardization sample, which was stratified to approximate the U.S. population on important demographic characteristics including race/ethnicity, parent education level, and regional representation (Pfeiffer & Jarosewich, 2003). However, cross-validation studies would corroborate the findings in support of the diagnostic accuracy and criterion-referenced validity of the GRS-P.

In addition to cross-validation studies with different samples, future research may want to use alternative criteria to operationally define intellectually gifted. We posited that the FSIQ was a reasonable proxy for intellectual giftedness; other investigators may want to use alternative criteria such as developmental measures of language ability (Maxwell, 1995), complexity of play (Kitano, 1985), early reading (Sankar-DeLeeuw, 2002), or nonverbal measures of intelligence (Bernal, 2002; Ford, Harris, Tyson, & Trotman, 2002; McCallum, Bracken, & Wasserman, 2001; Shaunessy, Karnes, & Cobb, 2004).

Future studies may also want to extend the present research by validating the other GRS-P scales. This will not be an easy task, because there are few gold standard criterion measures for creativity or artistic ability. Small-scale studies using ratings by a panel of artists of the artwork of preschool and kindergarten students and scores on the Figural Form B of the Torrence Test of Creative Thinking provided preliminary validation for the Artistic Talent and Creativity scales (Pfeiffer & Jarosewich, 2003). However, considerably more work is needed to validate these two gifted scales.

Although infrequently reported in psychology or education journals, the validity of a test should include reporting its diagnostic performance (Cohen, 1988; Streiner, 2003). Test sensitivity, specificity, and overall correct classification are now regularly reported in medical journals (Verhulst & Koot, 1992). It is important to remember that test sensitivity can only be increased at the expense of test specificity, and test specificity can only be increased at the expense of test sensitivity (Swets, 1988). Unfortunately, there is no way that one can improve *both* the sensitivity and the specificity at the same time. The score selected to represent the dividing line between gifted and not gifted (the cut score) is, we recognize, somewhat arbitrary. We act as if giftedness is a truly dichotomous state, when, in fact, giftedness lies on a continuum, varying from clear absence of giftedness to obvious presence of giftedness (like many diseases in medicine, e.g., hypertension, hyperlipedemia, obesity, spina bifida; Brenner & Gefeller, 1997; Chinchilli, 1983). Different practical, philosophical, political, and even economic factors argue for different cutoff points, and changes in the cutoff score will affect the sensitivity of any screening test.

The predictive value of a test is affected by the particular cut score that is used and also by the *base rate* of the disorder or condition of interest in the population (Glaros & Kline, 1988). When the base rate is low, the predictive value of a negative test result will be greater than the predictive value of a positive test result. That is, when the disorder or condition of interest is relatively rare, a positive test finding is typically not useful in confirming its presence (Cronbach, 1984; Meehl & Kline, 1988). For example, for the sake of illustration, assume that giftedness is a truly dichotomous phenomenon and, further, imagine that we know that the base rate in the school population for intellectual giftedness is 5%. The GRS-P has a sensitivity of .73 and a specificity of .65 with a cut score of $T \geq 60$ (again, assuming a 5% prevalence rate). Consider that the population of a school district is 10,000. In this example, with a prevalence rate of 5%, we can assume that 500 students are intellectually gifted and 9,500 are not. Of the 500 gifted students, the GRS-P will identify correctly 365 (73%) using a cut score of $T \geq 60$. Similarly, of the 9,500 not-gifted students in the school district, 6,175 (65%) will obtain a “normal range” *T* score on the GRS-P (in this instance, $T \leq 60$). Now let us use a different cut score but assume the same 5% prevalence rate. The GRS-P has a sensitivity of .27 and a specificity of .92 when using a cut score of $T \geq 70$ (and the same 5% prevalence rate). Considering the same 500 gifted students, the GRS-P will now identify correctly only 135 (27%); however, 8,740 (92%) of the same 9,500 not-gifted students will now obtain a “normal range” *T* score on the GRS-P. As this example illustrates, the sensitivity and specificity (and overall predictive value) of the GRS-P or any test varies with the cut score and prevalence rate.

As mentioned earlier, a number of authors suggest that teachers may not be able to make valid judgments about a young student’s intellectual abilities (Ehrlich, 1986; Fatouros, 1986; Hadaway & Marek-Schroer, 1992). The present study lends empirical support to the view that teachers *are* able to provide valid ratings on the academic and intellectual abilities of young students. Furthermore, recent research indicates that there are no age differences across the 3-year age span 4:0–6:11 on any of the GRS-P scales and gender differences on only one of the five scales (Artistic Talent). This same research found small but statistically significant race/ethnicity differences with Asian American students rated, on average, 1.5 scale score points higher than Whites and Native Americans and, on average, 7 points higher than African American and Hispanic students (Pfeiffer et al., 2005). Of course, every test is culturally loaded to some extent (Barona & Pfeiffer, 1992; Flanagan, McGrew, & Ortiz, 2000; Rushton & Jensen, 2005). The GRS-P does *not* appear to be strongly culturally loaded with average differences across racial/ethnic groups in the 1.5–7.0 point range, all less than 1 *SD*.

One limitation of the study is the small number of typically underrepresented gifted minority group students in the sample (African American, n=16; Asian, $n = 4$; Hispanic, $n = 8$; Native American, $n = 1$). It would be imprudent to draw any conclusions in terms of race/ethnicity. We hope that future studies explore the impact of race/ethnicity and other relevant sociocultural and family factors on the effectiveness of the GRS-P as a gifted screening test.

The present research underscores that the GRS-P holds potential as a new screening test that can assist in the identification of gifted preschool and kindergarten students. The need to identify and intervene with gifted children at an early age is critical if we hope to improve their chances for optimal development. The gifted field is rightfully concerned with fair and equitable identification practices, in part because of a history of underrepresentation of African American, Hispanic, and Native American students in gifted education programs

(Ford, 1998; Ford & Whiting, in press; Gridley, Mucha, & Hatfield, 1995; Pfeiffer, 2002). When used as part of a comprehensive gifted assessment, the GRS-P appears to hold promise in ensuring that *no* young child with promise will inadvertently be excluded from being identified in terms of his or her probability of being intellectually or academically gifted.

References

Barona A, Pfeiffer SI. Effects of test administration procedures and acculturation level on achievement test scores. Journal of Psychoeducational Assessment. 1992; 10:124–132.

Bernal EM. Three ways to achieve a more equitable representation of culturally and linguistically different students in GT programs. Roeper Review. 2002; 24:82–88.

Bloom, BS. Developing talent in young people. New York: Ballantine Books; 1985.

Borland JH, Wright L. Identifying young, potentially gifted, economically disadvantaged students. Gifted Child Quarterly. 1994; 38:164–171.

Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Statistics and Medicine. 1997; 16:981–991.

Chinchilli VM. Estimates of sensitivity and specificity in a multistage screen for medical diagnosis. Biometrics. 1983; 39:333–340.

Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Lawrence Erlbaum; 1988.

Cronbach, LJ. Essentials of psychological testing. 4. New York: Harper & Row; 1984.

Ehrlich VZ. Recognizing superior cognitive abilities in disadvantaged, minority, and other diverse populations. Journal of Children in Contemporary Society. 1986; 18:55–70.

Fatouros C. Early identification of gifted children is crucial … but how should we go about it? Gifted Education International. 1986; 4:24–28.

Flanagan, DP.; McGrew, KS.; Ortiz, SO. The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation. Boston: Allyn & Bacon; 2000.

Ford DY. The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. Journal of Special Education. 1998; 32:4–14.

Ford DY, Harris JJ III, Tyson CA, Trotman MF. Beyond deficit thinking: Providing access for gifted African American students. Roeper Review. 2002; 24:52–58.

Ford, DY.; Whiting, GW. Recruiting and retaining under-represented gifted students. In: Pfeiffer, SI., editor. Handbook of giftedness in children: Psycho-educational theory, research, and best practices. New York: Springer Publishers; in press

Glaros AG, Kline RB. Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. Journal of Clinical Psychology. 1988; 44:1013–1023. [PubMed: 3216006]

Gridley, BE.; Mucha, L.; Hatfield, BB. Best practices in preschool screening. In: Thomas, A.; Grimes, J., editors. Best practices in school psychology–III. Washington, DC: National Association of School Psychologists; 1995. p. 213-225.

Guralnick, MJ.; Bennett, FC. The effectiveness of early intervention for at-risk and handicapped children. Orlando, FL: Academic Press; 1987.

Hadaway N, Marek-Schroer MF. Multidimensional assessment of the gifted minority student. Roeper Review. 1992; 15:73–77.

Harrison C. Giftedness in early childhood: The search for complexity and connection. Roeper Review. 2004; 26:78–84.

Hodge KA, Kemp CR. Exploring the nature of giftedness in preschool children. Journal for the Education of the Gifted. 2000; 24:46–73.

Hotulainen RHE, Schofield NJ. Identified preschool potential giftedness and its relation to academic achievement and self-concept at the end of Finnish comprehensive school. High Ability Studies. 2003; 14:55–70.

Hsu LM. Diagnostic validity statistics and the MCMIIII. Psychological Assessment. 2002; 14:410–422. [PubMed: 12501567]

Jackson, NE. Young gifted children. In: Colangelo, N.; Davis, GA., editors. Handbook of gifted education. 3. Boston: Allyn & Bacon; 2003. p. 470-482.

Jarosewich T, Pfeiffer SI, Morris J. Identifying gifted students using teacher rating scales: A review of existing instruments. Journal of Psychoeducational Assessment. 2002; 20:322–336.

Kaplan CH. Predictive validity of the WPPSI-R: A four year follow-up study. Psychology in the Schools. 1996; 33:211–220.

Kaplan CH, Fox LM, Paxton L. Bright children and the revised WPPSI: Concurrent validity. Journal of Psychoeducational Assessment. 1991; 9:240–246.

Kitano M. Ethnography of a preschool for the gifted: What gifted young children actually do. Gifted Child Quarterly. 1985; 29:67–71.

Margulies AS, Floyd RG. Test review: The Gifted Rating Scales. Journal of Psychoeducational Assessment. 2004; 22:275–282.

Maxwell, E. The changing developmental needs of the gifted: Birth to maturity. In: Genshaft, JL.; Bireley, M.; Hollinger, CL., editors. Serving gifted and talented students: A resource for school personnel. Austin, TX: Pro-ed; 1995. p. 17-30.

McCallum, RS.; Bracken, BA.; Wasserman, J. Essentials of nonverbal assessment. New York: John Wiley; 2001.

Meehl PE, Kline RB. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin. 1988; 52:194–216. [PubMed: 14371890]

Meisels SJ. Can developmental screening tests identify children who are developmentally at-risk? Pediatrics. 1989; 83:578–585. [PubMed: 2467250]

Morelock, MJ.; Feldman, DH. The assessment of giftedness in preschool children. In: Nuttall, EV.; Romero, I.; Kalesnik, J., editors. Assessing and screening preschoolers: Psychological and educational dimensions. Needham Heights, MA: Allyn & Bacon; 1992. p. 301-309.

Pfeiffer SI. Identifying gifted and talented students: Recurring issues and promising solutions. Journal of Applied School Psychology. 2002; 1:31–50.

Pfeiffer SI. Challenges and opportunities for students who are gifted: What the experts say. Gifted Child Quarterly. 2003; 47:161–169.

Pfeiffer, SI.; Jarosewich, T. Gifted Rating Scales. San Antonio, TX: The Psychological Corporation; 2003.

Pfeiffer SI, Jarosewich T. The Gifted Rating Scales–School Form. An analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. Gifted Child Quarterly. 2007; 51:39–50.

Pfeiffer SI, Petscher Y, Jarosewich T. The Gifted Rating Scale–Kindergarten form: An analysis of the standardization sample based on age, gender, and race. Roeper Review. 2007; 29:206–211.

Pfeiffer SI, Reddy LA. School-based mental health programs: Present status and blueprint for the future. School Psychology Review. 1998; 27:84–96.

Pfeiffer SI, Stocking VB. Vulnerabilities of academically gifted students. Special Services in the Schools. 2000; 16:83–93.

Ross, PO. National excellence: A case for developing America's talent. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement; 1993.

Rushton JP, Jensen AR. Thirty years of research on race differences in cognitive ability. Psychology, Public Policy, and Law. 2005; 11:235–294.

Sankar-DeLeeuw N. Gifted preschoolers: Parent and teacher views on identification, early admission, and programming. Roeper Review. 2002; 24:172–177.

Sattler, JM. Assessment of children: Cognitive applications. 4. San Diego, CA: Jerome M. Sattler; 2001.

Schneider BH, Gervais MD. Identifying gifted kindergarten students with brief screening measures and the WPPSI-R. Journal of Psychoeducational Assessment. 1991; 9:201–208.

Shaunessy E, Karnes FA, Cobb Y. Assessing culturally diverse potentially gifted students with nonverbal measures of intelligence. The School Psychologist. 2004; 58:99–102.

Sparrow, SS.; Pfeiffer, SI.; Newman, TM. Assessment of children who are gifted with the WISC-IV. In: Prifitera, A.; Saklofske, DH.; Weiss, LG., editors. WISC-IV: Clinical use and interpretation. Burlington, MA: Elsevier Academic Press; 2005. p. 282-299.

Stephens KR, Karnes FA. State definitions for the gifted and talented revisited. Exceptional Children. 2000; 66:219–238.

Streiner DL. Diagnosing tests: Using and misusing diagnostic screening tests. Journal of Personality Assessment. 2003; 81:209–219. [PubMed: 14638445]

Swets JA. Measuring the diagnostic accuracy of diagnostic systems. Science. 1988; 240:1285–1293. [PubMed: 3287615]

U.S. Bureau of the Census. Current population survey, October 2000: School Enrollment Supplement File [CDROM]. Washington, DC: Author; 2000.

Verhulst, FC.; Koot, HM. Child psychiatric epidemiology. Newbury Park, CA: Sage; 1992.

Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence–. 3. San Antonio, TX: The Psychological Corporation; 2002.

Zigler, A.; Heller, KA. Conceptions of giftedness from a meta-theoretical perspective. In: Heller, KA.; Mönks, FJ.; Sternberg, RJ.; Subotnik, RF., editors. International handbook of giftedness and talent. 2. Amsterdam: Elsevier; 2000. p. 3-21.

Putting the Research to Use

Identifying young gifted children is critical if we hope to develop educational interventions at the earliest possible time that support their optimal psychoeducational development. The present article reports on the diagnostic validity of a new teacher rating scale developed to assist in the early identification of preschool and kindergarten gifted students. The Gifted Rating Scales–Preschool/Kindergarten Form (GRS-P) is a 60-item teacher rating scale based on a multidimensional model of giftedness that is designed to be used as a screening tool or as part of a comprehensive assessment protocol. The GRS-P is very effective as a diagnostic instrument in identifying intellectual giftedness, irrespective of whether an IQ cutoff score of 115, 120, 125, or 130 is used to demarcate intellectual giftedness. The GRS-P holds promise as a *first-stage screening test* in the early identification of preschool and kindergarten students of high potential. Early childhood, preschool, and kindergarten teachers could use the GRS-P to screen entire classrooms of students or cohorts of students selected for screening with the GRS-P based on early reading, precocious language development, complex play, or other signs of early giftedness.

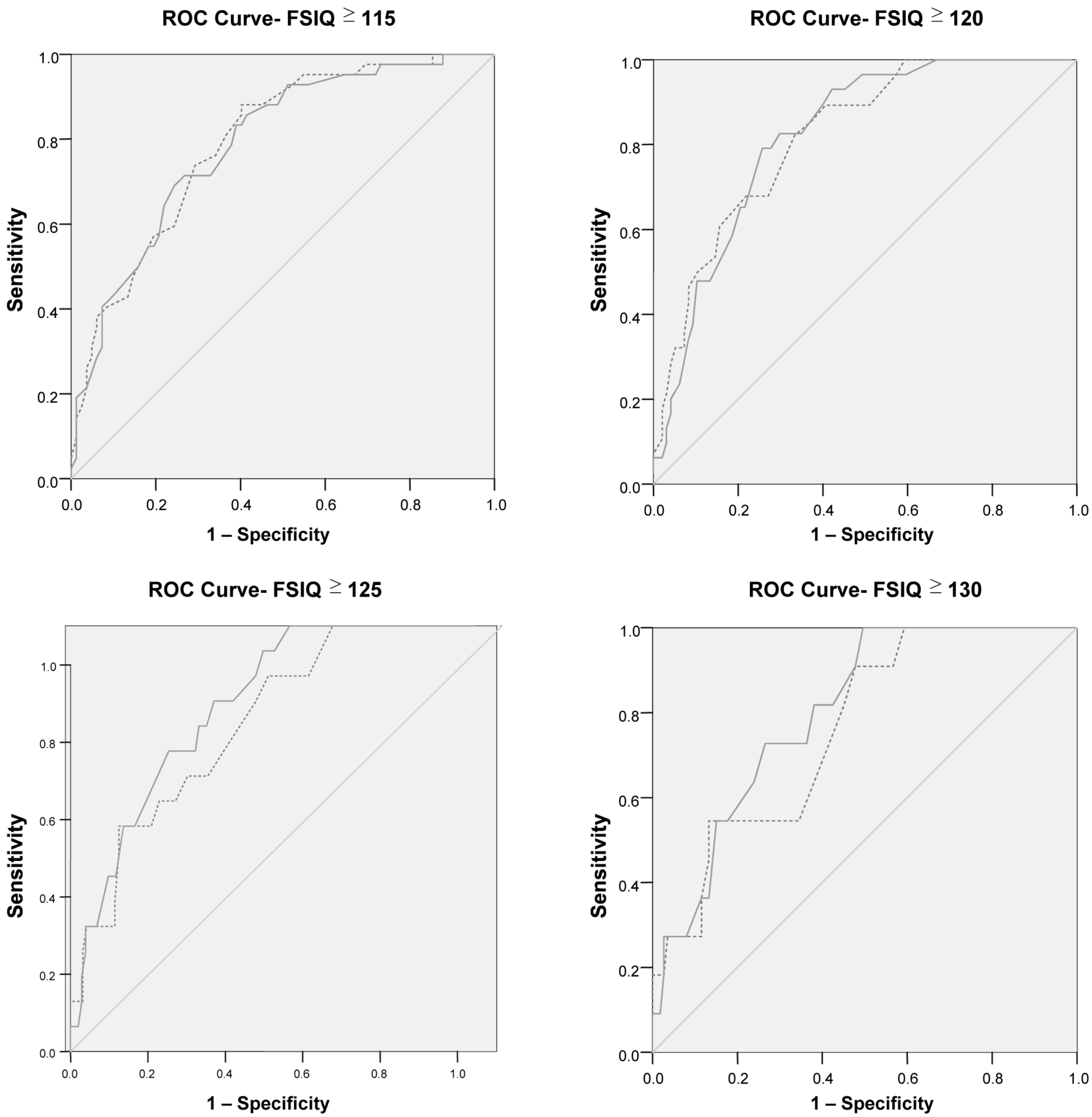

Figure 1.

ROC for WPPSI-III FSIQ and GRS-P Intellectual and Academic Ability Scales

Note: Diagonal line represents chance-level prediction. The dotted line represents the Academic Ability scale, and the solid line represents the Intellectual Ability scale.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

GRS-P Sample Demographics by Race/Ethnicity and Age Band

Age Band	Group	Caucasian	African American	Hispanic	Asian	Other
4:0–4:5	U.S. Population	59.79	15.28	19.51	3.96	1.47
	Sample	77.80	22.20	0	0	0
4:6–4:11	U.S. Population	59.79	15.28	19.51	3.96	1.47
	Sample	69.40	16.70	8.30	2.80	2.80
5:0–5:5	U.S. Population	62.79	16.01	16.15	3.79	1.25
	Sample	80.80	11.50	7.70	0	0
5:6–5:11	U.S. Population	62.79	16.01	16.15	3.79	1.28
	Sample	82.10	7.10	7.10	3.60	0
6:0–6:11	U.S. Population	62.79	16.01	16.15	3.79	1.28
	Sample	78.60	10.70	3.60	7.10	0

Source: U.S. Bureau of the Census. (2000).

Note: $n = 75$ for each age band.

Table 2

Results of Diagnostic Efficiency Statistics: WPPSI-III IQ ≥ 115, ≥ 120, ≥ 125, and ≥ 130

	Sensitivity[a]	**Specificity**[b]	**OCC**[c]	**LR+**[d]	**LR−**[d]
WPPSI Full Scale (FS) IQ ≥ 115					
Intellectual Ability Scale					
GRS *T* score 55	.83	.61	.68	2.13	3.59
GRS T score 60	.67	.87	.73	5.15	2.63
GRS T score 70	.21	.96	.72	5.25	1.22
Academic Ability Scale					
GRS *T* score 55	.81	.63	.69	2.25	3.32
GRS T score 60	.57	.81	.70	3.00	1.88
GRS T score 70	.21	.96	.73	5.68	1.22
WPPSI FSIQ ≥ 120					
Intellectual Ability Scale					
GRS *T* score 55	1.00	.50	.54	2.00	0.50
GRS T score 60	.73	.65	.65	2.09	2.41
GRS T score 70	.27	.92	.84	3.38	1.26
Academic Ability Scale					
GRS *T* score 55	.91	.52	.52	1.90	5.78
GRS T score 60	.55	.70	.65	1.83	1.56
GRS T score 70	.27	.92	.85	3.38	1.26
WPPSI FSIQ ≥ 125					
Intellectual Ability Scale					
GRS *T* score 55	.94	.52	.57	1.96	8.67
GRS T score 60	.77	.68	.68	2.41	2.96
GRS T score 70	.29	.93	.84	4.14	1.31
Academic Ability Scale					
GRS *T* score 55	.88	.54	.55	1.91	4.50
GRS T score 60	.65	.73	.68	2.41	2.09
GRS T score 70	.29	.93	.83	4.14	1.31
WPPSI FSIQ ≥ 130					

	Sensitivity[a]	Specificity[b]	OCC[c]	LR+[d]	LR−[d]
Intellectual Ability Scale					
GRS *T* score 55	1.00	.50	.54	2.00	0.50
GRS T score 60	.73	.65	.65	2.09	2.41
GRS T score 70	.27	.92	.84	3.38	1.26
Academic Ability Scale					
GRS *T* score 55	.91	.52	.52	1.90	5.78
GRS T score 60	.55	.70	.65	1.83	1.56
GRS T score 70	.27	.92	.85	3.38	1.26

Note: WPPSI-III = Wechsler Preschool and Primary Scale of Intelligence–Third Edition; GRS = Gifted Rating Scales.

[a]Sensitivity = the ability of a test to yield a positive test result for a person who has the attribute or condition.

[b]Specificity = the ability of a test to yield a negative result for a person who does *not* have the attribute or condition.

[c]OCC = overall correct classification; a composite index of the diagnostic accuracy of the test.

[d]LR = likelihood ratio; depiction of what the odds are that a positive test result comes from a person with the attribute (LR+) and the odds that a negative test result comes from a person *without* the attribute (LR−).