PRACTICE FORUM

Gifted Identification: Measuring Change in a Student's Profile of Abilities Using the Gifted Rating Scales

Steven I. Pfeiffer, Alper Kumtepe and Javier Rosado
Florida State University

Abstract

This article reports on the application of a methodology to measure reliable student change over time using the *Gifted Rating Scale*. The *Gifted Rating Scale*s is a new teacher rating scale based on a multidimensional model of giftedness. The methodology incorporates the standard error of prediction (SE*p*) model that takes into consideration regression effects and the reliability of the scales of the *Gifted Rating Scales*. Tables bracket *Gifted Rating Scales* posttest scores at the .90 and .95 confidence level to afford practitioners a choice with how certain they want to be that a student has demonstrated true change over time with one or more of the scales on the *Gifted Rating Scales*. A case example illustrates the use of the SE*p* table for the *GRS-Preschool/Kindergarten Form*. Implications are provided for best practices in gifted screening and evaluation.

"One important element in serving the gifted is being able to accurately identify gifted students."

One important element in serving the gifted is being able to accurately identify gifted students. However, a number of issues compromise practitioners' ability to identify gifted students.

Three such issues are varying state definitions of giftedness (Stephens & Karnes, 2000), a lack of consensus about how to conceptualize giftedness (Pfeiffer, 2003), and few technically sound screening instruments (Jarosewich, Pfeiffer, & Morris, 2002).

The lack of technically sound screening instruments is a continuing problem in gifted identification. The ubiquitous IQ test is routinely irrespective of state definitions or expert opinion on how to define giftedness to determine whether a student qualifies for gifted placement (Sparrow, Pfeiffer, & Newman, 2005). There are few screening instruments available to complement an IQ test in providing a comprehensive picture of a student's abilities.

Recently, a new screening instrument has been developed for use in gifted identification, the *Gifted Rating Scales* (*GRS*; Pfeiffer & Jarosewich, 2003). 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. The *GRS-P* consists of five scales with 12 items each; the *GRS-S* consists of six scales with 12 items each. Developmental considerations were taken into account in designing the scale; the two forms share a similar format but item overlap is only 29%. The *GRS-S* includes a sixth, leadership scale.

The *GRS* is based on a multi-dimensional model of giftedness. The test 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). The scale is intended to complement an IQ test and other procedures used in gifted identification (e.g., auditions, portfolio review, nonverbal tests). Standardization of the *GRS* was co-linked to standardization of the new Wechsler Intelligence Scale for Children-Fourth Edition and Wechsler Preschool and Primary Scale of Intelligence-Third Edition.

Development of the *GRS* followed a carefully prescribed set of steps. Final item selection was guided by factor structure, item mean scores, consideration of parent education level, gender and ethnicity, inter-rater and test-retest reliability, and expert opinion.

What follows is a brief description of each of the scales: The *Intellectual Ability* scale measures a teacher's perception of a student's verbal and/or nonverbal mental skills, capabilities, and intellectual competence. The *Academic Ability* scale measures a teacher's perception of a student's skill in dealing with factual and/or school-related material. The *Creativity* scale measures a teacher's perception of a student's ability to think, act, and/or produce unique, original, novel or innovative thoughts or products. The *Artistic Talent* scale measures a student's potential for, or evidence of ability in

CONTINUED ON PAGE 107

CONTINUED FROM PAGE 106

Measuring Change with the GRS

drama, dance, drawing, singing, and/or playing a musical instrument. The *Leadership Ability* scale measures a student's ability to motivate others toward a common or shared goal. Items rate understanding social dynamics and displaying strong interpersonal communication and conflict resolution skills. The *Motivation* scale refers to a student's drive or persistence and ability to work well without encouragement. The motivation scale is *not* viewed as a measure of giftedness.

Each item is rated on a 9-point scale divided into three ranges: 1-3 *Below Average*, 4-6 *Average*, and 7-9 *Above Average*. The authors provide a classification system that indicates the *likelihood* that a student might be gifted, based on their *T* scores. The *GRS* does *not* presume to determine whether the child is gifted or not. The higher the student's *T* score on one or more gifted scales, the higher the *probability* that they are gifted. A *T* score below 55 (below 69%) indicates a low probability of being gifted, a score between 55-59 (69-83%) moderate probability, a score between 60-69 (84-97%) high probability, and a score above 70 (98+%) a very high probability. The manual emphasizes that test users should *always* look for other evidence to corroborate a classification of gifted, consistent with the guidelines in *Standards for Educational and Psychological Testing* (AERA, 1999). An analysis of the national standardization sample employing diagnostic efficiency and receiver operating curve statistics supports the validity of this classificatory scheme, particularly the intellectual and academic ability scales (Pfeiffer & Jarosewich, in press; Pfeiffer, Petscher, & Jarosewich, in press).

Purpose of the Study

In screening and evaluating gifted students, school psychologists frequently find that one or more academically precocious students come close but do not quite make the district or state cut-score. In other instances, a student may be *on the bubble* with a lack of corroborating evidence to support classifying the student as gifted. Quite often, typically underrepresented groups of gifted students—those who come from families where English is not the primary language spoken in the home, those from rural and/or low income families, and those of color— seem to show exceptional promise and yet do not score high enough on traditional measures to qualify for gifted programs (Ford, 1998; Naglieri & Ford, 2003; Pfeiffer, 2001).

The above diagnostic issues illustrate the dilemma that school psychologists often face with the uncertain decision of what to do when a bright student does not qualify for a gifted classification. In addition, school psychologists are uncertain how to measure change resulting from attending a gifted program. For these reasons, we developed statistically *and* conceptually sound, and easy-to-use, tables to augment clinical decision making in those instances when one is evaluating change using the *GRS* scales.

Method

The methodology that we adopted is based on comparing a student's original *GRS* scale score(s) with a range of scores that take into account the variability expected by both regression to the mean *and* measurement error. We calculated standard error of prediction (SE*p*) scores to create confidence bands for *T* scores so that practitioners could compare a second set of *GRS T* scores with an original set of *T* scores. We used the SE*p* rather than the *SEM* because it is preferable in providing an unbiased estimate of population measurement error (Atkinson, 1991).

Technically speaking, a student's obtained score on any measure is not the best estimate of her/his true score because of the phenomenon of regression toward the mean (Lord, 1956). SE*p*-based confidence bands are bracketed around a student's *predicted true score* and *not* around her/his obtained score because of the pernicious phenomenon of regression toward the mean. This phenomenon is common in gifted assessment because students typically score at or near the ceiling. Following the formula provided by Atkinson (1991), we calculated the SE*p* for *GRS T* scores,

$$SEp = SD\sqrt{1 - r}$$

where the *SD* is 10 and *r* is the internal reliability coefficient (Cronbach & Furby, 1970) for each *GRS* scale. The *GRS* reports high levels of internal consistencies across both forms, with r's ranging from .97-.99 for all *GRS-P* and *GRS-S* scales across the entire age range (Pfeiffer & Jarosewich, 2003). We then multiplied the SE*p* by 1.64 (90% level of significance) and by 1.96 (95% level of significance) to obtain a range of *T* scores that bracket the estimated true score at the .10 and .05 levels of significance. Posttest confidence ranges were calculated for scores ranging from a *T* score at the mean (50) to a *T* score three *SD* above the mean (80). These values appear in Table 1 for the *GRS-P*.

CONTINUED ON PAGE 108

CONTINUED FROM PAGE 107

Measuring Change with the GRS

Participants

Data used to generate Table 1 was obtained from the *GRS* standardization sample. The Psychological Corporation provided the authors with data files that included the data for the entire *GRS* standardization sample. The sample consisted of 975 children selected from across the country to match the U.S. census by ethnicity and by parent education level (U.S. Bureau of the Census, 2000). More detailed information on the standardization sample is available in the user manual (Pfeiffer & Jarosewich, 2003).

Using the *GRS* to Measure Change

Table 1 provides posttest confidence ranges for *GRS-P* pretest-posttest *T* score comparisons.[1] The Table provides initial or pretest *T* scores ranging from the mean to 3 SD above the mean. As mentioned above, the *GRS* reports high levels of internal consistencies across both forms, with r's ranging from .97-.99 for *all GRS-P* and *GRS-S* scales across the entire age span, 4:0-13:11 (*SEM's* ranged from 1.0-1.73) (Pfeiffer & Jarosewich, 2003, pp. 29-30). In creating the posttest confidence ranges, we used the decision rule of providing the most conservative confidence range (i.e., we always used the lowest reported internal reliability coefficient for a given scale when the reliability coefficient varied by age).

To use the Table, first determine which scale(s) the child's teacher rated. For example, assume that you are working with a kindergarten student. In early fall, her teacher completes a *GRS-S* record form, and she obtains a *T* score of 65 on the Academic Ability scale. Assume further that the school district has agreed that a *GRS* cut score of *T* ≥ 70 on the Academic Ability, Intellectual Ability or Creativity scale is the minimal criterion to 'trigger' a comprehensive gifted evaluation. The teacher completes a second *GRS-P* in the spring and the student now obtains a *T* score of 70 on Academic Ability. Table 1 allows the user to determine whether this student's second score demonstrates real change. Table 1 brackets posttest score confidence ranges at two levels of confidence (.95 and .90). The far left and right columns provide initial *T* scores. Recall that the student originally obtained a *T* score of 65 on Academic Ability. We first find the *T* score of 65, and then locate the column with the confidence range we want to use. Assume that we have decided *a priori* to use a .95 confidence level for gifted screenings. For an initial *T* score of 65, the posttest confidence interval for Academic Ability is 60-69. In this instance, the student's most recent *T* score of 70 on Academic Ability falls *outside* the SE*p* range. We can conclude, with 95% confidence, that the student's second score is different from her original score. In addition, the student's new score meets the criterion to trigger a more comprehensive gifted evaluation.

The following hypothetical case illustrates the application of the SE*p* methodology. The case is fictitious; however, it is based on the use of this methodology in an actual school district. This case highlights the use of the *GRS* as a screening tool. The gifted coordinator of an urban school district was concerned that, over the past few years, the numbers of students referred for gifted consideration had declined. Of particular concern was the fact that very few minority group students were referred for the gifted program. In consultation with a university faculty member (the first author), the gifted coordinator invited the kindergarten teachers to complete the *GRS-P* on all their students. In addition to the *GRS-P*, teachers were asked to submit one classroom product per student that depicted each student's "academic or intellectual ability." A gifted screening initiative committee that consisted of a school psychologist, two first grade teachers, the university faculty member, and the district gifted coordinator was formed.

The committee developed a simple rubric to rate each of the student products on a scale of 1-4. It was decided that any student who obtained a 3 or 4 on the rubric *and* a *T* score ≥ 60 on either the *GRS-P* Intellectual or Academic Ability scales ("high probability" *GRS* classification) would be recommended for a more in-depth gifted evaluation (a group nonverbal test and an individual, short form IQ test). The committee predicted that 10-20% of kindergarteners would be identified by this screening procedure.

Thirteen percent of the kindergarteners qualified as a result of the screening procedure. However, a small group of students obtained 4's on

[1]A Table providing posttest confidence ranges for the GRS-S is available from the first author.

CONTINUED ON PAGE 109

Table 1
GRS-P Posttest Confidence Ranges for Pretest-Posttest T-score Comparisons

	Intellectual		Academic		Creativity		Artistic		Motivation		
	2nd score range		2nd score range		2nd score range		2nd score range		2nd score range		
Pre-test Score	90% CI	95% CI	90% CI	95% CI	90% CI	95% CI	90% CI	95% CI	90% CI	95% CI	Pre-test Score
80	76 – 83	75 – 83	75 – 83	74 – 84	75 – 83	74 – 84	76 – 83	75 – 83	76 – 83	75 – 83	80
79	75 – 82	75 – 82	74 – 82	73 – 83	74 – 82	73 – 83	75 - 82	75 – 82	75 – 82	75 – 82	79
78	74 – 81	74 – 81	73 – 81	72 – 82	73 – 81	72 – 82	74 – 81	74 – 81	74 – 81	74 – 81	78
77	73 – 80	73 – 80	72 – 80	71 – 81	72 – 80	71 – 81	73 – 80	73 – 80	73 – 80	73 – 80	77
76	72 – 79	72 – 79	71 – 79	70 – 80	71 – 79	70 – 80	72 – 79	72 – 79	72 – 79	72 – 79	76
75	71 – 78	71 – 78	70 – 78	69 – 79	70 – 78	69 – 79	71 – 78	71 – 78	71 – 78	71 – 78	75
74	70 – 77	70 – 77	69 – 77	69 – 78	69 – 77	69 – 78	70 – 77	70 – 77	70 – 77	70 – 77	74
73	69 – 76	69 – 76	68 – 76	68 – 77	68 – 76	68 – 77	69 – 76	69 – 76	69 – 76	69 – 76	73
72	68 – 75	68 – 75	67 – 75	67 – 76	67 – 75	67 – 76	68 – 75	68 – 75	68 – 75	68 – 75	72
71	67 – 74	67 – 74	66 – 74	66 – 75	66 – 74	66 – 75	67 – 74	67 – 74	67 – 74	67 – 74	71
70	66 – 73	66 – 74	65 – 73	65 – 74	65 – 73	65 – 74	66 – 73	66 – 74	66 – 73	66 – 74	70
69	65 – 72	65 – 73	64 – 72	64 – 73	64 – 72	64 – 73	65 – 72	65 – 73	65 – 72	65 – 73	69
68	64 – 71	64 – 72	63 – 71	63 – 72	63 – 71	63 – 72	64 – 71	64 – 72	64 – 71	64 – 72	68
67	63 – 70	63 – 71	63 – 70	62 – 71	63 – 70	62 – 71	63 – 70	63 – 71	63 – 70	63 – 71	67
66	62 – 69	62 – 70	62 – 70	61 – 70	62 – 70	61 – 70	62 – 69	62 – 70	62 – 69	62 – 70	66
65	61 – 68	61 – 69	61 – 69	60 – 69	61 – 69	60 – 69	61 – 68	61 – 69	61 – 68	61 – 69	65
64	60 – 67	60 – 68	60 – 68	59 – 68	60 – 68	59 – 68	60 – 67	60 – 68	60 – 67	60 – 68	64
63	59 – 66	59 – 67	59 – 67	58 – 67	59 – 67	58 – 67	59 – 66	59 – 67	59 – 66	59 – 67	63
62	58 – 65	58 – 66	58 – 66	57 – 66	58 – 66	57 – 66	58 – 65	58 – 66	58 – 65	58 – 66	62
61	58 – 64	57 – 65	57 – 65	56 - 65	57 – 65	56 - 65	58 – 64	57 – 65	58 – 64	57 – 65	61
60	57 – 63	56 – 64	56 – 64	55 – 64	56 – 64	55 – 64	57 – 63	56 – 64	57 – 63	56 – 64	60
59	56 – 62	55 – 63	55 – 63	54 – 63	55 – 63	54 – 63	56 – 62	55 – 63	56 – 62	55 – 63	59
58	55 – 61	54 – 62	54 – 62	53 – 63	54 – 62	53 – 63	55 – 61	54 – 62	55 – 61	54 – 62	58
57	54 – 60	53 – 61	53 – 61	52 – 62	53 – 61	52 – 62	54 – 60	53 – 61	54 – 60	53 – 61	57
56	53 – 59	52 – 60	52 – 60	51 – 61	52 – 60	51 – 61	53 – 59	52 – 60	53 – 59	52 – 60	56
55	52 – 58	51 – 59	51 – 59	50 – 60	51 – 59	50 – 60	52 – 58	51 – 59	52 – 58	51 – 59	55
54	51 – 57	50 – 58	50 – 58	49 – 59	50 – 58	49 – 59	51 – 57	50 – 58	51 – 57	50 – 58	54
53	50 – 56	49 – 57	49 – 57	48 – 58	49 – 57	48 – 58	50 – 56	49 – 57	50 – 56	49 – 57	53
52	49 – 55	48 – 56	48 – 56	47 – 57	48 – 56	47 – 57	49 – 55	48 – 56	49 – 55	48 – 56	52
51	48 – 54	47 – 55	47 – 55	46 – 56	47 – 55	46 – 56	48 – 54	47 – 55	48 – 54	47 – 55	51
50	47 - 53	46 – 54	46 – 54	45 – 55	46 – 54	45 – 55	47 - 53	46 – 54	47 - 53	46 – 54	50

Note: 90% CI = 90% confidence interval; 95% CI = 95% confidence interval

CONTINUED ON PAGE 110

CONTINUED FROM PAGE 109

Measuring Change with the GRS

their rubric but *GRS-P T* scores less than 60. A large proportion of these *almost qualified* were minority group students and students from families where English was not the primary language spoken in the home. The committee decided to look at this *almost qualified* group of students again in 3 months, after they had more time in kindergarten and more opportunity to acclimate to the kindergarten learning environment.

One such student, Javier, was a six-year-old, 1st generation Hispanic male. Javier's parents spoke Spanish in the home and neither parent completed high school. Prior to kindergarten, Javier did not attend a preschool or daycare. During the gifted screening, Javier obtained *T* scores of 55 on both the *GRS-P* Intellectual and Academic ability scales. However, his drawing and accompanying narrative to his picture that was transcribed by the teacher aide was rated 4 on the rubric. Javier's teacher was asked to complete a second *GRS-P* three months following the first ratings.

Table 1 indicates that if Javier originally obtained a *T* score of 55, then he would need to obtain a second *T* score ≥ 60 on the Intellectual Ability scale and a second *T* score ≥ 61 on the Academic Ability scale for the screening committee to conclude, with 95% certainty, that the more recent ratings were different from his first ratings. Three months later, Javier obtained *T* scores of 60 on both the *GRS-P* Intellectual and Academic ability scales. Javier's second set of scores met the criteria set by the gifted screening committee. Equally important, his second Intellectual Ability rating exceeded the respective SE*p* range of 51-59 at the .95 confidence level. The committee legitimately (and ethically) concluded that Javier qualified for a gifted evaluation.

Discussion

An issue that has generated considerable discussion in the measurement literature is the topic of just how much change is needed to be considered meaningful. The question addresses the practical importance of statistical effects. The child therapy literature has grappled with this conceptual issue–even large effect sizes can be clinically insignificant (Jacobson, Roberts, Berns, & McGlinchey, 1999).

Gifted classification systems are not nearly as refined or researched as the multi-axial DSM classification system. It remains difficult demarcating exactly where normal ends and abnormal begins (one reason for the category *sub-clinical*). Presently, the distinction between gifted and not gifted is even more blurry. This is an important challenge facing the gifted field (Gagné, 1998; Pfeiffer, 2003).

The *GRS* classification system indicates the *likelihood* that a student might be gifted. The higher the student's *T* score on one or more of the scales, the higher the *probability* that they are gifted in that domain. The classification system proposes that a *T* score below 55 (below 69%) is unlikely to reflect giftedness; a score between 55-59 (69-83%) suggests moderate probability; a score between 60-69 (84-97%) high probability; and a score above 70 (98%) very high probability. To be considered *diagnostically meaningful*, the second *T* score should exceed the posttest confidence range and be ≥ 60 (high probability) or ≥ 70 (very high probability of gifted). Recent validity studies provide preliminary validation for this gifted classification system (Pfeiffer & Jarosewich, in press; Pfeiffer, Petscher & Jaorosewich, in press; Ward, 2005).

A single test score should *never* be used alone in making any diagnostic or classificatory decision (Pfeiffer, 2002). The case of Javier illustrates the obvious benefit of using multiple screening measures. Overall predictive accuracy is increased with the use of technically sound multiple measures (Pfeiffer, 2002). Interestingly enough, only four states, Hawaii, Pennsylvania, Virginia and Washington, refer to the use of *multiple criteria* for gifted identification (Stephens & Karnes, 2000). There is *no* one best test or test battery for screening or classifying gifted. Authentic assessment, review of portfolio material, auditions, tryouts and interviews complement teacher ratings in providing important data when used as part of a comprehensive gifted evaluation.

"A single test score should never be used alone in making any diagnostic or classificatory decision (Pfeiffer, 2002). The case of Javier illustrates the obvious benefit of using multiple screening measures."

Please e-mail all submissions for The Commentary Section to: LReddy2271@aol.com

CONTINUED ON PAGE 111

CONTINUED FROM PAGE 110

Measuring Change with the GRS

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Reseach Association.

Atkinson, L. (1991). Three standard errors of measurement and the Wechsler Memory Scale-Revisited. *Psychological Assessment, 3,* 136-138.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—or should we? *Psychological Bulletin, 74*, 68-80.

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.

Gagné, F. (1998). A proposal for subcategories within gifted or talented populations. *Gifted Child Quarterly, 42*, 87-95.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67*, 300-307.

Jarosewich, T., Pfeiffer, S. I., & Morris, J. A. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment, 20*, 322-336.

Lord, F. M. (1956). The measurement of growth. *Educational and Psychological Measurement, 16,* 421-437.

Margolis, A. S., & Floyd, R. G. (2004). Test review: The Gifted Rating Scales. *Journal of Psychoeducational Assessment, 22*, 275-282.

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.

Pfeiffer, S. I. (2001). Professional psychology and the gifted: Emerging practice opportunities. *Professional Psychology: Research and Practice, 32*, 175-180.

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: The Psychological Corporation.

Pfeiffer, S. I., & Jarosewich, T. (in press). The Gifted Rating Scales-School Form: An analysis of the standardization sample based on age, gender and race and diagnostic efficiency. *Gifted Child Quarterly.*

Pfeiffer, S. I., Petscher, Y., & Jarosewich, T. (in press). The GRS-Preschool/Kindergarten From: An analysis of the standardization sample based on age, gender, race. *Roeper Review*.

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.

Sparrow, S. S., Pfeiffer, S. I., & Newman, T. M. (2005). Assessment of children who are gifted with the WISC-IV. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.). WISC-IV: *Clinical use and interpretation* (pp. 282-299). Burlington, MA: Elsevier Academic Press.

Stephens, K. R., & Karnes, F. A. (2000). State definitions for the gifted and talented revisited. *Exceptional Children, 66,* 219-238.

U.S. Bureau of Census. (2000). *Current population survey, October 2000: School Enrollment Supplement File* [CD-ROM]. Washington, DC: U.S. Bureau of Census (Producer/Distributor).

Ward, S. A. (2005). Review of the Gifted Rating Scales. In B. A. Spies & B. S. Plake (Eds.). *The Sixteenth Mental Measurements Yearbook* (pp. 404-407). Lincoln, NE: Buros Institute of Mental Measurements.

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 Edition* (pp. 3-21). Amsterdam: Elsevier.