Journal of Psychoeducational Assessment
http://jpa.sagepub.com/

Quantifying the "Degree of Linguistic Demand" in Spoken Intelligence Test Directions

Damien C. Cormier, Kevin S. McGrew and Jeffrey J. Evans

Journal of Psychoeducational Assessment published online 19 May 2011
DOI: 10.1177/0734282911405962

The online version of this article can be found at:
http://jpa.sagepub.com/content/early/2011/04/19/0734282911405962

Published by:
SAGE
http://www.sagepublications.com

Journal of Psychoeducational Assessment
XX(X) 1–19

Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0734282911405962
http://jpa.sagepub.com

Quantifying the "Degree of Linguistic Demand" in Spoken Intelligence Test Directions

Damien C. Cormier[1], **Kevin S. McGrew**[1], **and Jeffrey J. Evans**[2]

Abstract

The linguistic demand of spoken instructions on individually administered norm-referenced psychological and educational tests is of concern when examining individuals who have varying levels of language processing ability or varying cultural backgrounds. The authors present a new method for analyzing the level of verbosity, complexity, and total demand of spoken directions for individually administered test batteries. This preliminary methodological investigation suggests it is possible, and relatively easy, to gather useful empirical information regarding the complexity of spoken test directions using existing (readily available) text readability programs. It is suggested that best practice for individually administered psychological and educational tests may benefit from including this information when tests are published and/or compared.

Keywords

intelligence, assessment, linguistic demand, test directions

There is little doubt that the diversity of the U.S. population has been increasing for decades (Reynolds & Lakin, 1987). As a result, it is likely that most psychological and educational assessment professionals will, at some point during their career, assess an individual who has not been acculturated to reflect the general U.S. population or whose primary language is not English. In response to this trend, professional organizations have produced special diversity and nondiscrimination guidelines (American Psychological Association [APA], National Council on Measurement in Education [NCME], & American Educational Research Association [AERA], 1999). Unfortunately, the issues involved in the assessment of individuals with limited English proficiency are complex. According to Flanagan, McGrew, and Ortiz (2000),

> Few things create as much confusion or cause as much consternation for applied psychologists as the attempt to validly measure the cognitive abilities of individuals from diverse cultural and linguistic backgrounds. There are important implications when this task, one of the most difficult facing psychologists today, is not accomplished or is accomplished poorly. (p. 290)

[1]University of Minnesota
[2]Evans Consulting

Corresponding Author:
Damien C. Cormier, Institute on Community Integration/National Center on Educational Outcomes, Room 218 Pattee Hall, 150 Pillsbury Drive SE, Minneapolis, MN 55455
Email: cormi013@umn.edu.

Accordingly, there is a need for practitioners to be provided with assessment tools that accommodate the diversity of cultural and linguistic characteristics of potential examinees and a means by which to evaluate the sensitivity of existing assessment instruments to these examinee characteristics. Although advancements have been made in the technical adequacy (construct validity) of cognitive measures (McGrew, 2005, 2009), a number of other test design components may influence the validity of test-based inferences from standardized testing, particularly with culturally and linguistically diverse (CLD) students. For example, the translation and adaptation of an English-based cognitive or achievement test battery to the student's native language is a practice that has received increased attention in the assessment literature (Rhodes, Ochoa, & Ortiz, 2005). Thus, test authors and publishers are likely to continue their efforts to develop reliable and valid measures of cognitive and achievement skills for CLD students. In the interim, it is important that researchers provide empirical methods by which to evaluate the cultural and linguistic influences on existing tests and provide comparative test battery information that allows practitioners to engage in empirically validated best practices.

CLD Students

The phrase *culturally and linguistically diverse* has been used interchangeably with other labels such as *limited English proficient*, *language minority*, or *English-language learners* (Garcia & Cuellar, 2006). In general, CLD designates a broad group of individuals, which includes all individuals from diverse backgrounds, whether they are proficient in English or not. This broad definition provides a multifaceted approach to diversity issues by including both culture and language, which may provide individually distinctive as well as interactive qualities in an individual's ability to function as a part of American society. This is particularly relevant to assessment practices, as an individual's unique combination of cultural and linguistic characteristics may have an impact on the results, interpretations, and decisions derived from the individual's performance on a norm-referenced test battery. More culturally and linguistically sensitive assessment procedures have been a part of the larger goal of a more culturally responsive educational system that better serves the increasing diversity of the U.S. student population (Klingner et al., 2005).

The emerging issues regarding how CLD individuals will influence the future of education and society in the United States is not surprising, given that the proportion of minority students enrolled in U.S. elementary and secondary schools has reached 40% (Garcia & Cuellar, 2006). Furthermore, according to year 2000 U.S. census data, there is close to 47 million people aged 5 years or older who speak a language other than English at home. Unfortunately, as has been the case for over 30 years, concerns related to overrepresentation of these groups in special education programs continues to be an important and complex issue for educational professionals (Harry & Klingner, 2006). Furthermore, it is believed that measures of cognitive abilities should be improved to take into account social and cultural learning or their use should be minimized and occur only within the context of expert clinical judgment when making educational placement decisions (Klingner et al., 2005). A fundamental variable that underlies the issues of cultural and linguistic diversity in assessment is the process of acculturation.

Acculturation. According to the *APA Dictionary of Psychology* (VandenBos, 2007), *acculturation* is "the process by which groups or individuals integrate the social and cultural values, ideas, beliefs, and behavioral patterns of their culture of origin with those of a different culture" (p. 8). Anthropologists have defined acculturation as a bidirectional, dynamic process that involves the interchange between two cultural groups that come into contact with each other (Sayegh & Lasry, 1993). The process of acculturation is recognized to be different for, and multidimensional across, different cultures and largely dependent on continuous contact occurring in order for it to progress (Padilla & Perez, 2003). A number of measurement tools have been developed in an attempt to quantify the degree of acculturation of a diverse group of individuals (Matsudaira, 2006). However, the influence of acculturation on assessment is often confused with assimilation, though it has been noted that it is common practice to qualify acculturation as the influence a dominant culture has on a minority culture (Cabassa, 2003). In the current context, *assimilation* is defined as "the process by which an immigrant to a new culture adopts the culture's beliefs and practices" (VandenBos, 2007, p. 76). Cultural assimilation models also typically describe a multilayered, largely unidirectional process, which include linguistic, social, economic, and legislative influences (Flannery, 2001).

Based on the anthropological definition, the process of acculturation should cause the entire population characteristics of the United States to change as a function of the fluctuating cultural composition of its society and the bidirectional influences between the dominant and smaller subcultures. Due to space constraints, a thorough discussion of the conceptual and sociopolitical issues surrounding acculturation and assimilation will not be covered in this article. Instead, in the interest of facilitating an understanding of the role of these cultural and linguistic variables in the context of individually norm-referenced testing, where instruments are typically designed as per the dominant culture in a population (e.g., American culture in the development of intelligence tests), the more common term—*acculturation*—will be used. In this article, the term *acculturation* describes the *unidirectional* process that involves a CLD individual becoming more familiar with the dominant American culture, as opposed to simultaneously considering the bidirectional influence that CLD individuals have on the dominant American culture.

Bias in Testing

Bias in cognitive assessment has been an active and controversial topic of discussion for decades. Psychometric test bias, as defined by the *Joint Test Standards* (AERA, APA, & NCME, 1999), is "a systematic error in a test score. In discussing test fairness, bias may refer to construct under-representation or construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers" (p. 172). Although a broad range of statistical and sociopolitical bias issues have been investigated and discussed in the context of individual intelligence or cognitive assessment (see Jensen, 1980), the changing landscape of education results in many of these issues reemerging with regularity. Potential cultural and linguistic bias in cognitive assessment is now being actively debated regarding Hispanic and Asian students, similar to earlier debates and investigations that focused primarily on differences between Caucasian and African American populations. Regardless of the specific cultural groups involved, test bias has typically focused on two broad areas—cultural and linguistic bias.

Cultural bias. Historically, a number of test construction issues have been suggested as contributing to inaccurate lower scores on cognitive measures for CLD individuals. Inappropriate content and standardization samples, examiners' and language bias, and differential predictive validity are select examples of components implicated in the inaccurate assessment of cognitive ability in nondominant cultural groups in the U.S. population (Brown, Reynolds, & Whitaker, 1999). However, empirical evidence has not always supported the culture-based face or content validity claims of biased cognitive tests or individual test items (Brown et al., 1999). In addition, some research has suggested that bias can work to increase the scores of students from a minority group (e.g., see Reynolds, 2000). A detailed accounting of the salient statistical and sociopolitical bias issues is well beyond the scope of the current article. Nevertheless, a clear outcome of test bias research and discussions has been the consensus that test developers need to consider both cultural and linguistic biases when testing CLD individuals. This is illustrated by the articulation of professional standards that indicate that assessment professionals (e.g., school psychologists) need to engage in culturally sensitive practice, which is a fundamental competency of school psychologists as reflected in the Blueprint for School Psychology Training and Practice (Ysseldyke et al., 2004).

Linguistic bias. Although cultural bias has been the focal point of much of the debate regarding the use of cognitive measures with CLD students (Reynolds, 2000), some research has suggested that linguistic bias may have a greater impact on cognitive test scores. Specifically, the influence of expressive and receptive language demands for certain individual tests in cognitive test batteries has been identified as a potentially important source of construct irrelevant invalidity in individual test scores (Cormier, Hansen, & McGrew, 2011). Therefore, considerations related to individual test selection, particularly for CLD students, should take into account the level of linguistic demand of each subtest to be administered.

Clearly, the assessment process involves a number of procedures that can introduce a variety of biases during the administration of individual tests or throughout the entire assessment. The ability of an individual to participate meaningfully in the assessment process may be most influenced by his or her level of language ability. Specifically, an individual's ability to "comprehend the instructions, formulate and verbalize the responses, or otherwise complete a given task or provide an appropriate response" (Rhodes et al., 2005, p. 157) is likely to have an effect on overall performance. The individual subtests in a cognitive test battery may display a wide range of diversity in the language demands placed on an examinee. As a result, the linguistic load of a given subtest's directions may be an important consideration in test administration, particularly with CLD students.

The Assessment of CLD Students

Practitioners must be sensitive to potential sources of cultural and linguistic biases that may occur when testing and making educational decisions given the increasingly diverse U.S. population. Unfortunately, it appears that a number of school psychologists do not feel comfortable with standardized cognitive test–based interpretations when assessing CLD students, which they attribute primarily to a lack of training in this area (Rhodes et al., 2005). The development of empirically validated procedures and measures that reflect the degree with which each individual test in a cognitive test battery may be influenced by cultural loading and linguistic demand would be useful to practitioners as they select, administer, and interpret norm-referenced measures of cognitive ability.

Test Directions and Linguistic Demand

The emergence of systematic investigation of CLD assessment issues with individually administered norm-referenced cognitive tests has been a recent phenomenon. Ortiz (in McGrew & Flanagan, 1998) first described a task analysis expert–based approach to determining the degree of linguistic demand placed on examinees by the various tests in intelligence batteries This approach, as well as the subsequent classification of the tests in all major intelligence batteries, was first presented by McGrew and Flanagan and was recently updated by Flanagan and colleagues (Flanagan & Ortiz, 2001; Flanagan, Ortiz, & Alfonso, 2007). According to Flanagan et al. (2000), tests were "evaluated on the basis of the level of language proficiency required by the examinee in order to comprehend the assessor's instructions and provide an appropriate response" (p. 116). The product was a three-category qualitative classification (high, medium, low) of tests in terms of degree of linguistic demand.

Despite the usefulness of these initial logical attempts to classify tests in intelligence batteries as per degree of linguistic demands, the lack of an empirical basis for these classifications represents a critical unmet need. Test directions appear to be central to the issue of linguistic demand. This could be conceptualized as either the listenability of the test directions from the perspective of the individual being tested or the readability of the test directions from the administrator's perspective. In previous studies (Beatty & Payne, 1984; Sawyer & Kosoff, 1981), listenability and readability have been assumed to be equivalent. In one case, this was based on the contention that this is an assumption commonly held in the field of education (Sawyer & Kosoff, 1981). Although Fang (1966-1967) stated that listenability is not necessarily readability, his easy listening formula had a correlation of .96 with the Flesch reading ease formula (Klare, 1974). Furthermore, within models of listening and reading there are "nearly identical components of reading; both involve sensory, attentional, decoding, comprehending and inferential activity" (Rubin, Hafer, & Arata, 2000). Thus, given the context of the current study, the term *readability* will be used to maintain consistency with the formulae used though it is assumed that it can be used interchangeably with listenability, as has been done in previous work.

The purpose of this article is to describe the development and empirical evaluation of a preliminary applied methodology for quantifying one dimension (degree of receptive linguistic demand placed on examinees during testing) of the multidimensional construct of degree of linguistic demand. The primary research question was whether useful empirical information regarding the complexity of spoken test directions can be obtained by adapting the component parts of existing text readability programs.

Method

Sample

The 20 tests from the Woodcock–Johnson Tests of Cognitive Ability—3rd ed. (WJ-III: Woodcock, McGrew, & Mather, 2001) and the 10 tests from the *Woodcock–Johnson Diagnostic Supplement* (Woodcock, McGrew, Mather, & Schrank, 2003) memory for names, visual closure, sound patterns—voice, number series, cross-out, memory for sentences, block rotation, sound patterns—music, and memory for names—delayed) were analyzed,[1] as well as cognitive components of the WJ-III Achievement Battery: Ga (sound awareness) test, and four oral language tests (understanding directions, oral comprehension, story recall, picture vocabulary). For each test, a minimum of two data files were created, each of which, when analyzed with the software program, provided the 11 test parameter variables for the text passages (see Table 1). The first file included the standard directions spoken by an examiner (to an examinee) for the complete test. This *standard* file represented the typical instructions heard by most examinees. The second file (*supplemental*) included the additional directions spoken by the examiner in response to queries and error or no response boxes. Finally, to facilitate interpretation of the results the scores for two of the variables (monosyllabic percent and syllables percent) were *reflected* (i.e., multiplied by −1). This ensured that a high value for all variables was associated with a greater degree of linguistic demand.

Procedure

A literature search identified more than 100 different predictive text readability formulas that have been reported to produce divergent estimates as a function of the type of text (technical, prose, etc.), length of text analyzed, age of the subjects, and so on. The Readability Calculations suite of software programs (Micro Power and Light Co., 2000) was selected for this investigation as it provided nine of the most frequently used readability formulae (e.g., Dale-Chall, Flesch Reading Ease, FOG, Spache). More important, the program reported the component parts used

Table 1. Individual Text Parameter Variables Produced by Reliability Calculation Software: Names and Definitions

Variable label	Software variable name	Definition
Monosyllabic Percent	MONOSYL100[a]	Total number of monosyllabic words per 100 words (easy words)
Average syllables	SYLWRD	Average syllables per word in text sample
Percent syllables	SYLWRD100	Average syllables per 100 words in text sample
Total syllables	SYL	Total number of syllables in text sample
Total words	WRD	Total number of words in text sample
Total monosyllabic	MONOSYL	Total number of monosyllabic words in text sample
Total sentences	SENT	Total number of sentences in text sample
Average words	WRDSNT	Average number of words per sentence in text sample
Sentence percent	SENT100[a]	Total number of sentences per 100 words in text sample
Polysyllabic words	POLYSYL	Total number of words of three or more syllables in the sample
Polysyllabic percent	POLYSYL100	Total number of polysyllabic words per 100 words (difficult words) in text sample

a. MONOSYL100 and SENT100 variables were each *reflected* (multiplied by −1) to produce scales with uniform interpretation

in the derivation of each readability estimate. Table 1 lists and defines the 11 individual text parameter variables provided by the software.

Analysis of the Standard Direction Test Files

The *standard* direction analysis was designated the benchmark analyses, given that it included data for all tests and it represented the most common set of directions different examinees may hear. Upon completion of the *standard* file analyses for each test all summary statistics (all readability estimates plus the individual text parameters) were entered into a file for analyses via standard statistical software.

Correlations of the individual text parameters. Simple Pearson product–moment correlations were calculated between all readability estimates across tests. An inspection of the correlation matrix revealed that many of the readability estimates were perfectly (or nearly perfectly) correlated. This finding was not unexpected, given that the different readability formula estimates shared many of the same individual text component parameters in their respective formula calculations. Given that readability estimates often produce disparate estimates (different grade levels scores), it was decided to eliminate the readability formulas from further analyses and to explore the data with a new lens. Although the readability formulas have the advantage of providing absolute and meaningful reference points (e.g., 3.5 grade level), the variability in the estimates, plus the surplus connotation from the term *readability*, argued for their elimination. Instead, the individual text parameters were used for all subsequent analyses.

Inspection of the individual correlations and two-variable scatterplots between all pairs of individual text variables revealed a number of variables that were more or less mathematically equivalent, which introduced multicolinearity into the correlation matrix. For example, monosyllabic percent correlated .92 with the average syllables and syllables percent variables. In turn, the syllables variable correlated .90 with polysyllabic percent. The total monosyllabic variable correlated .97 with total syllables and .99 with total words. Finally, the sentence percent and total words variables correlated .91.[2] As a result, six variables were dropped from the analyses (monosyllabic percent, syllables percent, total syllables, total monosyllabic, sentence percent, and polysyllabic percent). The variance accounted for by these variables was almost completed accounted for by the remaining five variables (average syllables, total words, total sentences, average words, polysyllabic words). Inspection of the correlation matrix of these five variables indicated that though related, each offered unique information regarding the characteristics of the test direction files that were analyzed.

Exploratory component analysis of individual text parameters. The five final individual text parameters were subjected to a principal component analysis with oblique (correlated) rotation of components. A principal component analysis produced Eigenvalues of 2.27, 1.09, 0.99, 0.61, and 0.04. The Eigenvalues were graphed as per the standard scree plot method. The scree plot suggested the possibility of two to three components. The third component was a singleton dimension defined by a .99 loading by the average syllables variable, with all other variable loadings on the component ranging from −.14 to +.14. Two- and three-factor common-factor and maximum-likelihood factor analysis solutions were then attempted, but both suffered from problems in numerical estimation. The common-factor solution required 69 iterations for convergence and produced a Heywood case on the first rotated factor (total words variable loaded at 1.00). The maximum-likelihood solution converged in 11 iterations but indicated that the maximum number of factors that could be extracted was two and produced a similar Heywood case on the first factor (total words variable factor loading of 1.01). It is most likely that the numerical estimation problems with the common-factor and maximum-likelihood factor solutions were due to the reduced variable correlation matrix still containing multicolinearity between some variables (or sets of variables). Given this difficulty, coupled with the fact that the research question was not attempting to isolate latent individual difference variables but instead account for the variance in the directions in test stimuli, it was decided that the two-component principal component solution was the best available solution. This solution was deemed acceptable, given the exploratory nature of the study and the interpretability of the two components. Although the common-factor and maximum-likelihood-factor solutions were not considered acceptable, the same −.20 correlation reported between the two factors in each solution does provide support for the two different dimensions being present in the data. Clearly the two components need to be interpreted with caution and be considered exploratory at this time. The two-component solution was retained and rotated to an oblique (correlated) solution. Table 2 presents the rotated two-component solution.

Interpretation of components. A review of the variables with salient loadings on each component in Table 2 suggests the following interpretations of the two components. Component 1 is defined primarily by the absolute total amount words (total words) and sentences (total sentences) in the directions of the tests and thus was labeled *verbosity*. Component 2 (*complexity*) was defined primarily by the variable average words (.96). Sentences with more words are assumed to be more complex. This hypothesis is supported by the second highest component loading (.34) for the variable that represented the average number of syllables per word in the text samples (total syllables). Thus, this second component appears to reflect how complex both the sentences (average number of words) and words in the sentences (how many syllables) are in a tests set of directions.

Construction of composite indices. To simplify subsequent data analyses, three different index scores were created based on the two-component model (one index for each component and a global average of the two index scores). The composition of the two primary component indexes

Table 2. Principal-Component Oblique-Rotated Pattern Matrix

Variable	Component 1	Component 2
WRD (total number of words in text sample)	0.99	0.27
SENT (total number of sentences in text sample)	0.89	−0.18
POLYSYL (total number of words of three or more syllables in the sample)	0.66	−0.15
WRDSNT (Average number of words per sentence in text sample)	0.07	0.96
SYLWRD (Average syllables per word in text sample)	−0.05	0.34

mirrored the composition of the components in Table 2. Given that it was not possible to generate factor scores due to the use of principle components analysis, index scores were calculated based on standardized *z* scores for each variable. All variables were first standardized to place them on a common *z*-score metric ($M = 0$, $SD = 1.0$). The respective variables for each index score were then summed and averaged by the number of variables. A total index score (total demand) was obtained from the average of the two index scores. The three final index variables were labeled *total direction demand index*, *direction verbosity index*, and *direction complexity index*.

Results

As noted previously,[2] the results reported here are only for the 20 tests from the WJ-III cognitive battery and the cognitive ability tests in the achievement battery.[3] As reported in Table 3, the WJ-III understanding directions test was rated highest on both the total verbosity (3.703) and total demand indices (1.708). This suggests that the instructions for the understanding directions test contains the most words per sentence and that the words and syllables were more numerous, longer, and more complex, relative to the spoken directions in the other WJ-III tests. For example, the task demands of the understandings test are similar to a test that would provide the examinee with an picture that contains a number of shapes[4] and began by saying, “Point to the red circle, then the green square,” and move on to more difficult items, such as “Point to the top right corner of the square, then point to the center of the blue circle, but start by touching the three corners of the yellow triangle.” In these type of tasks, the directions given to the test taker increase not only in length but also in the complexity of the task demand that is requested by the test administrator. The words are not difficult (e.g., *point to*, *circle*, etc.), but the length and complexity of each item increases in directional complexity.

As reported in Table 4, the spatial relations test rated highest on the total complexity index (3.248). This suggests that spatial relations is highest in the number of words per sentence and the number of syllables per word in the verbal instructions set. Spatial relations was also rated highly (second highest) on the total demand index (1.389; see Table 5). The instructions for spatial relations and understanding directions were rated higher than all other tests in the WJ-III Standard Battery. These two tests were more than 3 times higher in their ratings on verbosity and complexity than the next highest tests and were approximately twice as high on the total demand index compared to the next highest rated set of instructions. A test like spatial relations, though using nonverbal visual-figural stimuli to solve a puzzle that requires the examinee to fit pieces together, requires examinees to first understand that the pieces are meant to go together to create a desired shape. Although the essential cognitive task demand is not difficult to comprehend, the verbal directions necessary to communicate to the examinee what they must do and how they

Table 3. WJ-III Test—Spoken Directions Verbosity Index

Standard directions		Supplemental directions	
Test name	Verbosity (z)	Test name	Verbosity (z)
Understanding directions	3.703	Concept formation	2.942
Concept formation	1.306	Picture vocabulary[b]	1.669
Story recall	1.099	Oral comprehension	0.816
General information	1.059	Antonymsb	0.692
Analysis-synthesis	0.404	Auditory attention	0.370
Picture vocabulary[b]	0.347	Synonyms[b]	0.300
Rhyming[a]	0.127	Visual matching	0.151
Antonyms[a]	0.079	Rhyming[a]	0.130
Auditory attention	0.069	Planning	0.062
Picture recognition	−0.002	Analysis-synthesis	0.041
Visual matching	−0.178	Auditory working memory	−0.006
Substitution[a]	−0.199	Substitution[a]	−0.205
Synonyms[b]	−0.247	Deletion[a]	−0.246
Reversal[a]	−0.251	General information	−0.300
Numbers reversed	−0.259		
Planning	−0.298	Incomplete words	−0.370
Decision speed	−0.344	Verbal analogies[b]	−0.559
Verbal analogies[b]	−0.352	Numbers reversed	−0.714
Pair cancellation	−0.389	Rapid picture naming	−0.718
Retrieval fluency	−0.395	Pair cancellation	−0.731
Auditory work memory	−0.436	Picture recognition	−0.764
Visual closure	−0.469	Story recall	−0.772
Spatial relations	−0.470	Reversal[a]	−0.790
		Sound blending	−0.866
Oral comprehension	−0.531	Retrieval fluency	−0.915
Sound blending	−0.585		
Deletion[a]	−0.602		
Memory for words	−0.654		
Visual-auditory Learning	−0.724		
Incomplete words	−0.792		
Rapid picture naming	−0.799		
VAL—delayed	−0.934		

Note: Gaps represented by white space in the columns represent the natural breaks in the empirical evidence gathered in the current study and were interpreted as appropriate, albeit arbitrary, categorizations of low, moderate, and high linguistic demand.
a. Subcomponents of sound-awareness subtest.
b. Subcomponents of verbal comprehension subtest.

must respond are rather lengthy. The task directions become increasingly difficult as more pieces are introduced and the puzzle must be put together keeping in mind certain guidelines such as piece rotation.

Among the WJ-III tests with supplemental directions, concept formation was highest on the total verbosity index (2.942) and the total demand index (1.579). This suggests that the supplemental test instructions for concept formation contain the most words per sentence and that the words and

Table 4. WJ-III Tests Spoken Directions Complexity Index

Standard directions		Supplemental directions	
Test name	Complexity (*z*)	Test name	Complexity (*z*)
Spatial relations	3.248	Pair cancellation	1.551
Auditory working memory	0.711	Planning	1.057
Decision speed	0.559	Deletion[a]	0.962
Analysis-synthesis	0.498	Numbers reversed	0.583
Planning	0.452	Sound blending	0.463
Memory for words	0.429	Oral comprehension	0.436
Story recall	0.408	Auditory working memory	0.387
Substitution[b]	0.388		
		Concept formation	0.215
Concept formation	0.167	Incomplete words	0.164
Rapid picture naming	0.113	Antonyms[b]	0.085
Pair cancellation	0.112	Picture recognition	0.005
Deletion[a]	0.088	Synonyms[b]	−0.022
Visual matching	0.000	Analysis-synthesis	−0.047
Sound blending	−0.041	Reversal[a]	−0.082
Verbal analogiesb	−0.070	Picture vocabulary[b]	−0.109
Numbers reversed	−0.081	Substitution[a]	−0.198
Incomplete words	−0.108	Retrieval fluency	−0.204
Reversal[a]	−0.186	General information	−0.261
Oral comprehension	−0.251	Rhyming[a]	−0.309
Retrieval fluency	−0.284	Visual closure	−0.320
Under directions	−0.287	Story recall	−0.372
Visual-auditory learning	−0.511	Verbal analogies[b]	−0.576
Picture vocabulary[b]	−0.522	Rapid picture naming	−0.692
Picture recognition	−0.550	Visual matching	−0.740
Rhyming[a]	−0.555	Memory for names	−1.101
VAL—delayed	−0.631	Auditory attention	−1.506
Synonyms[b]	−0.825		
Antonyms[b]	−0.836		
General information	−0.847		
Auditory	−0.919		

Note: Gaps represented by white space in the columns represent the natural breaks in the empirical evidence gathered in the current study and were interpreted as appropriate, albeit arbitrary, categorizations of low, moderate, and high linguistic demand. VAL—delayed = Visual-auditory learning—delayed.
a. Subcomponents of sound-awareness subtest.
b. Subcomponents of verbal comprehension subtest.

syllables are more numerous, longer, and more complex, creating the greatest linguistic demand of all the WJ-III tests analyzed. Tests similar to concept formation involve asking the examinee why objects are grouped together (what is the underlying conceptual rule), and supplemental directions are often necessary to explain the relationships and rules when an examinee fails to comprehend an item. For example, an analogous task would require an examinee, when presented with three cats and three dogs of the same color, to induce the underlying conceptual rule that explains why the different species of animals are contained in each respective box. The examinee is asked what the relationship is

Table 5. WJ-III Test Spoken Directions Total Demand Index

Standard directions		Supplemental directions	
Test name	Total demand (z)	Test name	Total demand (z)
Understanding directions	1.708	Concept formation	1.579
Spatial relations	1.389	Picture vocabulary[b]	0.780
Story recall	0.753	Oral comprehension	0.626
Concept formation	0.736	Planning	0.560
Analysis-synthesis	0.451	Pair cancellation	0.410
		Antonyms[b]	0.389
Auditory working Memory	0.138	Deletion[a]	0.358
Decision speed	0.108		
General information	0.106	Auditory working Memory	0.191
Substitution[b]	0.095	Synonyms[b]	0.139
Planning	0.077	Analysis-synthesis	−0.003
Picture vocabulary[b]	−0.088	Numbers reversed	−0.066
Visual matching	−0.089	Rhyming[a]	−0.089
Memory for words	−0.112	Incomplete words	−0.103
Pair cancellation	−0.139	Sound blending	−0.201
Numbers reversed	−0.170	Substitution[a]	−0.202
Verbal analogies[b]	−0.211	General information	−0.281
Rhyming[a]	−0.214	Visual matching	−0.295
Reversal[a]	−0.218	Picture recognition	−0.379
Deletion[a]	−0.257	Reversal[a]	−0.436
Picture recognition	−0.276		
Sound blending	−0.313	Retrieval fluency	−0.559
Retrieval fluency	−0.339	Verbal analogies[b]	−0.568
Rapid picture naming	−0.343	Auditory attention	−0.568
Antonyms[b]	−0.379	Story recall	−0.572
Oral comprehension	−0.391	Rapid picture naming	−0.705
Auditory attention	−0.425		
Incomplete words	−0.450		
Synonyms[b]	−0.536		
Visual-auditory learning	−0.617		
VAL—delayed	−0.783		

Note: VAL—Delayed = Visual-auditory learning—delayed. Gaps represented by white space in the columns represent the natural breaks in the empirical evidence gathered in the current study and were interpreted as appropriate, albeit arbitrary, categorizations of low, moderate, and high linguistic demand.
a. Subcomponents of sound-awareness subtest.
b. Subcomponents of verbal comprehension subtest.

between the animals in the box. The task can become increasingly difficult as animals are grouped together and other characteristics of the animals are varied (e.g., by color, size, quantity). Since the concept formation task is a miniature conceptual rule–learning task where the examiner provides corrective feedback and reinforcement, when a participant fails an item, the supplemental corrective directions become increasingly complex and verbose as the relationships between the objects are explained. For the direction complexity index on supplemental instructions, pair cancellation was rated highest (1.551), followed closely by planning (1.057). Test directions for a cancellation task (similar to pair cancellation) may involve a number of rules such as circling a specified sequence of three items while working through columns of items under the constrain to "work as quickly as you can." For example, a page containing a seemingly random sequence of apples and oranges organized in 12 columns with 40 objects per column is presented to the examinee. The examinee is instructed to only circle groups of three oranges in a row and is instructed to do so as efficiently and effectively as possible. Supplemental directions would involve a repetition of the directions followed by a visual demonstration of what is expected of the test taker, in addition to corrective instructions that are often lengthy.

Operational classification of low, medium, and high categories. To allow comparison with the Flanagan and colleagues three-category high/medium/low system of the 20 reported tests, the z values in Tables 3 and 4 were inspected for natural breaks in the values. For example, in Table 3 there appeared to be a "gap" between the group of highest verbosity tests (standard direction condition) that had picture vocabulary as its lowest boundary (0.347 or above) and the next grouping of tests starting with rhyming (0.127 to −0.470) subtests (for standard directions). The lowest grouping appeared to start with oral comprehension (−0.531). These subjective gaps are represented in each respective column by white space.

Correlations Between Indices

A Pearson correlation analysis was conducted to determine the general strength of the relationship between the verbosity and complexity of subtests directions. As seen in Figure 1, the Pearson correlation for the standard test directions is $r = -.129$ ($p = .50$). Figure 2 shows the Pearson correlation with the inclusion of supplemental test direction. The correlation between verbosity and complexity with supplemental test directions is $r = -.041$ ($p = .85$). Taken together, these results indicate the verbosity and complexity dimensions provide different information.

Discussion

The dearth of information available on the topic of degree of linguistic demand in the spoken directions for intelligence batteries prompted the current investigation. The purpose of this article was to describe and report the results of the development of a preliminary applied methodology for quantifying one dimension (degree of receptive linguistic demand placed on examinees during testing) of the multidimensional construct of degree of linguistic demand. Our primary research question was whether useful empirical information regarding the complexity of spoken test directions could be obtained by adapting the component parts of existing text readability programs. We conclude that it is possible, and relatively easy, to gather useful empirical information regarding the complexity of spoken test directions using existing (readily available) readability programs. Therefore, this may prompt future research that could be potentially helpful to practitioners when selecting subtests to be administered to students, particularly to those with limited language proficiency. It becomes evident that these types of practice may be especially useful when working with CLD students, as their culture and linguistic backgrounds have a high probability of influencing their ability to understand directions, particularly if those directions use complex language and are rather lengthy.

Comparison to Previous Classifications

A classification of the Woodcock–Johnson Test of Cognitive Abilities was presented in Flanagan et al. as a representation of the cultural loading and linguistic demand of each subtest on a

Table 6. Linguistic Demand Classification Comparison

	Classification						
Test name	C-LIM linguistic demand	Verbosity standard	Verbosity supplemental	Complexity standard	Complexity supplemental	Total demand standard	Total demand supplemental
Analysis-synthesis	High	High	Moderate	High	Moderate	High	Moderate
Auditory attention	High	Moderate	High	Low	Low	Moderate	Low
Auditory working Memory	High	Moderate	Moderate	High	High	Moderate	Moderate
Concept formation	High	High	High	Moderate	Moderate	High	High
Decision speed	High	Moderate		High		Moderate	
General information	High	High	Moderate	Low	Moderate	Moderate	Moderate
Incomplete words	High	Low	Low	Moderate	Moderate	Moderate	Moderate
Memory for words	High	Low		High	Low	Moderate	
Sound blending	High	Low	Low	Moderate	High	Moderate	Moderate
Story recall	High	High	Low	High	Moderate	High	Low
Numbers reversed	Moderate	Moderate	Low	Moderate	High	Moderate	Moderate
Rapid picture naming	Moderate	Low	Low	Moderate	Low	Moderate	Low
Retrieval fluency	Moderate	Moderate	Low	Moderate	Moderate	Moderate	Low
VAL—delayed	Moderate	Low		Low		Low	
Visual-auditory learning	Moderate	Low		Low		Low	
Visual matching	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Pair cancellation	Low	Moderate	Low	Moderate	High	Moderate	High
Picture recognition	Low	Moderate	Low	Low	Moderate	Moderate	Moderate
Planning	Low	Moderate	Moderate	High	High	Moderate	High
Spatial relations	Low	Moderate		High		High	

Note: C-LIM = culture-language interpretive matrix. Only tests included in the C-LIM classification are presented in the table.

continuum of low, medium, and high for both these dimension in the culture-language interpretive matrix (C-LIM). The classification of the linguistic demand of the subtests suggests that when tests have an increase in linguistic demand, there is a pattern of decline in subtest scores that emerges (Flanagan et al., 2007). Thus, we compare the results presented here to the classification seen in the C-LIM.

As seen in Table 6, there is considerable variation between the current WJ-III linguistic demand classifications (based on verbosity and complexity indices) classification and the Flanagan and colleagues' classifications. For example, Flanagan and colleagues classified the WJ-III spatial relations subtest as a low-linguistic-demand measure. In contrast, as reported in Table 6, the spatial relations subtest is classified moderate in terms of verbosity and high on directional complexity for the standard directions and is not classified for supplemental

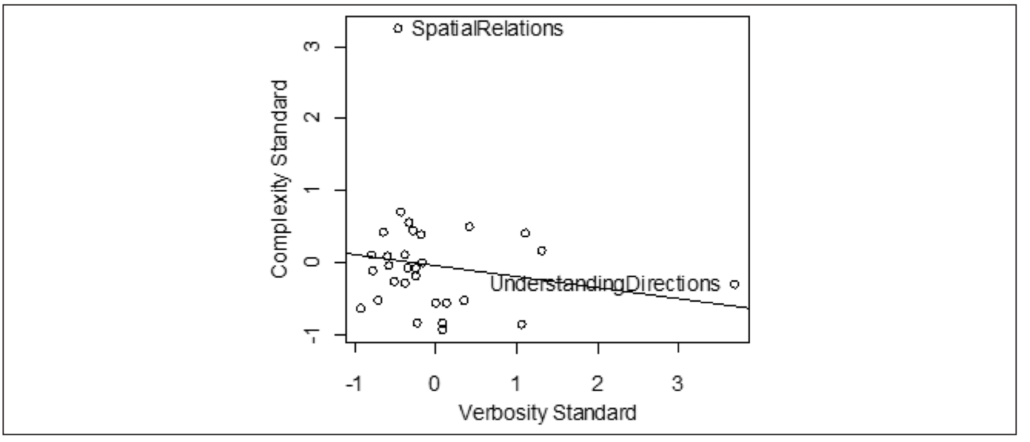

Figure 1. Pearson correlation between complexity and verbosity indices for standard test directions.

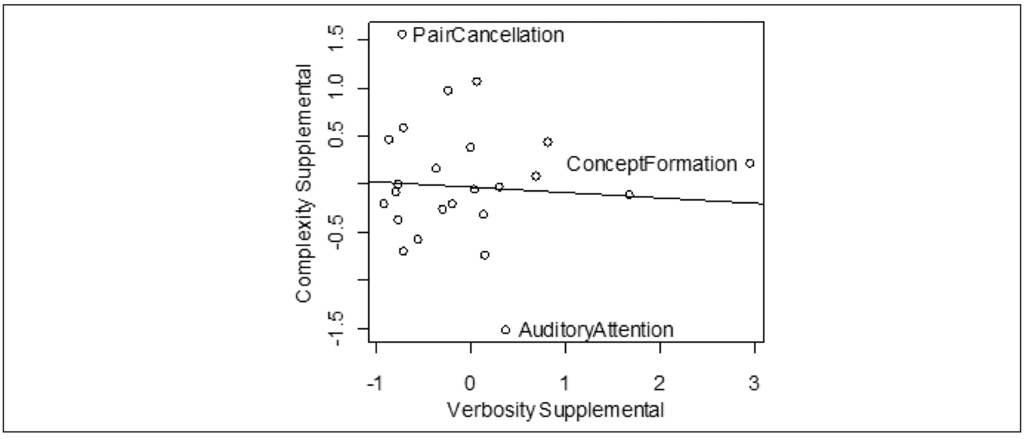

Figure 2. Pearson correlation between complexity and verbosity indices with supplemental test directions.

directions due to lack of sufficient supplemental directions to analyze. Thus, the current analysis suggests that the WJ-III spatial relations test may be more linguistically demanding than suggested by Flanagan and colleagues' low classification. Another example is the concept formation test. Flanagan and colleagues classify concept formation as high in linguistic demand. The current empirically based classification system suggests this high classification may be partially correct. As reported in Table 6, concept formation is indeed a high-linguistic-demand test when defined by verbosity. However, though the directions may be lengthy and verbose, the complexity of the directions is only moderate.

On the basis of the information seen in Tables 3 and 4, we see that the classification varies depending on the index used (verbosity or complexity) and whether or not the supplemental directions are included. It should be noted that not a single test was categorized at the same degree of linguistic demand for all four indices represented in Tables 3 and 4 (e.g., verbosity standard, verbosity supplemental, complexity standard, and complexity supplemental). For example, the C-LIM categorizes pair cancellation as being low in linguistic demand, but in the current analysis this test is only categorized as low only when considering the verbosity of the supplemental directions. If an examinee requires only the standard directions, then the linguistic demands for pair cancellation are moderate for both verbosity and complexity. The above three examples (spatial relations, concept formation, and pair cancellation) indicate that the receptive language linguistic demand of tests is more nuanced and complex than implied by the Flanagan and colleagues' singular low-, medium-, and high-category system. The classification of tests varies as a function of linguistic demand dimension (verbosity vs. complexity) and whether an examinee requires frequent exposure to supplemental directions. Thus, in general, the classification of the linguistic demand of tests appears to be far more complex than suggested in previous classification systems. Although a relatively simple three-category system has intuitive appeal it appears to mask an important finding that emerged from the current investigation. Namely, degree of linguistic demand of direction (in this case receptive linguistic demand) is more of a continuum and not a simple three-category scheme.

Limitations

It was not possible in this study to gauge and compare the WJ-III tests to tests from other batteries in their linguistic demand dimensions. The focus was not on rendering a judgment about the WJ-III tests, but rather, the WJ III tests served only as a medium to develop and demonstrate the methodology described here. It would only be possible to compare tests (on this dimension) across different batteries if the other batteries were also included in the analysis. Then relative (not absolute) evaluation comparisons could be made and broader recommendations could be made to practitioners, as a variety of tests are used in practice, depending on what is available.

In addition, the current analysis only focused on the receptive language demands placed on examinee's and did not incorporate information regarding overt oral expressive demands or any internal symbolic language that may transpire when an examinee works to perform a task. The singular focus on receptive linguistic demands may explain, in part, some of the discrepancy between the linguistic demand classifications reported here and those reported by Flanagan and colleagues.

Finally, the inability to produce appropriate factor analysis solutions (vs. principal component solutions) suggests the current interpretations need to be tempered by the caveat that the identified components included measurement error variance in addition to latent variable variance (that could not be isolated). The verbosity and complexity dimensions should not be interpreted as representing latent *individual difference* factors. Rather, they represent the totality of complete variance present in the WJ-III test directions.

Implications

Research. The approach to quantifying the degree of demand in test directions used in the current article provides information about a potential influence on a student's ability to answer prompts from a cognitive measure in a way that best represents his or her actual cognitive abilities. Furthermore, the results presented here demonstrate the variability in test characteristics independent from the student's characteristics, which may be an area explored in future research. Specifically, it may be important to determine the extent to which the variation in the complexity of directions interacts with a student's language ability. Moreover, this is yet another area that may be of particular interest when testing CLD individuals and future research may also uncover that linguistic complexity of test directions has a particularly differential influence on certain language or cultural backgrounds. Additional research should also focus on subjecting a large pool of tests across common intelligence batteries (e.g., WISC-IV, SB5, WJ-III, KABC-II, DAS-II) to provide potentially useful information to practitioners across cognitive test batteries. Finally, similar to the WJ-III-based C-LIM research findings of Kranzler, Flores, and Coady (2010), the current results raise questions regarding the accuracy of all linguistic demand classifications based on expert consensus as per the published C-LIM matrices for most cognitive test batteries. Although the classifications provided in this article suffer from limitations (particularly only a focus on receptive linguistic demands of directions), they do suggest that a simple high/medium/low linguistic demand classification is too simplistic and fails to capture the multivariate complexity of this continuous (not categorical) dimension—demand may vary by whether an examinee is subjected to both standard and supplemental directions and whether an examinee's difficulties focus more on the verbosity or complexity of receptive language.

Practice. The C-LIM proposed by Flanagan and colleagues has served an extremely important function of drawing attention to the linguistic demands placed on examinee's during cognitive testing. It has clearly served as the spark plug to direct practitioners to examine this characteristic of tests when examining individuals from different cultural and linguistic backgrounds. However, the current results suggest that improvements in the analysis of tests as per linguistic demand can be improved if future research and applied writings recognize that the linguistic demand dimension of tests is a continuum and is a complicated interaction of at least two linguistic dimensions (verbosity and complexity) and whether examinee's are primarily exposed to standard directions while other examinees require administration of both standard and supplemental directions on tests.

One potential practical implication of the methodology described and implemented in this study is the possibility that assessment personnel could determine which test batteries, individual tests, or individual items within a test battery may pose relatively more difficulty for students who may struggle with receptively processing lengthy or complex verbal instructions. Information about a test's linguistic difficulty may be particularly important when working with students who are English-language learners, from diverse cultural backgrounds, or when testing children with known or suspected language processing disorders. Requiring, or presuming a given level of English-language proficiency of students, is not considered best practice. It is incumbent upon assessment specialists to sufficiently understand the characteristics of the tools they use to assess individuals, particularly when an individual is from a diverse cultural background or has language processing difficulties. This is important given that a student's performance on a test may be hindered by poor receptive English-language competence and not the specific ability construct intended to be measured by the test. The optimal selection of the most appropriate tests, and subsequent clinical interpretation of test performance, might be facilitated by consulting the verbosity, complexity, and total demand indices proposed in this preliminary study. The relative degree of linguistic demand of tests of intelligence batteries (and other test batteries) needs to be a better known quantity to better inform the selection of subtests, much like nutrition facts on food packaging in the United States.

Acknowledgment

The authors wish to thank (posthumously) Dr. John "Jack" Carroll for initially suggesting this relatively simple preliminary approach to the analysis of cognitive test directions. Aside from being a seminal scholar regarding the nature and structure of human intelligence (see Carroll, 1993), his expertise and passion for psycholinguistics is often not recognized. When the authors first consulted with Dr. Carroll, we had expected that he would suggest a complex psycholinguistic discourse analysis method based on computer-intensive text-based algorithms (e.g., corpus-based analysis; see Gregg, Coleman, Stennett, & Davis, 2002). Instead, he suggested that before investing in time-intensive research one should first explore the feasibility of developing a method for quantifying the receptive language load of directions via simple reading text readability formulas.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interests with respect to the authorship and/or publication of this article: Dr. Kevin S. McGrew wishes to provide a potential conflict of interest notice. He has a financial (royalty) interest in the WJ-III Battery as a WJ-III coauthor.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

Notes

1. Descriptions of the 30 WJ-III tests used in this analysis can be found in the WJ-III *Technical, Cognitive, Achievement, and Diagnostic Supplement Test Manuals* (e.g. McGrew & Woodcock, 2001). Although the 10 tests from the *Diagnostic Supplement* were included in the analysis, only the results for the 20 tests in the WJ-III tests of cognitive abilities (the cognitive-related tests in the WJ-III tests of achievement) are reported and discussed in this paper. The WJ-III *Diagnostic Supplement* tests are not used frequently in practice and had not been classified by Flanagan and colleagues (Flanagan, Ortiz, & Alfonso, 2007) at the time the current research was completed.
2. A copy of the complete set of variable intercorrelations can be obtained by contacting the second author: Kevin McGrew, iap@earthlink.net
3. The results for all 30 tests, inclusive of the WJ-III *Diagnostic Supplement* tests, can be obtained by contacting the second author: Kevin McGrew, iap@earthlink.net
4. The understanding directions test uses pictures of common scenes as test stimuli. In this manuscript, when the task demands of WJ-III tests are explained, the actual test items are not described. Instead, analogous fictitious tasks are described that convey the same WJ-III task demand. The use of fictitious test descriptions is to protect the content of the WJ-III tests described for test security purposes.

References

American Psychological Association, National Council on Measurement in Education, & American Educational Research Association. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

Beatty, M. J., & Payne, S. K. (1984). Listening comprehension as a function of cognitive complexity: A research note. *Communication Monographs, 51*, 85-90.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since bias in mental testing. *School Psychology Quarterly, 14*(3), 208-238.

Cabassa, L. J. (2003). Measuring acculturation: Where we are and where we need to go. *Hispanic Journal of Behavioral Sciences, 25*(2), 127.

Cormier, D. C., Hansen, A., & McGrew, K. (2011). Cultural and linguistic biases in cognitive assessment. Paper presented at the February, 2011 conference of the National Association of School Psychologists, San Francisco.

Carroll, J.B. (1993). Human Cognitive Abilities. Cambridge University Press: Cambridge.

Fang, I. E. (1966-1967). The “Easy Listening Formula”. *Journal of Broadcasting, 11*, 63-68.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler Intelligence Scales and Gf-Gc Theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. (2001). *Essentials of cross-battery assessment*. New York, NY: John Wiley.

Flanagan, D. P., Ortiz, S., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). New York, NY: John Wiley.

Flannery, W. (2001). An empirical comparison of acculturation models. *Personality and Social Psychology Bulletin, 27*(8), 1035.

Garcia, E., & Cuellar, D. (2006). Who are these linguistically and culturally diverse students? *The Teachers College Record, 108*(11), 2220-2246.

Gregg, N., Coleman, C., Stennett, R. B., & Davis, M. (2002). Discourse complexity of college writers with and without disabilities: A multidimensional analysis. *Journal of Learning Disabilities, 35*(11), 23-38.

Harry, B., & Klingner, J. (2006). *Why are so many minority students in special education?*. New York: Teacher College Press.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: Free Press.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 1*, 62-102.

Klingner, J. K., Artiles, A. J., Kozleski, E., Harry, B., Zion, S., Tate, W., . . . Riley, D. (2005). Addressing the disproportionate representation of culturally and linguistically diverse students in special education through culturally responsive educational systems. *Education Policy Analysis Archives, 13*, 43.

Kranzler, J. H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review, 39*(3), 431-446.

Matsudaira, T. (2006). Measures of psychological acculturation: A review. *Transcultural Psychiatry, 43*(3), 462.

McGrew, K. S. (2005). The Cattell-Horn-Carroll (CHC) theory of cognitive abilities. Past, present and future. In D. Flanagan & Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 136-202). New York, NY: Guilford.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross battery assessment*. Boston: Allyn & Bacon.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual. Woodcock-Johnson-III*. Itasca, IL: Riverside Publishing.

Micro Power and Light Co. (1995). *Readability calculations (readability calculations)*. Dallas, TX: Author.

Padilla, A. M., & Perez, W. (2003). Acculturation, social identity, and social cognition: A new perspective. *Hispanic Journal of Behavioral Sciences, 25*(1), 35.

Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored?. *Psychology, Public Policy, and Law, 6*(1), 144-150.

Reynolds, M. C., & Lakin, K. C. (1987). Noncategorial special education: Models for research and practice. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), *Learner characteristics and adaptive education* (pp. 331-356). New York, NY: Pergamon.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. A. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. New York, NY: Guilford.

Rubin, D. L., Hafer, R., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education, 49*(2), 121-133.

Sawyer, D. J., & Kosoff, T. O. (1981). Accommodating the learning needs of reading disabled adolescents: A language-processing issue. *Learning Disability Quarterly, 4*(1), 61-68.

Sayegh, L., & Lasry, J.-C. (1993). Immigrants' adaptation in Canada: Assimilation, acculturation, and orthogonal cultural identification. *Canadian Psychology, 34*(1), 98-109.

VandenBos, G. R. (2007). *APA dictionary of psychology*. Washington, DC: American Psychological Association.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Woodcock-Johnson III diagnostic supplement to the tests of cognitive abilities*. Itasca, IL: Riverside Publishing.

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., . . . Telzrow, C. (2006). School psychology: A blueprint for training and practice III. Bethesda, MD: National Association of School Psychologists.

Bios

Damien C. Cormier, Department of Educational Psychology, University of Minnesota.

Kevin S. McGrew, Department of Educational Psychology, University of Minnesota.

Jeffrey J. Evans, Evans Consulting, Bloomington, MN.