DRAFT DRAFT DRAFT

A SIMPLE PROCEDURE FOR COMBINING ABILITY TEST SCORES, ACHIEVEMENT TEST SCORES, AND TEACHER RATINGS TO IDENTIFY ACADEMICALLY TALENTED CHILDREN

David Lohman & Joseph Renzulli

April 18, 2007

In January of 2007, Circuit Judge Michael Nowakowski ruled that the Wisconsin Department of Public Instruction must prepare specific rules for school districts to use in identifying gifted and talented students. As in many states, the current Wisconsin rules for identifying students in need of gifted and talented (G&T) services require only that school districts use "multiple criteria that are appropriate for the category of gifted including intelligence, achievement, leadership, creativity, product evaluations, and nominations." Responding to a suit brought by a parent whose child was denied services in one school but had received them in another school, Judge Nowakowski observed that rules for identifying these children ought to be consistent across schools.

In this paper we describe a relatively simple procedure for using the primary identification measures – ability test scores, achievement test scores, and teacher ratings – to help identify those children most in need of academic acceleration or enrichment. The procedures that many schools use for identifying gifted children evolved in an earlier era when IQ scores were the sole criteria of giftedness. Most G&T programs were designed to serve only those children who clearly exhibited well-developed academic and cognitive abilities when compared with all other children in the nation – those children whose academic giftedness was, as Callahan (2005) put it, "signed, sealed, and delivered."

However, if programs are to become more inclusive, then the goal must be to identify and develop extraordinary talent in all children. Emphasizing talent identification and development rather than giftedness changes the focus of the identification process. Identification becomes an ongoing search for students who show talent for competencies that can be developed in schools rather than a fruitless effort to measure ability in a way that is uninfluenced by culture, education, or opportunity to learn.

Schools that hope to develop more inclusive G&T programs must provide a wider range of educational opportunities than can be offered in a single pull-out class. Diverse services require diverse selection criteria. Indeed, the golden rule of identification is that "there must be congruence between the criteria used in the identification process and the goals and types of services that constitute the day-to-day activities that students will pursue (Renzulli, 2005, p. 11). Some children will be ready to receive some or all of their instruction with children several years their senior, or participate in special classes, or attend schools that can provide appropriate levels of challenge. Others need focused instruction in particular domains – such as learning to speak fluently the dialect of the language they are expected to read and to write – but can profit from on-grade level instruction in other areas. And some students will thrive from special projects that allow them to pursue their interests in topics, develop relationships with mentors, or engage in other enrichment activities. The development of creative-productive abilities in domains that elicit the student's interest and strengthen their commitment to the pursuit of excellence should be the goal of every program. One size does not fit all unless the identification system reduces children to one size.

As in the schools that Judge Nowakowski reviewed, it is common practice to collect many different kinds of information about students, arrange this information in a matrix, and then combine it in some way to decide which children to admit to the G&T program. Invariably, however, the rules for weighting and combining this information are somewhat arbitrary and, even when well-intended, can have unintended consequences.

For example, in an effort to increase the number of underrepresented minority students, some schools use only nonverbal tests to screen all children. However, nonverbal tests exclude many of the most

academically talented children in all ethnic groups – especially among African American children. Other programs use more valid ability and achievement tests but have broadened the criteria for admission to include teacher ratings, behavioral checklists, and other observational scales. Adding points for these measures to points earned for test scores can increase the diversity of the population of students who are admitted, especially when ratings are weighted heavily. However, unless raters are well-trained, the information that they provide can be quite unreliable, invalid, or both.

Further, when the number of children who are served by the program is fixed, for every child who is admitted, another child is denied access to the program. When evaluating the efficacy of an identification system, it is important to consider these children as well. If schools do not do this, then disgruntled parents and their lawyers are increasingly likely to do it for them. For both minority and non-minority children, the critical question is whether the system actually identifies and properly assists the students whose talents can best be developed by the program. If the identification criteria lack reliability and validity, many children who should be served will be excluded, and many in the group that are selected will not retain their status a year or two later. In other words, an identification system that seems to be achieving the goal of inclusion can be fundamentally flawed.

When challenged in court, good intentions are not enough. For example, although one could certainly justify providing enrichment opportunities to a child whom a teacher rates as highly creative, it would be difficult to defend a procedure that denied the opportunity for advanced instruction to a child who received lower ratings on creativity but obtained much higher ability and achievement test scores. Yet this can easily happen when points are summed across diverse measures. The system we propose allows schools to make good use of teacher ratings without the problems that ensue when points for teacher ratings are simply added to the mix. The weights that we use to combine measures are well-grounded in research on talent development. Further, we show how to use different kinds of information in a way that can inform fundamental decisions about instruction – especially the need for advanced instruction or for enrichment.

The identification procedures described here were developed using Form 6 of the Cognitive Abilities Test (*CogAT*; Lohman & Hagan, 2001), the Iowa Test of Basic Skills (*ITBS*; Hoover, Dunbar, & Frisbie, 2001), and the Scales for Rating the Behavioral Characteristics of Superior Students (*SRBCSS* ; Renzulli et al., 2002). The procedures assume that ability and, if possible, achievement test scores are available for all students, but that only those students who have been nominated for the program are rated by their teachers. We recommend the use of both ability and achievement test scores because the best prediction of subsequent success in school is given not by ability scores alone, or by achievement test scores alone, but by appropriate combinations of the two. But before scores can be combined, they must be put on the same score scale. The best way to do this is to put scale scores for all tests into a spreadsheet, convert the scale scores to standard (or z) scores, and then combine the appropriately weighted standard scores.[1] Directions for doing this are given in Lohman (2005) and in the accompanying sample data set that is available on the web.

However, if this is not possible, then a simpler approximation is to assign point values to percentile ranks (national or local) using the tables below. The point system transforms percentile scores to a common metric so that they can be combined.[2] We first illustrate the case in which achievement test scores are not available for all children. This is commonly the case prior to spring of third grade.

---

[1] Averaging test scores that are not on the same scale weighs the test on which scores differ the most. Percentile ranks (PR) should generally not be averaged. The PR of the average scale score is generally not the same as the average of two PRs.

[2] Information is lost whenever several values map on to a single value. For example, rounding age measured in years and months to the nearest year discards information. Rounding to the nearest decade discards even more.

Worksheet 1: CogAT and SRBCSS only

CogAT	PR	Points (Table 1)		Enter Figure 1
Verbal	_____	_____		
Quantitative	_____	_____	_____ QN Average	_____ Maximum
Nonverbal	_____	_____		

SRBCSS	Rating	Local Average	Above Average?		
Learning Ability	_____	_____	yes/no		
Creativity	_____	_____	yes/no	Any "yes"?	_____
Motivation	_____	_____	yes/no		

Procedure 1: CogAT and SRBCSS only

1. Enter percentile ranks (PRs) from the three CogAT batteries (Verbal, Quantitative, and Nonverbal) in the first column of the worksheet. [3]
2. Convert Percentile Ranks (PRs) to points using Table 1 (see next page). Enter these points in Worksheet 1.

Table 1
Use to convert PR from any test or CogAT SAS scores to points

Points	PR	SAS
1	80-83	113-115
2	84-88	116-119
3	89-92	120-123
4	93-95	124-127
5	96-97	128-131
6	98	132-135
7	99	136-139
8	99+	140+

3. Average the points for the Quantitative and Nonverbal batteries. Enter this value on the worksheet.

[3] You can use either Local or National PR's (SAS scores are based on national age norms). However, the same comparison group (local or national) must be used for all scores.

4. Enter ratings for the three SRBCSS scales.
5. Compute the average teacher rating on each of the three *SRBCSS* scale for the group of students who were nominated for the program. (If the on-line version of the SRBCSS is used, this information will be generated automatically). If ratings are available for fewer than 15 – 20 children, then use your best judgment to identify those children who obtain strong ratings on one or more of the *SRBCSS* scales from those who receive ratings that are not as strong.

For each child, then, one will have three pieces of information: (1) points for the CogAT Verbal PR, (2) points for the average of the CogAT Quantitative and Nonverbal PRs, and (3) a "yes" or "no" indicating whether any of the three ratings was above average. These three pieces of information are then combined using the identification scheme shown in Figure 1.

The vertical dimension distinguishes children who exhibit superior reasoning abilities in either the verbal domain or in the quantitative-nonverbal domain from those who exhibit strong but less stellar reasoning abilities in these domains. We have set two cut scores. One identifies those students who score at or above the 96th percentile rank; the other identifies those students who score at or above the 80th percentile rank (but below the 96th PR) on either verbal reasoning or quantitative-nonverbal reasoning. These percentile-rank criteria are commonly used in gifted programs. Although national norms can be used for this purpose, we strongly recommend that schools use local norms. Local norms can be requested from the publisher when ordering CogAT score reports and can output from the on-line version of the SRBCSS. We recommend local norms because the need for special services depends primarily on the disparity between children's cognitive and academic development and that of the other children in the classes that they attend, not all other children in the nation at the time that the test was normed.[4]

Figure 1

CogAT and SRBCSS Ratings Only

		Teacher Rating on Learning Ability, Motivation, or Creativity	
		Below average teacher ratings	Above average teacher ratings
***CogAT* Verbal OR Quantitative-Nonverbal Reasoning**	5 or more points (≥96th PR)	II	I
	1 – 4 points (80th – 95th PR)	IV	III

The horizontal dimension of the matrix distinguishes between children who, when compared to other children nominated for the program, obtain above average teacher ratings and students who obtain average or below average teacher ratings. Research with the *SRBCSS* shows that each of its three main scales provides unique information. Therefore, the decision rule we suggest is that teacher ratings be

[4] Further, norms for many ability tests (e.g., the Raven Progressive Matrices) are out of date and thus far too easy.

considered high if any of the three ratings is high. Note that, for ratings, the average is computed only on the subset of the student population who are nominated for inclusion in the program. If there is no variability in these ratings, then procedures for nominating children are too restrictive or raters are poorly trained. Renzulli (2005) shows how to ensure that an appropriately wide net is cast when forming the initial talent pool and how to train teachers to perform the difficult but important task of the rating students. Of course, schools can implement a rule that is either more stringent or more lenient than *above (*or *below) average*.

Combining these two criteria gives four categories of assessment results.

- Children in Category I exhibit superior reasoning abilities on *CogAT* and are rated as highly capable, motivated, or creative by their teachers.
- Children in Category II also exhibit superior reasoning abilities but, when compared to other children who were nominated, are not rated as highly by their teachers on any one of the three major scales of the SRBCSS. Programs that follow a traditional identification scheme (e.g., self-contained classrooms or schools) would accept children in Category I. Most would also accept children in Category II, given the difficulty of defending rejections on the basis of low teacher ratings. However, the progress of children in Category II should be monitored more closely. For example, once they are available, achievement test scores should be considered as well (see Procedure 2).
- Children in Category III exhibit somewhat lower but strong reasoning abilities (80^{th} to 95^{th} PR) on *CogAT*, and are rated as highly capable, motivated, or creative by their teachers. These children would be included in school-wide enrichment programs that aim to serve a broader range of children than are served by traditional programs (Renzulli, 2005). Schools that serve many poor children would find that many of their best students would fall in this category, especially when using national rather than local (i.e., school) test norms. Combining test scores and ratings in this way would enable these schools to identify the students most likely to benefit from curriculum compacting or enrichment programs, including instruction at a higher level than that received by most other students in the school.
- Finally, children in Category IV exhibit good but not exceptional reasoning abilities (between 80^{th} and 95^{th} PR), and are not rated as unusually capable, motivated, or creative by their teachers. Although good students, these children would not be provided with special programming on the basis of either their *CogAT* scores or teacher ratings. However, they should be reconsidered when information on achievement is available (see Procedure 2). Indeed, given the inevitability of regression to the mean in all status scores (e.g., percentile ranks and teacher ratings) (Lohman & Korb, 2005; Lohman, 2006), the level of participation of all children in the program should be routinely re-evaluated (Renzulli, 2005). This can be made more palatable if children (and their parents) are told that they are being identified for participation in a talent development program in a particular domain rather than being identified as "gifted."

Procedure 2: Incorporating Achievement Test Scores

The preferred procedure for identifying students most likely to need (and to profit from) higher levels of academic challenge is to use both achievement and ability test scores. How best to combine scores from the three *CogAT* batteries with Reading and Mathematics achievement test scores when predicting future academic success is well documented in the research literature. Importantly, the weights that should be applied to each test battery in making these predictions are the same for all ethnic groups that have been studied (Lohman, 2005). Competence in a broad range of verbal domains (e.g., social studies and literary arts) is best predicted by an equally weighted average of the CogAT Verbal score and the Reading Total

score from the achievement test. On the other hand, success in mathematics and domains of study that demand quantitative thinking is best predicted by a combination of the CogAT Quantitative and Nonverbal Reasoning Batteries and the Mathematics Total score from the achievement test..

Here are steps for implementing this procedure.

1. Enter percentile ranks (PRs) from the three CogAT batteries (Verbal, Quantitative, and Nonverbal) in the first column of the worksheet. [5]
2. Convert Percentile Ranks (PRs) to points using Table 1 (see next page). Enter these points in Worksheet 1.
3. Average the points for the Quantitative and Nonverbal batteries. Enter this value on the worksheet.
4. Sum the points for CogAT Verbal and Reading Total.
5. Sum the points for the CogAT QN Composite (from step 3) and Mathematics Total.
6. Enter ratings for the three SRBCSS scales.
7. Compute the average teacher rating on each of the three *SRBCSS* scale for the group of students who were nominated for the program. (If the on-line version of the SRBCSS is used, this information will be generated automatically). If ratings are available for fewer than 15 – 20 children, then use your best judgment to identify those children who obtain strong ratings on one or more of the *SRBCSS* scales from those who receive ratings that are not as strong.

Worksheet for Procedure 2: CogAT, SRBCSS, and Achievement

Test	PR	Points (Table 1)			Enter Figure 2
CogAT V	_____	_____			
Reading Total	_____	_____	_____ V-RT Sum		
CogAT Q	_____	_____			_____ Max
CogAT NV	_____	_____	_____ QN Avg		
Math Total	_____	_____	_____ QN-M Sum		

SRBCSS	Rating	Local Average	Above Average?		
Learning Ability	_____	_____	yes/no		
Creativity	_____	_____	yes/no	Any "yes"?	_____
Motivation	_____	_____	yes/no		

[5] You can use either Local or National PR's (SAS scores are based on national age norms). However, the same comparison group (local or national) must be used for all scores.

The point totals for the composite verbal/reading total and the composite quantitative/nonverbal/mathematics total can now be used to identify students. Figure 2 assumes that the cut points are set at the 80th and 96th PRs. Approximate point totals that correspond to other percentile ranks are given in Table 2 (see next page).

Figure 2

Using CogAT, Achievement test scores, and Teacher Ratings

		Teacher Rating on Learning Ability, Motivation, or Creativity	
		Below average teacher ratings	Above average teacher ratings
***CogAT* Verbal + Reading T. OR *CogAT* QN + Math T.**	8 or more points (≥96th PR)	II	I
	2 – 7 points (80th – 95th PR)	IV	III

Categories I – IV are interpreted as in Figure 1.

One of the main advantages of using both ability and achievement test scores is that there will be considerably less regression to the mean – especially out of groups I and II -- when students are retested the following year (see Lohman & Korb, 2006). Keep in mind, however, that these procedures are but a helpful first step in creating a Talent Pool (Renzulli, 2005). They show how to make effective use of some of the most important information that should be gathered. But they do not show how to incorporate all of the other information that can usefully inform how best to assist students in developing their talents. For example, student interests should always be assessed and used to help students direct their choice of educational activities. For suggestions on next steps, see Renzulli (2005) or www.renzullilearning.com

References

Callahan, C. M. (2005). Identifying gifted students from underrepresented populations. *Theory into Practice, 44,* 98-104.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Test of Basic Skills: Form A.* Itasca, IL: Riverside.

Lohman, D. F. (2005). The role of nonverbal ability tests in the identification of academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49,* 111-138.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.

Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in *ITBS* and *CogAT* scores during elementary school. *Journal for the Education of the Gifted*, 29, 451-484.

Renzulli, J. S. (2005). *Equity, excellence, and economy in a system for identifying students in gifted education: A guidebook* (RM05208). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. L. (2002). *Scales for Rating the Behavioral Characteristics of Superior Students*. Mansfield Center, CT: Creative Learning Press.

Table 2

Approximate PR values for the sum of two scores
(use only if PR values other than those show in Figure 2 are desired)

Sum of 2 scores	PR
0	72
1	75
2	80
3	84
4	87
5	90
6	92
7	94
8	96
9	97
10	98
11	99
12	99
13	100
14	100