TCRE	782	B	Dispatch: 6.9.06 Journal: TCRE	CE: Blackwell
Journal Name	Manuscript No.		Author Received: No. of pages: 22	PE: Rashmi/HVA

Psychometric Issues in the ELL Assessment and Special Education Eligibility

JAMAL ABEDI
University of California, Davis

Assessments in English that are constructed and normed for native English speakers may not provide valid inferences about the achievement of English language learners (ELLs). The linguistic complexity of the test items that are not related to the content of the assessment may increase the measurement error, thus reducing the reliability of the assessment. Language factors that are not relevant to the content being assessed may also be a source of construct-irrelevant variance and negatively impact the validity of the assessment. More important, the results of these tests used as the criteria for identification and classification of ELL students, particularly those at the lower end of the English proficiency spectrum, may be misleading. Caution must be exercised when the results of these tests are used for special education eligibility, particularly in placing ELL students with lower English language proficiency in the learning/reading disability category. This article discusses psychometric issues in the assessment of English language learners and examines the validity of classifying ELL students, with a focus on the possibility of misclassifying ELL students as students with learning disabilities.

The policy of including of English language learners (ELLs) and students with disabilities (SD) is not only a necessary for reliable, valid, and fair assessments for all, but it is also a congressionally mandated policy. The recent No Child Left Behind Act (2001), which is the reauthorization of the Elementary and Secondary Education Act (the Improving America's Schools Act), and the amendments to the Individuals with Disabilities Education Act (IDEA) in 1997 require that all students be included in national and state assessments. The term *all* in the legislation means *every student*—including those with disabilities and limited English proficiency. The intent is to provide appropriate assessment based on the same high standards and ensure that all students are part of the indicators used for school accountability (Thurlow & Liu, 2001).

Many instructional decisions that will be made could have grave consequences for ELLs if their knowledge and skills are not appropriately assessed. Although the increasing level of attention to the inclusion and assessment of these students is encouraging, not enough work has been

Teachers College Record Volume 108, Number 11, November 2006, pp. 2282–2303

0161-4681

done to examine the issues and improve the quality of instruction and assessment for these students. For example, Ortiz (2002) indicated that "students learning English are likely to fail when they do not have access to effective bilingual education of English as a second language (ESL) program" (p. 41). Lack of access to effective education will also affect their assessment results. Research has clearly demonstrated that assessments designed and normed mainly for native English speakers may not as reliable and valid for ELL students.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999) elaborated on this issue:

> For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. Test use with individuals who have not sufficiently acquired the language of the test may introduce construct-irrelevant components to the testing process. In such instances, test results may not reflect accurately the qualities and competencies intended to be measured. . . . Therefore it is important to consider language background in developing, selecting, and administering tests and in interpreting test performance. (p. 91)

This article will discuss the impact of linguistic factors on assessment and classification of ELL students. Among the major threats to the validity of classifying ELL students is the indistinct line between ELL students at the lower levels of English proficiency and students with learning disabilities.

We will first explain the psychometric issues in the assessment of ELLs and then discuss the possibility of misclassification—due to the use of inappropriate assessment tools—of ELL students as having a learning disability. Jenkins and O'Connor (2002) indicated that students with reading and/or learning disabilities are not proficient in reading and writing skills. ELL students, particularly those at the lower level of English proficiency spectrum, also suffer from such lack of proficiency in English. Although a comprehensive and valid diagnostic approach can distinguish students with reading/learning disabilities from ELL students, distinguishing between these two groups of students can be a daunting task.

PSYCHOMETRIC ISSUES IN THE ASSESSMENT OF ELL STUDENTS

Literature on the assessment of English language learners suggests that students' content-based assessment outcomes may be confounded by lan-

guage background variables. ELLs generally perform lower than non-ELLs on content-based assessments such as math, science, and social sciences—a strong indication that English language proficiency affects instruction and assessment. Research also shows that ELL students' assessment outcomes suffer from lower reliability and validity; that is, language factors may be a source of measurement error in the content-based assessment of ELL students and may impact the reliability of the test. Language factors may also be a source not relevant to the construct of such assessments (Messick, 1994) and may affect the test's construct validity.

In the assessment of ELL students, the results of studies suggested that unnecessary linguistic complexity is a nuisance variable that introduces a source of measurement error and is considered a construct-irrelevant factor in the assessment (Abedi & Leon, 1999; Abedi, Leon, & Mirocha, 2003). Such nuisance variables may seriously confound the outcome of assessment in content-based areas (Abedi, Lord, & Hofstetter, 1998; Cocking & Chipman, 1988; De Corte, Verschaffel, & DeWin, 1985; Kintsch & Greeno, 1985; Trenholme, Larsen, & Parker, 1978; Lepik, 1990; Mestre, 1988; Munro, 1979; Noonan, 1990; Orr, 1987; Rothman & Cohen, 1989; Spanos, Rhodes, Dale, & Crandall, 1988).

The results of a series of experimentally controlled studies by researchers at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) on the impact of language on test performance of ELLs have clearly demonstrated that (1) ELLs have difficulty with linguistically complex test items, and (2) reducing linguistic complexity of test items narrows the performance gap between ELL and non-ELL students in content-based areas such as math and science (see for example, Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi et al., 1998; Abedi, Lord, Kim-Boscardin, & Miyoshi, 2000). Summarized below are studies that demonstrate the impact of language on the assessment outcome of ELLs and suggest that (1) ELLs had more difficulty with test items that were more linguistically complex, and (2) modifying test items to reduce the level of their linguistic complexity reduced the performance gap between ELL and non-ELL students in content-based areas.

Analyses of test data from four locations nationwide (Abedi et al., 2003) found a large performance gap between ELL and non-ELL students in reading and writing, areas that have a substantial amount of language demand. The performance gap was lower for science and lowest for math problem solving, for which the test items were less linguistically challenging for ELL students. The performance gap virtually disappeared in math computation, for which the language demands of the test items were minimal.

By reducing the impact of language factors on content-based test performance, the validity and reliability of assessments can be improved and can result in fairer assessments for all students—including ELLs and stu-

dents with disabilities. To minimize the impact of language factors and consequently reduce the performance gap between ELL and other students, language modification of assessment tools may be a viable option. For example, when math test items were modified to reduce the level of linguistic complexity, over 80% of middle school students who were interviewed preferred the linguistically modified over the original English version of the test items (see Abedi et al., 1997). Abedi et al. (1998), in a study of 1,394 eighth-grade students in schools with high enrollments of Spanish speakers, showed that modification of language of the items contributed to improved performance on 49% of the items; the students generally scored higher on shorter problem statements.

Another study (Abedi & Lord, 2001) of 1,031 eighth-grade students in Southern California found small but significant differences in the scores of students in low- and average-level math classes taking the linguistically modified version of the test. Among the linguistic features that appeared to contribute to the differences were low-frequency vocabulary, conditional clauses, and passive-voice verb constructions (for a description of the linguistic features below and for a discussion of the nature of and rationale for the modifications, see Abedi et al., 1997).

Beattie, Grise, and Algozzine (1983) found positive results in modifying tests for students with a learning disability. Math, reading, and writing tests were modified in the following ways: hierarchical progression of difficulty; unjustified arrangement of sentences; vertical arrangement of bubbles; placement of passages in shaded boxes; examples set off from test items; and arrows and stop signs in the corner of pages to indicate continuing and ending pages.

Tindal, Anderson, Helwig, Miller, and Glasgow (2000) used ''simplified language'' as a test accommodation for students with a learning disability, which they argued could also be used for ELLs. Results indicated that simplifying the language did not affect the test. However, the authors noted that their simplification process was perhaps too limited, which suggested the need for future studies to expand the simplification process.

Another study consisting of 422 students in eighth-grade science classes (Abedi, Lord, Kim-Boscardin, & Miyoshi, 2000) compared performance on National Assessment of Educational Progress (NAEP) science items in three test formats: one booklet in original format (no accommodation); one booklet with English glosses and Spanish translations in the margins; and one booklet with a customized English dictionary at the end of the test booklet. The customized dictionary, which was used for the first time by Abedi and his colleagues, included only the non-content words that appeared in the test items. By helping students with their language needs, English learners scored highest on the customized dictionary accommodation (their mean scores for the customized dictionary was 10.18 on a 20-item test as compared with means of 8.36 and 8.51 for the other two accommodations).

In a study on the impact of accommodation on eighth-grade students in math, Abedi, Lord, Hofstetter, and Baker (2000) applied four different types of accommodation: linguistically modified English version of the test; standard NAEP items with glossary only; extra time only; and glossary plus extra time. Students were also tested using standard NAEP items with no accommodation. Among these accommodations, linguistic modification of test items was the only accommodation that reduced the performance gap between ELL students and non-ELL students. Because the non-ELL students in this study, who are among the low-performing student population, also benefited from linguistic modification of test items, this suggests that clarifying the language of assessment may be helpful not only to ELL students but to SD students as well.

The summary of research above suggests that reducing the linguistic complexity of assessment materials helps ELL students and low-performing native English speakers to provide a more valid picture of what they know and can do. All students can benefit from instructional materials that are easier to understand (i.e., material with unnecessary linguistic complexity). Similarly, all students can better understand assessments that clearly convey the message related to the concept being assessed.

LINGUISTIC FEATURES THAT MAY IMPACT COMPREHENSION

The results of CRESST research led to identification of linguistic features that have greater effects on ELL student performance. These features slow down students, make misinterpretation more likely, and add to students' cognitive load, thus interfering with concurrent tasks. Indexes of language difficulty include unfamiliar words, long phrases in questions, complex sentences, and conditional and adverbial clauses. Other linguistic features that may cause difficulty for readers include long noun phrases, relative clauses, prepositional phrases, abstract versus concrete presentation of problem, passive voice, and negation. Below is a brief description of some of these features, along with some illustrative examples. A few references are added for each of the features. For a detailed description of these features, see Abedi, Courtney, and Goldberg (2000).

UNFAMILIAR WORDS

Assessments containing unfamiliar words are more difficult for ELL students than those with more familiar words. Some words, word pairs, or groups of words still unfamiliar to ELLs might be used in a test item. They are unnecessary if they are not essential to the concept being tested.

Idioms are words, phrases, or sentences that cannot be understood literally. Many proverbs, slang phrases, phrasal verbs, and common sayings

cannot be decoded by ELLs. On the other hand, words that are high on a general frequency list for English are likely to be familiar to most readers because they are often encountered. Following are a few examples of unfamiliar words used in assessments (Abedi et al., 1997; Adams, 1990; Chall, Jacobs, & Baldwin, 1990; Dale & Chall, 1948; Gathercole & Baddeley, 1993; Klare, 1974;).

Circle the clumps of eggs in the illustration.
Patty expects that each tomato plant in her garden will bear 24 tomatoes.
In the last census, 80% of the households had one or more wage-earners.

LONG PHRASES IN QUESTIONS

Long questions are used less than short questions. Complex question types might have an opening phrase or clause that either replaces or postpones the question word (Adams, 1990).

At which of the following times should Ed feed the parking meter?
Of the following bar graphs, which represents the data?
According to the passage above, where do sea turtles lay their eggs?

COMPLEX SENTENCES

A complex sentence contains a main clause and one or more subordinating (dependent) clauses. Subordinating words include *because, when, after, although, if,* and *since* (more on *if* under Conditional Clauses; Botel & Granowsky, 1974; Hunt, 1965, 1977; Wang, 1970).

Because she wants to stay in touch, Peggy frequently _____.
When she came home, he _____ the letter.
Although the ship was _____, she was calm.

LOGICAL CONNECTORS: CONDITIONAL/ADVERBIAL CLAUSES

Conditional clauses and adverbial clauses are among the sources contributing to the linguistic complexity of assessments. Logical connectors are adverbial expressions that allow a listener or reader to infer connections between two structures. They include dependent words (subordinating conjunctions; see above). In mathematics, they often include conditional "if-then" statements. Some take the form of complex sentences (Celce-Murcia & Larsen-Freeman, 1983; Haiman, 1985; Shuard & Rothery, 1984; Spanos et al., 1988).

Adverbial clauses:

When the barber was finished with the haircut, he took the customer's money.
While he was listening to music, he did his homework.

Conditional clauses:

As long as you bring your own bedding, you can stay with us.
Given that *a* is a positive number, what is–*a?*
If one pint will fill 2 cups, how many cups can be filled from 8 pints? (vs. One pint will fill 2 cups. Eight pints will fill _____ cups).
In Jean's class, there are twice as many boys as girls. If there are 10 girls in the class, how many boys and girls are there in the class?

LONG NOUN PHRASES

Nouns sometimes work together to form one concept, such as *pie chart* or *bar graph*. Sometimes adjectives and nouns work together to create meaning: *high school diploma, income tax return*. To further complicate interpretation, strings of adjectives and nouns create subjects and objects: *freshwater pond, long-term investment, new word processing program* (Celce-Murcia & Larsen-Freeman, 1983; Halliday & Martin, 1993; Just & Carpenter, 1980; King & Just, 1991; MacDonald, 1993; Spanos et al., 1988).

A loaded trailer truck weighs 26,643 kilograms. When the trailer truck is . . .
Of the following number pairs, which is the dimension of a 100-square-foot room?
To become next year's tennis team captain, how many votes will Sandra need?

RELATIVE CLAUSES

A relative clause is an embedded clause that provides additional information about the subject or object it follows. Because relative clauses are less frequent in spoken English than in written English, some students may have had limited exposure to them. Words that lead a relative clause include *that, who*, and *which*. Note: Often *that* is omitted from a relative clause. When possible, relative clauses should be removed or recast (Pauley & Syder, 1983; Schachter, 1983).

A bag that contains 25 marbles . . . (vs. One bag has 25 marbles. A second . . .)
Joe found the student who had loaned him the book.

PREPOSITIONAL PHRASES

Prepositional phrases work as adjectives or adverbs to modify nouns, pronouns, verbs, adverbs, or adjectives. When they occur before question words, between the subject and the verb, or in strings, they can be especially confusing to ELLs (Orr, 1987; Slobin, 1968; Spanos et al., 1988).

Which of the following is the best approximation of the area of the shaded rectangle in the figure above if the shaded square represents one unit of area?

ABSTRACT (VS. CONCRETE) PRESENTATION OF PROBLEM

Respondents show better performance when assessment questions are presented in concrete rather than abstract terms. Information presented in narrative structures tends to be understood and remembered better than information presented in expository text (Cummins, Kintsch, Reusser, & Weimer, 1988; Lemke, 1986).

The weights of two objects were measured *vs*. The clerk weighed two suitcases.

PASSIVE VOICE

Assessments containing passive-voice construction are more difficult for ELL students to follow. In active voice, the subject is the one performing an action. In passive voice, the one receiving the action is in the subject position. Often the "actor" is not stated (Abedi et al, 1997; Celce-Murcia & Larsen-Freeman, 1983; Forster & Olbrei, 1973)

He was given a ticket *vs*. The officer gave him a ticket.
Girls' ears were pierced in infancy *vs*. Parents pierced infant girls' ears.
When comparisons were made, the amounts in each jar had been reduced.

NEGATION

Studies suggest that a sentence containing negations (e.g., *no*, *not*, *none*, *never*) are harder to comprehend than affirmative sentences. Several types of negative forms are confusing to ELLs (Mestre, 1988):

Proper double negative:

Not all the workers at the factory are not male.
It's not true that all the workers at the factory are not male.

POSSIBLE IMPACT OF LANGUAGE FACTORS ON RELIABILITY OF ASSESSMENT

The unnecessary linguistic complexity of test items can be a source of measurement error and can reduce the reliability of the tests. Because true-score variance (σ^2_T) in classical test theory is defined as the observed-score variance (σ^2_X) minus the error variance (σ^2_E), any increase in the size of error variance directly affects (reduces) the size of true score variance (Allen & Yen, 1979; Linn & Gronlund, 1995; Salvia & Ysseldyke, 1998) and consequently decreases the reliability of the assessment. In a perfectly reliable test, the error variance (σ^2_E) would be zero; therefore, the true-score variance (σ^2_T) would be equal to the observed-score variance.

However, in measurement involving human subjects, there is always an error component. Appropriate evaluation of the measurement error is important in any type of assessment, whether in the traditional multiple-choice approach or in performance-based assessments (Linn, 1995; see also AERA, APA, & NCME, 1999). Many different sources (e.g., occasion, task, test administration conditions) may contribute to measurement error in traditional assessment instruments for all students. It is important to note, however, that the unnecessary linguistic complexity of test items as a source of measurement error differentially impacts performance of different groups of students with different levels of English proficiency. The linguistic complexity factor affects the performance of ELL students because the common characteristic of these students is their needs in the area of English language. Thus, there is an interaction between students' ELL status and their underlying measurement model.

In the classical approach to estimating reliability of assessment tools, the level of contribution of different sources to measurement error may be indeterminable. Through the generalizability approach, one would be able to determine the extent of the variance that each individual source contributes (such as occasion, task, item, scorer, and language factors) to the overall measurement error (see Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991).

To estimate reliability of the standardized achievement tests and to investigate their measurement error across different subgroups of students (e.g., ELLs versus non-ELLs), we considered different approaches. Because parallel forms or test-retest data were not available in many districts and

states across the nation, we decided to use an internal consistency approach for estimating the reliability of some of the commonly used standardized achievement tests. As mentioned, the linguistic complexity of test items may introduce another dimension into the assessment, the dimension of language. If this is the case, and because the internal consistency approach is extremely sensitive to multidimensionality of test items (see, for example, Abedi, 1996; Cortina, 1993), then the estimated reliability for ELL students may be considerably lower than the reliability for native speakers of English. One may also argue that language factors may create a restriction of range on content-based outcome measures, and the restriction of range may reduce the internal consistency coefficient. In other words, unnecessary linguistic complexity makes assessment more difficult for ELL students, thereby reducing their performance level. Lower performance level then creates a restriction of range for the ELL performance distribution, and that in turn results in lower reliability for ELL students. These two explanations both relate to the impact of language factors on the assessment of ELLs, and both indicate the negative effects of linguistic complexity of content-based assessment on ELL student performance. To illustrate this phenomenon, we present results of some analyses performed by CRESST researchers on the reliability of test items.[1]

Table 1 presents internal consistency (alpha) coefficients for the Stanford 9 subscale scores for second- and ninth-grade students from one of our data sites. Data in Table 1 show two interesting trends: (1) as expected, the internal consistency coefficients for non-ELL students were higher than those for ELL students, and (2) the difference between ELL and non-ELL students was higher in the higher grades, in which content-based language is more complex.

For students in second grade, the alpha coefficient was .914 for non-ELL students as compared with .856 for ELL students. In math, the alpha for

Table 1. Internal Consistency Coefficients for ELL/Non-ELL Students in Grades 2 and 9

	Grade 2		Grade 9	
Content Area	Non-ELL	ELL	Non-ELL	ELL
Reading	.914 (n = 234,505)	.856 (n = 101,399)	.876 (n = 181,202)	.750 (n = 52,720)
Math	.894 (n = 249,000)	.881 (n = 118,740)	.898 (n = 183,262)	.802 (n = 54,815)
Science			.805 (n = 144,821)	.597 (n = 40,255)
Social science			.806 (n = 181,078)	.530 (n = 53,925)

non-ELL students was .894, as compared with .881 for ELL students. As these data show, although alpha coefficients for ELL students were lower, the gap between ELL and non-ELL students was not large in second grade. For students in ninth grade, however, there was a larger gap between ELL and non-ELL students. In reading, for non-ELL students, the alpha coefficient was .876, as compared with .750 for ELL students. In math, the alpha for ninth grade non-ELL students was .898, as compared with .802 for ELL students. In science, the alpha for non-ELLs was .805, as compared with .597 for ELLs. Finally, in social science, the alpha for non-ELLs was .806, as compared with .530 for ELLs.

As these data suggest, the difference between internal consistency coefficients for ELL and non-ELL subscale scores were substantially larger for students in higher grades (ninth grade) than for students in lower grades (second grade). These differences were statistically significant. The average alpha for students in second grade over all subject areas was .904 for non-ELL students, as compared with .869 for ELL students—a small difference of 4%. For students in ninth grade, however, the average alpha for non-ELL was .846, as compared with an average alpha of .670, a difference of 26%. Comparing the percent alpha difference of 26% for ninth-grade students with the 4% of second-grade students once again suggests that in a more linguistically complex environment, the difference between ELL and non-ELL students is more apparent.

VALIDITY

In content-based assessments such as in math and science, the linguistic complexity of test items may introduce another dimension or construct in addition to the construct that is the target of assessments. This may be the case particularly for ELL students. The linguistic complexity factors in content-based assessment may be considered a source of construct-irrelevant variance because it is not conceptually related to the content being measured (Messick, 1994):

> With respect to distortion of task performance, some aspects of the task may require skills or other attributes having nothing to do with the focal constructs in question, so that deficiencies in the construct-irrelevant skills might prevent some students from demonstrating the focal competencies. (p. 14)

The concept of "construct-irrelevant" applies to the situations in which a construct other than the construct targeted for assessment is involved. For example, when the linguistic structure of an assessment in a content area (e.g., math or science) is so complex that ELL students cannot adequately

understand the question, then the English language becomes another construct that is measured by the test but is not relevant to the content being assessed. In other words, language interferes with the targeted content in the assessment. Linguistic complexity of test items as a possible source of construct-irrelevant variance may be a threat to the validity of achievement tests because it could be a source of measurement error in estimating the reliability of the tests. The construct-irrelevant variance may change the structural relationships among test items, between subscale scores and the fit of the structural model. Because linguistic factors may have more influence on the assessment outcomes for ELL students, then the structural relationships of ELL assessment outcomes may be different with those of English-only students.

To examine the hypothesis of differences between ELL and non-ELL students on the structural relationship of test items, a series of structural equation models was created. Fit indices were compared across ELL and non-ELL groups. The results generally indicated that the relationships between individual items, items with the total test score, and items with the external criteria were stronger for non-ELL students than for ELL students.

Item parcels in each content area (e.g., reading, science, and math) were created. Each parcel was constructed to systematically contain items with varying degrees of item difficulty. Through this process, each parcel contained three categories of difficulty: easy, difficult, and moderately difficult items. The main reason for creating item parcels was to provide multiple measures for each of the content areas. For example, rather than having a single score of math as the sum of all items in a math test, items were divided into several groups, or parcels. Each parcel of items provided a measure of math; therefore, we obtained as many measures of math as the number of parcels (for a detailed description of item parcels and ways to create them, see Cattell & Burdsal, 1975). A reading latent variable was constructed based on these four parcels. Similarly, item parcels and latent variables for science and math were created from the 48 math items and 40 science items by the same process. The correlations between the reading, math, and science latent variables were estimated. Models were tested on randomly selected subsamples to demonstrate the cross-validation of the results.

Table 2 shows the results of the structural models for ninth-grade students from a large state. To examine the consistency of the results of these analyses over independent samples, students were randomly divided into two cross-validation sample, namely sample 1 and sample 2. The results obtained under these two independent samples were consistent. For example, the average factor loading across all four item parcels in each content area and across all three content areas was .795 for sample 1 and

Table 2. Grade 9 Students From Data Site, Stanford 9 Reading, Math, and Science Structural Modeling Results (df = 51)

	Non-ELL (N = 22,782)		ELL (N = 4,872)	
	Sample 1	Sample 2	Sample 1	Sample 2
Average factor loadings Across the four parcels				
Reading comprehension	.846	.846	.746	.749
Math	.830	.830	.711	.724
Science	.708	.707	.541	.539
Average factor correlation across the three content areas	.830	.827	.794	.755
Goodness of fit				
Chi square	488	446	152	158
NFI	.997	.998	.992	.992
NNFI	.997	.997	.993	.993
CFI	.998	.998	.995	.995

Note. There was significant invariance for all constraints tested with the multiple group model (Non-ELL/ELL).

.794 for sample 2 for non-ELL students. For ELL students, the average factor loading was .666 for sample 1 and .671 for sample 2. Similarly, there is a high level of consistency on average factor correlations across the two independent samples. The average factor correlation for non-ELL was .830 for sample 1 and .827 for sample 2. For ELLs, the average factor correlation was .794 for sample 1 and .755 for sample 2. Once again, these data suggest high level of consistency across the cross-validation samples.

Table 2 also compares the structural relationships of test items across the categories of ELL. The data show major differences between ELL and non-ELL students. As data in Table 2 suggest, correlations of item parcels with the latent factors (factor loadings) were consistently lower for ELL students than they were for non-ELL students. For example, the average factor loadings across different content areas and multiple samples was .795 for non-ELLs as compared with .668 for ELLs, a substantial difference. Similarly, there was a large difference between ELL and non-ELLs on average factor correlations. For non-ELLs, the average factor correlation was .829, as compared with .774 for ELLs—once again, a large difference.

In term of fit indices, the structural models showed a good fit of the model to the data for both ELLs and non-ELLs. However, the trend of differences between ELL and non-ELLs was also seen here, even though the difference was small. Models for non-ELLs had slightly higher fit as compared with the models for ELLs.

The hypotheses of invariance of factor loadings and factor correlations between the ELL and non-ELL groups were tested. Specifically, we tested the invariance of (1) the correlations between parcel scores and a reading latent variable; (2) correlations between parcel scores and a science latent variable; (3) correlations between parcel scores and a math latent variable; and (4) correlations between content-based latent variables across the ELL and non-ELL groups.

The null hypotheses for all these tests of invariance were rejected, suggesting that ELL and non-ELL students responded differently to the test items.

ISSUES CONCERNING ELL CLASSIFICATION AND SPECIAL EDUCATION ELIGIBILITY

Researchers have expressed concerns over the validity of classification for ELL students. Because of the lack of a commonly accepted operational definition of the term *ELL* or *LEP* (limited English proficiency) and because of validity issues in the criteria used for such classification, large discrepancies have been reported in the ELL classification practices across the nation (see, for example, Abedi, 2004, 2005; Abedi et al., 1997). Although problems in the classification of ELL students is very serious and affects both instruction and assessment for these students, a more serious problem is the possibility of ELL students at the lower level of English proficiency distribution being misidentified as students with disabilities because students' limitations in English may be interpreted as a sign of learning (or reading) disability.

Artiles, Rueda, Salazar, and Higareda (2005) found that ELL students with lower levels of proficiency in L1 and L2 (first and second language, respectively) showed the highest rate of identification in the special education categories. The authors also indicated that more ELL students tend to be placed in the "learning disability" category than in "language and speech impairment." Similarly, Artiles and Ortiz (2002) found a differential rate of overrepresentation of ELL students in special education programs. For example, based on their data, 26.5% of ELLs in Massachusetts, 25.3% in South Dakota, and 20.1% in New Mexico were placed in special education programs as compared with less than 1% of ELLs in Colorado, Maryland, and North Carolina placed in similar programs. Rueda, Artiles, Salazar, and Higareda (2002) reported that in a 5-year period—1993–1994 to 1998–1999—the placement rate of Latino ELLs increased by 345%, while their overall population in the district increased by only 12% during this period of time.

To examine this complex issue of classification of ELL students when related to eligibility for special education, we first discuss issues concerning

validity of classification for ELL students and then elaborate on the criteria for placing ELL students in learning disability category. It must be noted at this point, however, that issues concerning classification of students with learning disability are beyond the scope of this article. The aim of this section is to discuss some of the technical issues concerning ELL students who are placed in the learning disability category. (For a thorough discussion of classification issues for students with learning disabilities, see Bradley, Danielson, & Hallahan, 2002).

Based on Title IX No. 25 in the No Child Left Behind Act (2001), an LEP student is defined as someone who (1) is aged 3–21; (2) is enrolled or preparing to enroll in an elementary or secondary school; (3) was not born in the United States or whose native language is not English; (4) is a Native American, Alaskan Native, or a resident of outlying areas; (5) comes from an environment in which a language other than English has had a significant impact on an individual's English language proficiency; (6) is migratory and comes from an environment where English is not the dominant language; and (7) has difficulties in speaking, reading, writing, or understanding the English language that may deny the individual the ability to meet the state's proficient level of achievement, to successfully achieve in classrooms where English is the language of instruction, or to participate fully in society (No Child Left Behind Act).

The above definition is based on information related to students' language background and their level of English proficiency. Research has expressed concerns about these sources of information. Information on students' language background is obtained from the Home Language Survey (HLS), and data on the students' level of English proficiency are based on existing English language proficiency tests. Unfortunately, the validity of the HLS is often threatened because parents may give inconsistent information for a variety of reasons, including concerns over equity of opportunity for their children, citizenship issues, and the literacy of the parent (see Abedi, 2005). Research has also raised concerns about the validity of existing English language proficiency scores as a criterion for ELL classification. Reviewers of these tests found major differences between the content that these tests measure and the alignment of the content of these tests to English as a Second Language (ESL) standards (see Abedi, 2005; Zehler, Hopstock, Fleischman, & Greniuk, 1994).

Let us assume that students are correctly being classified across the categories of ELL (ELL/non-ELL). The next question would be to look into the validity of criteria used for special education eligibility for these students. There are many different forms of disabilities with different needs and different characteristics. Based on the data from the U.S. Department of Education, Office of Special Education and Rehabilitative Services, there are at least 12 categories of disabilities. Among these categories, however,

students with a learning disability are the largest group, constituting 46% of all students with disabilities (National Center for Education Statistics, 2000).

Jenkins and O'Connor (2002) summarized some of the techniques that have been used for identifying students with reading disabilities (RD). Students with reading and/or learning disabilities often leave elementary school with deficient reading and writing skills, which makes early identification and prevention important. The authors defined the foundation of reading as "the ability to read words; the ability to comprehend language; and the ability to access background and topical knowledge relevant to specific texts" (p. 100). A student with RD, they argued, has a weakness in one or more of these three foundation areas. However, during the early developmental stages of reading, "word-level reading skill" is the most salient characteristic, and difficulties in this area can signal an RD.

Jenkins and O'Connor (2002) provided instructions as what they consider "sensible actions" to identify children with reading disability based on research. Among the criteria suggested are assessment of the prerequisite skills of letter naming and phonemic awareness early in kindergarten; watching children as they attempt to write or spell words for clues into their understanding of the alphabetic principle; and recording progress in letter and phonemic knowledge in ways that encourage closer monitoring of children who appear most at risk. It is therefore clear from the discussion above that language factors are among the most important criteria for classifying a student as having a learning disability.

Students with learning disabilities and ELL students (particularly those at the lower levels of English proficiency distribution) may have more difficulty with test items that have unfamiliar words and/or a complex linguistic structure. Thus, language factors that affect the performance of ELLs may also influence the performance of students with a learning disability. These similarities between the language background characteristics and the level of English proficiency may make ELL students with lower level of English particularly vulnerable for misclassification as students with learning and/or reading disability.

Earlier in this article, we presented results of data analyses that showed larger performance gaps between ELL and non-ELLs in areas with greater levels of language demand. A similar trend can be seen for students with disabilities in general, and students with learning disabilities in particular. To demonstrate this trend, we present results of analyses of achievement test data for students identified as having disabilities/learning disability. These results show that these students also have difficulty in content areas with higher level of language demands.

A 4-year trend of student performance in reading and math was examined on the New York State Pupil Evaluation Program (PEP) test for the third and sixth grades from 1995 to 1998 (New York State Education De-

partment, 1999). In all 4 years and for both grade levels, the performance gap between the SD and non-SD students[2] was much larger on the reading assessment than on the math assessment. For example, in 1995, the gap between the percentage of third-grade SD and non-SD students scoring above the state reference point was 51.6 percentage points on the reading assessment and 35.0 points on the math assessment. In 1998, the gap between the percentage of SD and non-SD students scoring above the state reference point was 46.6 percentage points on the reading assessment and 27.2 points on the math assessment. Similar PEP test performance gaps between SD students and non-SD students were seen in sixth grade. It is interesting to note that on a separate state assessment (Regents Competency Test), sixth-grade performance gaps between SD and non-SD students for reading and math were much smaller. Many SD students were tested on the Regents Competency tests as compared with the PEP tests. This discrepancy highlights the effect that testing only a small proportion of the SD population can have on the interpretation of results.

As part of a recent CRESST study, we examined the 1998 reading and math Stanford 9 Test data for grades 3 and 11 for a state with a large student population. In each grade level, the gap in performance between the SD students and non-SD students was larger on the reading assessment than on the math assessment. For example, in grade 11, the gap between the mean normal curve equivalent (NCE) of SD/non-ELL and non-SD/non-ELL students was 23.2 on the reading assessment and 18.3 points on the math assessment. The gap between the mean NCE of SD/non-ELL students and non-SD/non-ELL students was 33.7 on the reading assessment and 23.1 points on the math assessment. The gaps were smaller in third grade, but again, the SD student population had more difficulty with the reading assessment than with the math assessment (Abedi et al., 2003).

In the same study, data from another state provided 1998 Stanford 9 test data for grades 3 and 10 in reading and math. In each grade level, the gap in performance between the SD students and non-SD/non-ELL students was larger on the reading assessment than on the math assessment. For example, in grade 10, the gap between the mean NCE of SD/non-ELL and non-SD/non-ELL students was 27.1 on the reading assessment and 21.2 points on the math assessment. The gap between the mean NCE of SD/non-ELL students and non-SD/non-ELL students was 39.8 on the reading assessment and 25.7 points on the math assessment. The gaps were smaller in third grade, but again, the SD population had more difficulty with the reading assessment than with the math assessment (Abedi et al., 2001).

An examination of New Jersey student performance in language arts, science, and math in 2001 on the Elementary and Grade School Proficiency Assessments (ESPA & GSPA) for fourth and eighth grades revealed that in each grade level, the gap in performance between the SD students and non-

SD students was larger on the language assessment than on the science and math assessments (New Jersey Department of Education, 2001). For example, in fourth grade, the gap between the percentage of students scoring in the partially proficient category of SD and non-SD/non-ELL students was 39.6 on the language arts assessment, compared with 20.1 on the science assessment and 33.2 points on the math assessment. In eighth grade, the gap between the percentage of students scoring in the partially proficient category of SD and non-SD/non-ELL students was 56.9 on the language arts assessment, compared with 42.5 on the science assessment and 52.7 points on the math assessment.

These findings once again clearly suggest that language factors not only influence the performance of ELL students, but they also affect the performance of students with disabilities, particularly those identified as having learning disability.

DISCUSSION

Federal and state legislation calls for equal educational opportunity and inclusion of all students in assessments. On the other hand, research on the assessment and accommodation of ELL students questions the fairness of assessments that are used for these students, particularly those assessments that are developed and normed for mainstream native English speakers.

Studies that were summarized in this article clearly show a large performance gap in content-based assessment outcomes between English language learners (ELLs) and native English speakers. However, this performance gap is not necessarily due to the lack of content knowledge; it may be due to students' lack of English proficiency. The confounding of language factors with the content knowledge has raised concerns about the validity and authenticity of the available high-stakes assessment and accountability systems for ELLs, particularly those at the lower level of English proficiency.

Assessment tools that have complex linguistic structures may provide poor achievement outcomes for ELLs and SDs. The results of such assessments may not be as reliable and valid for ELLs and SDs as for non-ELL/non-SD students. Consequently, decisions made based on the results of these assessments may be problematic for ELL students and other subgroups of students with language barriers. In this article, based on the findings of experimentally controlled studies, we illustrated that the reliability of commonly used standardized assessments in content-based areas may be negatively affected by the complex linguistic structure of test items, a construct that is not the target of assessment. We have also discussed the influence of linguistic complexity of test items as a source of construct-irrelevant variance in influencing the validity of assessment.

As we demonstrated in this article, there is a larger performance gap between ELL and non-ELLs in areas with greater levels of language demand. We also showed a similar trend for students with disabilities in general, and students with learning disabilities in particular. Therefore, language factors that affect the performance of ELLs may also influence the performance of students with a learning disability. These similarities between the language background characteristics and the level of English proficiency may make ELL students with lower levels of English particularly vulnerable to misclassification as students with learning and/or reading disability.

Thus, assessment results that are influenced by linguistic factors as construct-irrelevant may not be valid criteria in the classification of ELL students. This situation becomes even more complex when ELL students are being assessed for eligibility in special education. Unfortunately, as we demonstrated in this article, the likelihood of misclassification of low-performing ELL students as students with a learning disability is not negligible. Care must be taken to increase the validity and authenticity of criteria used for eligibility of ELL students for special education. Misclassification of ELL students, particularly misidentifying them as students with learning disabilities, may have very serious consequences for these students. They may be placed in an inappropriate educational system and subsequently receive inappropriate curriculum.

Based on the results of multiple studies, cited in this article, that focus on the impact of language factors on assessment of the special needs student population, we believe that if the education community truly wants no child left behind, serious attention must be given to the current assessment and classification system for English language learners and students with disabilities, particularly ELL students with lower levels of English proficiency.

Notes

1 Data were obtained from four different locations in the nation. For further detail regarding these sites, please see Abedi, Leon, and Mirocha (2003).

2 In the studies mentioned in the rest of this subsection, the population referred to as "non-SD students" does not include English language learners (ELLs); thus, the comparison group is less likely struggling with language.

References

Abedi, J. (1996). The interrater/test reliability system (ITRS). *Multivariate Behavioral Research*, *31*, 409–417.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*, 4–14.

Abedi, J. (2005). Issues and consequences for English language learners. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing* (pp.). Malden, MA: Blackwell.

Abedi, J., Courtney, M., & Goldberg, J. (2000). *Language modification of reading, science and math test items*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Leon, S. (1999). *Impact of students' language background on content-based performance: Analyses of extant data*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Leon, S., & Mirocha, J. (2001). *Examining ELL and non-ELL student performance differences and their relationship to background factors: Continued analyses of extant data*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*, 219–234.

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Tech. Rep. No. 478). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, *19*(3), 16–26.

Abedi, J., Lord, C., Kim-Boscardin, C., & Miyoshi, J. (2000). *The effects of accommodations on the assessment of LEP students in NAEP* (CSE Tech. Rep. No. 537). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., & Plummer, J. (1997). *Language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No. 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

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

Artiles, A. J., & Ortiz, A. (Eds.). (2002). *English language learners with special needs: Identification, placement, and instruction*. Washington, DC: Center for Applied Linguistics.

Artiles, A. J., Rueda, R., Salazar, J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, *7*(1), 283–300.

Beattie, S., Grise, P., & Algozzing, B. (1983). Effects of test modifications on the minimum competency of learning disabled students. *Learning Disabled Quarterly*, *6*, 75–77.

Bradley, R., Danielson, L., & Hallahan, D. P. (Eds.). (2002). *Identification of learning disabilities: Research to practice*. Hillsdale, NJ: Erlbaum.

Botel, M., & Granowsky, A. (1974). A formula for measuring syntactic complexity: A directional effort. *Elementary English*, *1*, 513–516.

Cattell, B. R., & Burdsal, A. C. (1975). The radial parcel double factoring design: A solution to the item-vs.-parcel controversy. *Multivariate Behavioral Research*, *10*, 165–179.

Celce-Murcia, M., & Larsen-Freeman, D. (1983). *The grammar book: An ESL/EFL teacher's book*. Rowley, MA: Newbury House.

Chall, J. S., Jacobs, V. S., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Cocking, R. R., & Chipman, S. (1988). Conceptual issues related to mathematics achievement of language minority children. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17–46). Hillsdale, NJ: Erlbaum.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*, 405–438.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, *27*, 11–20 *28*, 37–54.

De Corte, E., Verschaffel, L., & DeWin, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology, 77*, 460–470.

Forster, K. I., & Olbrei, I. (1973). Semantic heuristics and syntactic trial. *Cognition*, *2*, 319–347.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.

Haiman, J. (1985). *Natural syntax: Iconicity and erosion*. New York: Cambridge University Press.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Pittsburgh, PA: University of Pittsburgh Press.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Rep. No. 3). Urbana, IL: National Council of Teachers of English.

Hunt, K. W. (1977). Early blooming and late blooming syntactic structures. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 69–90). Urbana, IL: National Council of Teachers of English.

Jenkins, J. R., & O'Connor, R. E. (2002). Chapter II: Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–149). Hillsdale, NJ: Erlbaum.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixation to comprehension. *Psychological Review, 87*, 329–354.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109–129.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*, 62–102.

Lemke, J. L. (1986). *Using language in classrooms*. Victoria, Australia: Deakin University Press.

Lepik, M. (1990). Algebraic word problems: Role of linguistic and structural variables. *Educational Studies in Mathematics*, *21*, 83–90.

Linn, R. L. (1995). *Assessment-based reform: Challenges to educational measurement*. Princeton, NJ: Educational Testing Service.

Linn, R. L., & Gronlund, N. E. (1995). *Measuring and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, *32*, 692–715.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Mestre, J. P. (1988). The role of language comprehension in mathematics and problem solving. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 200–220). Hillsdale, NJ: Erlbaum.

Munro, J. (1979). Language abilities and math performance. *Reading Teacher, 32*, 900–915.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

National Center for Education Statistics. (2000). *Digest of education statistics, 2000*. Retrieved July 12, 2002, from http://nces.ed.gov/pubs2001/digest/dt053.html

New Jersey Department of Education. (2001). *New Jersey statewide assessment reports*. Retrieved July 11, 2002, from http://www.state.nj.us/njded/schools/achievement/2002/

New York State Education Department. (1999). *1999 pocketbook of goals and results for individuals with disabilities*. Retrieved July 11, 2002, from http://www.vesid.nysed.gov/pocketbook/1999/

Noonan, J. (1990). Readability problems presented by mathematics text. *Early Child Development and Care*, *54*, 57–81.

Orr, E. W. (1987). *Twice as less: Black English and the performance of black students in mathematics and science*. New York: W. W. Norton.

Ortiz, A. A. (2002). Prevention of school failure and early intervention for English language learners. In A. J. Artiles & A. A. Ortiz (Eds.), *English language learners with special education needs* (pp. 40–58). Washington, DC: Center for Applied Linguistics.

Pauley, A., & Syder, F. H. (1983). Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics*, *7*, 551–579.

Rothman, R. W., & Cohen, J. (1989). The language of math needs to be taught. *Academic Therapy*, *25*, 133–142.

Rueda, A., Artiles, A. J., Salazar, J., & Higareda, I. (2002). An analysis of special education as a response to the diminished academic achievement of Chicano/Latino students: An update. In R. R. Valenica (Ed.), *Chicano school failure and success: Past, present, and future* (2nd ed., pp. 310–332). London: Routledge/Falmer.

Salvia, J., & Ysseldyke, J. (1998). *Assessment*. Boston: Houghton Mifflin.

Schachter, P. (1983). *On syntactic categories*. Bloomington: Indiana University Linguistics Club.

Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shuard, H., & Rothery, A. (Eds.). (1984). *Children reading mathematics*. London: J. Murray.

Slobin, D. I. (1968). Recall of full and truncated passive sentences in connected discourse. *Journal of Verbal Learning and Verbal Behavior*, *7*, 876–881.

Spanos, G., Rhodes, N. C., Dale, T. C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 221–240). Hillsdale, NJ: Erlbaum.

Thurlow, M., & Liu, K. (2001). *State and district assessments as an avenue to equity and excellence for English language learners with disabilities* (LEP Projects Report 2). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Tindal, G., Anderson, L., Helwig, R., Miller, S., & Glasgow, A. (2000). *Accommodating students with learning disabilities on math tests using language simplication*. Eugene: University of Oregon, Research, Consultation, and Teaching Program.

Trenholme, B., Larsen, S. C., & Parker, R. M. (1978). The effects of syntactic complexity upon arithmetic performance. *Learning Disabilities Quarterly*, *1*, 81–85.

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, *9*, 398–404.

Zehler, A. M., Hopstock, P. J., Fleischman, H. L., & Greniuk, C. (1994). *An examination of assessment of limited English proficient students*. Arlington, VA: Development Associates, Special Issues Analysis Center.

JAMAL ABEDI is a professor at the Graduate School of Education of the University of California, Davis, and a research partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). His interests include studies in the area of psychometrics focusing on the validity of assessment and accommodation for English language learners (ELL), and research on the opportunity to learn for ELLs.

Author Query Form

Journal	**TCRE**
Article	**782**

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers clearly on the query sheet if there is insufficient space on the page proofs. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Query No.	Description	Author Response
Q1	Please provide page range for Abedi, 2005 entry.	