Lohman, D. F. (in press). Nontraditional uses of traditional measures. In C. M. Callahan & H. Hertberg-Davis (Eds.) *Fundamentals of gifted education.* NY: Taylor & Francis/Routledge.

Identifying Gifted Students: Nontraditional Uses of Traditional Measures

by David F. Lohman
In C. M. Callahan & H. Hertberg-Davis (Eds.) *Fundamentals of gifted education.*
NY: Taylor & Francis/Routledge (in press)

How can schools make better use of information they routinely collect that can help them identify academically talented students? How can they use the same kinds and diversity of information they use for talent identification with middle-class, English-speaking students to identify academically talented English language learners (ELL) or students with a low socioeconomic status (SES)? How can educators avoid being misled by poorly normed tests or misusing scores from state achievement tests that have no norms?

At some point, these questions and underlying measurement issues must be considered when discussing talent identification and development programs (Lohman, in press; Lohman & Foley Nicpon, 2012). Developing a good identification system requires attention to much more than whether a student's test scores exceed some preordained cut score. Fortunately, there are several easy-to-implement procedures that can dramatically increase the effectiveness of talent identification. The goals of this chapter are to explain and illustrate these simple procedures. The common thread is the use of test scores and other information that schools commonly collect (or could collect). However, this information will be interpreted (or reinterpreted) in the local context. This means asking for or developing local norms on ability and achievement test scores. It also means taking into account the student's opportunity to learn when making inferences about talent. Neither of these steps is difficult, but each requires moving beyond entrenched beliefs about test score use. We begin with a discussion of test score scaling and its use in above-grade-level testing. Although this is an unconventional use of test scores, it is only a small step off the beaten path. Next, we introduce local norms: why they are important and how they can be obtained from testing companies or developed using a spreadsheet. Third, we go one step beyond local norms to subgroup norms or comparisons. These procedures offer the most defensible method for identifying academically talented poor, ELL, and minority students. Finally, we show how to integrate these strands into a simple scheme for combining ability test scores, achievement test scores, and teacher ratings.

Above-Grade-Level Testing

Above-grade-level testing refers to the practice of assessing a gifted child with a test designed for older children. Sometimes gifted children are administered a higher level of a test in order to obtain a better estimate of their abilities than can be derived from an on-grade-level test that is too easy for them. At other times, however, the goal is to make a decision about acceleration for the child. In such cases, it can be helpful to know how the child's achievement in a domain compares with the achievement of older children. For example, a third grade child may be administered a mathematics achievement test at the level that is designed for seventh grade children. Here the goal is not to obtain a better estimate of the child's rank compared to other third grade children, but rather to estimate her rank compared to seventh graders on the seventh grade test.

Scores on achievement and ability tests that span multiple age or grade groups are typically placed on a common, developmental score scale. This is most easily done when the test consists of one long string of items ordered by difficulty. The scale score that the child obtains can be interpreted using norms for children at different age or grade levels. On the *Cognitive Abilities Test* (*CogAT*) (Lohman, 2011), scale scores are called "Universal Scale Scores" (USS); on the *Iowa Tests of Basic Skills (ITBS)* they are called "Scale Scores" (SS). Both of these scales are constructed by comparing the performance of adjacent age (or grade) groups on common subsets of items.

Whether above-grade-level achievement testing is appropriate or inappropriate depends on how the scores will be interpreted. For example, is the goal to obtain a better estimate of the child's ability when compared to other children of the same age? Inferences about *ability* assume that the child's opportunity to acquire the knowledge or skills presented in the test is similar to other children in the norm group. If the child is less familiar with some of the test content, then the above-grade-level achievement test could underestimate the child's ability. Or is the goal to decide whether a third grade child is likely to succeed in a seventh grade mathematics class? In this case, the appropriate norm group consists of seventh grade students in the class that the child might attend. The test that gives the best information on ability compared to age-mates may not provide the best guidance for acceleration. Similarly, the test that best guides acceleration may not provide the best estimate of ability compared to age-mates.

Using Local Norms

Most test users rely exclusively on national norms when interpreting scores on ability and achievement tests. Often, they are not aware of the limitations of national norms and the extent to which other normative comparisons can assist them. In the field of gifted education, the exclusive use of national norms stems, in part, from early definitions of giftedness based on *Stanford-Binet* IQ scores. In the early years of ability testing, psychologists did not understand the extent to which IQ scores change with age and experience, and how the methods used to scale the Binet tests impacted the range of scores that the test produced (see Lohman & Foley Nicpon, 2012). Psychologists also did not appreciate the extent to which performance on ability tests was improving across decades especially for young children on the nonverbal items (Flynn, 1999; Thorndike, 1975).

Advantages of National Norms

Well-developed national norms on ability tests have a number of desirable characteristics. When national norms are used, the scores of test-takers are compared to a common standard defined by the performance of a representative national sample of children of the same age or grade. When the same test is administered to a new class of students and interpreted using the same national norms, variation in the abilities of different cohorts of students who are being considered for talent development programs are readily documented.

When national norms are developed, vagaries of performance due to sampling are typically removed by careful smoothing of score distributions across age or grade groups (for an explanation of the process, see Lohman, 2009). The accuracy and stability of scores obtained using good national norms are particularly important for interpreting score profiles across subtests or batteries within the test.

Limitations of National Norms

However, national norms can be misleading, especially on ability tests that do not require a high level of professional certification to administer. National norms on many group ability tests used in schools are seriously deficient. Some widely used tests have never been properly normed and give IQ scores that are 10 to 17 points too high; others have been normed but not recently, and on at least one widely used test, normative scores were incorrectly computed, vastly over-identifying the number of very high- and very low-scoring children (for details, see Lohman, Korb, & Lakin, 2008). In such cases, national norms are more harmful than helpful.

Advantages of Local Norms

The primary limitation of national norms is the failure to take into account local variations in ability or achievement. However, the need for special programming at the local level depends on the discrepancy between the individual student's current levels of cognitive or academic development and that of his classmates not that of all other students in the nation. In some schools, the average student scores at the 20th national percentile (NPR). In such a school, a student who scores at the 70th NPR is probably significantly mismatched with her peers. Conversely, in some very high-achieving schools, a child who scores at the 95th NPR may not be seriously mismatched with the instructional challenges in the classroom. Because schools vary widely in the average ability and achievement of their students, policies that require all students in the district or state to attain the same level of excellence on a nationally normed test result in some schools in which no children are served by the program, and other schools in which a substantial fraction of the children are labeled "gifted." Local norms eliminate both of these problems.

Local norms also allow users to focus on how best to measure academic talent rather than on finding a test with national norms that give the desired percentage of "gifted" students overall or within subgroups of the population. Often, tests that achieve these goals either have out-of-date (or inaccurate) norms or they measure only a limited aspect of scholastic aptitude. Poorly normed nonverbal tests that measure only figural reasoning abilities suffer on both counts.

Use of local norms, especially when presented as percentiles, also helps educators to avoid many of the problems that attend IQ scores and test scores merely reported on an IQ-like scale.[1] Chief among these problems is the common but erroneous belief that IQ tests measure innate abilities that should remain constant as the child matures. It is much easier for parents and teachers to believe that the child's local percentile rank on a reasoning test can be expected to change with experience and opportunity than for them to understand that IQ scores are similarly changeable.

Limitations of Local Norms

Local norms typically represent the performance of a particular sample of students only for the year in which the test was administered. Thus, they require census testing (i.e., testing all second grade children rather than testing only those children nominated for the program). Furthermore, because the normative scores are based on a small sample, the local percentile rank (PR) scores are less stable than national PRs. However, this may not be an issue if the goal is simply to identify and serve the top X percent of students in the school. Indeed, simple ranks (rather than percentile ranks) may be sufficient. Later discussion in this paper will elaborate on this approach.

CogAT®

LIST OF STUDENT SCORES

Cognitive Abilities Test™ (CogAT®)

Class: Halladay
Building: Oak Elementary
District: Elm ISD
Form-Level: 7-9
Test Date: 10/2012
Norms: Fall 2011
Grade: 3
Page: 1

STUDENT NAME / I.D. Number / F-1 F-2 F-3 Code / ABCDEF G H I J K L M NOPZ	Birth Date / Age / Program	Level / Form	(Gender)		No. of Items	No. Att	Raw Score	USS	AGE SCORES SAS	AGE SCORES PR	AGE SCORES S	GRADE SCORES PR	GRADE SCORES S	LOCAL SCORES PR	LOCAL SCORES S	Student Profile APR Graph (1 10 25 50 75 90 99)	Profile
Bagsby, Aiden	12/03	9	(F)	Verbal	62	62	57	190	115	79	7	82	7	95	8	79	7A
0000152607	09-02	7		Quantitative	52	52	46	188	114	77	7	80	7	94	8	77	
				Nonverbal	56	56	47	188	114	76	6	79	7	93	8	76	
				Composite (VQN)				**189**	**114**	**78**	**7**	**81**	**7**	**95**	**8**		
Brigerton, Ryan	07/04	9	(F)	Verbal	62	62	19	139	79	6	2	9	2	23	3	6	2A
0000131198	08-03	7		Quantitative	52	52	17	144	83	10	2	13	3	27	4	10	
				Nonverbal	56	56	13	147	85	13	3	16	3	30	4	13	
				Composite (VQN)				**143**	**82**	**9**	**2**	**12**	**3**	**26**	**4**		

Figure 12.1 Portion of a report showing national age scores, national grade scores, and locally normed scores on Form 7 of the Cognitive Abilities Test. USS = Universal Scale Score; SAS = Standard Age Score; PR = Percentile Rank; S = Stanine.

Multiple Perspectives

Fortunately, one does not have to interpret test scores using a single normative perspective. For example, some group-administered ability tests report both age and grade norms as well as local norms that are calculated from the data submitted by a given school or school district. A score that is not unusual from one perspective may be unusual when viewed from another perspective. Figure 12.1 shows an example from Form 7 of the *Cognitive Abilities Test (CogAT)* (Lohman, 2011). The columns for “Age Scores” and “Grade Scores” use national norms. The “Local Norms” column reports percentile ranks for the distribution of Standard Age Scores (SAS) in the local group that this tested. The value of each perspective depends on the inferences that will be made from the test scores. If the goal is to identify the most talented students in the school or district, then local norms provide critical information. For example, scores on the Verbal Battery for the first student (Alden Bagsby) would not be considered remarkable when compared to all other children in the nation. Alden’s national age percentile rank (APR) for the Verbal Battery is only 79 and the national grade percentile rank is only slightly higher at 82. However, the local percentile rank for the Verbal score is 95. He may well benefit from a greater challenge than he is currently experiencing in his classes.

Computing Local Norms

Rank Orders

Precise local norms are not needed for many school-based talent development programs. The problem is akin to identifying athletes for the varsity basketball or track team. All that is needed is some way to identify the most talented students in each domain in which programs are offered. For example, many programs distinguish between verbal and quantitative/spatial abilities. A rank ordering of students on each of these dimensions is all that is needed. Ranking works both when all students in a grade are tested (i.e., census testing) and when only a subset of students who were nominated for the program are tested. The middle panel of Table 12.1 shows an example in which scores for 20 students were rank ordered by sorting the data using a spreadsheet.

Table 12.1

Getting Ranks

1. Get the data into an excel spreadsheet.
   a. SAS scores
   b. Potential grouping variable (e.g. ELL)

ID	CogAT	ELL
1	92	N
2	85	N
3	111	N
4	90	Y
5	105	N
6	102	Y
7	72	Y
8	121	N
9	95	Y
10	114	N
11	100	N
12	74	N
13	81	Y
14	88	Y
15	107	N
16	97	N
17	86	N
18	93	N
19	84	Y
20	78	N

2. To get local ranks, sort (rank order) the data by CogAT scores.

ID	CogAT	ELL
8	121	N
10	114	N
3	111	N
15	107	N
5	105	N
6	102	Y
11	100	N
16	97	N
9	95	Y
18	93	N
1	92	N
4	90	Y
14	88	Y
17	86	N
2	85	N
19	84	Y
13	81	Y
20	78	N
12	74	N
7	72	Y

3. To get separate ranks for each ELL group, sort by ELL and then SAS.

ID	CogAT	ELL
8	121	N
10	114	N
3	111	N
15	107	N
5	105	N
11	100	N
16	97	N
18	93	N
1	92	N
17	86	N
2	85	N
20	78	N
12	74	N
6	102	Y
9	95	Y
4	90	Y
14	88	Y
19	84	Y
13	81	Y
7	72	Y

Most test publishers offer reports that provide this information. Figure 12.2 shows a portion of one such report for scores on the *CogAT* Verbal Battery. In the full report, separate rankings for all students in each class (or building) are given for each of the three *CogAT* batteries.

CogAT®

CLASS LIST RANKED BY TESTS

Cognitive Abilities Test™ (CogAT®)

Class/Group: Ortega
Building: Central Elementary
System: Spring Lake
Form: 6
Test Date: 11/2002
Norms: Fall 2000
Order No.: 000493608
Page: 1
Grade: 3

Verbal	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name
	139	99	99	Regis, Clare	109	71	78	Delarosa, Amanda	101	52	53	Washington, Shanika	84	16	26	Peters, Matt
	137	99	98	Russell, Jalen	109	71	70	Maclean, Darnell	99	48	53	Freed, Jenna				Card, Susan
	121	91	91	Chavez, Natalia	107	67	65	Dukes, Sanetra	95	38	40	Brown, Nara				
	115	83	86	Hwang, Jung	101	52	55	Atsushi, Eri	94	35	48	Fry, Michelle				
	110	73	78	Lee, Samuel	101	52	58	Garcia, Felipe	90	27	27	Hogan, Ryan				

Quantitative	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name	SAS	APR	GPR	Student Name
	137	99	98	Russell, Jalen	112	77	81	Atsushi, Eri	94	35	40	Freed, Jenna	73	5	4	Card, Susan
	131	97	98	Hwang, Jung	111	75	81	Delarosa, Amanda	94	35	36	Maclean, Darnell				Garcia, Felipe
	121	91	92	Brown, Nara	102	55	70	Peters, Matt	90	27	26	Dukes, Sanetra				
	117	86	87	Regis, Clare	99	48	50	Chavez, Natalia	89	25	30	Lee, Samuel				
	115	83	89	Fry, Michelle	95	38	40	Hogan, Ryan	85	17	20	Washington, Shanika				

Figure 12.2 Portion of a class report showing student ranks on the Verbal Battery of *CogAT*. SAS = Standard Age Score; APR = national Age Percentile Rank; GPR = national Grade Percentile Rank.

If particular groups of students differ markedly in opportunity to develop the abilities measured by the test, then ranks should be computed separately within these groups. The panel on the right in Table 12.1 shows how ranks within a grouping variable (here ELL status) can be obtained by sorting on two variables: first on ELL status and second on the *CogAT* score of interest. Within-group ranks also can easily be obtained using the test publisher's online test analysis and interpretation tools.

Percentile Ranks (PRs)

Although a simple rank order may suffice for many purposes, true local norms require the estimation of percentile ranks. Two scores that have adjacent ranks may differ substantially in percentile rank. For example, suppose the three highest scores in a data set are IQs of 140, 120, and 119. Clearly, the difference between the first and second student is larger than the distance between the second and third students. Simple ranks discard this information; percentile ranks can preserve it.

The critical difference between ranks and percentile ranks (PRs) is that the PRs take into account where each score falls in the overall distribution of scores. The PR of any score is simply the percentage of cases in the distribution with the same or lower value. Percentile ranks will be unstable, however, unless the distribution has many cases or has known characteristics. If, as is the case with ability test scores, one can assume that scores are approximately normally distributed, then the distribution is well described by the mean (M) and standard deviation (SD). In this case, the computation of local percentile ranks requires only that one know the mean and

SD of scores within the local group. Although one can use different kinds of test scores in this computation (e.g., on *CogAT*, one could use raw scores, USS scores, or SAS scores), it is helpful to use the nationally normed age-based scores (such as *CogAT* SAS scores) in these calculations. By design, SAS scores are made to be normally distributed in the population. This makes the assumption of a normal distribution for the local sample more plausible. Standard Age Scores (SAS) also use the power of the national norms to control for the effects of age. Then the percentile ranks of SAS scores are computed for the local population using the local mean and SD of SAS scores.

If all students in a grade are tested, then the mean and standard deviation are easily estimated. Table 12.2 shows an example that uses the "AVERAGE," "STDEV," and "NORMDIST" functions in Microsoft Excel to compute local PRs. The example has only 20 cases, but at least 50 cases approximately normally distributed are needed before computing local norms in this way. Administering the same test each year to a different cohort of students (e.g., all second grade students) will provide increasingly stable local norms by cumulating cases across years (recalculating the local mean and SD each year). Thus, the sample that has only 50 cases the first year would have approximately 100 the second year, 150 cases the third year, and so on. Note that if one uses the national mean and SD for SAS scores (M = 100; SD = 16), rather than the local mean and SD, then one will obtain the National Age PRs that correspond to each SAS score. Indeed, performing this calculation provides a good check on the accuracy of the procedures.

Although local PRs can be computed in a spreadsheet such as Excel (as illustrated in Table 12.2), one can also obtain local norms from some test publishers simply by asking for them at the time the tests are scored (see Figure 12.1).

The Importance of Opportunity to Learn (OTL)

Individual differences in rate (or depth) of learning can indicate talent. In any domain, children with an aptitude or talent for a particular kind of learning or performance will typically learn in a few trials what otherwise similar children take many trials to learn. Inferences about intellectual ability from test scores, classroom activities, projects, and other behavioral evidence are thus always judged relative to some larger group of individuals that we assume have had similar opportunities to develop the knowledge, skills, or other observed characteristics. For most rating scales, the comparison group is usually only roughly defined (e.g., "When compared to other second grade children that you have taught, how able is this child?"). On ability and achievement tests, however, the norm groups consist of narrowly defined samples of children of the same age or in the same grade. On ability tests, age groups typically consist of other children in the norming sample who differ from the examinee by no more than a few months. The child who is six months older than another child is expected to perform somewhat better on the same set of tasks. Some tests use even more narrowly defined comparison groups. For example, normative scores on the *ITBS* depend on the number of weeks that the child has been in a particular grade. This means that a given raw score maps onto different age or grade percentile ranks, depending on the child's age in years and months or the number of weeks he or she has been enrolled in a particular grade in school.

Table 12.2

Getting Local PRs

1. Get the data into Excel
2. Get the mean and standard deviation for all SAS scores (ignores ELL)
   a. Use "AVERAGE" function for the mean
   b. Use "STDEV" function for SD

ID	*CogAT*	*ELL*
8	121	N
10	114	N
3	111	N
15	107	N
5	105	N
6	102	Y
11	100	N
16	97	N
9	95	Y
18	93	N
1	92	N
4	90	Y
14	88	Y
17	86	N
2	85	N
19	84	Y
13	81	Y
20	78	N
12	74	N
7	72	Y
Mean =	13.4	
SD =	93.8	

3. To get local Percentile Ranks (PRs), use Excel function "NORMDIST"
   a. Insert local mean and SD (see column 1)
   b. NORMDIST (X, 93.8, 13.4, TRUE) where "X" = *CogAT* score

ID	*CogAT*	*ELL*	*LPR*
8	121	N	98
10	114	N	93
3	111	N	90
15	107	N	84
5	105	N	80
6	102	Y	73
11	100	N	68
16	97	N	59
9	95	Y	54
18	93	N	48
1	92	N	45
4	90	Y	39
14	88	Y	33
17	86	N	28
2	85	N	26
19	84	Y	23
13	81	Y	17
20	78	N	12
12	74	N	7
7	72	Y	5

4. To get within-ELL local PRs, use mean and SD for each group (M = 93.8, SD = 13.4 for non-ELL; and M = 87.4, SD = 9.7 for ELL). Using Mean= 100 and SD = 16 for all students gives National PRs for *CogAT* (last column)

ID	*CogAT*	*ELL*	*LPR*	*Within ELL LPR*	*NPR*
8	121	N	98	95	91
10	114	N	93	88	81
3	111	N	90	83	75
15	107	N	84	75	67
5	105	N	80	71	62
11	100	N	68	58	50
16	97	N	59	49	43
18	93	N	48	38	33
1	92	N	45	36	31
17	86	N	28	22	19
2	85	N	26	20	17
20	78	N	12	9	8
12	74	N	7	5	5
6	102	Y	73	93	55
9	95	Y	54	78	38
4	90	Y	39	61	27
14	88	Y	33	52	23
19	84	Y	23	36	16
13	81	Y	17	25	12
7	72	Y	5	6	4

However, if for any reason the individual's experiences differ markedly from those of other students who are the same age or in the same grade, then these normative comparisons will either underestimate or overestimate the individual's ability to learn. Clearly, the intellectual abilities of students who live in poverty, who have irregular or poor schooling, who have less experience with the language of instruction (or testing) than the students they are being compared to are often underestimated when their behavior is compared with that of other students who are the same age or in the same grade.

It is important to understand that regardless of the adequacy of the norm group for making inferences about talent, performance on the test tells something useful about the student's current level of development of the knowledge and skills measured by the test. Thus, we may rightly say that a fourth grade ELL child is reading at the first-grade level on an English-language reading test. For many instructional decisions, this interpretation may be the most important. However, the same child's reading ability may be at a much higher level in another language. *Reading ability* is a broader construct than *reading ability in English*. And the child's *verbal reasoning ability* may be higher than his ability to read in any language. This is the broadest construct and as such requires the largest inference. Judgments about intellectual talent rest on this third type of inference. Such inferences are valid only when the child's performance on some set of tasks can be compared to the performance of others who have had similar opportunities to develop the abilities, knowledge, or skills required by those tasks.

Since the earliest days of mental testing, psychologists have struggled with the problem of accounting for differences in opportunity to learn, especially those differences moderated by exposure to the language of testing. Two fundamentally different approaches have been taken:

1. adjustment or redevelopment of norms so that students' scores can be compared with the scores of other children who have had similar opportunities to learn the language in which the test is presented or the knowledge it presumes; or
2. attempts to reduce or eliminate the impact of language or culture on the test itself.

The second approach has long been the preferred option. The use of culture- and language-reduced or so-called "non-verbal" tests stretches from the form boards of Itard through the Army Beta to the performance battery of the Wechsler scales, the Progressive Matrices test (Raven, 1938), the Nonverbal Battery of the *Cognitive Abilities Test* (Thorndike & Hagen, 1963), and the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998). The most important disadvantage of this approach is that the abilities measured by nonverbal tests—especially those that use only figural reasoning items—underrepresent the construct of intelligence. The most salient advantage of this approach is that the scores of all students can be interpreted using the same set of norms. However, using common norms assumes that the effects of language and culture have indeed been eliminated.

In recent years, nonverbal tests have been widely administered to ELL children being considered for inclusion in talent development programs. To understand how ELL children perform on such tests, it is helpful to distinguish between the *language loading* and the *cultural loading* of a test. Paradoxically, reducing the language demands may actually increase the cultural loading of the test. In a comparison of three of the most widely used group-administered nonverbal tests, Laing, Castellano, and Buss (2006), trained examiners, administered the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1996), the Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1983), and Form 6 of the *Cognitive Abilities Test* (*CogAT*; Lohman &

Hagan, 2001) to over 1,200 students (approximately half ELL) in grades K–6. Directions were given in English or Spanish, as appropriate. When scores for the three nonverbal tests were placed on a common scale (M = 100, SD = 16), ELL students scored an average of ten points lower than non-ELL students on the NNAT, nine points lower on the *CogAT* Nonverbal Battery, and eight points lower on the SPM (Lohman et al., 2008). Restricting the analysis to Hispanic students eligible for free/reduced-price school lunches still showed ELL/non-ELL differences of approximately eight points. Thus, in spite of controls for ethnicity, location, and socio-economic status (SES), Spanish-speaking ELL children performed substantially lower than their non-ELL Hispanic classmates on all three nonverbal reasoning tests. The overt demands for language on all three nonverbal tests were minimal, yet familiarity with the culture still had a substantial effect.

The second example comes from the standardization of a Spanish adaptation of the *WISC-IV* (Wechsler, 2004) intended for Spanish-speaking children with no more than five years in the U.S. educational system. Since the target population for the test was Spanish-speaking children in the United States, the sample of bilingual children used to norm the test was selected to represent that population. Normative data were later analyzed by the number of years each child had attended U.S. schools relative to his or her total education. The perceptual reasoning tests (Block Design, Picture Concepts, and Matrix Reasoning) measure nonverbal/fluid reasoning. The verbal comprehension tests (Similarities, Vocabulary, and Comprehension) measure verbal, crystallized abilities using items that were translated into Spanish. Surprisingly, exposure to the U.S. educational system had large effects on the perceptual reasoning index (10 points) and only a small effect on the verbal comprehension index (2.2 points). Those students with the most exposure to U.S. schooling performed significantly better on the perceptual reasoning tests. As Weiss, Saklofske, Prifitera, and Holdnack (2006) note, "[T]his is an interesting finding because it is widely assumed that the lack of an adaptation and acculturation primarily affects crystallized knowledge" (p. 47). However, the assumption that typical nonverbal tests provide a culture-fair measure of innate ability is not supported by research—recent or dated. Indeed, as Anastasi and Urbina (1997) concluded in their summary of seventy years of research on the topic, "nonverbal tests are often more culturally loaded than verbal tests" (p. 344).

Note that in both of these examples presented above, the nonverbal tests measured primarily or exclusively figural reasoning abilities. However, some nonverbal tests use other kinds of picture-based tasks in order to represent a broader reasoning construct. For example, picture-based analogies can measure verbal or quantitative reasoning in addition to figural–spatial reasoning. Two recent examples are the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998) and the primary-level tests on Form 7 of *CogAT* (Lohman, 2011). Items in these tests not only reduce the demands of language, but they also measure the ability to reason with pictorially represented verbal and quantitative concepts. The broader measure of ability they provide predicts success in school better than unidimensional nonverbal tests that use only figural–spatial content (McCallum, Bracken, & Wasserman, 2001). Furthermore, ELL, low SES, and minority children in grades K–2 perform as well or better on picture–verbal and picture–quantitative tests than on figural reasoning tests (Lohman & Gambrell, in press).

Measuring Opportunity to Learn

Measuring opportunity to learn requires finding some variable (or set of variables) that not only captures learning opportunities but that can be unambiguously coded for all students. In the *WISC IV–Spanish* study (Wechsler, 2004) for example, the number of years attending U.S. schools was used. Some researchers also add home language. More refined measures of ELL status are available in schools in which all ELL students are administered the same English Language Proficiency test each year. These tests allow schools to create groups that distinguish between students who have different levels of familiarity with, and competence in, American English. To determine economic opportunity, the most accessible measure for schools is whether the child qualifies for a free or reduced-price school lunch.

If children can be grouped using one or more measures of opportunity to learn, separate rank orders are easily created within the different OTL groups. If there are many students in each group, then separate percentile ranks (PRs) can also be calculated within each group. This is easily done if one can estimate the mean and SD of test scores for each group. For relatively small samples (fewer than 50 students), one can estimate different means for each OTL group, but then use the same SD for all groups. Again, accumulating cases across years allows one to estimate increasingly stable local PR scores for the students in each OTL group.

Using Within-OTL Scores

The need for precise estimates of ability is a direct consequence of trying to determine whether or not a child is truly "gifted." However, if the goal is merely to identify poor or ELL children who might profit from special encouragement, projects, or enrichment, then there is no need for such precision.

One of the major stumbling blocks for effective talent identification among poor and ELL children is the presumption that all talented students must receive the same kind of special instruction. In athletics, we would expect that some children who had little experience swimming might have talent for the sport. But we would not think it reasonable to immediately expect them to swim at the same pace as children who had had many years of practice in the sport. Clearly, the inference of talent is distinguishable from a judgment about the current level of development of that talent. Thus, any attempt to identify talent within OTL groups must also be accompanied by a redesign of the programs that serve the children who will be identified as talented.

In considering how this might be done, it is helpful to keep in mind that encouraging interest and persistence in the pursuit of excellence is as important for talent development as the acquisition of academic knowledge and skills. Further, unlike their classmates whose parents may have greater resources, children from poor and immigrant families often must rely on their school to provide special services and opportunities for talent development. If the school cannot make it happen, then it does not happen. Therefore, some form of enrichment may be most appropriate for many of these students whose academic development is similar to that of their classmates but who exhibit undeveloped talent. The scheme outlined in the next section is specifically designed for schools that must serve both those students who are considerably in advance of their peers and those who exhibit talent but are not currently mismatched with their peers.

Combining Ability Tests, Achievement Tests, and Teacher Ratings

Many schools use multiple criteria to identify academically talented students. However, the various sources of evidence must be combined in some way. Often this is done by converting test scores, teacher ratings, and other information to point values. Points are then summed and students admitted accordingly. Although these methods are often easy to understand, they can easily mislead. Some of the potential problems follow:

1. At best, converting a continuous score with many values to a point scale with few values discards information. Often, the information that is retained is distorted by the way points map onto the original score scale.
2. When adding or averaging scores, the final rank order is determined by the score with the greatest variability in point values, not the score that on average contributes the most points. The former may not be the score one wishes to emphasize.
3. Unless different rankings are determined for different domains (e.g., verbal versus quantitative), students with uneven score profiles are often excluded. This is because adding or averaging diverse information discards the information in the student's highest score.
4. Rating-scale data and other sources of information can be helpful. However, even when raters are well trained, such measures are usually less reliable and less valid than the test scores. Hence, even assigning ratings a lesser weight (or point value) can be problematic. For example, although one could certainly justify providing enrichment opportunities to a child rated (by a teacher) as highly creative, it would be difficult to defend a decision to deny the opportunity for advanced instruction to a child who received lower ratings on creativity but obtained high ability and achievement test scores. Yet for every child who gains admission because of high teacher ratings, another child with equally high achievement or ability test scores is denied admission because of lower teacher ratings. An effective way to overcome this dilemma is to use ratings (and other measures that are potentially less reliable and valid) to provide opportunity, but never to remove it.

Figure 12.3 shows a scheme that accomplishes this goal.[2] This particular version uses only *CogAT* scores and teacher ratings from the Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli et al., 2004). The vertical dimension of Figure 12.3 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 obtained using the procedures outlined earlier in this chapter.

		Teacher Rating on Learning Ability, Motivation, or Creativity	
		Low teacher ratings	High teacher ratings
CogAT Verbal OR Quantitative-Nonverbal Reasoning	($\geq 96^{th}$ PR)	II	I
	(80^{th} – 95^{th} PR)	IV	III

Figure 12.3 Combing ability (CogAT Verbal or Quantitative-Nonverbal) and teacher ratings (SRBCSS ratings of ability, motivation, or creativity). Two levels of ability and two levels of ratings are combined to provide four categories (see text for explanation).

The horizontal dimension of the matrix distinguishes between children who, when compared to other children nominated for the program, obtain relatively high teacher ratings and students who obtain lower teacher ratings. Teacher ratings are considered high if *any one* of the ratings for learning ability, motivation, or creativity is high. If teachers rate all students, then a high rating might be one that was obtained by the top 5 percent or 10 percent of students. However, it is often difficult to obtain reliable ratings for classes with many students. Another option, then, is to obtain teacher ratings on the much smaller subset of students whom teachers or others have nominated for the program or whose test scores are above the 80th PR. When only a subset of students is rated, then a much more lenient standard should be set. For example, one could distinguish between those students with ratings above the average of the group of nominated students and those with ratings that are below the average of this group. If there is no variability in the ratings of nominated students, then procedures for nominating children are too restrictive or raters are poorly trained. Of course, schools can implement a rule that is either more stringent or more lenient than *above* (or *below*) *average*.

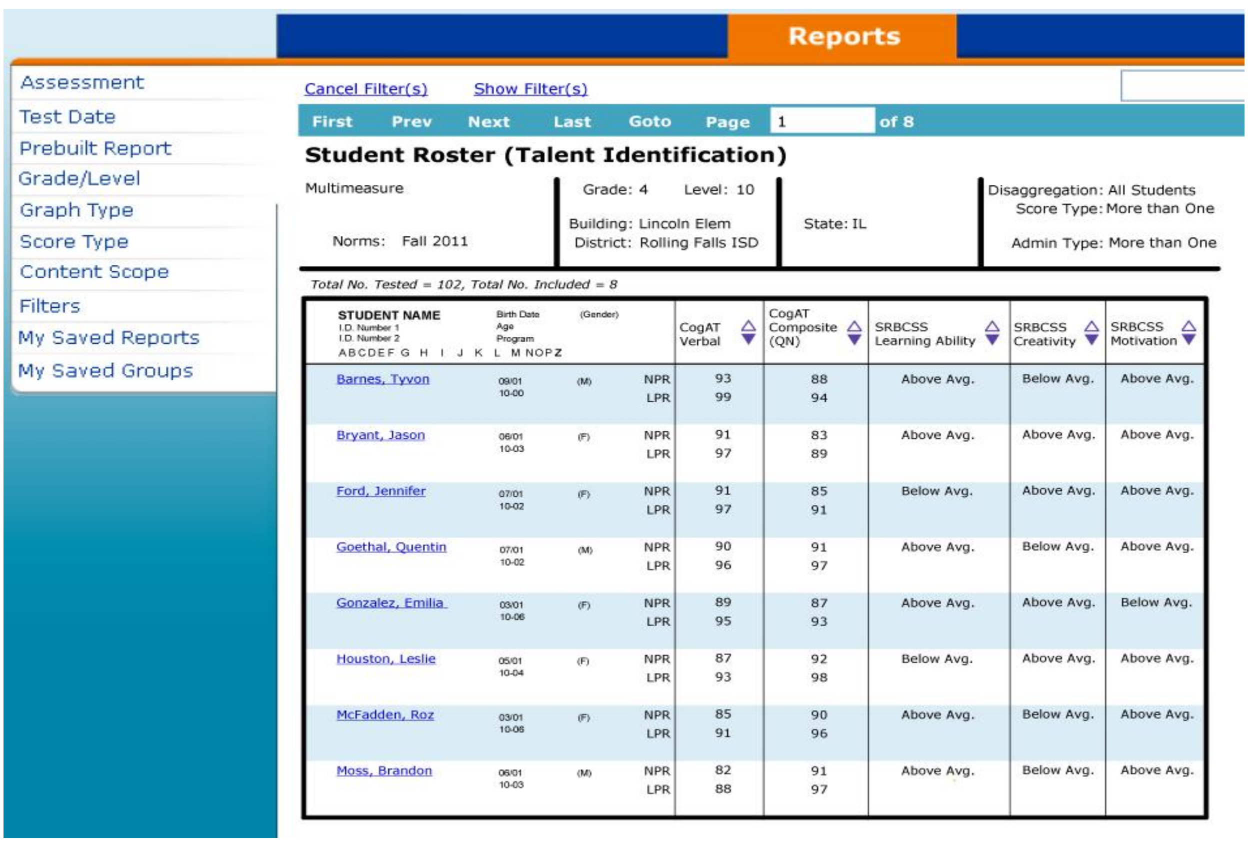
Combining these two criteria gives four categories of assessment results.

- 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, 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) accept children in both Category I and Category II. However, the progress of children in Category II should be monitored closely.

- Children in Category III exhibit somewhat lower but still strong reasoning abilities (80th to 95th 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 find that many of their best students would fall into this category, especially when using national rather than local (i.e., school) test norms.

- Finally, children in Category IV exhibit good but not exceptional reasoning abilities (between the 80th and 95th 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.



STUDENT NAME / I.D. Number 1 / I.D. Number 2 / ABCDEF G H I J K L M NOPZ	Birth Date / Age / Program	(Gender)		CogAT Verbal	CogAT Composite (QN)	SRBCSS Learning Ability	SRBCSS Creativity	SRBCSS Motivation
Barnes, Tyvon	09/01 10-00	(M)	NPR LPR	93 99	88 94	Above Avg.	Below Avg.	Above Avg.
Bryant, Jason	06/01 10-03	(F)	NPR LPR	91 97	83 89	Above Avg.	Above Avg.	Above Avg.
Ford, Jennifer	07/01 10-02	(F)	NPR LPR	91 97	85 91	Below Avg.	Above Avg.	Above Avg.
Goethal, Quentin	07/01 10-02	(M)	NPR LPR	90 96	91 97	Above Avg.	Below Avg.	Above Avg.
Gonzalez, Emilia	03/01 10-06	(F)	NPR LPR	89 95	87 93	Above Avg.	Above Avg.	Below Avg.
Houston, Leslie	05/01 10-04	(F)	NPR LPR	87 93	92 98	Below Avg.	Above Avg.	Above Avg.
McFadden, Roz	03/01 10-06	(F)	NPR LPR	85 91	90 96	Above Avg.	Below Avg.	Above Avg.
Moss, Brandon	06/01 10-03	(M)	NPR LPR	82 88	91 97	Above Avg.	Below Avg.	Above Avg.

Figure 12.4 Sample student roster in Data Director showing *CogAT* Verbal and Quantitative-Nonverbal national and local percentile ranks and whether teacher ratings for Learning Ability, Creativity, or Motivation are above or below the average rating of the group of students that were rated.

These procedures can be carried out on a spreadsheet or similar data management tool such as Riverside Publishing's Interactive Results Manager (iRM). Figure 12.4 shows a sample student roster in iRM that has *CogAT* Verbal (V) and Quantitative–Nonverbal (QN) local percentile ranks (LPRs), and three SRBCSS rating scales that have been classified as above or below the average of the group that was rated. Typically, ratings would be gathered only for students who had test scores above the 80th LPR. Figure 12.5 shows the number of students in each category, which is linked to a roster with names and scores.

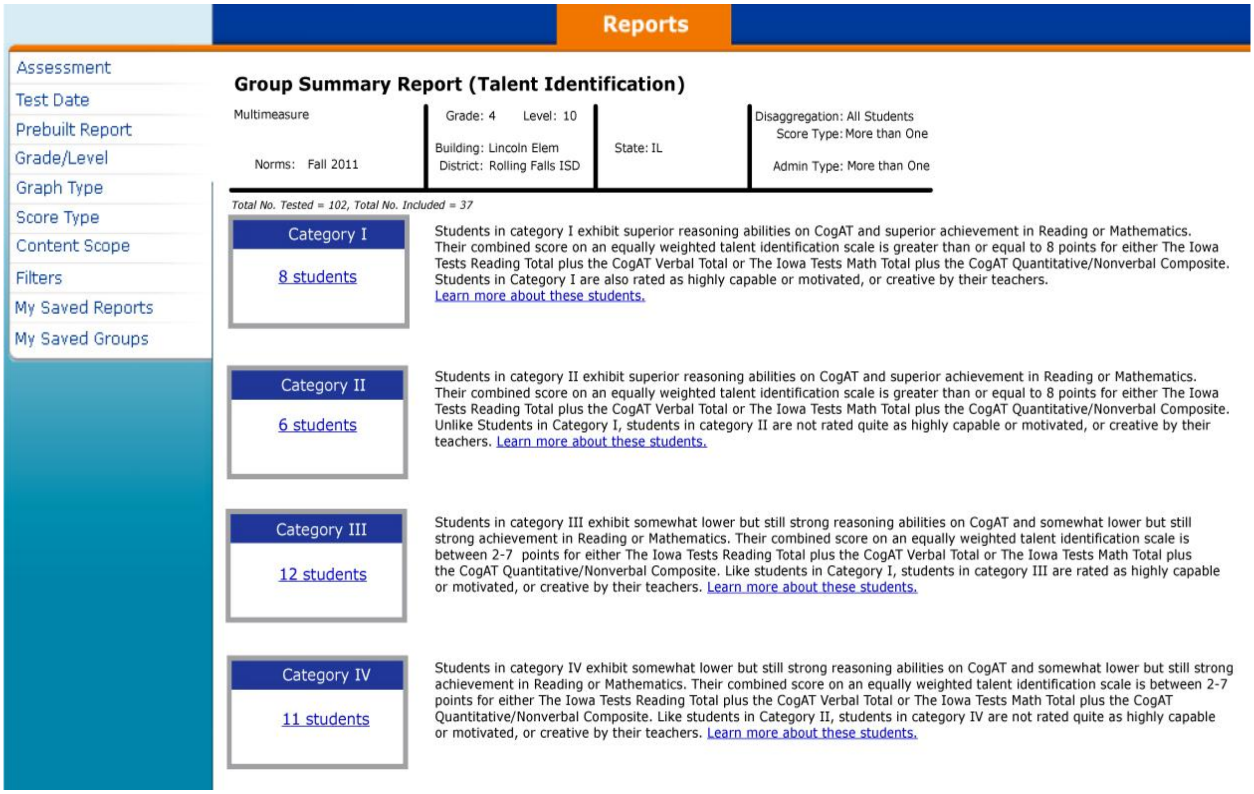


Figure 12.5 Number of students in each of the four categories of the ability by teacher ratings scheme shown in Figure 12.3. Names, tests scores, and rating data for each group of students are linked to each category.

Summary

Although conventional reliance on national test norms can be very helpful, effective talent identification at the local level often requires the use of local and subgroup normative perspectives. The value of these alternative perspectives increases as test scores for the student group, the school, or the district depart from the national mean. The need for alternative instruction depends primarily on the mismatch between a student's current level of cognitive and academic development and that of his classmates. Local norms for ability and achievement tests can often be obtained from the test publisher, but they can also easily be developed using commonly available spreadsheet programs. The simplest comparisons report local rank orders of students—for all students or for subgroups defined by some measure of opportunity to learn.

Local percentile ranks can also be computed providing one has sufficient data to estimate them. Increasingly accurate local norms can be obtained by accumulating data across years.

Opportunity to learn (OTL) can be estimated by ELL status or by an index of family income such as eligibility for a free/reduced-price school lunch. By accounting for opportunity to learn, schools are better able to identify their most academically talented students than is possible using only figural reasoning, nonverbal tests. Poor and minority children often score well below other children on nonverbal tests, especially when they have had limited exposure to tests that present unfamiliar tasks. More importantly, verbal and quantitative reasoning abilities are essential aspects of academic talent for all children. By accounting for opportunity to learn, one can measure these abilities for all children, thereby better identifying the most academically talented children in all groups. Providing appropriately challenging instruction for all children who might benefit from enrichment or acceleration requires a rethinking of the purpose of the gifted program—and perhaps beginning with a renaming of the program so that it encourages talent identification and development rather than simple certification of giftedness (Callahan, 2005).

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Bracken, B. A., & McCallum, S. (1998). *Universal Nonverbal Intelligence Test—UNIT*. Itasca, IL: Riverside.

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

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5–20.

Laing, P., Castellano, J., & Buss, R. (2006). Project Bright Horizon. Phoenix, AZ: Washington Elementary School District.

Lohman, D. F. (2012). Decision strategies. In S. L. Hunsaker (Ed.), I*dentification: The Theory and Practice of Identifying Students for Gifted and Talented Education Services.* (pp. 217–248). Mansfield Center, CT: Creative Learning Press.

Lohman, D. F. (2009). The contextual assessment of talent. In MacFarlane, B., & Stambaugh, T. (Eds.), *Leading change in gifted education: The* festschrift *of Dr. Joyce VanTassel-Baska* (pp. 229–242). Waco, TX: Prufrock Press.

Lohman, D. F. (2011). *Cognitive Abilities Test (Form 7)*. Rolling Meadows, IL: Riverside.

Lohman, D. F., & Foley Nicpon, M. (2012). Ability testing and talent identification. In S. L. Hunsaker (Ed.), *Identification: The Theory and Practice of Identifying Students for Gifted and Talented Education Services*. (pp. 283-335). Mansfield Center, CT: Creative Learning Press.

Lohman, D. F., & Gambrell, J. L. (in press). Use of nonverbal measures in gifted identification. *Journal of Psychoeducational Assessment.*

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

Lohman, D. F., Korb, K., & Lakin, J. (2008). Identifying academically gifted English language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*, 275–296.

Lohman, D. F., & Renzulli, J. S. (2007). *A simple procedure for combining ability test scores, achievement test scores, and teacher ratings to identify academically talented children.* Retrieved from http://faculty.education.uiowa.edu/dlohman/

McCallum, R. S., Bracken, B. A., & Wasserman, J. D. (2001). *Essentials of nonverbal testing*. NY: Wiley.

Naglieri, J. A. (1996). *Naglieri Nonverbal Ability Test.* San Antonio, TX: Harcourt Brace Educational Measurement.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London, UK**:** Lewis.

Raven, J. C., Court, J. H., & Raven, J. (1983). *Manual for Raven's Progressive Matrices and vocabulary scales, section 4: Advanced Progressive Matrices, sets I and II.* London, UK: Lewis.

Renzulli, J. S. (2005). *Equity, excellence, and economy in a system for identifying students in gifted education: A guidebook.* (RM05208). Storrs: 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, M. (2004). *Scales for Rating the Behavioral Characteristics of Superior Students*. Mansfield Center, CT: Creative Learning Press.

Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4,* 3–7.

Thorndike, R. L., & Hagen, E. (1963). *Cognitive Abilities Test*. New York: Houghton Mifflin.

Wechsler, D. (2004). *WISC IV–Spanish.* San Antonio, TX: Harcourt.

Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV advanced clinical interpretation*. Burlington, MA: Elsevier.

Notes

1. Examples include *CogAT* Standard Age Scores (SAS); Otis–Lennon School Ability Index (SAI); Naglieri Nonverbal Index scores (NAI).
2. Elsewhere (Lohman & Renzulli, 2007), we show how to include achievement test scores as well.