Cognitively Speaking

Introducing CogAT Form 7

By Dr. David F. Lohman, University of Iowa

This issue of *Cognitively Speaking* is unlike prior issues. Previous newsletters have discussed common questions about ability testing or summarized research on abilities that can help schools use test scores more effectively. This issue has a different goal. I want to share the exciting news about the new *Cognitive Abilities Test™* (*CogAT®*). For the past nine years, my research team at the University of Iowa and the staff at Riverside have worked diligently on a major revision of the test. And although there is still much work to do, last month Form 7 of the *Cognitive Abilities Test* was published. The new form incorporates the most substantial enhancements to *CogAT* since the test was first published in 1968; first however, I want to reassure current users that we have preserved those features of the *Cognitive Abilities Test* that have made it the most trusted and widely used group abilities test in the U.S and abroad. Second, I want to explain how the test has been revised and why Form 7 is even better than its predecessors. And third, I want to share some of the research that demonstrates the fairness of the new tests.

How Does Form 7 Preserve the Important Features of *CogAT*?

In designing Form 7, we wanted not only to make major enhancements but also to preserve essential features of the test. The new Form 7 retains many characteristics that current users of previous editions of *CogAT* will recognize:

- All levels **have three independent batteries**: Verbal, Quantitative, and Nonverbal. These batteries measure the student's learned reasoning abilities in the three areas most closely related to success in school. Users can administer one, two, or all three batteries, depending on their needs and the needs of their students.

- Each battery has subtests that make use of three **different test formats**. Measuring with multiple formats rather than the same format on all items increases both the fairness and the validity of the scores students obtain.

- Educators who administer all three batteries of *CogAT* obtain an **ability profile** for each student. This code summarizes the pattern and level of the student's scores on the three batteries. The ability profile code is explained and linked to specific instructional suggestions in the Interactive Profile Interpretation System at www.CogAT.com.

- Finally, the testing time remains the same, so the tests are easily administered within a typical school schedule.

What's New in Form 7?

We revised the test to meet the many needs of educators in today's schools. The major changes in Form 7 are:

- As on the *Iowa Assessments*, **test levels are now designated by age** rather than by grade (K, 1, 2) or alphabetic code (A-H). This reduces confusion in achievement-ability testing.

- **Levels 5/6–8** (grades K to 2) **are completely revised to be a bilingual primary battery**. They now contain three subtests in each battery. Each subtest blends seamlessly with its counterpart at

the upper levels (grades 3 to 12). Each subtest at grades k to 2 is bilingual/ELL-friendly. With one exception, all subtests use only pictures; language is used only in the directions. The one exception is an optional English/Spanish Sentence Completion subtest that requires students to listen to a sentence that the teacher reads aloud. Students then select the picture that best answers the question.

- For students whose native language is Spanish, the complete test can be administered using the *Spanish Directions for Administration*. These directions allow the test administrator to switch back and forth between Spanish and English as necessary. For other ELL students, the optional English/Spanish Sentence Completion test can be omitted.

- Primary-level tests (levels 5/6 – 8) now contain **three subtests** in each battery, just like at the upper levels (levels 9 – 17/18). Each primary-level subtest blends seamlessly with its counterpart at the upper levels. This increases consistency in the abilities measured with no increase in administration time.

- A short, ELL-friendly *CogAT* **7 Screening Form** was created for schools that use *CogAT* only for a preliminary screening of students who should be considered for inclusion in talent-development programs,

- Levels 9–17/18 (grades 3 to 12) have two **new quantitative subtests** that better measure quantitative reasoning for today's students.

- Levels 9–17/18 (grades 3 to 12) also have a new **figure matrices subtest** that increases the ceiling on the Nonverbal Battery to more adequately measure the abilities of the most capable students.
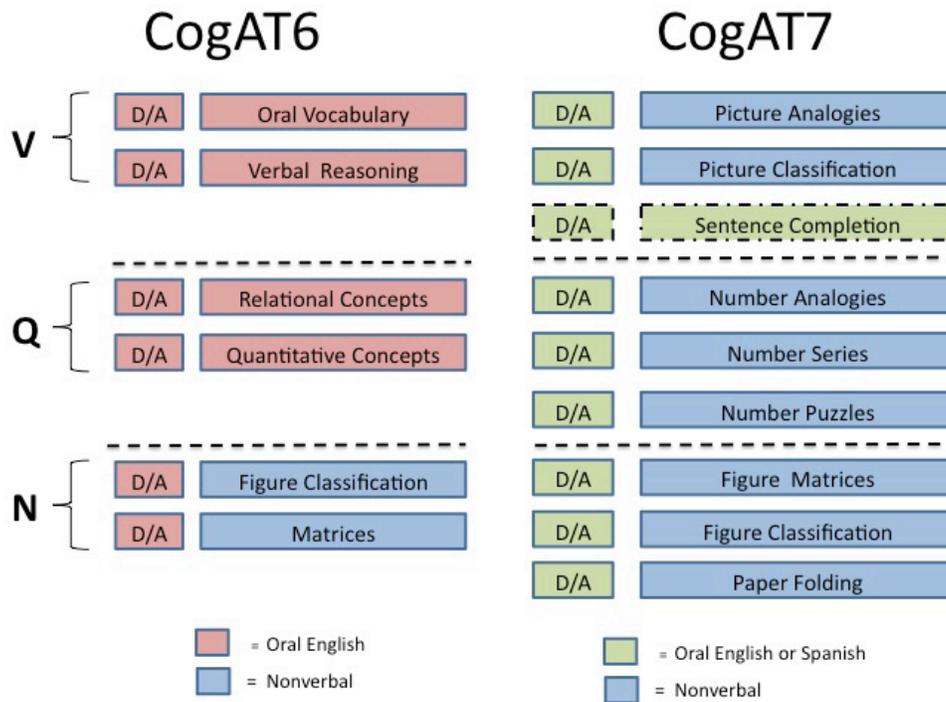
- For schools that test students frequently, **level-to-level item overlap was reduced** from five to two levels so that every-other test level now presents a completely new set of items.

- Form 7 has **enhanced data management and score-reporting services**. For example, when electronic score reports are ordered, educators can more readily access the instructional suggestions linked to each student's ability profile score.

- An **online edition** of the test will be available in the Fall of 2012.

- **Free Practice Activities** with teacher guides can be downloaded from a secure Riverside website. These activities help level the playing field by ensuring that all students are prepared to do their best on the tests.

- Form 7 Standard Age Scores (SAS) **extend up to 160** for better discrimination among the highest levels of ability.

New Primary- Level Subtests

The greatest structural changes in Form 7 were made to the tests designed for students in kindergarten through second grade. The primary goal of the revisions was to make the tests more accessible to ELL students. Figure 1 on the preceding page shows a comparison of the primary-level subtests in Form 6 and Form 7.

Until recently, the vast majority of students in U.S. schools were native speakers of English. They could listen to a question the teacher read aloud and then select the picture that best answered that question. But understanding a question that the teacher reads aloud can be difficult for students who are not native speakers of English. By 2009, 21 percent of school-age children in the U.S. spoke a language other than English at home, and 5 percent spoke a language other than English and had difficulty speaking English.[i]

Primary Battery

CogAT6

V	
D/A	Oral Vocabulary
D/A	Verbal Reasoning
Q	
D/A	Relational Concepts
D/A	Quantitative Concepts
N	
D/A	Figure Classification
D/A	Matrices

☐ = Oral English
☐ = Nonverbal

CogAT7

D/A	Picture Analogies
D/A	Picture Classification
D/A	Sentence Completion
D/A	Number Analogies
D/A	Number Series
D/A	Number Puzzles
D/A	Figure Matrices
D/A	Figure Classification
D/A	Paper Folding

☐ = Oral English or Spanish
☐ = Nonverbal

Figure 1. *Comparison of content and language requirements of Forms 6 and 7 of CogAT. V = Verbal Battery, Q = Quantitative Battery, N = Nonverbal Battery, and D/A = Direction for Administration.*

In earlier times, the handful of primary-level ELL students could be tested using the Nonverbal Battery. Unfortunately, figural-spatial reasoning tests provide only a rough estimate of a student's academic abilities. Academic learning places much greater demands on a student's verbal and quantitative reasoning abilities. Our goal should be to measure these abilities for all students.

As we constructed items for the primary levels, the largest problem we faced was how to measure students' verbal and quantitative reasoning abilities without having the teacher read aloud a question for each item. Our solution was to create picture-based items similar to the text-based items that are used with older children. The teacher's task at the primary level is limited to providing directions and pacing children through the test. In order to ensure that the test items would be fair, we developed picture-based items that accurately measured ability for different cultural groups of students and then selected only those items that worked well in **all** groups. This required much time and effort, the cooperation of many school administrators, teachers, professional item reviewers, and hundreds of students who diligently worked on the "puzzles" we asked them to solve.

We also wanted picture-based items that measure verbal and quantitative reasoning, not just general ability. Measuring verbal reasoning using only pictures was particularly challenging. Our statistical analyses of the Form 7 picture-based verbal subtests show that they require the same verbal abilities as the corresponding text-based verbal subtests used at higher grades.

		Picture Format (Levels 5/6-8)	Text/Standard Format (Levels 9-17/18)
VERBAL BATTERY	Picture/Verbal Analogies		TV → watch : newspaper → J deliver K comics L read M magazine N listen
	Sentence Completion	"Which one swims in the ocean?"	The fastest runner _____ the race. A loses B wins C watches D starts E makes
	Picture/Verbal Classification		apple orange pear A fruit B carrot C pea D lemon E onion
QUANTITATIVE BATTERY	Number Analogies		$[1 \to 2]$ $[3 \to 4]$ $[5 \to ?]$ A 2 B 4 C 6 D 8 E 12
	Number Puzzles		$\boxed{?} = 2 + 3$ A 2 B 3 C 4 D 5 E 6
	Number Series		1 2 4 5 7 8 → A 7 B 8 C 9 D 10 E 11
NONVERBAL BATTERY	Figure Matrices		
	Paper Folding		
	Figure Classification		

Figure 2 shows the item types used in Form 7. The left column shows examples of how some of the items appear in the level 5/6 – 8 tests and the right column shows how the same kind of items appear in the corresponding level 9 – 17/18 tests.

Because verbal abilities are so important for success in school, an **optional** Sentence Completion subtest is included in the Form 7 Verbal Battery to measure these abilities for students who speak English or Spanish. This subtest follows the traditional format in which the teacher reads aloud a question and the student chooses the picture that answers the question. Again, we developed the Sentence Completion items simultaneously within each language/cultural group, tried out all of the items on both groups (and in other ELL and minority populations), and then built the test using only those items that worked well in **all** groups. This required the collaboration of bilingual teachers to help us develop hundreds of potential items, and then the cooperation of many schools with large ELL populations to administer the potential items to their students.

Lastly, the verbal and quantitative subtests for grades K to 2 on previous forms of *CogAT* have used different formats than the upper-level verbal and quantitative subtests. These changes in test format can sometimes be a factor in larger-than-expected changes in test scores when students who were initially tested in grades K to 2 are retested in grade 3 or higher.[ii]

In Form 7 we wanted the primary-level tests to blend as seamlessly as possible with the upper-level tests. Therefore, we created three subtests for each primary-level battery rather than the two subtests in previous editions of *CogAT*. However, because the new subtests do not require unique oral prompts, each of the three primary-level batteries in Form 7 actually takes less time to administer than the corresponding batteries in Form 6.

In summary, the completely revised primary levels now use subtests that blend seamlessly with the tests used at later grades. This increases the consistency in the abilities measured across test levels. The primary-level tests measure verbal, quantitative, and nonverbal reasoning for ELL and non-ELL students using only pictures. One optional subtest (Sentence Completion) requiring oral prompts can be omitted for students who are not fluent speakers of English or Spanish.

Test Fairness

A fair test is one in which test takers have an unobstructed opportunity to demonstrate their abilities. The test, then, does not advantage or disadvantage students because of characteristics that are irrelevant to the abilities being measured. Although this definition of fairness seems straightforward, implementation requires careful monitoring. For example, some items may, on average, be easier for males while other items in the same test may be easier for females. However, as long as there are approximately equal numbers of both kinds of items, the total score on the test will not be biased toward one gender.

Well-trained raters can give invaluable assistance in detecting items that might be unfair. Therefore, the entire pool of potential items for the new Verbal Battery were reviewed by an external team of twenty-two fairness reviewers that included two bilingual psychologists from Mexico who have worked extensively on adapting/translating ability tests. Members of the review panel identified content that might unfairly advantage or disadvantage students because of their race, ethnicity, gender, geographic region, socioeconomic background, or other characteristics not related to reasoning ability. The input from the Fairness Review Panel supplemented the statistical analyses for fairness and multiple rounds of independent reviews by the author and Riverside's test development staff.

Once the final tests were assembled, we compared the overall test performance of ELL and non-ELL students in the national standardization sample. We did this in two different ways. First, we compared the average score of all ELL students with the average score of all non-ELL students. However, ELL students differ from non-ELL students in more ways than their fluency in English. For example, ELL students are more likely to live in poverty than non-ELL students. This means that some of the differences between the scores of ELL and non-ELL students reflect the influence of poverty, not ELL status.

The second procedure for comparing ELL and non-ELL students was first to control statistically for the effects of poverty, ethnicity, and other variables. The advantage of this procedure is that it better estimates the unique effects of ELL status. The disadvantage is that the statistically adjusted group differences are often much smaller than the differences test users are likely to see, which can be misleading.[iii] Thus, we report both the observed differences and the statistically-adjusted differences for users' consideration.

Some of the key differences in Standard Age Scores (SAS)[iv] between ELL and non-ELL students are shown in Table 1. Note that at levels 5 - 8 findings are based on the Alternate Verbal Battery score that omits the optional Sentence Completion test. This Alternate Verbal Battery score can be ordered for specific students in any school or district that administers Form 7.

Table 1
Observed and Statistically-Adjusted
SAS Score Differences between ELL and
Non-ELL Students on *CogAT*, Form 7

CogAT Battery	Observed	Adjusted
	Levels 5 - 8	
Alt Verbal	-2.3	-1.5
Picture Quant	-2.8	-0.1
Nonverbal	-4.6	-2.4
	Levels 9 - 12	
Verbal	-11.2	-7.1
Quantitative	-3.9	-2.5
Nonverbal	-4.3	-2.7

In the new primary-level tests (levels 5 - 8) which use only pictures, the average score for ELL students was only 2.3 SAS points lower on the Alternate Verbal Battery and 2.8 points lower on the new Quantitative Battery. Note that both of these differences were smaller than the 4.6 point difference between ELL and non-ELL students on the Nonverbal Battery. After statistically controlling for poverty (estimated by free/reduced-price lunch) and ethnicity, these differences faded to about 1 SAS point on the Verbal tests and

0 SAS points on the Quantitative tests. We can conclude that the new primary-level Verbal and Quantitative batteries are fair to ELL students, who perform even better on these tests than on the traditional nonverbal reasoning tests that are administered to these children.[v]

The bottom half of the table shows the same data for students in grades 3 to 6 who took levels 9 to 12. At these levels, the subtests in the Verbal Battery are in English. As would be expected, the ELL/non-ELL difference is now much larger. This 11.2 point difference is the standard against which the Alternate (Picture) Verbal tests at the primary levels should be judged. Clearly, the Form 7 picture-based tests for young students are remarkably effective in reducing the impact of language on test scores.

Note that even at levels 9 to 12, differences between ELL and non-ELL students on the Quantitative Battery are once again **smaller** than the differences on the Nonverbal Battery.[vi] Ironically, the near-universal exposure of students to basic quantitative concepts both at home and at school, combined with the very limited universe of early mathematics concepts, may actually make these quantitative tests less sensitive to cultural differences at home than the nonverbal (figural) tests, especially for bilingual students attending U.S. schools. Combining the Quantitative and Nonverbal scores into a QN partial Composite often provides a much better estimate of academic ability for ELL students than either battery alone. However, if the Verbal Battery is administered to ELL students at these grades, it may be best to compare their Verbal Battery scores to those of other ELL students. This is not difficult to do and can provide important information on the ELL students' verbal reasoning abilities that is not apparent when national or even local norms are used.[vii]

Gifted Identification

When *CogAT* is administered to help identify academically talented students, users are often concerned with the relative proportion of ELL students among those who receive high scores on the test.

Table 2 shows the unadjusted percentages of students from various groups who scored in the top 5 percent of each score distribution in the national standardization. This table includes scores for both the complete Verbal Battery score and the Alternate Verbal score that excludes the optional Sentence Completion subtest. The table also includes scores on the new *CogAT* Screening Form.

The top row of the table shows the percentage of students in the national standardization sample who belonged to each group. For example, 6 percent were ELL students and 23 percent were eligible for free/reduced-price lunch (FRL). The number at the top of each column is the standard to use when interpreting the remaining numbers in that column.

Table 2
Percentage of Test Takers in Grades K to 2 Scoring in the Top 5 Percent of Each Test Battery Belonging to Various Subgroups

	ELL	FRL	Asian	Hispanic	Black
Sample percent	6	23	4	20	15
Verbal	3	11	8	15	7
Alternate Verbal	5	14	9	18	8
Quantitative	4	10	11	11	6
Nonverbal	2	11	8	14	8
Screening Form	3	12	9	13	7

The first entry shows that 3 percent of the students with the highest scores on the full Verbal Battery were ELL students. The entry below for the Alternate Verbal scores shows that if the English version of the Sentence Completion test were omitted, then 5 percent of the top scorers would be ELL students. This is quite favorable given that 6 percent of the students in the standardization sample were ELL students. Other entries for the Alternate Verbal score are especially important for the representation of ELL students. For example, 18 percent of top scorers

on the Alternate Verbal score were Hispanic. This is only slightly less than the percentage of Hispanic students in the sample (20 percent).

Other entries show that Blacks and students eligible for free/reduced-price lunch were included in the top 5 percent at about 50 percent of their representation in the standardization population. Although this is much higher than is seen with most tests, it makes the important point that the new tests are not a panacea that will suddenly achieve the goal of proportional representation of low socioeconomic status (SES) and minority students in talent development programs. Schools with large proportions of low SES, minority, or ELL students often consider other strategies to diversify their talent development programs. For example, schools in which students typically perform below the national average should always request local norms on *CogAT* and provide enrichment or accelerated programs for their highest-scoring students. (See also footnote.[viii])

CogAT 7 Screening Form

Some educators need a fair test to help them identify students from diverse backgrounds who should be considered for talent-development programs, but their schools cannot administer the complete *CogAT* to all students. The *CogAT* 7 Screening Form is designed to help. The test provides a fair (and more comprehensive) measure of ability than either the *CogAT* Nonverbal Battery or other nonverbal tests. The Screening Form consists of one subtest from each of the three *CogAT* batteries: the Picture/Verbal Analogies test, the Number Analogies test, and the Figure Matrices test. At the primary levels, all three tests are entirely pictorial. At grades 3 and up, only the optional Verbal Analogies subtest uses words, and this subtest may be omitted or not scored for ELL students.

The Screening Form provides a single score that estimates students' general reasoning abilities. Separate verbal, quantitative, and nonverbal scores are not

provided because they would not be reliable. The score on Screening Form score is highly correlated with both the *CogAT* Verbal Battery score and the Quantitative-Nonverbal partial composite. This makes the Screening Form an effective way to identify students who are expected to score well when administered the complete *CogAT*. However, the cut score on the Screening Form should always be more generous than the cut score on the complete *CogAT* or other placement test. See the footnote for additional information regarding the cut score for the Screening Form.[ix]

Free Practice Activities

Research shows that young children sometimes do not fully understand what they are supposed to do on tests, especially when tasks are unfamiliar and test directions are brief. Unsurprisingly, teacher-led practice can increase the validity and stability of the scores that children obtain on the tests. Therefore, we have developed practice materials that schools or districts that have purchased Form 7 can download and review with students sometime before administering either the complete *CogAT* or the Screening Form. Teachers who take the time to use these materials can help ensure that all students understand what they are supposed to do. This helps level the playing field.

Proper use of the practice materials can also help children acquire thinking strategies that can assist them in other domains, especially if they are encouraged to notice the similarities between the test-tasks and other school tasks. *CogAT* scores predict success in school because they require the same kinds of reasoning abilities that school-learning tasks require. Improving one improves the other.

Topics of forthcoming Newsletters

New features of Form 7 that are not discussed in this newsletter include: an online test (Fall 2012), new data analysis and talent-identification tools, and score reports that integrate *CogAT* scores with other information, such as achievement test scores and teacher ratings. How best to use these new features of Form 7 will be discussed in future editions of *Cognitively Speaking*. Copies of these newsletters can always be obtained from your local Riverside Assessment Consultant, from the *CogAT* website (www.CogAT.com), or from my website http://faculty.education.uiowa.edu/dlohman.

[i] Federal Interagency Forum on Child and Family Statistics. (2011). *America's Children: Key National Indicators of Well-Being*. Washington, DC: U.S. Government Printing Office.

[ii] Note that most of the changes in test scores reflect other influences – errors of measurement and different rates of development for different children are particularly important. See Lohman, D. F., & Korb K. (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. (paper on my website).

[iii] See, for example, Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri Nonverbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, 54(2), 75.

[iv] SAS scores have a mean of 100 and standard deviation of 16 in the population.

[v] Much other research shows that when compared to verbal tests, figural-reasoning nonverbal tests reduce but do not eliminate score the differences between ELL and non-ELL children. See Vol. 6 of *Cognitively Speaking* or 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. (both papers on my web site)

[vi] This is also observed on other nonverbal tests such as the NNAT. See, Naglieri, J. A., & Ronning, M. E. (2000). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, 12, 328–334.

[vii] See Lohman, D. F. (in press). Nontraditional uses of traditional measures. In C. M. Callahan & H. Hertberg-Davis (Eds.) *Fundamentals of Gifted Education*. This chapter shows how one can create local norms using some simple spreadsheet procedures. Examples are illustrated using talent identification reports from Form 7 of the *Cognitive Abilities Test*. (Paper on my website).

[viii] For details, see Lohman, D. F. (in press). Decision strategies. In S. L. Hunsaker (Ed.), Identification of students for gifted and talented services: Theory into practice. (Paper on my website).

[ix] If the same cut score is used on both tests, then many students who would have been admitted had they been allowed to take the complete *CogAT*, will be excluded. A useful rule of thumb is that in order to successfully identify eight out of ten students who would score high on the complete *CogAT*, the cut score on the Screening Form should be about three times as large as the cut score on the complete *CogAT*. For example, if the goal is to identify the top 3 percent of students on V or QN, then the cut score on the Screening Form should select the top 9 to 10 percent of the students for additional testing. Setting the proper cut score on the Screening Form is most easily accomplished if schools use local norms.