Assessment of English Language Learners:

Evidence-based evaluation and best practice.

Texas Association of School Psychologists

Georgetown, TX

November 4, 2021

Samuel O. Ortiz, Ph.D.
St. John's University

Most common language spoken at home other than English or Spanish

Source: Business Insider tabluations of 2017 American Community Survey IPUMS data

BUSINESS INSIDER

Evaluation of English Learners

The majority of school-based referrals for evaluation are based primarily on poor academic attainment, especially in the areas of reading and writing.

Unfortunately, ELs are at greater risk than native English speakers for poor academic achievement because:

ELs face the double challenge of learning academic content ***and*** *the language of instruction simultaneously.*

Even more unfortunate is that this means that ELs are also at greater risk for referral for school-based evaluation.

Source: Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. American Educator, 32 (2) pp. 8-23, 42-44.

Understanding First and Second Language Acquisition

Basic Interpersonal Communication Skills (BICS)

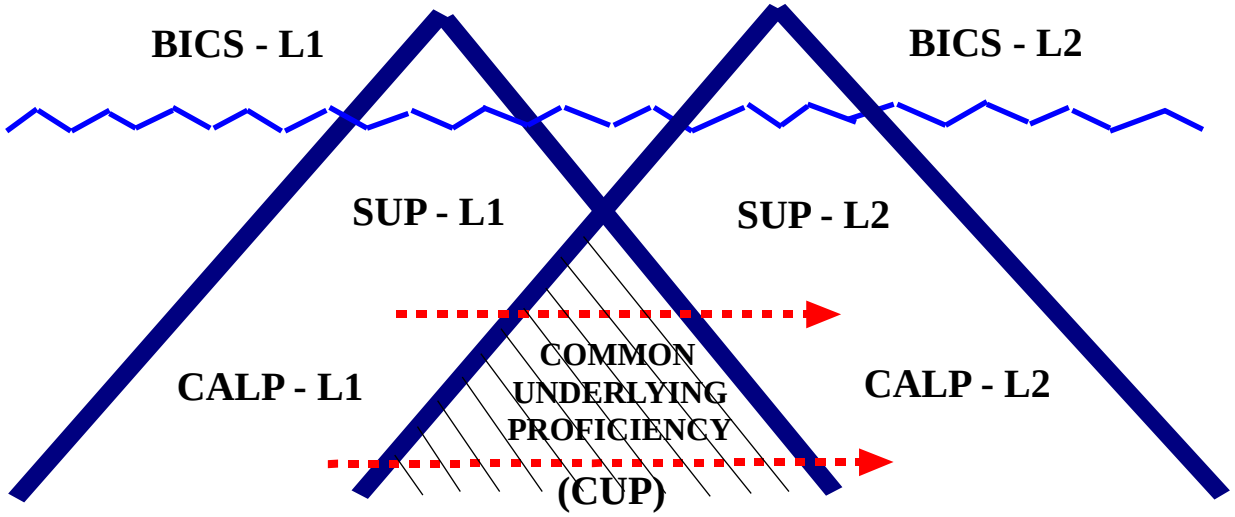
- *ability to communicate basic needs and wants, and ability to carry on basic interpersonal conversations*
- *takes 1 - 3 years to develop and is insufficient to facilitate academic success*

Cognitive Academic Language Proficiency (CALP)

- *ability to communicate thoughts and ideas with clarity and efficiency*
- *ability to carry on advanced interpersonal conversations*
- *takes at least 5-7 years to develop, possibly longer and is required for academic success*

Cummins' Developmental Interdependence Hypothesis ("Iceberg Model")

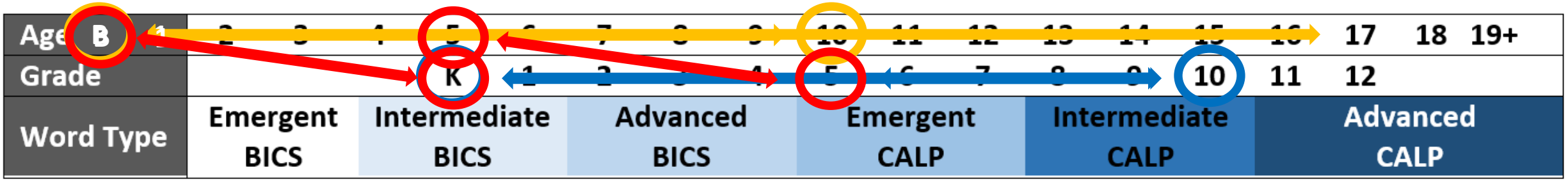
- *BICS is the small visible, surface level of language, CALP is the larger, hidden, deeper structure of language*
- *each language has a unique and Separate Underlying Proficiency (SUP)*
- *proficiency in L1 is required to develop proficiency in L2,*
- *Common Underlying Proficiency (CUP) facilitates transfer of cognitive skills*



Source: Illustration adapted from Cummins (1984) *Bilingual And Special Education: Issues In Assessment and Pedagogy*.

For Els, Validity is Not Established by Age Alone

Approximation between Age, Grade, and Word Type for Native English Speakers



The chronological age of an EL, by definition, does not indicate how long the individual has been learning English. Exposure to English can vary considerably among ELs of the same age or grade.

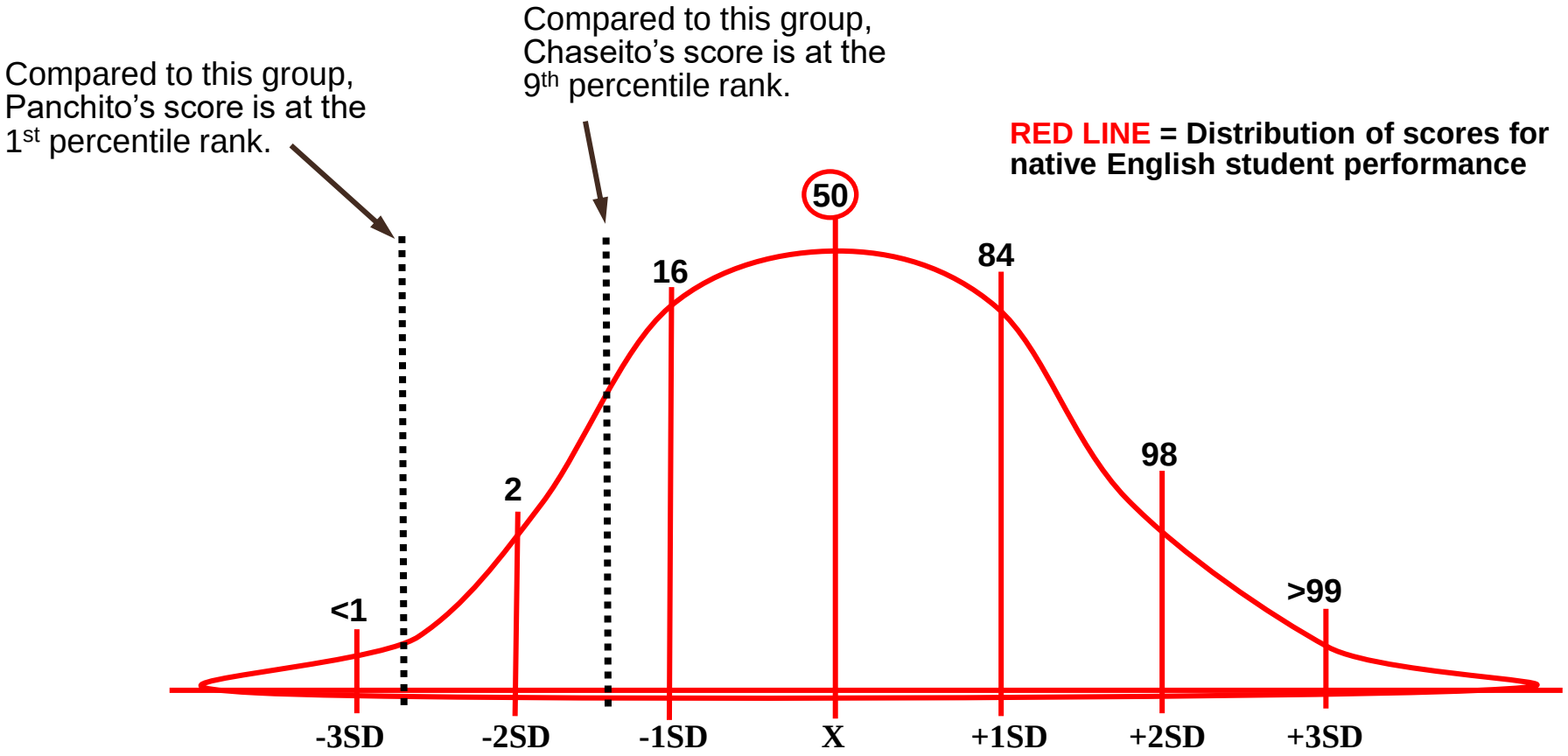
Some ELs may start learning English upon school entry at the age of 5.

Other ELs may start learning English upon school entry that occurs at a much later date, such as at the age of 10 or 5th grade.

A 17 year old EL may have been learning English for as long as 16+ years, or a 17 year old may have been learning English for as little as 1 month.

Comparing ELs by age alone, will not control or provide fairness regarding the wide range of variability in their respective exposures to English and the amount of time they each may have been learning English across their lifetimes.

Does L2 cognitive performance suggest difference or disorder?



For diagnostic purposes in determining whether a disability exists, use of a monolingual English speaking comparison group is discriminatory and makes it appear incorrectly that both students might have some type of disability that is actually due to differences in linguistic/cultural development.

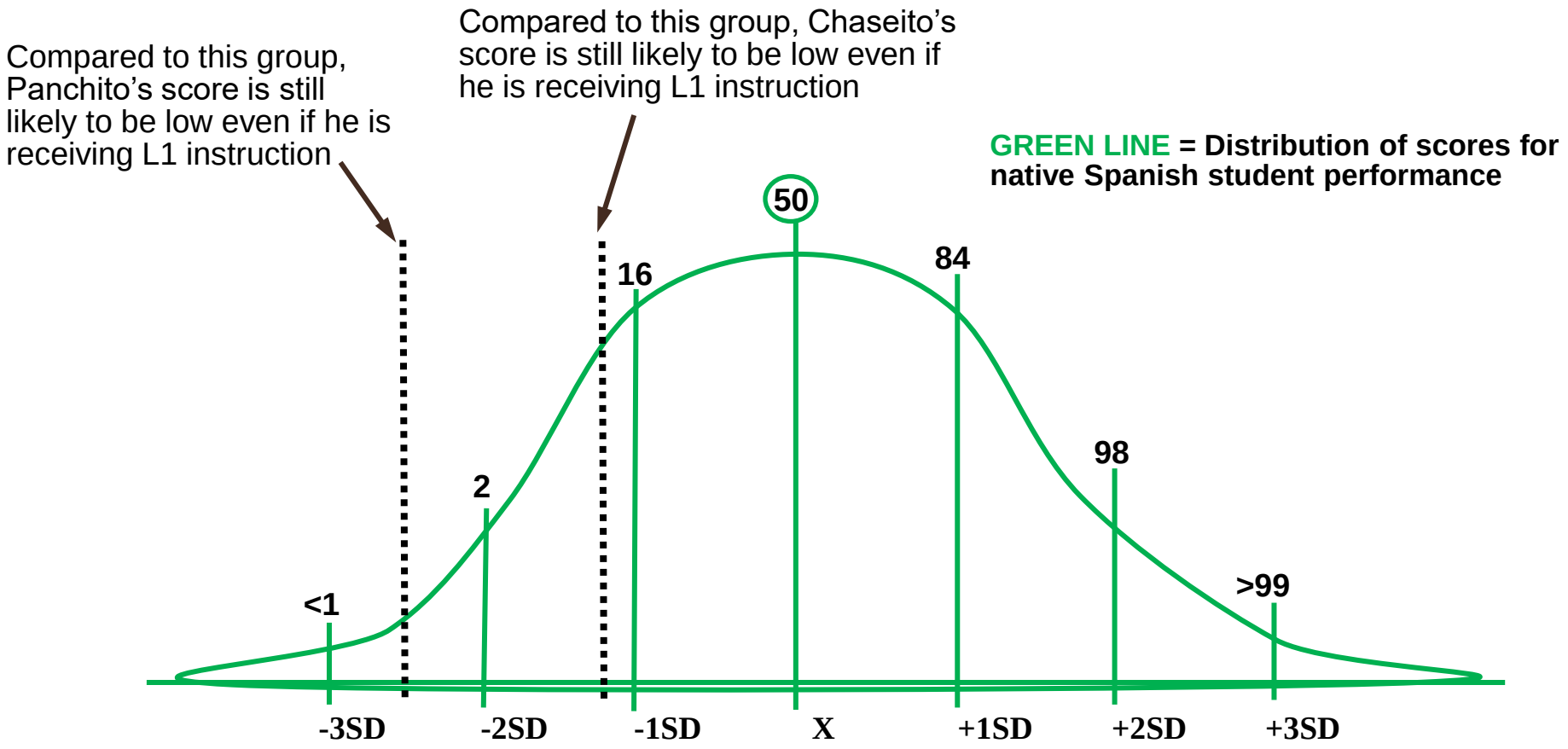
Isn't Language-Matching the Most Important Factor of Testing with ELs?

No. For native English speakers, growth of cognitive abilities and knowledge acquisition are tied closely to age and assumes normal educational experiences. Thus, age-based norms effectively control for variation in development and provide an appropriate basis for comparison. However, this is not true for English learners who may neither live in a "mainstream" culture nor benefit to an equivalent degree from formal education as native English speakers.

Development Varies by Experience – Not necessarily by race or ethnicity

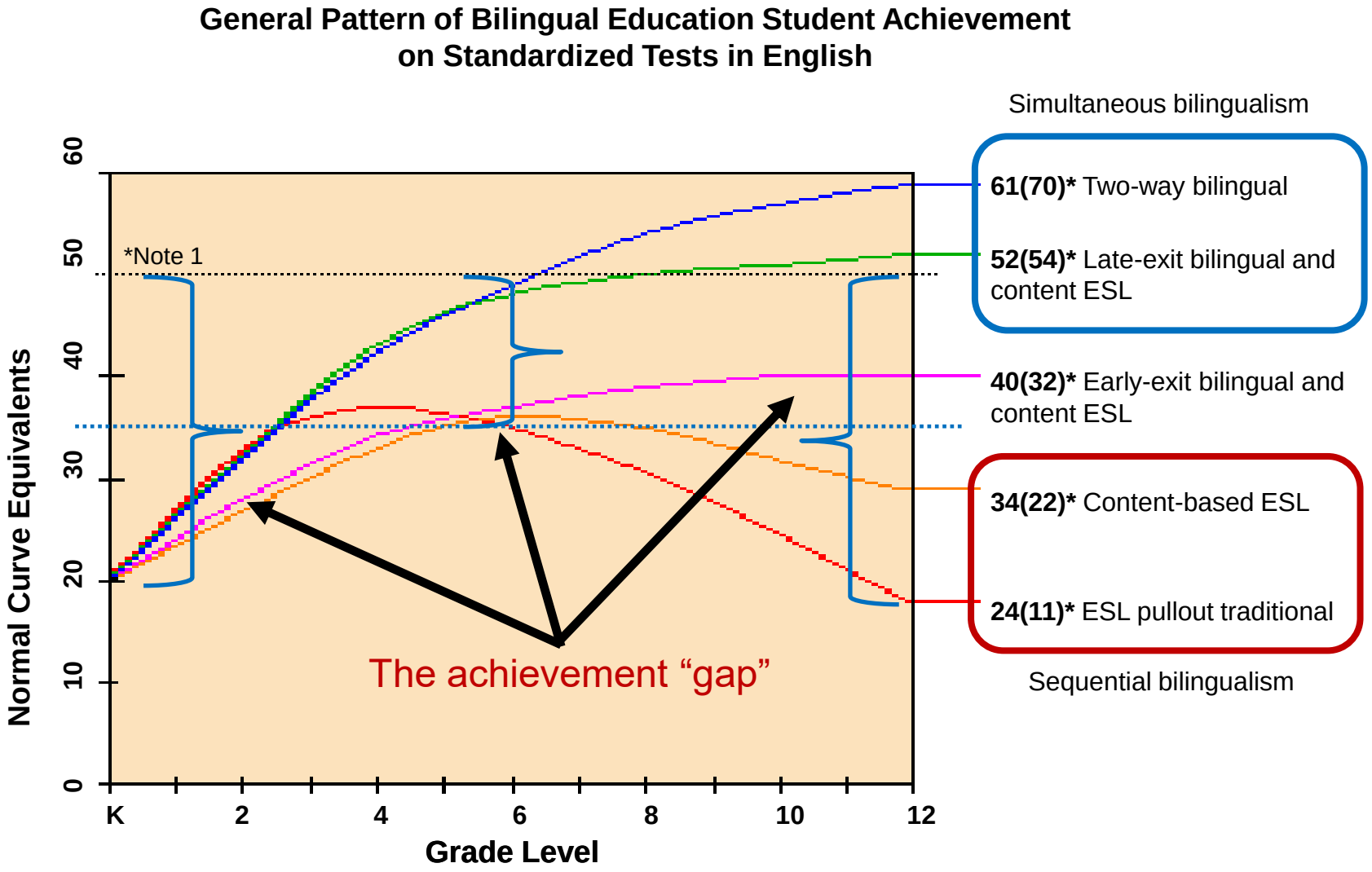
> *"The key consideration in distinguishing between a difference and a disorder is whether the child's performance differs significantly from peers with similar experiences." (p. 105)*
>
> - Wolfram, Adger & Christian, 1999

Does L1 cognitive performance suggest difference or disorder?



Similarly, use of a monolingual, native-language speaking group remains discriminatory because neither student is monolingual anymore (even when receiving native language instruction) and for diagnostic purposes, these differences in development continue to make it appear incorrectly that both have some type of disability.

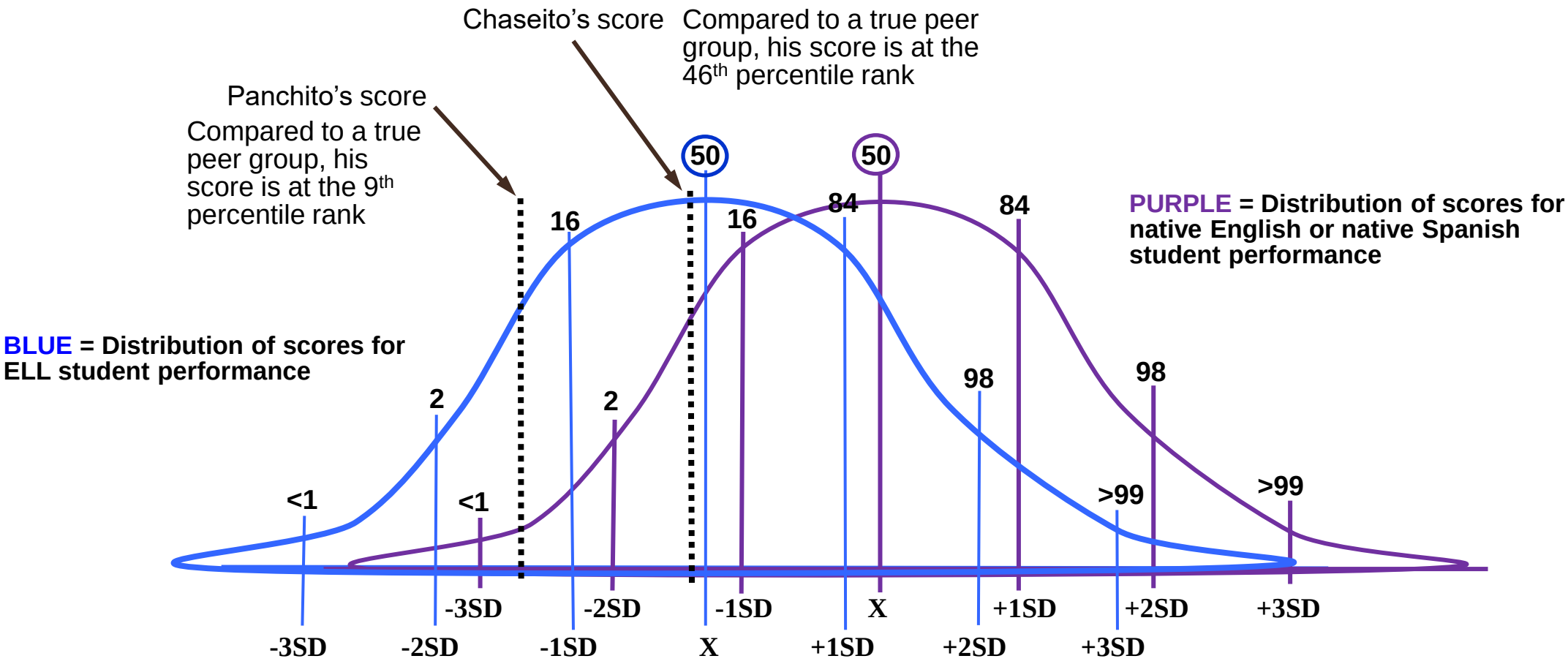
Does L2 academic performance suggest difference or disorder?



*Note 1: Average performance of native-English speakers making one year's progress in each grade. Scores in parentheses are percentile ranks converted from NCEs.

Adapted from: Thomas, W. & Collier, V. (1997). Language Minority Student Achievement and Program Effectiveness. Washington DC: National Clearinghouse for Bilingual Education.

Only true peer comparison can address difference or disorder



Whether conducted through RTI/MTSS or testing, only use of a "true peer" comparison group provides the basis for making non-discriminatory diagnostic decisions as long as there is control for developmental language differences between English learners and English speakers and among English learners and other English learners.

What's the Problem with Tests and Testing with ELs?

For native English speakers, growth of language-related abilities are tied closely to age because the process of learning a language begins at birth and is fostered by formal schooling. Thus, age-based norms effectively control for variation in development and provide an appropriate basis for comparison. However, this is not true for English learners who may begin learning English at various points after birth and who may receive vastly different types of formal education from each other. It is their experiences that differ, not merely their heritage languages and cultures.

Development Varies by Experience – Not necessarily by race or ethnicity

> *"It is unlikely that a second-grade English learner at the early intermediate phase of language development is going to have the same achievement profile as the native English-speaking classmate sitting next to her. The norms established to measure fluency, for instance, are not able to account for the language development differences between the two girls. A second analysis of the student's progress compared to linguistically similar students is warranted." (p. 40)*
>
> - Fisher & Frey, 2012

For ELs the Problem in Testing is Test Score Validity

NO BIAS

- ***Items***
  (content, novelty)
- ***Structure***
  (sequence, order, difficulty)
- ***Reliability***
  (measurement error/accuracy)
- ***Factor structure***
  (theoretical structure, relationship of variables to each other)
- ***Predictive Validity***
  (correlation with academic success or achievement)
- ***Differential Item Functioning***
  (DIF is not often found)

POTENTIAL BIAS

- ***Construct Validity***
  (nature and specificity of the intended/measured constructs)

Even when the intended variable is measured, inferences and interpretation may not be valid if comparability in development is lacking…

- ***Interpretive Invalidity***
  (it can undermine the validity of evaluative judgments and meaning assigned to scores)

"As long as tests do not at least sample in equal degree a state of saturation [assimilation of fundamental experiences and activities] that is equal for the 'norm children' and the particular bilingual child it cannot be assumed that the test is a valid one for the child."
Sanchez, 1934

Test Score Validity and Defensible Interpretation Requires "True Peer" Comparison

Example of Potential Construct Invalidity:

"Assemble these blocks together in the correct manner so they appear identical to this illustration."

A test designed to measure visual processing (Gv) in ELs must avoid over-reliance on language ability (Gc) or else measurement of visual processing may be confounded with language ability.

Example of Potential Interpretive Invalidity:

"After putting a blue block on top of a purple one, put the green block on the blue one."

A test designed to measure English language ability (Gc) is valid for EL's ability ***in English***, but poor performance cannot be ascribed to a potential disability unless developmental differences in English have been controlled.

Fundamental Requirements for Evaluation of ELs

Questions and concerns: If typical methods are insufficient, then what can we do? Where are the legal requirements, professional standards, ethical guidelines, etc., that specify best practice? Is there a difference between a bilingual evaluation and an evaluation of an EL? Do we all need to be bilingual to evaluate ELs? What are the "best" tests to use? Etc.

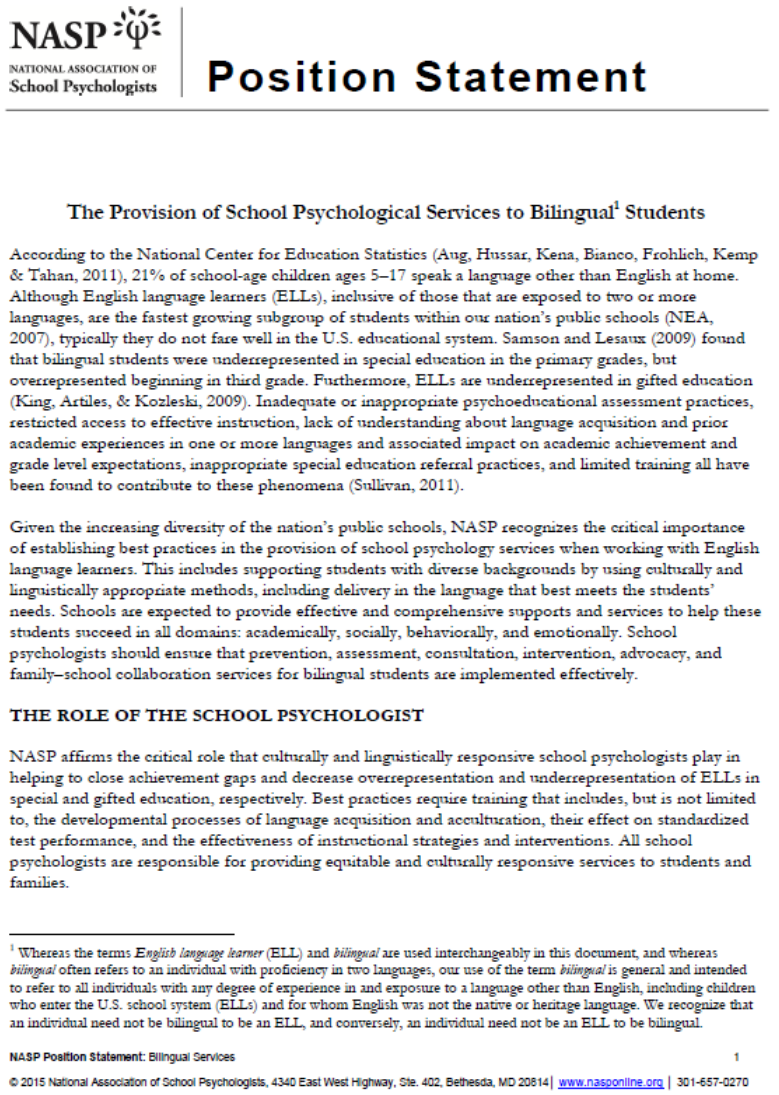
> "*…few national or state standards exist that define basic competencies as to what constitutes a "bilingual" psychologist. Mere possession of the capacity to communicate in an individual's native language does not ensure appropriate, non-discriminatory assessment of that individual. Traditional assessment practices and all their inherent biases can be quite easily replicated in any number of languages*" (p. 291).
>
> Flanagan, McGrew & Ortiz, 2000

Source: Flanagan, D.P., McGrew, K.S., & Ortiz, S.O. (2000). The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary interpretive approach. Boston, MA: Allyn & Bacon.

Fundamental Requirements for Evaluation of ELs

This document represents the very first official position by NASP on school psychology services to bilingual students was adopted in 2015.

It serves as official policy of NASP and is ***applicable to ALL school psychologists***, whether or not they are bilingual themselves.



NASP
NATIONAL ASSOCIATION OF School Psychologists

Position Statement

The Provision of School Psychological Services to Bilingual[1] Students

According to the National Center for Education Statistics (Aud, Hussar, Kena, Bianco, Frohlich, Kemp & Tahan, 2011), 21% of school-age children ages 5–17 speak a language other than English at home. Although English language learners (ELLs), inclusive of those that are exposed to two or more languages, are the fastest growing subgroup of students within our nation's public schools (NEA, 2007), typically they do not fare well in the U.S. educational system. Samson and Lesaux (2009) found that bilingual students were underrepresented in special education in the primary grades, but overrepresented beginning in third grade. Furthermore, ELLs are underrepresented in gifted education (King, Artiles, & Kozleski, 2009). Inadequate or inappropriate psychoeducational assessment practices, restricted access to effective instruction, lack of understanding about language acquisition and prior academic experiences in one or more languages and associated impact on academic achievement and grade level expectations, inappropriate special education referral practices, and limited training all have been found to contribute to these phenomena (Sullivan, 2011).

Given the increasing diversity of the nation's public schools, NASP recognizes the critical importance of establishing best practices in the provision of school psychology services when working with English language learners. This includes supporting students with diverse backgrounds by using culturally and linguistically appropriate methods, including delivery in the language that best meets the students' needs. Schools are expected to provide effective and comprehensive supports and services to help these students succeed in all domains: academically, socially, behaviorally, and emotionally. School psychologists should ensure that prevention, assessment, consultation, intervention, advocacy, and family–school collaboration services for bilingual students are implemented effectively.

THE ROLE OF THE SCHOOL PSYCHOLOGIST

NASP affirms the critical role that culturally and linguistically responsive school psychologists play in helping to close achievement gaps and decrease overrepresentation and underrepresentation of ELLs in special and gifted education, respectively. Best practices require training that includes, but is not limited to, the developmental processes of language acquisition and acculturation, their effect on standardized test performance, and the effectiveness of instructional strategies and interventions. All school psychologists are responsible for providing equitable and culturally responsive services to students and families.

[1] Whereas the terms *English language learner* (ELL) and *bilingual* are used interchangeably in this document, and whereas *bilingual* often refers to an individual with proficiency in two languages, our use of the term *bilingual* is general and intended to refer to all individuals with any degree of experience in and exposure to a language other than English, including children who enter the U.S. school system (ELLs) and for whom English was not the native or heritage language. We recognize that an individual need not be bilingual to be an ELL, and conversely, an individual need not be an ELL to be bilingual.

Fundamental Requirements for Evaluation of ELs

According to the NASP Position Statement:

"Given the dearth of bilingual school psychologists, particularly in languages other than Spanish, it is important to recognize that monolingual, English-speaking school psychologists will likely conduct the vast majority of evaluations with bilingual students. Therefore, ***proper training in the requisite knowledge and skills for culturally and linguistically responsive assessment is necessary for all school psychologists***." (p. 2; emphasis added).

NASP (2015). Position Statement: The Provision of School Psychological Services to Bilingual Students. Retrieved from http://www.nasponline.org/x32086.xml

Score Validity Requires Construct Validity Not "Caution"

According to the APA, evidence-based practice is defined as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences (p. 273)

In what manner is evidence-based assessment of ELs conducted and to what extent is there any research to support the use of any of the following methods as being capable of establishing test score validity?

- **Modified Methods of Evaluation**
  - *Working around the language by modifying/altering the assessment*
- **Nonverbal Methods of Evaluation**
  - *Avoiding the language by evaluating areas unrelated to language*
- **Dominant Language Evaluation**
  - *Choosing a language based simply on relative proficiency*

Whatever method or approach may be employed in evaluation of EL's, the fundamental obstacle to unbiased interpretation rests on the degree to which the examiner is able to defend claims of validity (construct and interpretive) that are being used to support conclusions. This idea is captured by and commonly referred to as a being a able to distinguish "difference vs. disorder."

Limited Methods for Addressing Test Score Validity

Example of Modified and Altered Assessment: The CCAST

The Checklist for Cross-cultural Administration of Standardized Tests (CCAST; Collier, 2016) is a tool for documenting cultural and linguistic modifications to standardized assessment instruments.

I. Analysis of Test Content

Score Range	Degree of Applicability
8-13	Not appropriate for use with this student.
14-20	Will need extensive modification to be valid.
21-27	Must modify most items and procedures.
28-34	Appropriate with specific modifications.
35-40	Appropriate for use with this student without modification.

Markedly Different

Moderately Different

Slightly Different

II. Modifications

Score Range	Degree of Applicability
8-13	Little to no modification was done on this test.
14-20	Few modifications were made.
21-27	Specific modifications were necessary.
28-34	Most items and procedures had to be modified.
35-40	Complete modifications were necessary.

III. Scoring and Interpretation

III. Scoring and Interpretation			
Assessment Item	Modifications	Standard Score	Modified Score

Limited Methods for Addressing Test Score Validity

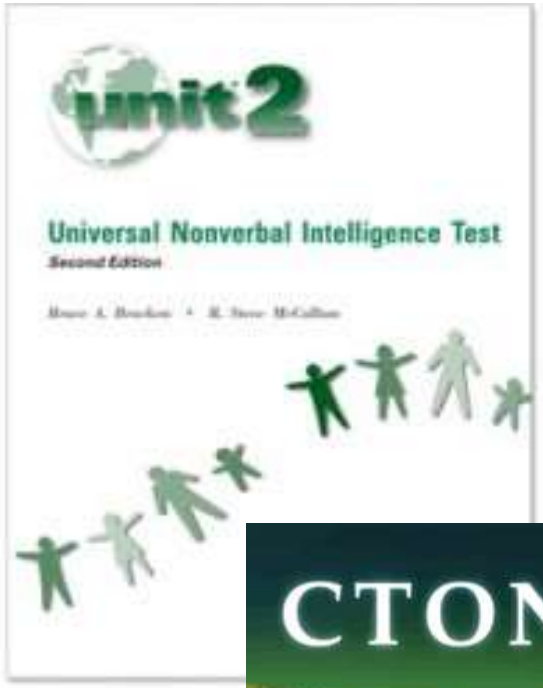
ISSUES IN MODIFIED METHODS OF EVALUATION

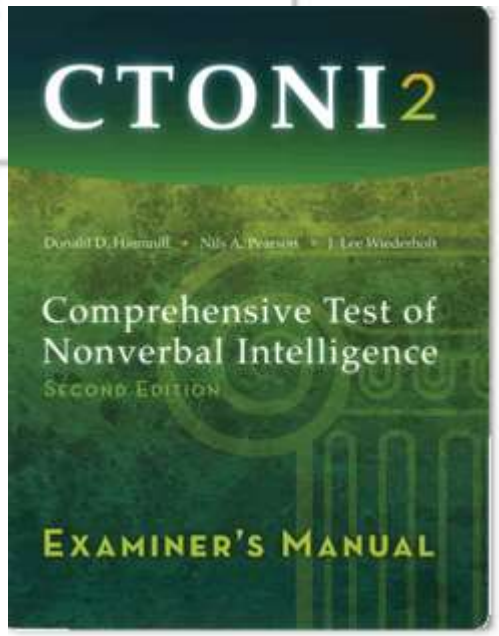
Modified and Altered Assessment: Just workaround the language.

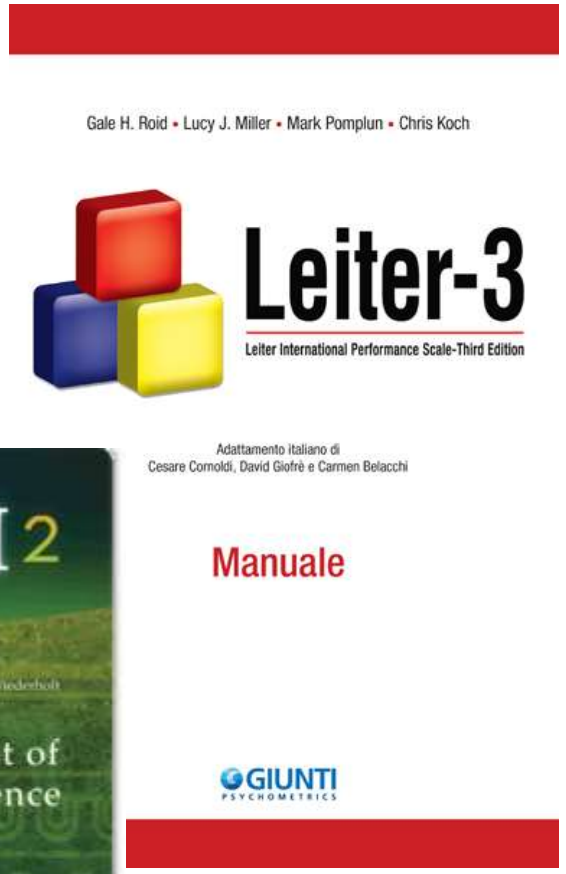
- use of a translator/interpreter for administration helps overcome the language barrier but is also a violation of standardization and undermines score validity, even when the interpreter is highly trained and experienced; tests are not usually normed in this manner
- in efforts to help the examinee perform to the best of his/her ability, any process involving "testing the limits" where there is alteration or modification of test items or content, mediation of task concepts prior to administration, repetition of instructions, acceptance of responses in either languages, or elimination/modification of time constraints, etc., violates standardization even when "permitted" by the test publisher except in cases where separate norms for such altered administration are provided
- any alteration of the testing process violates standardization and effectively invalidates the scores which precludes interpretation or the assignment of meaning by undermining the psychometric properties of the test
- alterations or modifications are perhaps most useful in deriving qualitative information—observing behavior, evaluating learning propensity, evaluating developmental capabilities, analyzing errors, etc.
- a recommended procedure would be to administer tests in a standardized manner first, which will potentially allow for later interpretation, and then consider any modifications or alterations that will further inform the referral questions
- because the violation of the standardized test protocol introduces error into the testing process, **it cannot be determined to what extent the procedures aided or hindered performance and thus the results cannot be defended as valid**

Limited Methods for Addressing Test Score Validity

Examples of Popular Nonverbal Tests







Limited Methods for Addressing Test Score Validity

ISSUES IN NONVERBAL METHODS OF EVALUATION

Language Reduced Assessment: Just avoid the language.

- "nonverbal testing:" use of language-reduced (or 'nonverbal') tests are helpful in overcoming the language obstacle, however:
- it is impossible to administer a test without some type of communication occurring between examinee and examiner, this is the purpose of gestures/pantomime
- some tests remain very culturally embedded—they do not become culture-free simply because language is not required for responding
- construct underrepresentation is common, especially on tests that measure fluid reasoning (Gf), and when viewed within the context of CHC theory, some batteries measure a narrower range of broad cognitive abilities/processes, particularly those related to verbal academic skills such as reading and writing (e.g., Ga and Gc) and mathematics (Gq)
- all nonverbal tests are subject to the same problems with norms and cultural content as verbal tests—that is, they do not control for differences in acculturation and language proficiency which may still affect performance, albeit less than with verbal tests
- language reduced tests are helpful in evaluation of diverse individuals and may provide better estimates of true functioning in certain areas, **but they are not a whole or completely satisfactory solution with respect to fairness and provide no mechanism for establishing whether the obtained test results are valid or not**

Limited Methods for Addressing Test Score Validity

Measuring Relative Language Proficiency

Limited Methods for Addressing Test Score Validity

ISSUES IN DOMINANT LANGUAGE EVALUATION

Determining the language of evaluation: Just choose a language.

- generally refers to the assessment of an EL after it has been determined that the examinee is more proficient ("dominant") in one language than the other
- being "dominant" in a language does not imply age-appropriate development in that language
- dominance does not inform instructional intervention, progress, growth, or expected test performance
- dominance is often affected by preferences that are shaped by social factors including identity development
- direct evaluation in the native language (L1) can only be conducted by a bilingual evaluator and is not an option available to monolingual English-speaking evaluators
- bilingual ability is no guarantee of nondiscriminatory assessment--native language assessment (L1) can be just as biased and inequitable as assessment in English (L2)
- in contrast to assessment in English, native language evaluation assessment is a relatively new idea without a substantive empirical base to guide or support standards of practice
- both L1 and L2 test norm samples fail to control for variability between and among ELs relative to their own amount of exposure to English and to that of monolingual, native English speaker
- **without a research base, there is no way to evaluate the validity of test results derived simply by testing in the dominant language** and any subsequent interpretations would be specious and amount to no more than a guess

An Evidence-based Approach Requires Evidence not "Caution"

Evaluation Issues and Methods	Norm sample representative of bilingual development	Measures a wider range of school-related abilities	Does not require the evaluator to be bilingual	Adheres to the test's standardized protocol	Substantial research base on bilingual performance	Sufficient to identify or diagnosis disability	Accounts for variation in bilingual development	Most likely to yield reliable and valid data and information	Provides extensive data regarding development
Modified or Altered Assessment	✗	✓	✓	✗	✗	✗	✗	✗	✗
Language Reduced Assessment	✗	✗	✓	✓	✗	✗	✗	✗	✗
Dominant Language Assessment in L1: native only	✗	✓	✗	✓	✗	✗	✗	✗	✗
Dominant Language Assessment in L2: English only	✗	✓	✓	✓	✓	✗	✗	✗	✗

Statements like, "scores should be interpreted with extreme caution" do nothing to establish validity of subsequent interpretation. Typical approaches for addressing test score validity are limited and have little, if any, research which indicates any of them effectively produce "valid" scores for English learners. Even nonverbal approaches fail to provide a satisfactory or comprehensive solution to the examination of exclusionary variables and test score validity.

Summary of Research on the Test Performance of English Language Learners

Research conducted over the past 100 years on ELs who are non-disabled, of average ability, possess moderate to high proficiency in English, and tested in English, has resulted in a basic and ubiquitous finding:

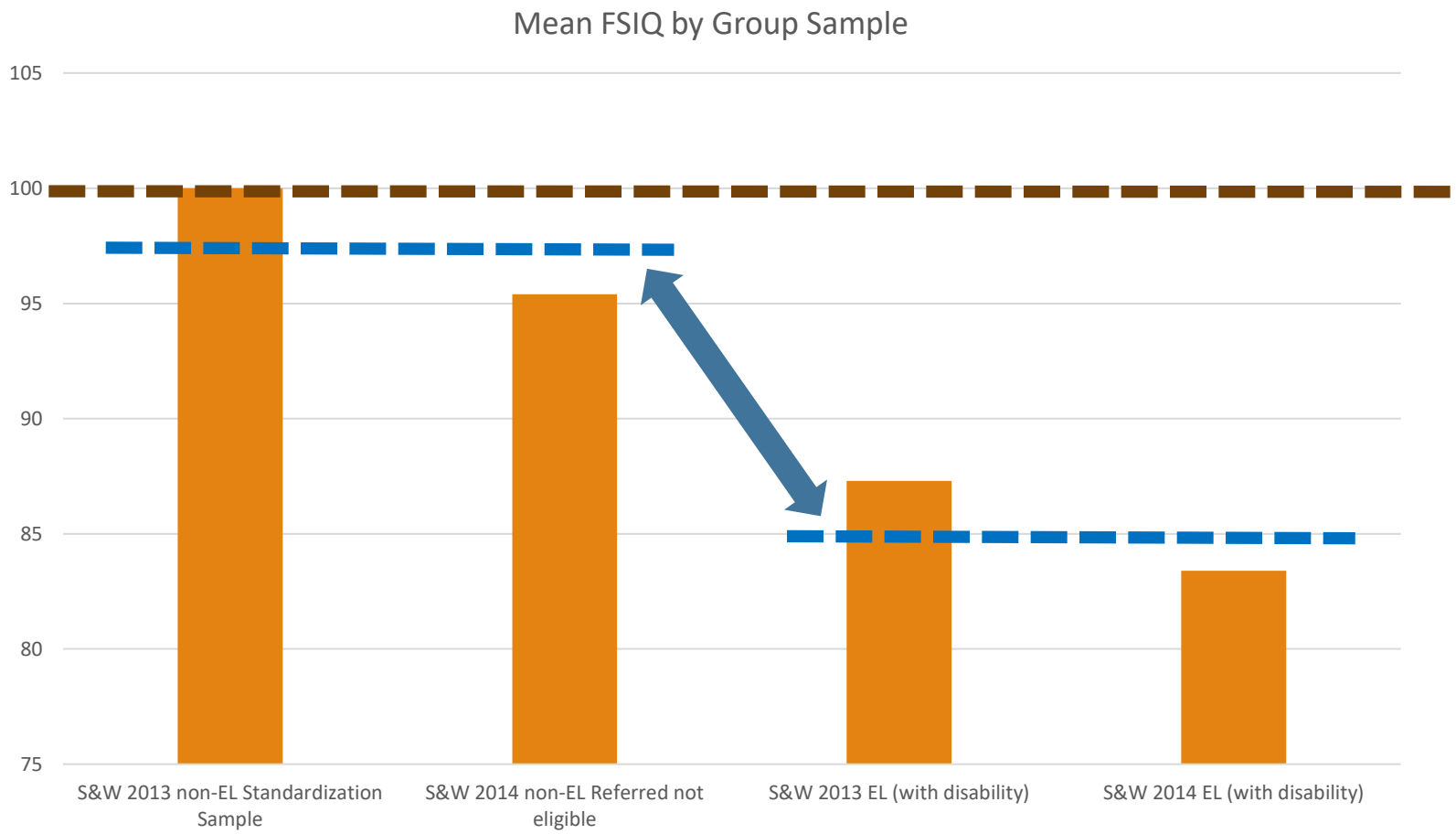
English Learners and Native English speakers tend to perform differently on standardized, norm-referenced tests of intelligence and general cognitive ability.

So what explains these findings? Early explanations relied on genetic differences attributed to racial inferiority. But even early researchers noticed that language differences (i.e., lack of proficiency) likely played a role in this difference, particularly because ELs also tended to perform better on nonverbal tests than on verbal tests (Ortiz, 2019).

Ortiz, S. O. (2019). On the Measurement of Cognitive Abilities in English Learners. Contemporary School Psychology, Vol. 23(1) 68-86. https://doi.org/10.1007/s40688-018-0208-8

Research Foundations for EL Evaluation

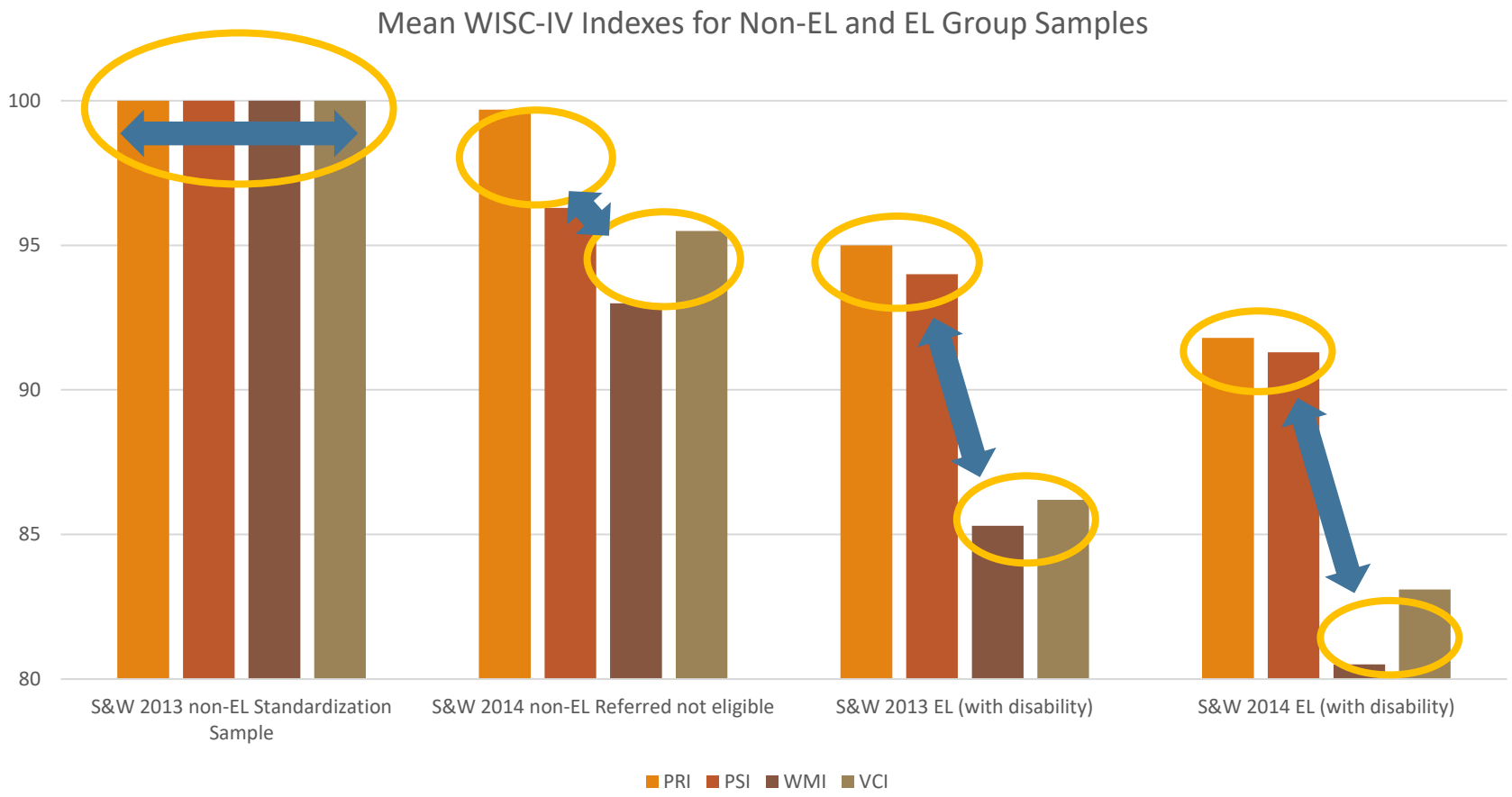
ELs and non-EL's perform differently: Broad ability level



Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.

Research Foundations for EL Evaluation

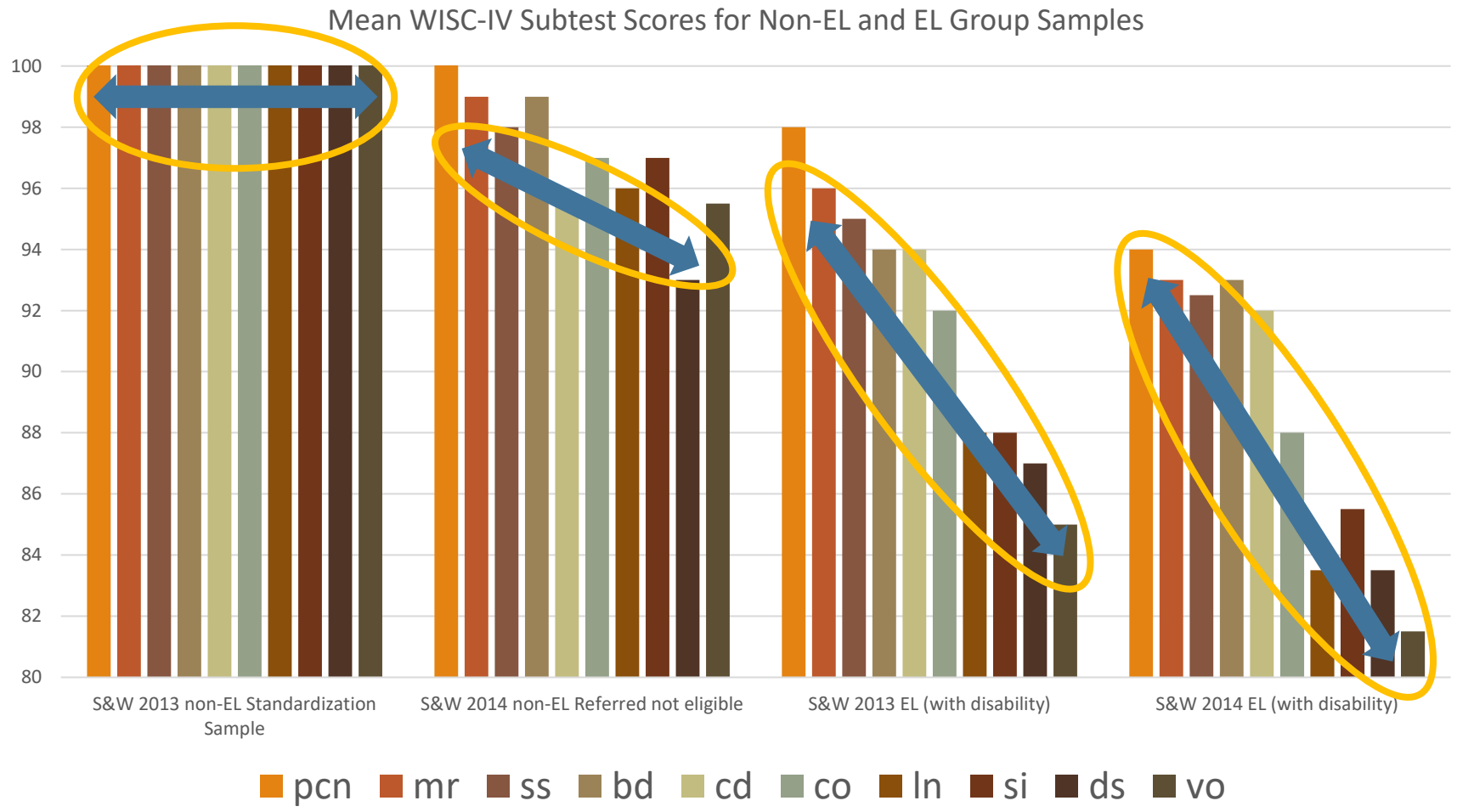
ELs and non-EL's perform differently: Index level



Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.

Research Foundations for EL Evaluation

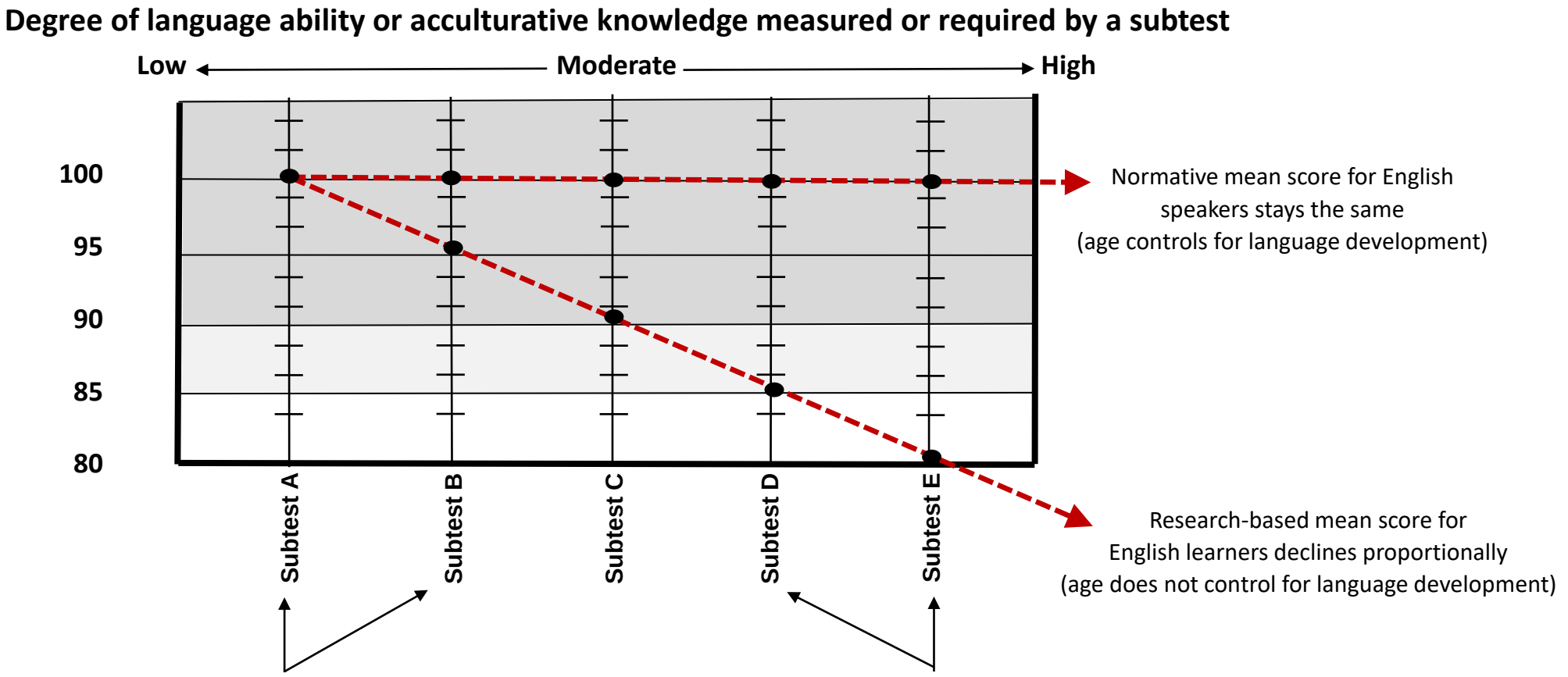
ELs and non-EL's perform differently: Subtest level



Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.

Research Foundations for EL Evaluation: EL to ES

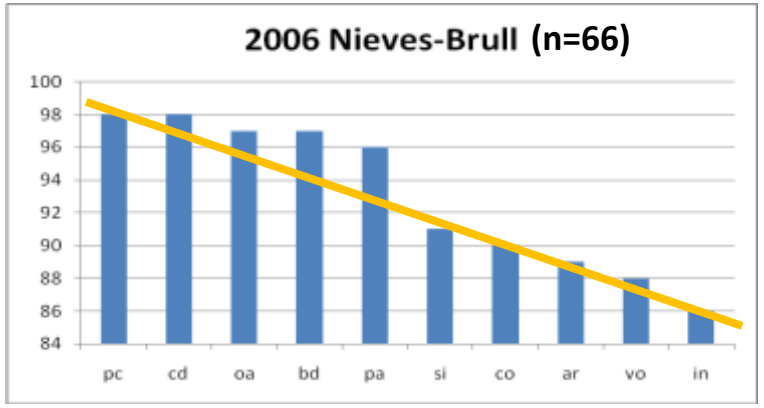
Language influences EL test performance in a linear, continuous manner, not dichotomously

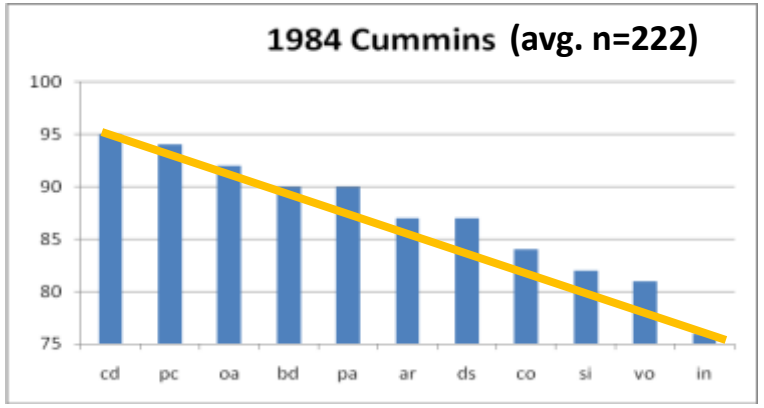


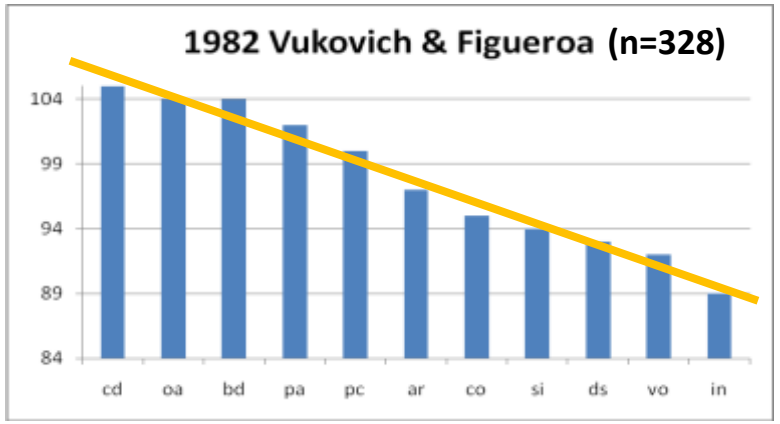
The more a test requires age-based developmental language proficiency and acculturative knowledge, the more the effect on test performance.

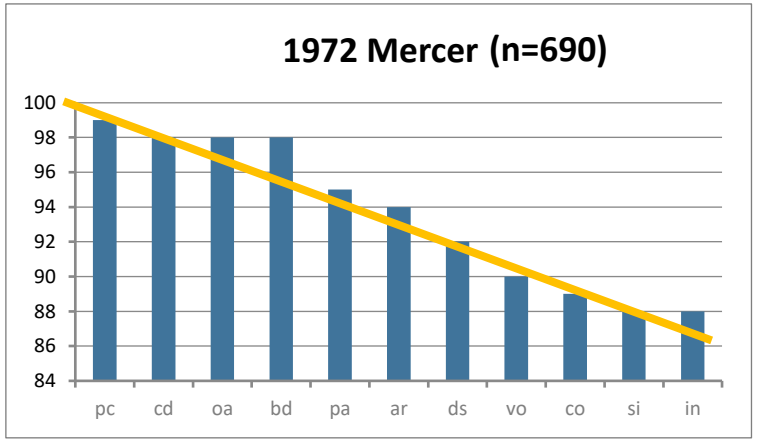
Research Foundations for EL Evaluation: EL to ES

Comparison of overall “average” test performance at the subtest level: EL to ES

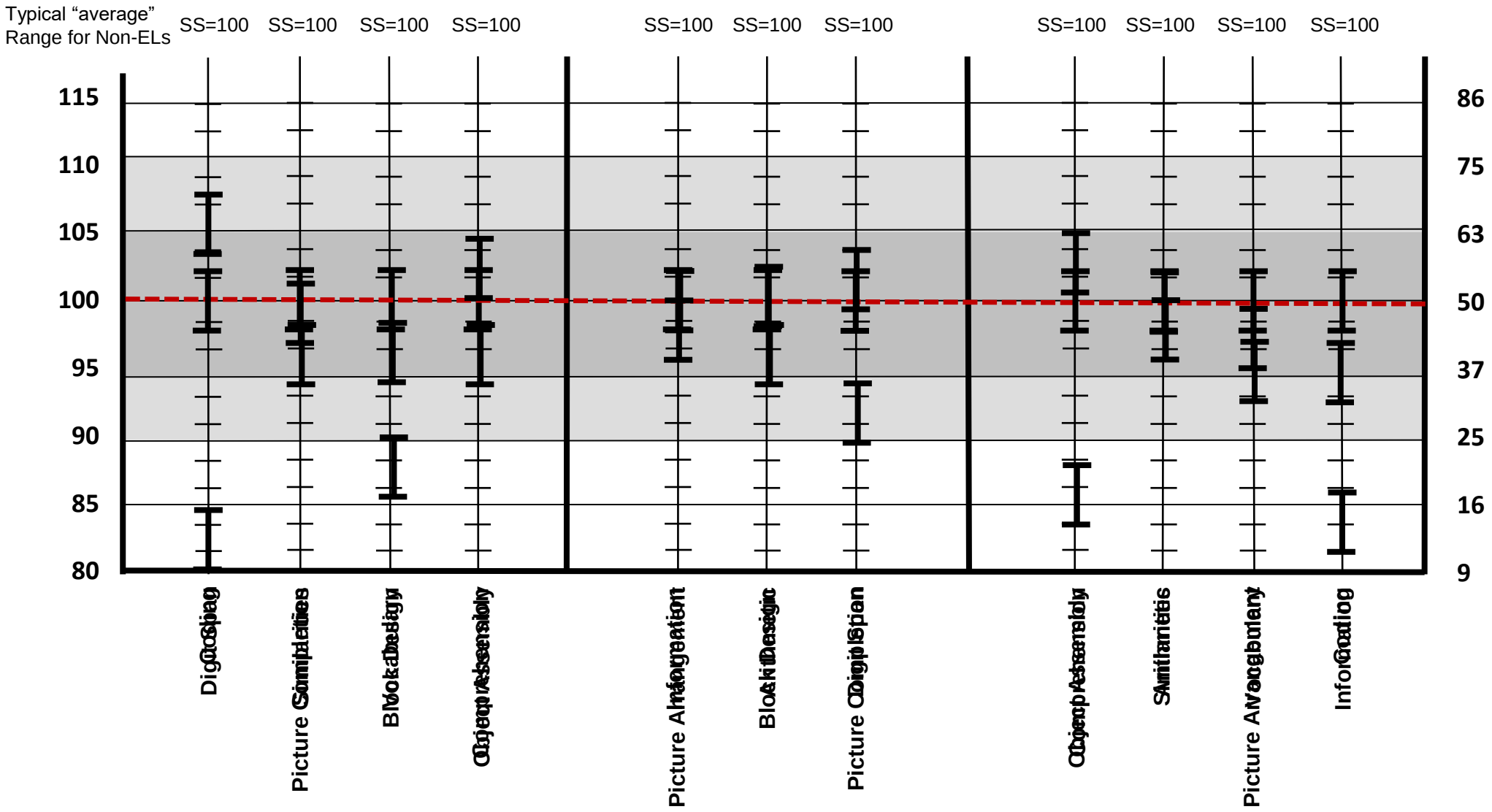








Research Foundations for EL Evaluation: ES vs. ES



Determining "average" performance for English Speakers

Research Foundations for EL Evaluation: EL to ES

Research-based mean performance of ELs on the WISC Subtests

Tests with **"high"** language demands

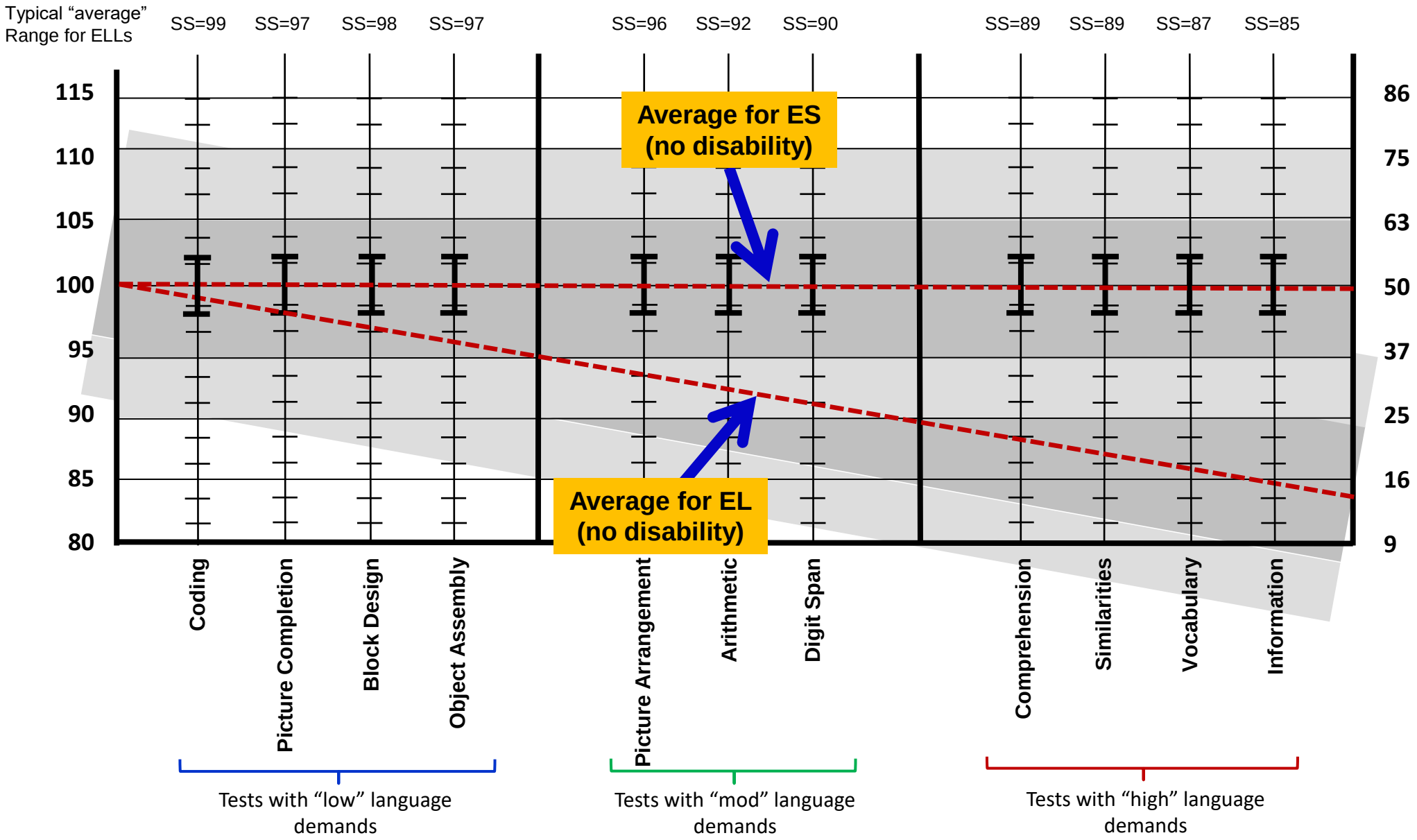
Tests with **"mod"** language demands

Tests with **"low"** language demands

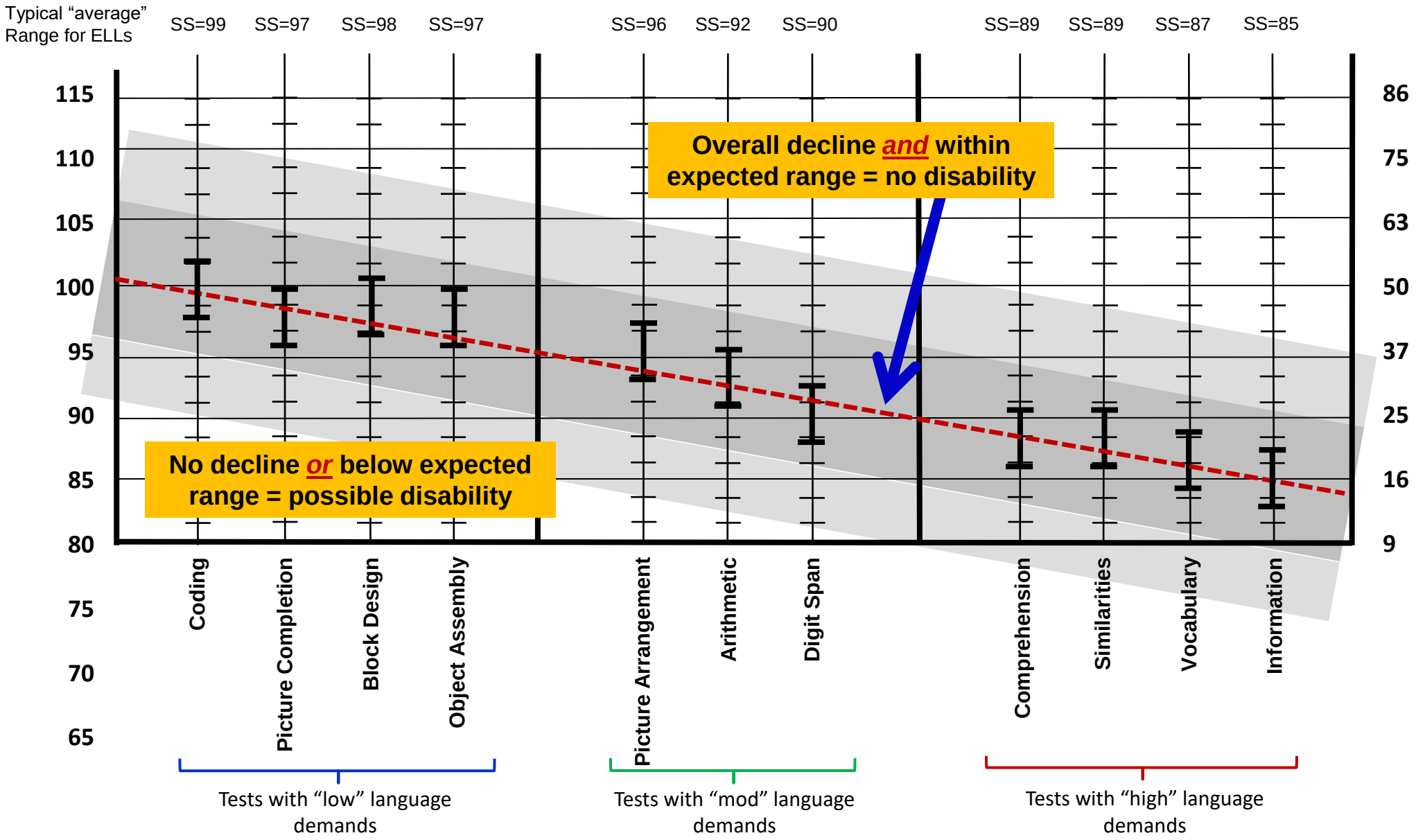
	Mercer 1972	Vukovich & Figueroa, 1982	Cummins 1982	Nieves-Brull 2006	
Subtest Name	**Scale Score**	**Scale Score**	**Scale Score**	**Scale Score**	**Grand Mean**
Information	7.5	7.8	5.1	7.2	85
Vocabulary	8.0	8.3	6.1	7.5	87
Similarities	7.6	8.8	6.4	8.2	89
Comprehension	7.8	9.0	6.7	8.0	89
Digit Span	8.3	8.5	7.3	*	90
Arithmetic	8.7	9.4	7.4	7.8	92
Picture Arrangement	9.0	10.3	8.0	9.2	96
Block Design	9.5	10.8	8.0	9.4	97
Object Assembly	9.6	10.7	8.4	9.3	98
Picture Completion	9.7	9.9	8.7	9.5	97
Coding	9.6	10.9	8.9	9.6	99

**Data for this subtest were not reported in the study.*

Fairness in Determining "Average" Performance: EL to ES



Fairness in Determining "Average" Performance: EL to ES



Foundations of the Culture-Language Interpretive Matrix

Research Foundations for EL Evaluation: EL to ES

The influence of language on subtest level performance in English speakers and English learners.

Highest Language Demands

Lowest Language Demands

Table 3. Variance Explained by Exogenous Variables (Individual Test Performance) by Age Group.

Individual test	Variance explained 7-10	11-14	15-18	
Verbal Comprehension	.79[c]	.86[c]	.81[c]	C-LIM Level 5
General Information	.71[c]	.85[c]	.86[c]	
Concept Formation	.67[c]	.71[c]	.67[c]	
Visual–Auditory Learning	.40[b]	.37[b]	.41[b]	C-LIM Level 4
Delayed Recall Visual–Auditory Learning	.39[b]	.32[b]	.37[b]	
Analysis Synthesis	.29[b]	.44[b]	.47[b]	
Sound Blending	.25[b]	.32[b]	.35[b]	
Auditory Working Memory	.22[b]	.44[b]	.32[b]	
Retrieval Fluency	.22[b]	.22[b]	.28[b]	
Memory for Words	.18[b]	.32[b]	.23[b]	C-LIM Level 3
Numbers Reversed	.17[b]	.26[b]	.30[b]	
Pair Cancelation	.17[b]	.11[b]	.11[b]	
Rapid Picture Naming	.16[b]	.07[a]	.16[b]	
Incomplete Words	.13[b]	.31[b]	.23[b]	
Visual Matching	.13[b]	.15[b]	.16[b]	C-LIM Level 2
Decision Speed	.12[b]	.15[b]	.19[b]	
Auditory Attention	.10[b]	.20[b]	.15[b]	
Spatial Relations	.08[a]	.16[b]	.16[b]	
Planning	.07[a]	.12[b]	.11[b]	C-LIM Level 1
Picture Recall	.02[a]	.06[a]	.10[b]	

**Source: Cormier, D.C., McGrew, K.S. & Ysseldyke, J. E. (2014). The Influences of Linguistic Demand and Cultural Loading on Cognitive Test Scores. Journal of Psychoeducational Assessment, 32(7), 610-623.*

The Culture-Language Interpretive Matrix (C-LIM)

A Systematic Framework for Organizing and Guiding Evidence-Based Practice

Translation of Research into Practice

1. Research on test performance of ELs establishes the foundations upon which the C-LIM is based and its only purpose is to assist in determining the extent to which obtained results are **likely valid (a minimal or only contributory influence of cultural and linguistic factors),** possibly valid (minimal or contributory influence of cultural and linguistic factors but which requires additional evidence from native language evaluation), or **likely invalid (a primary influence of cultural and linguistic factors)**.

2. Organization of the C-LIM as a matrix and graph, are simply visual organizers of this research and serve as a de facto "EL norm sample" for the purposes of examining ***cognitive, linguistic, and neuropsychological*** test results relative to exclusionary factors (i.e., cultural and linguistic differences). Achievement tests require a vastly different evidentiary base.

3. Because it relies on empirical studies that used standardized, English-language administration, norms, and scoring with non-disabled EL populations, the C-LIM can only be used if tests are also administered in English and without any form of modification to the administration or scoring protocols.

4. Although some native-language tests (e.g., WISC Spanish, Bateria) are included in the C-LIM, examination of those results should be accomplished independent of results from tests administered in English. Moreover, there is some, but likely insufficient research to promote the use of the C-LIM as being valid for ELs who are given native-language tests and such use should be viewed as exploratory and informational only.

Free version of C-LIM and other materials available at: http://facpub.stjohns.edu/~ortizs/CLIM/

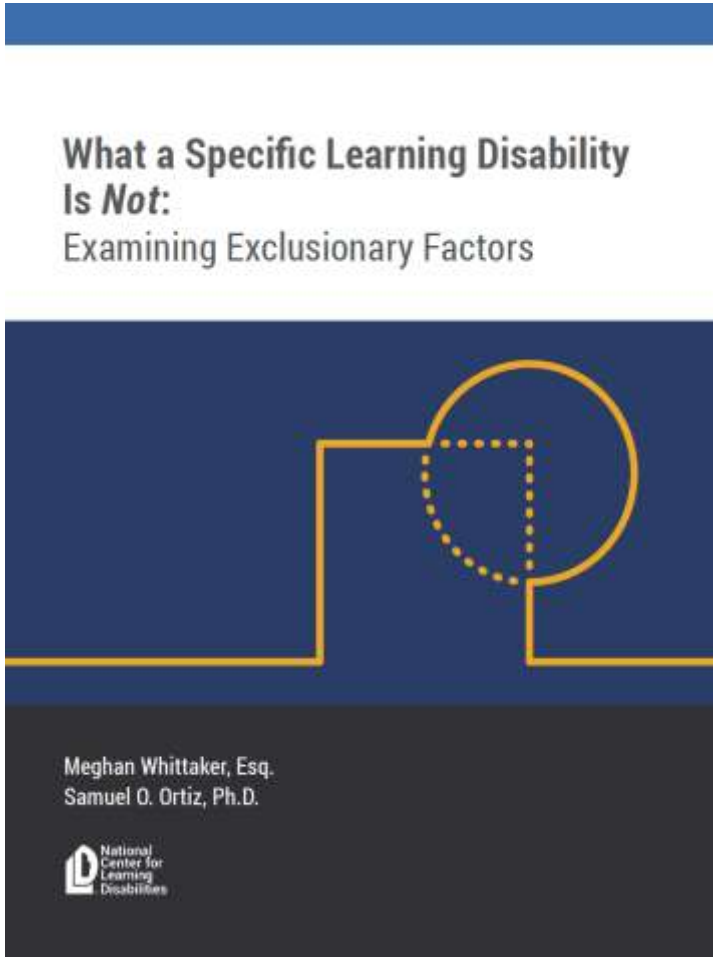
The Culture-Language Interpretive Matrix (C-LIM)

A Systematic Framework for Organizing and Guiding Evidence-Based Practice

Translation of research into practice

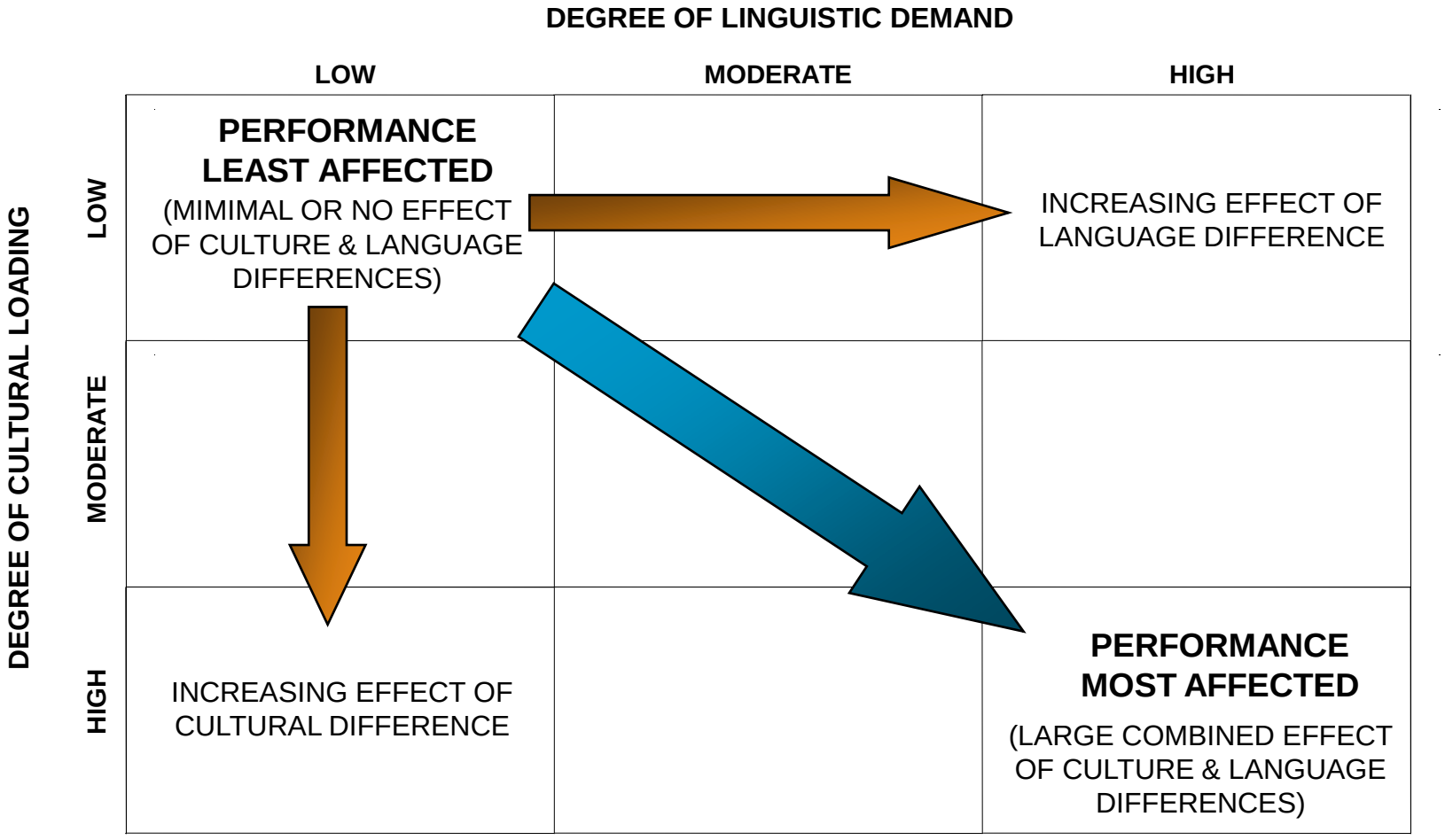
"To properly meet the definition and its exclusionary criteria, LEAs would first have to identify the primary cause(s) of a student's low achievement. For instance, if a child has limited English language proficiency, and it influences behavior and learning, it could appear as though the child has SLD. During an evaluation, it would be incumbent upon the school to determine whether the behavior or learning issues are ***primarily*** caused by one or more of these exclusionary factors. In the example above, the process of ruling out exclusionary factors would likely result in the determination of the child needing linguistic interventions and/or instructional support based on their limited English proficiency. Thus, the appropriateness of considering SLD will have been "ruled out" for this child and disability identification would not be appropriate.

Importantly, however, SLD can coexist with other disabilities, including limited English proficiency, sensory impairments, motor difficulties, emotional problems, etc. Any such factors may well be seen as ***contributory*** to the observed learning problems in the classroom and do not rule out a learning disability as long as they are not the ***primary*** reason for such difficulties." (p. 6)



Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and Culture-Language Interpretive Matrix

Matrix arrangement of expected subtest level performance for ELs vs. ES



Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and Culture-Language Interpretive Matrix

Research-based mean performance of ELs on the WISC Subtests

Tests with "high" language demands

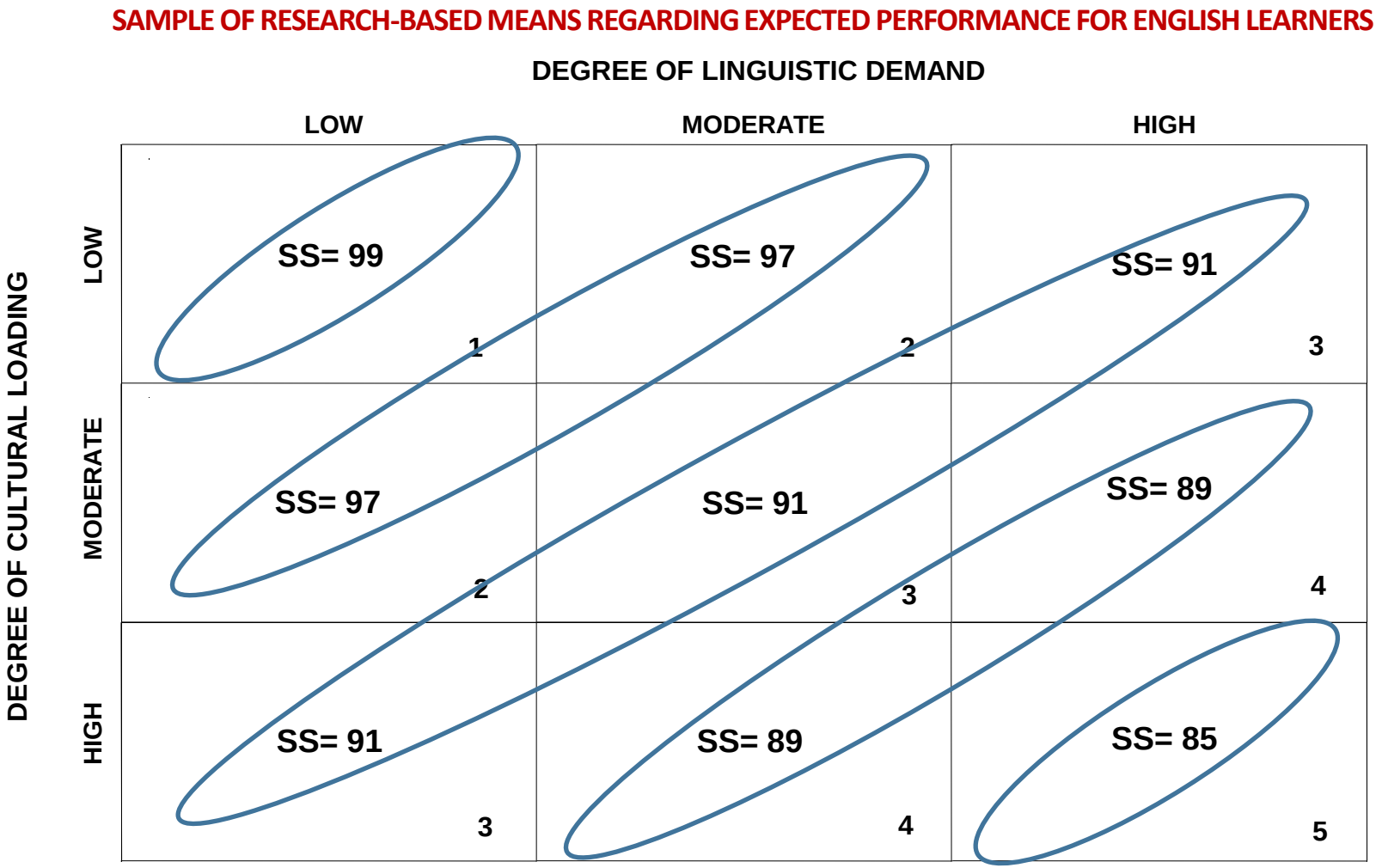
Tests with "mod" language demands

Tests with "low" language demands

	Mercer 1972	Vukovich & Figueroa, 1982	Cummins 1982	Nieves-Brull 2006		
Subtest Name	**Scale Score**	**Scale Score**	**Scale Score**	**Scale Score**	**Grand Mean**	**C-LIM Level**
Information	7.5	7.8	5.1	7.2	85 →	5
Vocabulary	8.0	8.3	6.1	7.5	87 →	5
Similarities	7.6	8.8	6.4	8.2	89 →	4
Comprehension	7.8	9.0	6.7	8.0	89 →	4
Digit Span	8.3	8.5	7.3	*	90 →	3
Arithmetic	8.7	9.4	7.4	7.8	92 →	3
Picture Arrangement	9.0	10.3	8.0	9.2	96 →	3
Block Design	9.5	10.8	8.0	9.4	97 →	2
Object Assembly	9.6	10.7	8.4	9.3	98 →	2
Picture Completion	9.7	9.9	8.7	9.5	97 →	1
Coding	9.6	10.9	8.9	9.6	99 →	1

**Data for this subtest were not reported in the study.*

Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and Culture-Language Interpretive Matrix



Because research is conducted with highly proficient ELs, these values represent performance only for "slightly different" individuals. Those with less English proficiency will score proportionally lower.

Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and Culture-Language Interpretive Matrix

Matrix of WISC subtest means arranged by EL vs. ES test performance

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING	LOW	MODERATE	HIGH
LOW	Coding Object Assembly Level 1 SS= 99	Block Design Level 2 SS= 97	Digit Span Level 3 SS= 91
MODERATE	Picture Completion Level 2 SS= 97	Arithmetic Level 3 SS= 91	Comprehension Level 4 SS= 89
HIGH	Picture Arrangement Level 3 SS= 91	Level 4 SS= 89	Information Similarities Vocabulary Level 5 SS= 85

The Culture-Language Interpretive Matrix – Basic Version 4.0

Clear ALL Data in Matrix

Culture-Language Interpretive Matrix - Basic Version 4.0

Clear Unused Subtests

WISC-V | WAIS-IV | WJ IV COG | DAS-II | CAS2 | RIAS-2 | LEITER-3 | CELF-5 | TAPS-3 | WRAML2 | NEPSY-II | Bateria III*
WPPSI-IV | KABC-II | WJ IV OL | SB5 | WNV | UNIT-2 | CELF-Pre2 | CASL-2 | CTOPP-2 | WMS-IV | D-KEFS | WISC Spanish*

Culture-Language Interpretive Matrix (Basic v4.0) - Analyzer and Data Entry

Print C-LIM Matrix

Name: _____ Age: _____ Grade: _____ Date: _____

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING

	LOW	Score	MODERATE	Score	HIGH	Score
LOW	WISC Coding	99	WISC Block Design	97	WISC Digit Span	91
	WISC Object Assembly	99				
	Cell Average =	99	Cell Average =	97	Cell Average =	91
MODERATE	WISC Picture Completion	97	WISC Arithmetic	91	WISC Comprehension	89
	Cell Average =	97	Cell Average =	91	Cell Average =	89
HIGH	WISC Picture Arrangement	91			WISC Information	85
					WISC Similarities	85
					WISC Vocabulary	85
	Cell Average =	91	Cell Average =	89	Cell Average =	91

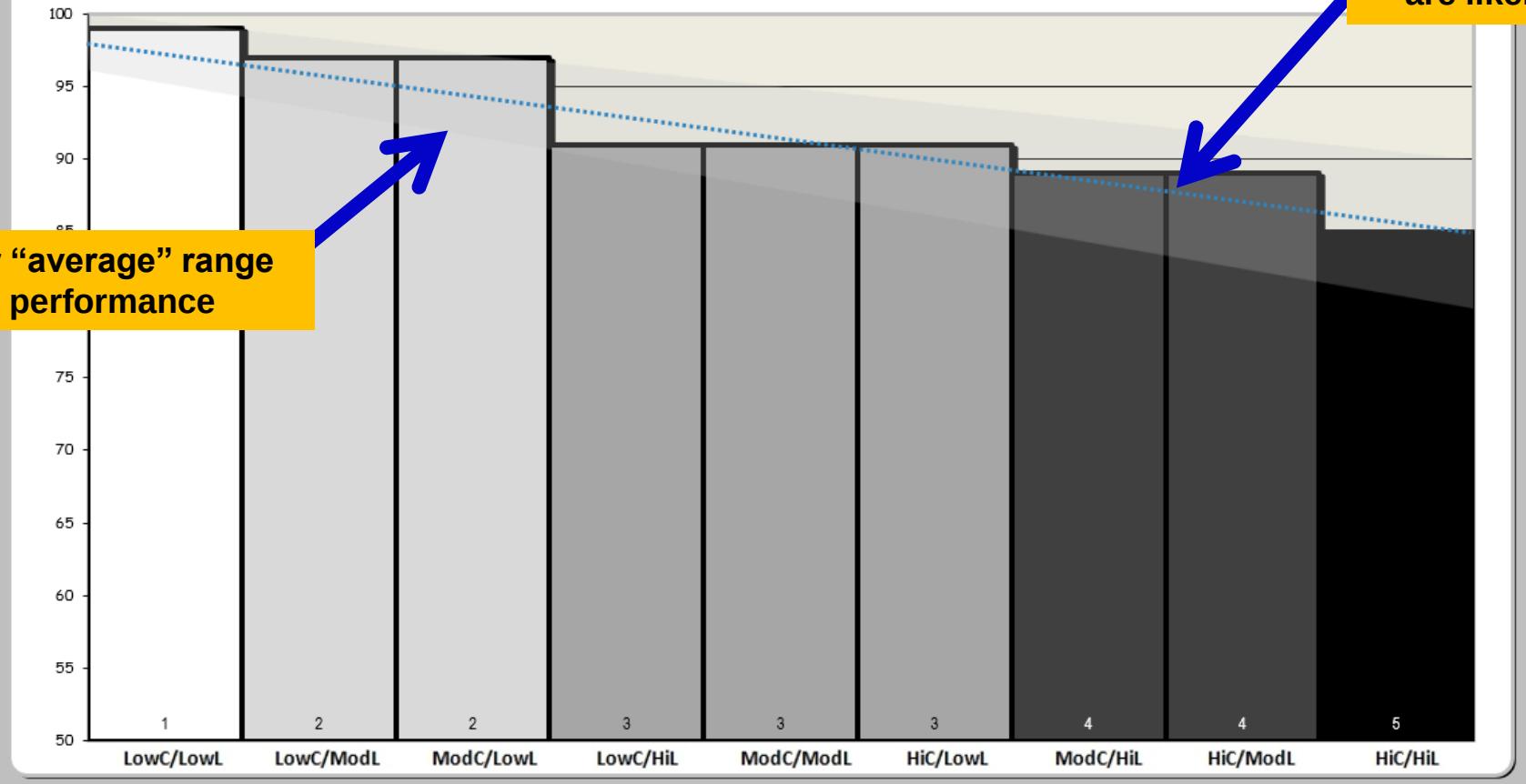
The Culture-Language Interpretive Matrix – Basic Version 4.0

Culture-Language Interpretive Matrix - Basic Version 4.0

Name: _____ Age: _____ Grade: _____ Date: ___

C-LIM Summary Graph: Evaluation of Cultural/Linguistic Influences

DEGREE OF DIFFERENCE FOR EVALUATION: (•) Slightly Different ( ) Moderately Different ( ) Markedly Different



Typical or "average" range for EL performance

Overall decline *and* within expected range *and* no significant variability: Indicates impact of culture and language is minimal or contributory and results are likely to be valid.

Fairness in Determining "Average" Performance: EL to EL

Although it has long been recognized that ***language*** likely account for the differences in test performance between English leaners and native English speakers, its influence has rarely been examined directly as a confounding variable and there has been a tendency instead to use "cultural" and "racial/ethnic" variables as proxies for language.

EL vs. ES: In general, research with ELs indicates that language (including acquisition of acculturative knowledge) has a powerful and significant effect on test performance that can be discerned at every level of testing, broad ability, index/composite, or subtest.

EL vs. EL: In addition, differences in exposure to and development in English varies among ELs such that the influence increases proportionally on tests that use, measure, and rely more on language and language-based abilities.

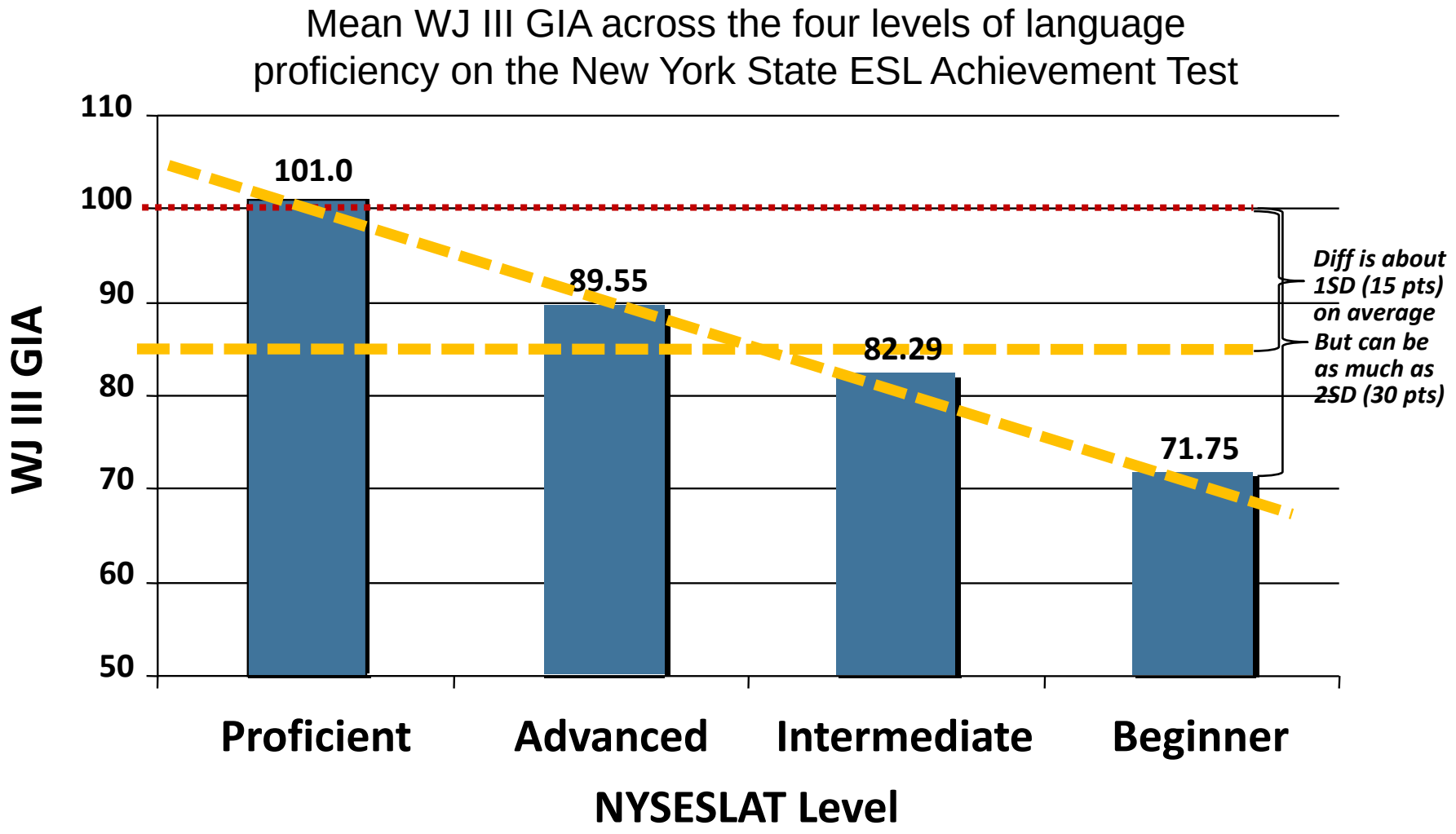
When understood as such, the impact of language on test performance of ELs is not seen to be a simple "verbal vs. nonverbal" dichotomy but rather ***a continuum formed by a linear and proportional attenuation of performance relative to both ESs and other ELs.***

Evaluation MUST account for EL vs. ES differences

Evaluation MUST also account for EL vs. EL differences

Research Foundations for EL Evaluation: EL to EL

General ability level performance as compared to other English learners

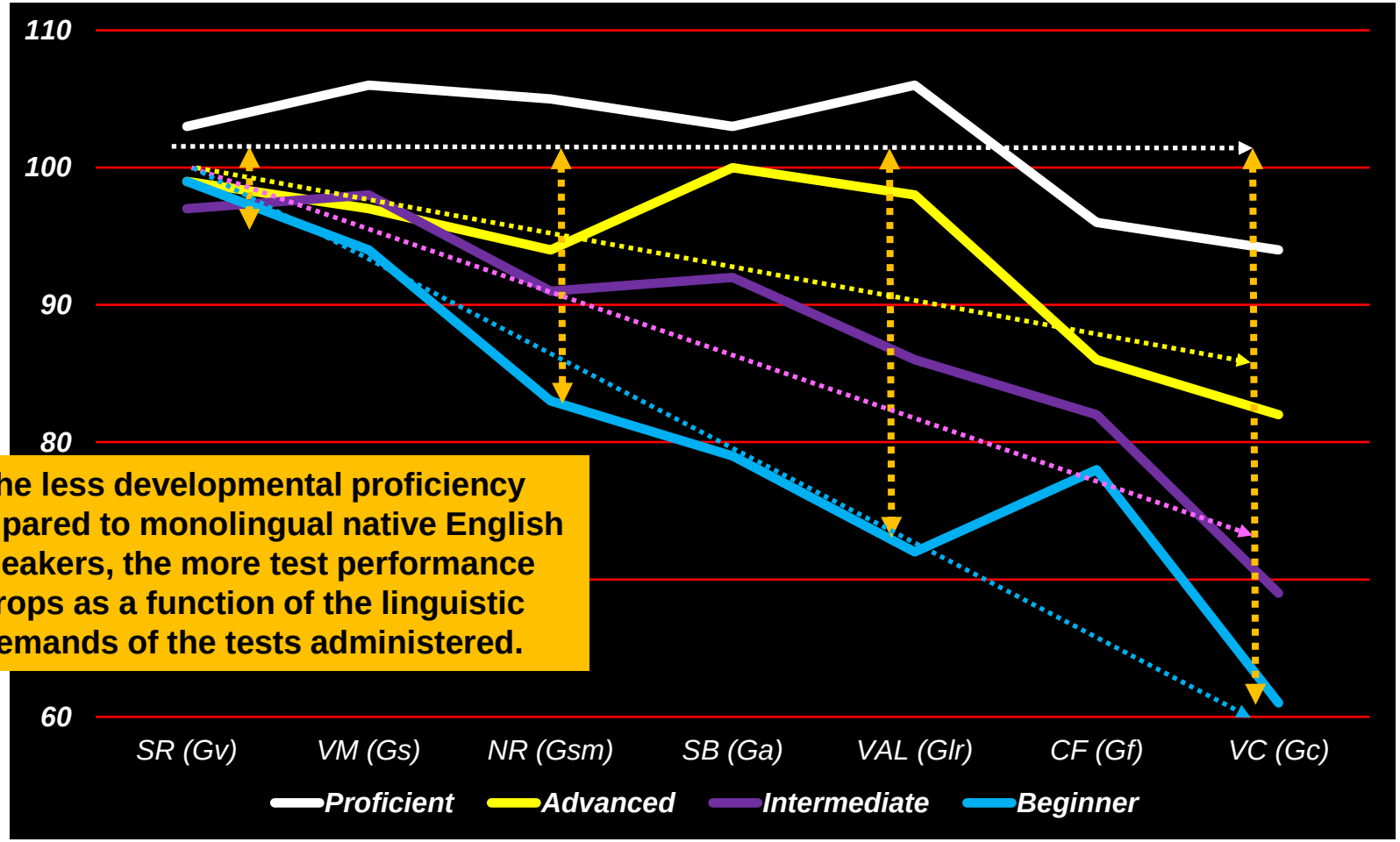


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013). English Language Proficiency and Test Performance: Evaluation of bilinguals with the Woodcock-Johnson III Tests of Cognitive Ability. Psychology in the Schools, Vol 50(8), pp. 781-797.

Research Foundations for EL Evaluation: EL to EL

Subtest level performance as compared to other English Learners

Domain specific scores across the seven WJ III subtests according to language proficiency level on the NYSESLAT



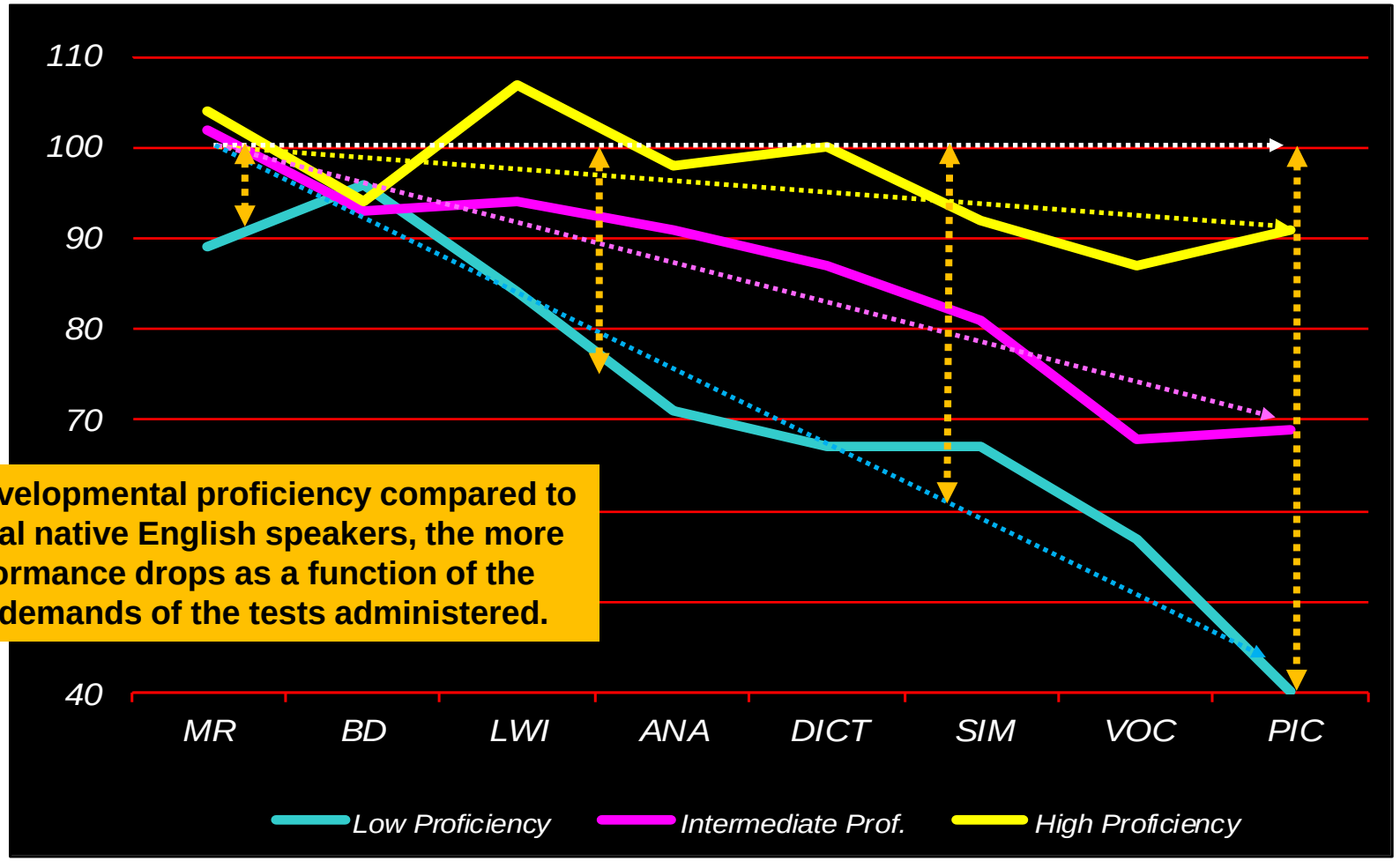
The less developmental proficiency compared to monolingual native English speakers, the more test performance drops as a function of the linguistic demands of the tests administered.

Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013). English Language Proficiency and Test Performance: Evaluation of bilinguals with the Woodcock-Johnson III Tests of Cognitive Ability. Psychology in the Schools, Vol 50(8), pp. 781-797.

Research Foundations for EL Evaluation: EL to EL

Subtest level performance as compared to other English Learners

Mean subtest scores across the four WASI subtests and four WMLS-R subtests according to language proficiency level



The less developmental proficiency compared to monolingual native English speakers, the more test performance drops as a function of the linguistic demands of the tests administered.

Source: Dynda, A. M. (2008). The relation between language proficiency and IQ test performance. Unpublished manuscript. St. John's University, NY.

Research Foundations for EL Evaluation: EL to EL

Research-based subtest means regarding expected test performance EL vs. EL

Degree of Linguistic Demand

Degree of Cultural Loading	Low	Moderate	High
Low	Slightly Different: 3-5 points Moderately Different: 5-7 points Markedly Different: 7-10 points	Slightly Different: 5-7 points Moderately Different: 7-10 points Markedly Different: 10-15 points	Slightly Different: 7-10 points Moderately Different: 10-15 points Markedly Different: 15-20 points
Moderate	Slightly Different: 5-7 points Moderately Different: 7-10 points Markedly Different: 10-15 points	Slightly Different: 7-10 points Moderately Different: 10-15 points Markedly Different: 15-20 points	Slightly Different: 10-15 points Moderately Different: 15-20 points Markedly Different: 20-25 points
High	Slightly Different: 7-10 points Moderately Different: 10-15 points Markedly Different: 15-20 points	Slightly Different: 10-15 points Moderately Different: 15-20 points Markedly Different: 20-25 points	Slightly Different: 15-20 points Moderately Different: 20-25 points Markedly Different: 25-35 points

***Slightly Different**: Includes individuals with very high levels of English language proficiency (e.g., CALP) and high acculturation, but still not entirely comparable to mainstream U.S. English speakers. Examples include individuals who are third generation in the U.S., have well educated/higher SES parents, have attended dual-language program for at least 6-7 years, or demonstrate native or near native-like proficiency in English language conversation and solid literacy skills. (Not a common category)*

***Moderately Different**: Includes individuals with moderate to higher levels of English language proficiency (e.g., advanced BICS/emerging CALP) and typical EL acculturative learning experiences. Examples include individuals who were born or came early to the U.S. with limited English speaking parents, usually from low to very low SES with parent's having low or limited literacy even in their own language, generally received formal education in English only or primarily in English since starting school.*

***Markedly Different**: Includes individuals with low to very low levels of English language proficiency (e.g., early BICS) or very limited acculturative learning experiences due to unusual influences on development. Examples include extremely low and limited parental SES and education, recently arrival in the U.S. or residence for in the U.S. 3 years or less, lack of prior formal education, exposure to trauma, violence, abuse, neglect, time spent in refugee or resettlement camps, changes in or multiple early languages.*

The Culture-Language Interpretive Matrix – Basic Version 4.0

Culture-Language Interpretive Matrix - Basic Version 4.0

Conceptualization by D. P. Flanagan, S. O. Ortiz, & V. C. Alfonso; Programming by S. O. Ortiz and A. M. Dynda.

Print C-LIM Graph

Name: _____ Age: _____ Grade: _____ Date: _____

C-LIM Summary Graph: Evaluation of Cultural/Linguistic Influences

DEGREE OF DIFFERENCE FOR EVALUATION: ○ Slightly Different ◉ Moderately Different ○ Markedly Different

"Slightly Different" The blue line represents the expected rate of decline for individuals considered to be slightly different

"Moderately Different" The green line represents the expected rate of decline for individuals considered to be moderately different (most typical)

"Markedly Different" The red line represents the expected rate of decline for individuals considered to be markedly different

Cross-Battery XBA Assessment

**NOTE: Tests marked with an asterisk (*) in the menu bar of the C-LIM Analyzer provide access to matrices that may be helpful in evaluating the validity of test scores when using Spanish-language tests (i.e., Bateria-III, WISC-IV Spanish). Use of these matrices should be considered EXPERIMENTAL ONLY as there is insufficient research at this time to support an evidence-based pattern of expected performance for ELs or firmly establish classification of subtests from such batteries. Use for qualitative analysis only.*

The Culture-Language Interpretive Matrix (C-LIM)

Guidelines for Addressing Exclusionary Factors via Examination of Test Score Validity

There are three basic criteria that, when all are met, provide evidence to suggest that test performance reflects the primary influence of cultural and linguistic factors and not actual ability, or lack thereof. These criteria are:

1. Overall Pattern of Decline: *There exists a general, overall pattern of decline in the scores from left to right and diagonally across the matrix where performance is highest on the less linguistically demanding/culturally loaded tests (low/low cells) and performance is lowest on the more linguistically demanding/culturally loaded tests (high/high cells), and;*

2. Within Expected Range: *The magnitude of the aggregate test scores across the matrix for all cells fall within or above the expected range of difference (shaded area around the line) determined to be most representative of the examinee's background and development relative to the sample on whom the test was normed.*

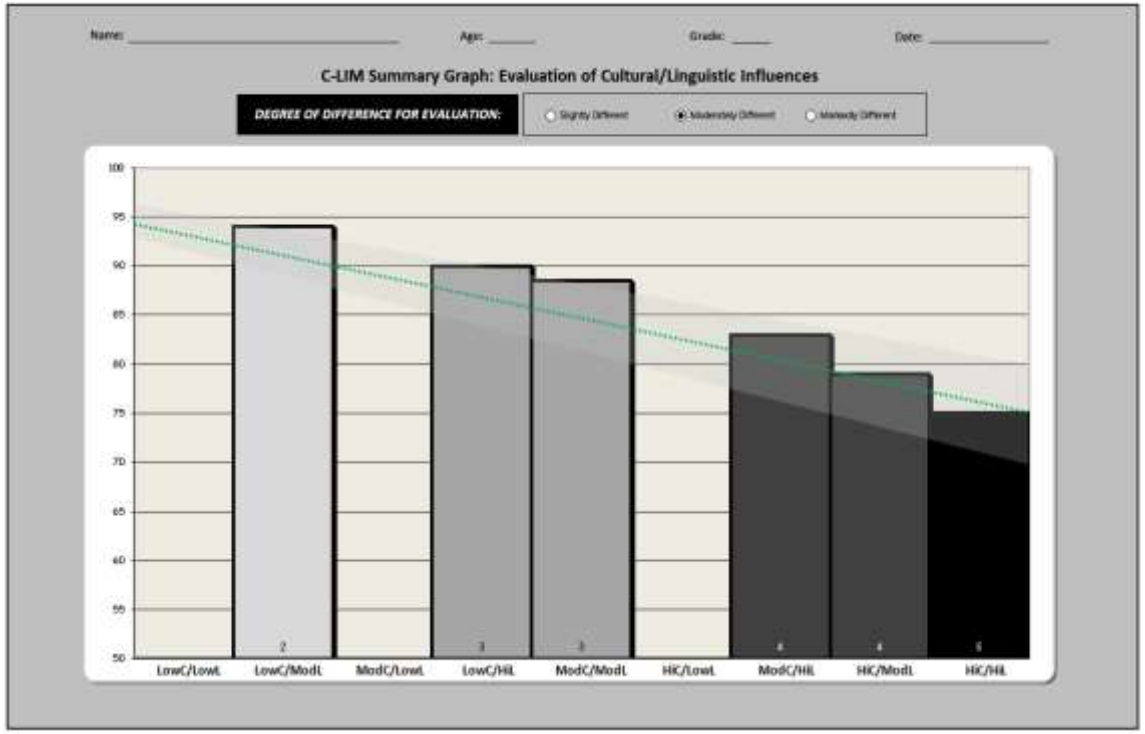
3. No Significant Score Variability: *There is no variability in the scores that form the aggregate in any one cell or any variability between or among cells in the same level where high score performance may be masking the presence of low performance. Variability is defined as one score below average AND below the expected range, and the next lowest score is 1SD (15 points) higher and within the expected range.*

Results are likely **INVALID** only if **ALL** conditions are **MET.**

Results are likely **VALID** when **ANY** condition is **NOT MET.**

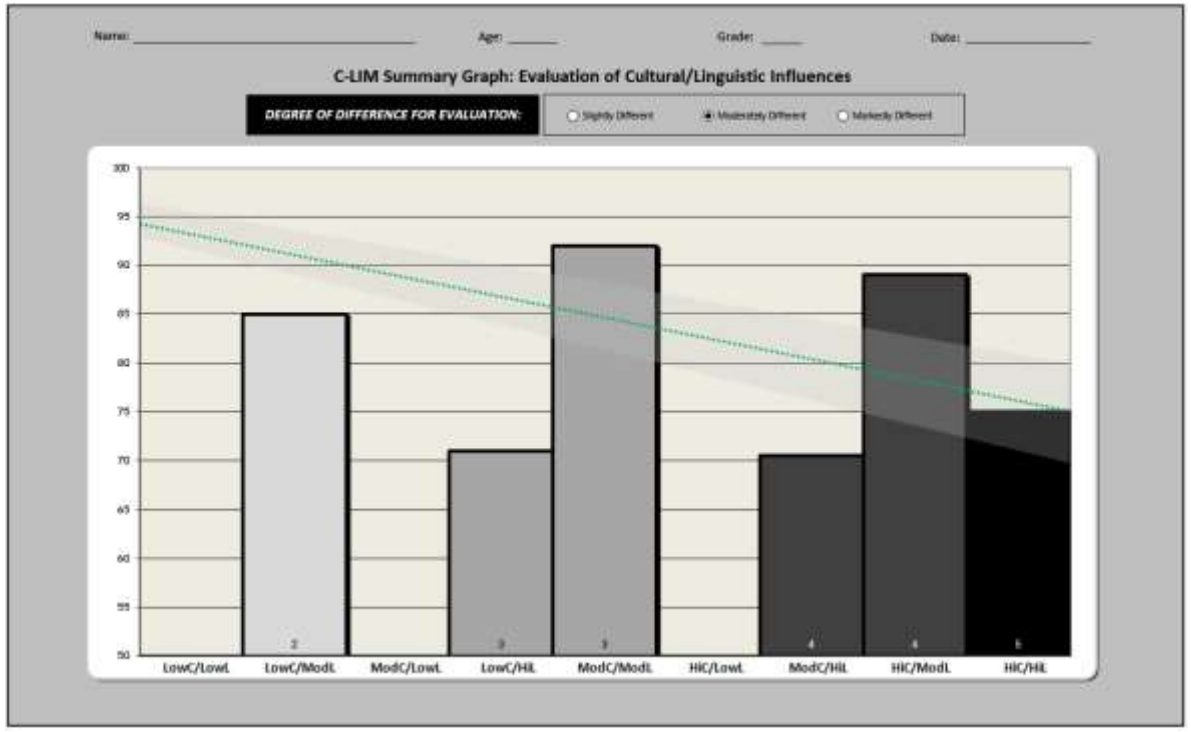
Interpretation: When **ALL** three criteria are **MET,** it may be concluded that the test scores are likely to be **INVALID** because they were influenced primarily by cultural/linguistic variables and cannot be interpreted. When any **ONE** criterion is **NOT MET**, the results can be assumed to be likely **VALID** and may be interpreted if further evidence is generated to support conclusions.

Interpreting Test Score Validity with the C-LIM



Example of "likely invalid" score pattern—overall general decline **AND** scores within or above expected ("average" or typical) range **AND** scores show no important variability.

Interpretation: Performance PRIMARILY due to linguistic and cultural factors, scores CANNOT be interpreted specifically, and provide no evidence to support disability.

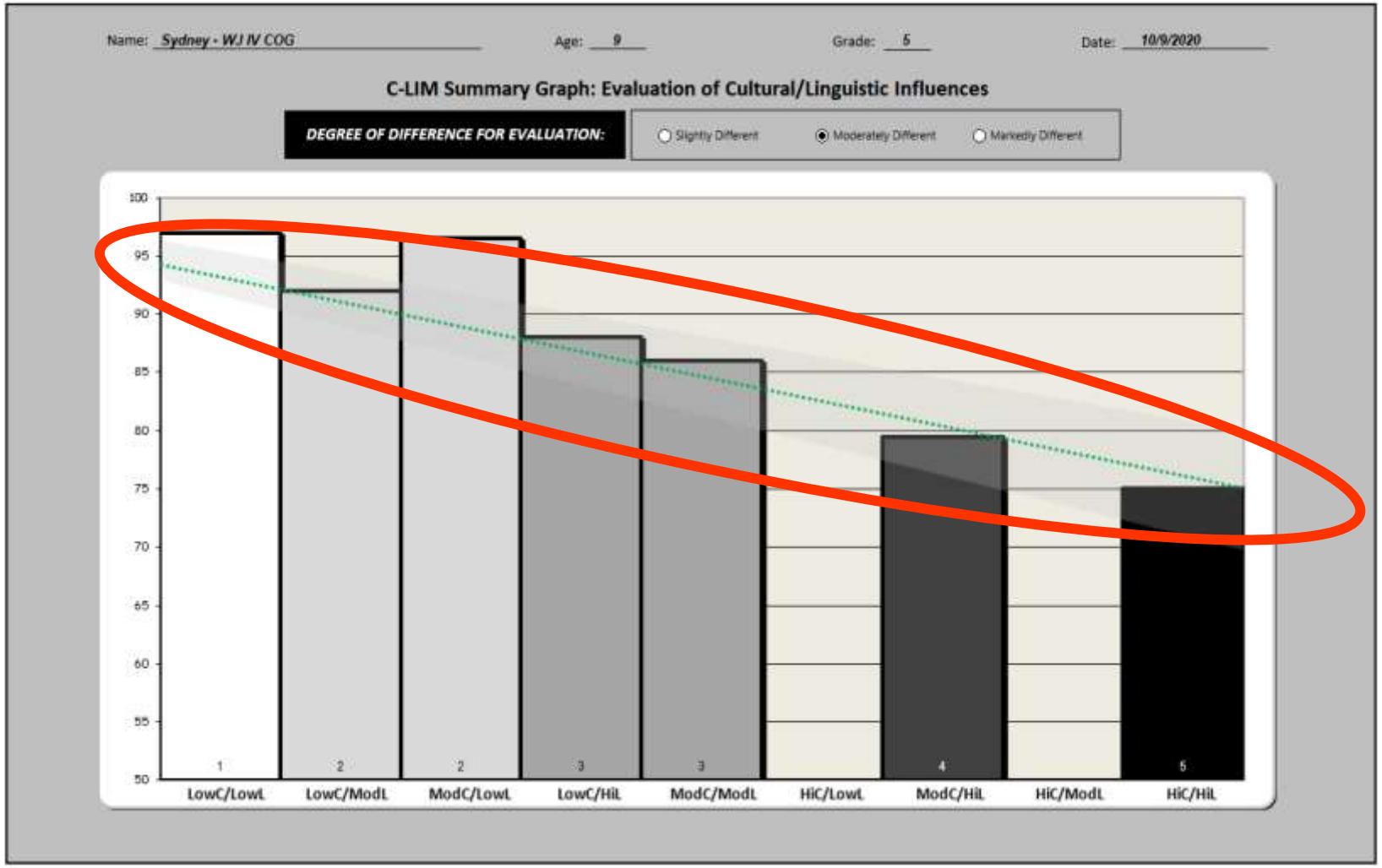


Example of "likely valid" score pattern—no overall decline **OR** scores below expected ("average" or typical) range **OR** scores show important variability.

Interpretation: Performance NOT PRIMARILY due to linguistic and cultural factors, scores CAN be interpreted but need further validation to provide evidence of possible disability.

C-LIM Guidelines for Evaluating Test Scores

General pattern of decline AND all scores within or above the expected range for ELs.

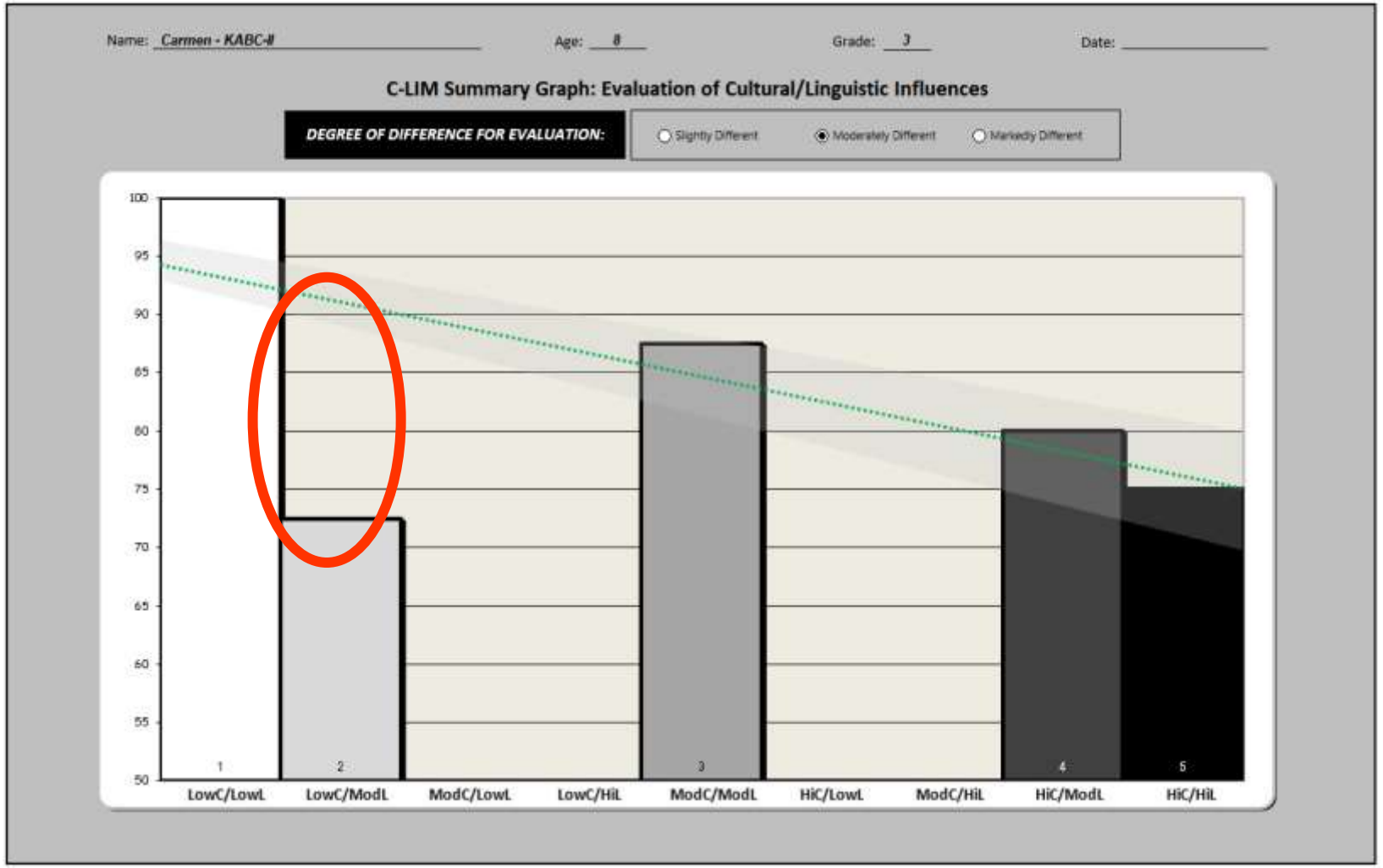


CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are LIKELY INVALID

Interpretation: "average" or typical functioning, no evidence to suggest cognitive or linguistic deficits that might support disability.

C-LIM Guidelines for Evaluating Test Scores

General pattern of decline OR one or more scores below expected range for ELs.



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are LIKELY VALID.

Interpretation: suggests possible evidence of cognitive or linguistic deficit that may be confirmed with additional testing and evaluation.

C-LIM Guidelines for Evaluating Test Scores

No general pattern of decline.

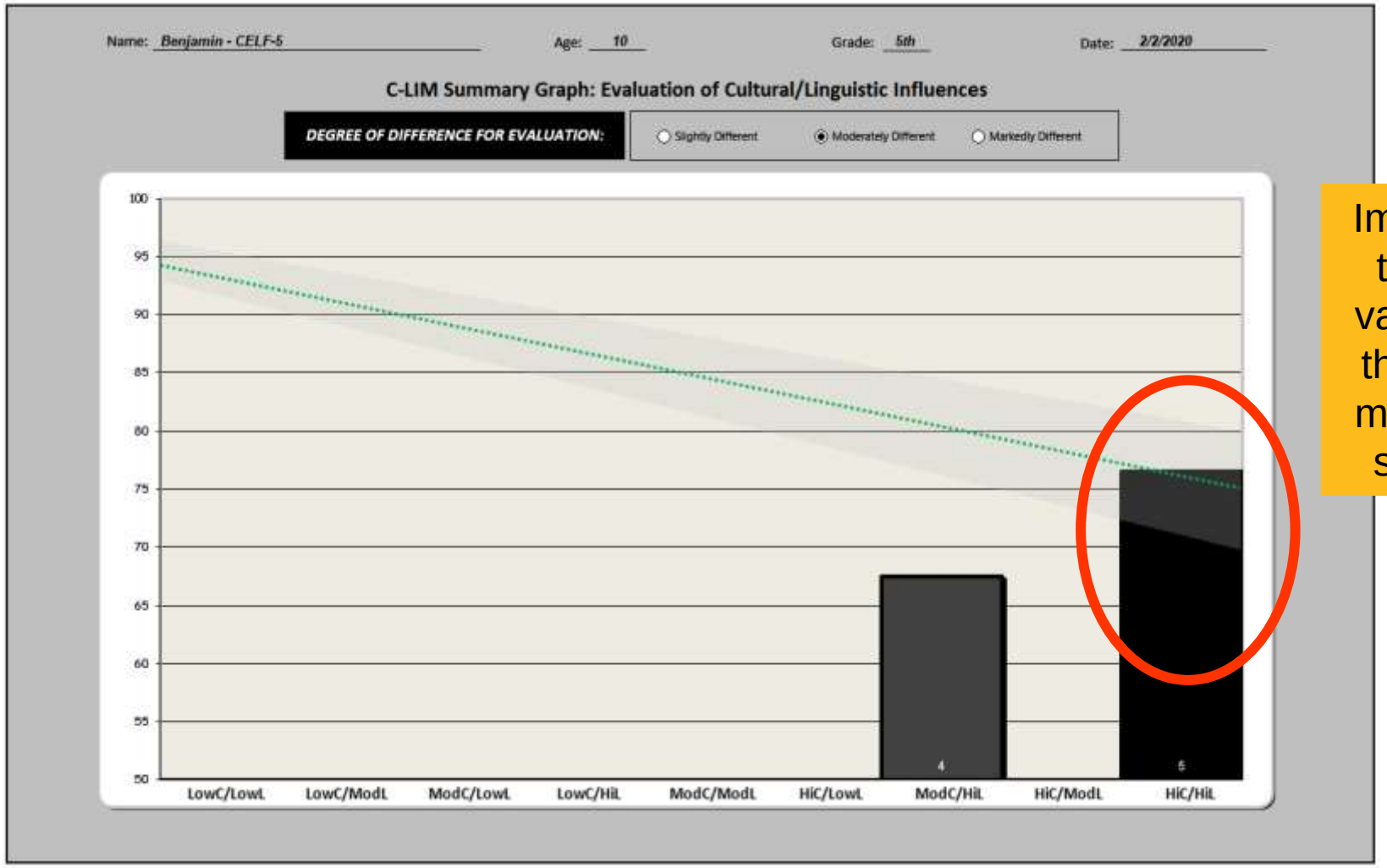
Important to note variability that may mask low scores.

CULTURE/LANGUAGE INFLUENCE: MINIMAL – test scores are LIKELY VALID.

Interpretation: suggests possible evidence of cognitive or linguistic deficit that may be confirmed with additional testing and evaluation.

C-LIM Guidelines for Evaluating Test Scores

No pattern of decline <u>BUT</u> at least one or more scores below expected range for ELs.



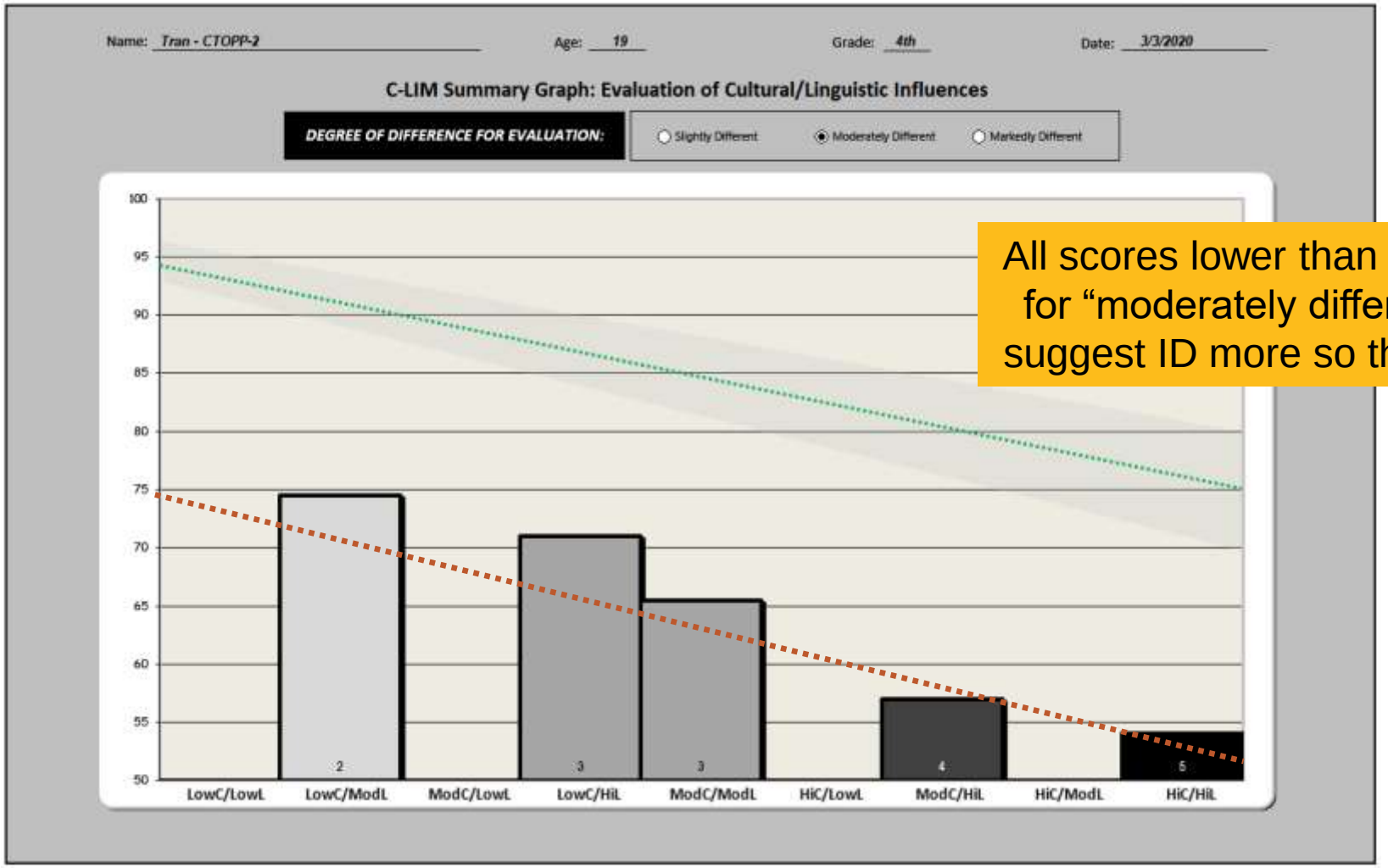
Important to note variability that may mask low scores.

CULTURE/LANGUAGE INFLUENCE: MINIMAL – test scores are LIKELY VALID.

Interpretation: suggests possible evidence of language impairment that may be confirmed with additional testing and evaluation.

C-LIM Additional Interpretive Issues

General pattern of decline, but all scores NOT within expected range



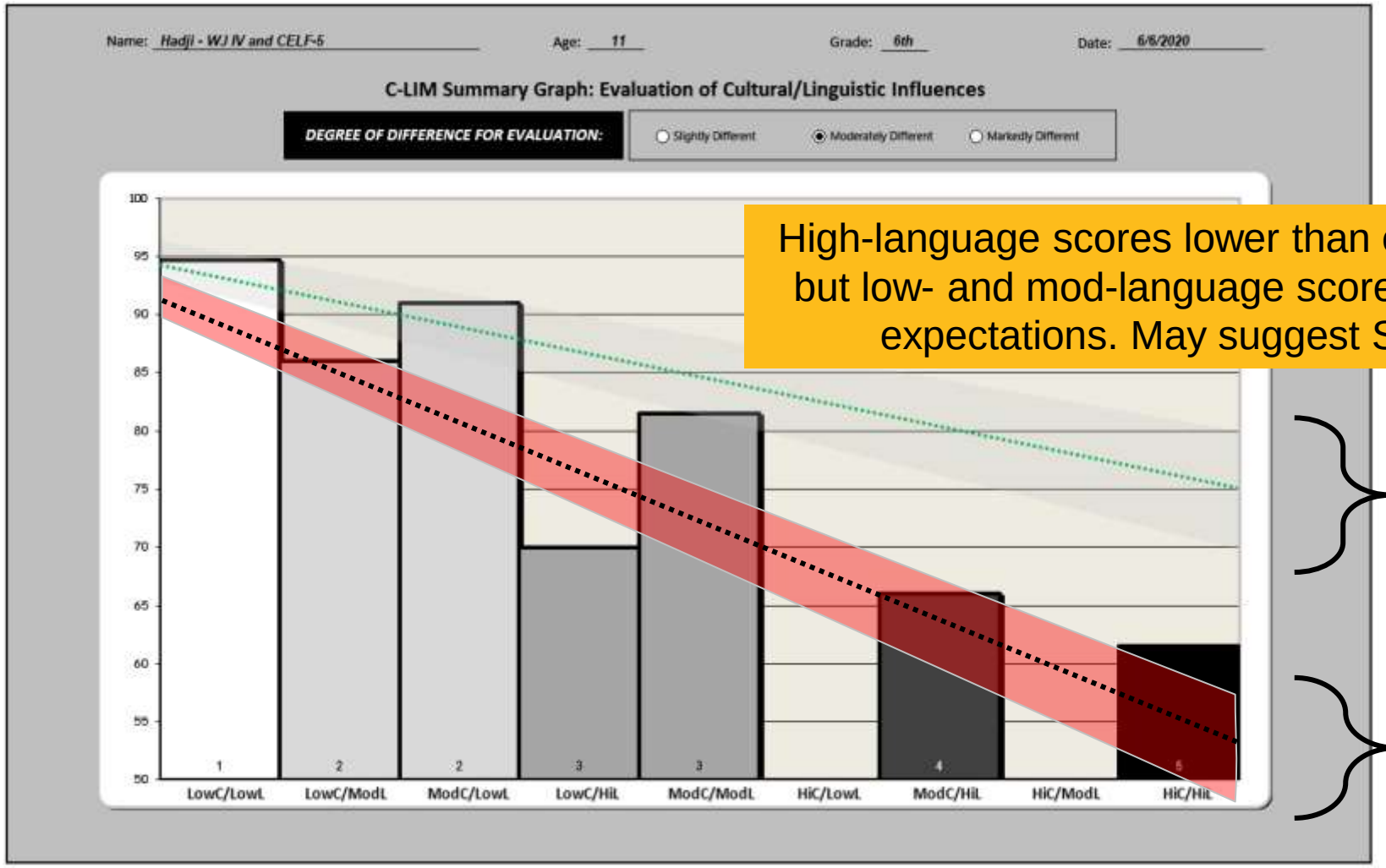
All scores lower than expected range for “moderately different” ELs. May suggest ID more so than SLD or SLI.

CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are LIKELY VALID.

Interpretation: suggests possible evidence of general cognitive deficit that may be confirmed with additional testing and evaluation.

C-LIM Additional Interpretive Issues

General pattern of decline BUT not all scores within expected range



High-language scores lower than expected but low- and mod-language scores within expectations. May suggest SLI.

Expected rate of decline

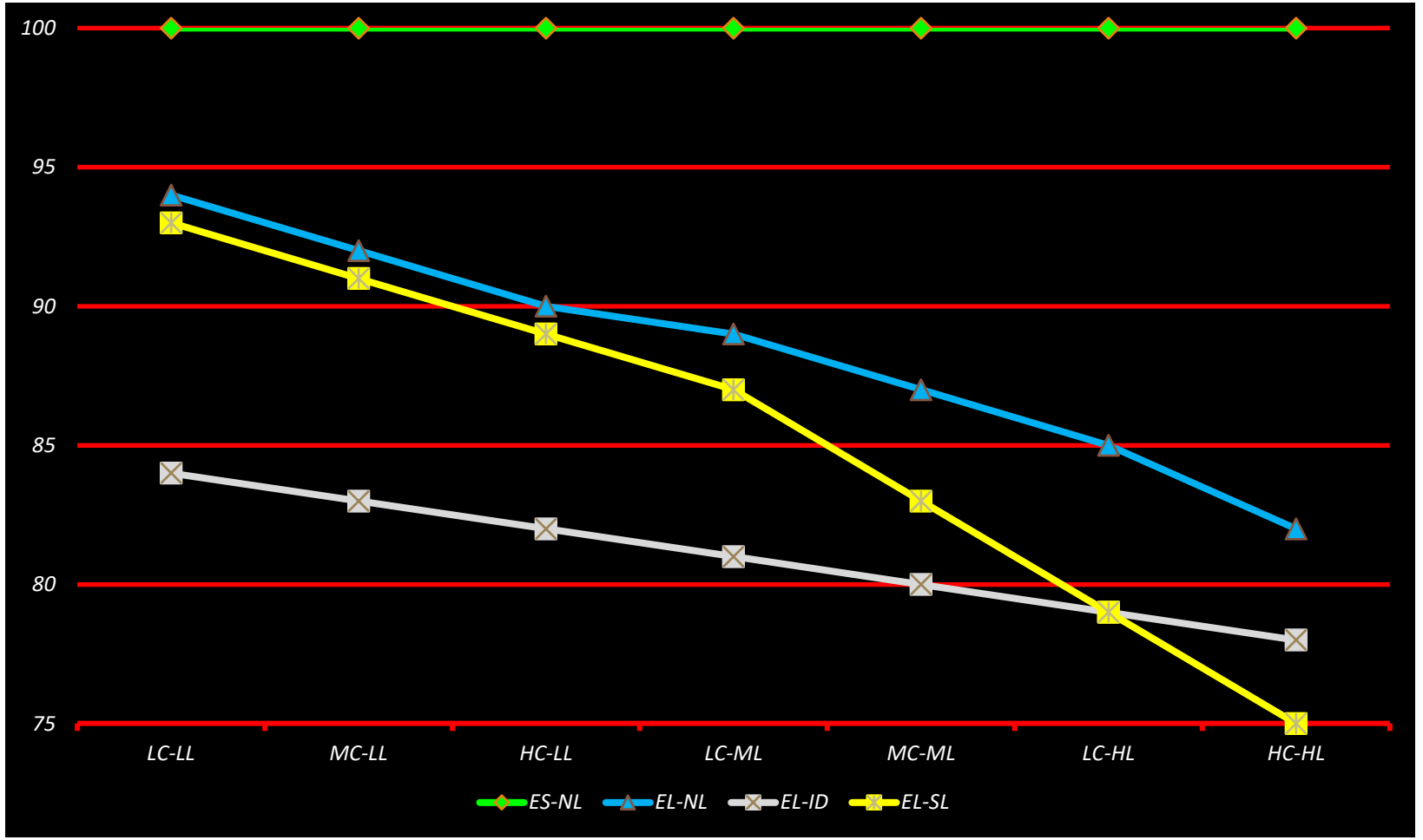
Steeper rate of decline

CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are LIKELY VALID.

Interpretation: suggests possible evidence of language-related learning deficit that may be confirmed with additional testing and evaluation.

C-LIM Additional Interpretive Issues

Mean C-LIM cell aggregates for WPPSI-III subtests arranged by degree of cultural loading and linguistic demand for ELs identified with language impairment, learning disability, and intellectual disability.



Source: Tychanska, J., Ortiz, S. O., Flanagan, D.P., & Terjesen, M. (2009), unpublished data..

C-LIM Additional Interpretive Issues

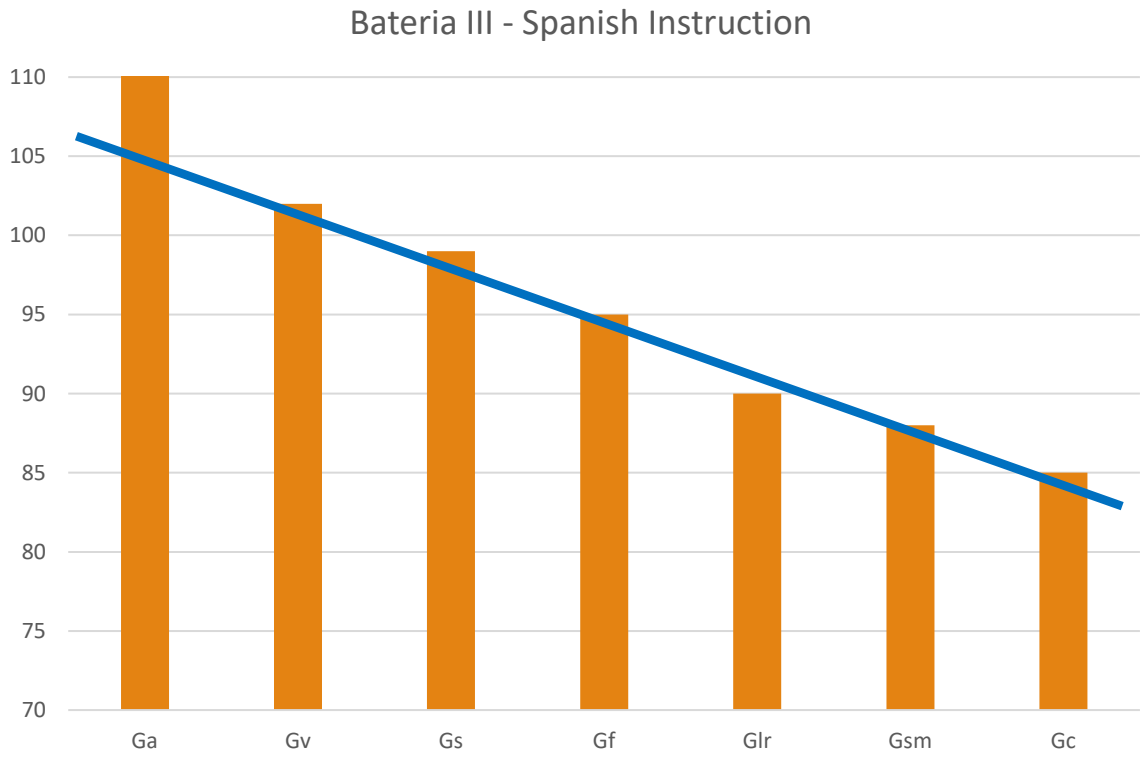
The Impact of Language Development on Native-Language Tests

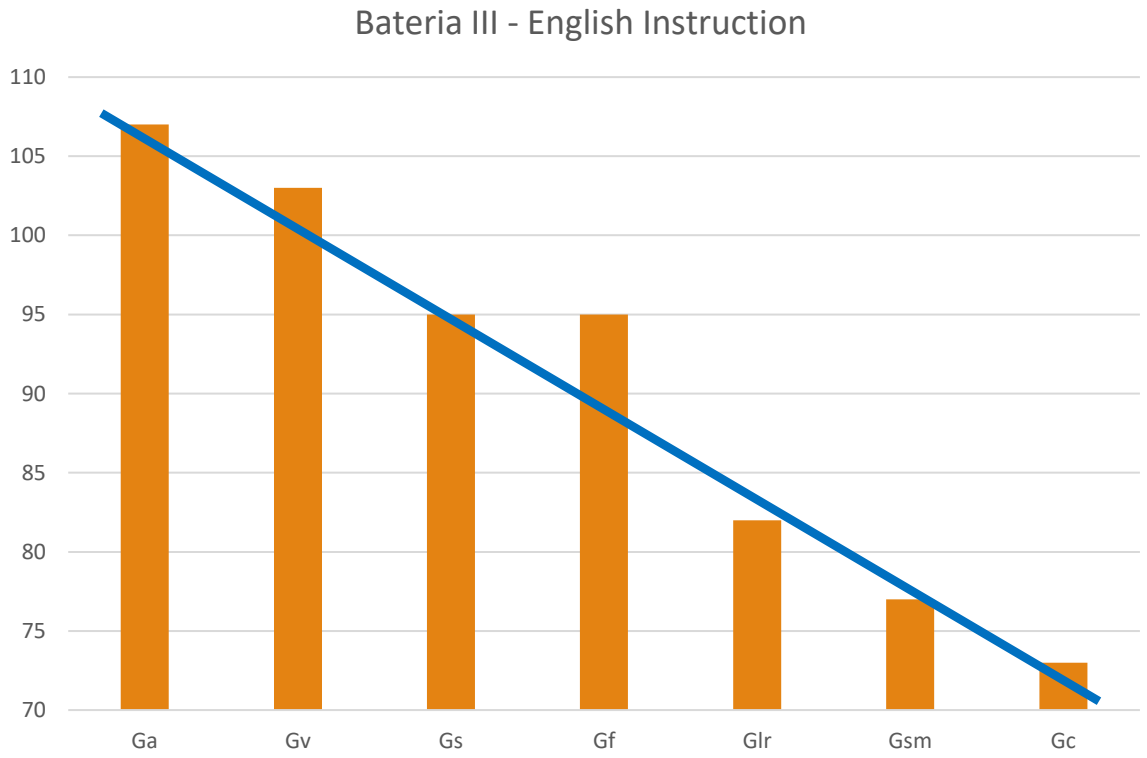
WJ III Classifications		Bateria III Classifications (NLD)		Bateria III Classifications (ELD)	
Mean	**Subtest**	**Mean**	**Subtest**	**Mean**	**Subtest**
98	Gv – Visual Processing	111	Ga – Auditory Processing	107	Ga – Auditory Processing
95	Gs – Processing Speed	102	Gv – Visual Processing	103	Gv – Visual Processing
95	Gsm – Short Term Memory	99	Gs – Processing Speed	95	Gs – Processing Speed
92	Gf – Fluid Reasoning	95	Gf – Fluid Reasoning	95	Gf – Fluid Reasoning
89	Ga – Auditory Processing	90	Glr – Long Term Memory	82	Gsm – Short Term Memory
89	Glr – Long Term Memory	88	Gsm – Short Term Memory	77	Glr – Long Term Memory
85	Gc – Crystallized Knowledge	85	Gc – Crystallized Knowledge	73	Gc – Crystallized Knowledge

**Source: Esparza Brown, J. (2008). The use and interpretation of the Bateria III with U.S. Bilinguals. Unpublished dissertation, Portland State University, Portland, OR.*

C-LIM Additional Interpretive Issues

Bateria III Performance Means Ordered by Broad Ability Domain





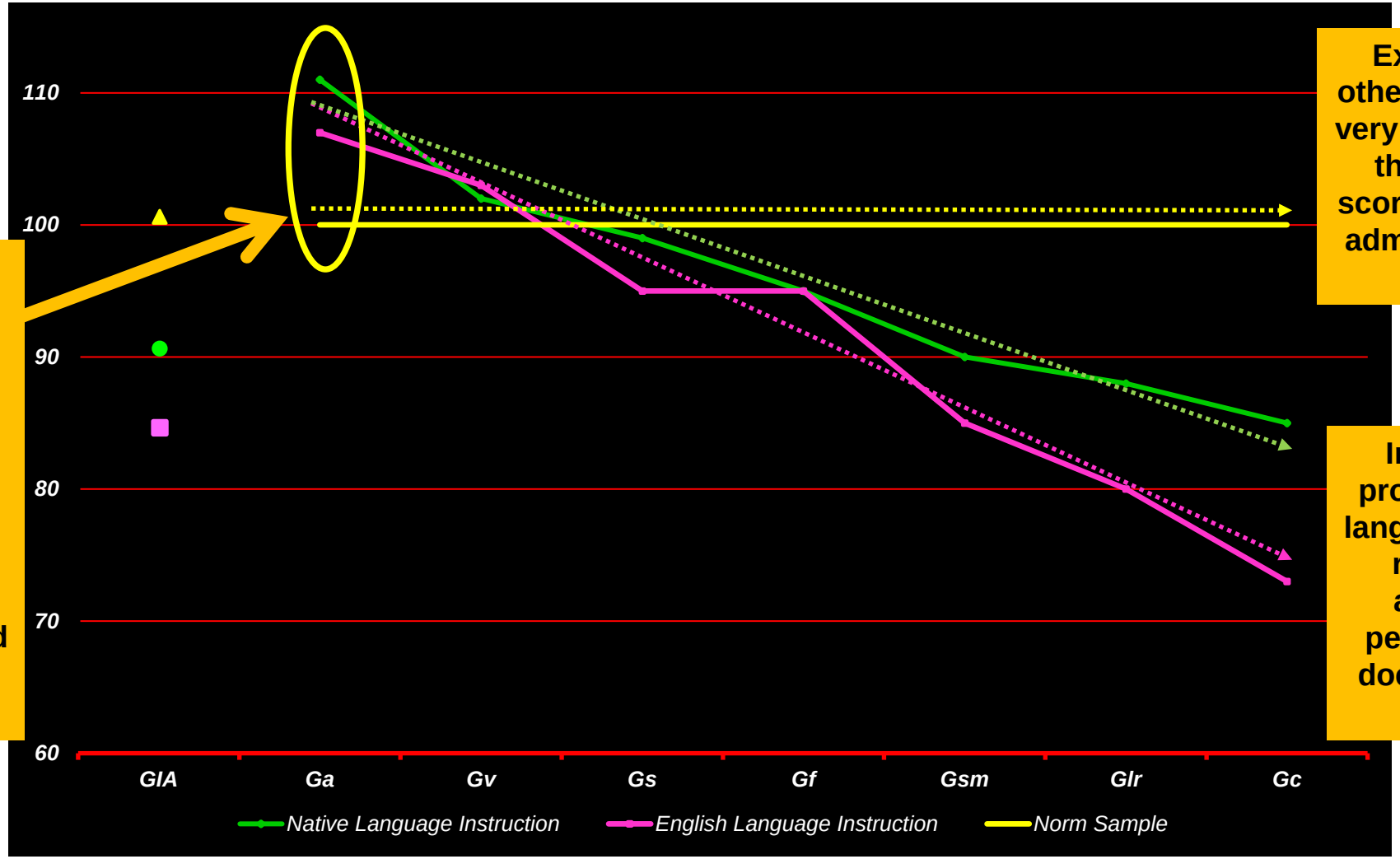
Native-language testing for students receiving bilingual instruction appears to result in a pattern of attenuation due to differences in native-language development. Students receiving bilingual instruction are less developed in Spanish than the monolingual, Spanish-speakers in the norm sample.

Native-language testing for students receiving ESL only appears to result a pattern of attenuation due to differences in native-language development. Students receiving ESL are even less developed in Spanish than the monolingual, Spanish-speakers in the norm sample.

**Source: Esparza Brown, J. (2008). The use and interpretation of the Bateria III with U.S. Bilinguals. Unpublished dissertation, Portland State University, Portland, OR.*

C-LIM Additional Interpretive Issues

The Impact of Language Development on Native-Language Tests



Spanish is a highly transparent language having very regular sound-symbol correspondence. English, in comparison is an opaque language where sound-symbol correspondence is significantly lower and therefore, more difficult.

Except for Ga, all other abilities follow a very similar pattern as that seen for test scores with ELs when administered tests in English.

In addition, the provision of native language instruction results in less attenuation of performance than does instruction in English only.

**Source: Esparza Brown, J. (2008). The use and interpretation of the Bateria III with U.S. Bilinguals. Unpublished dissertation, Portland State University, Portland, OR.*

Interpretive Errors in C-LIM Studies: Styck & Watkins

So why didn't the Styck & Watkins studies support the C-LIM? At the group level, the scores for ELs appeared to show a clear pattern of decline, yet they concluded otherwise:

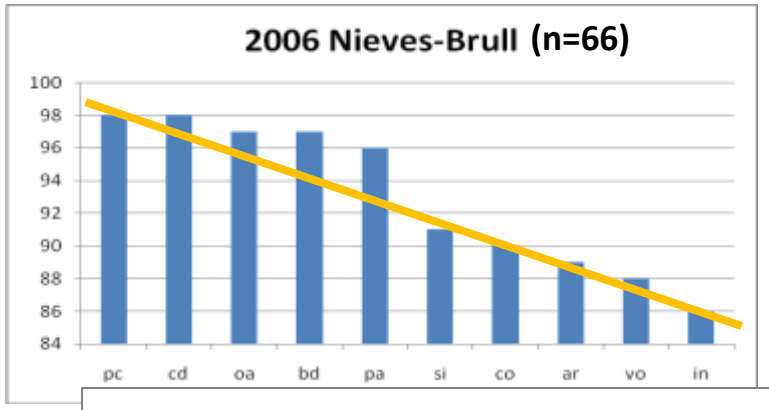
> *"The valid C-LIM profile (i.e., cell means did not decline) emerged in the mean WISC-IV normative sample and the ELL sample. Thus, neither sample of children exhibited the invalid C-LIM profile when group mean scores were considered" (p. 374) (emphasis added).*

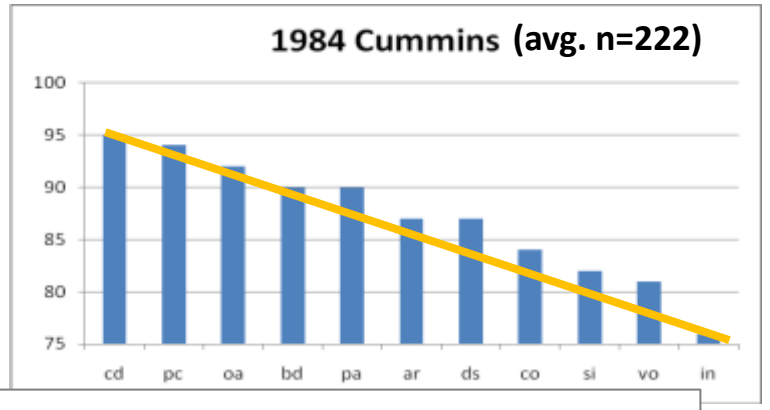
The normative sample should not and rightly "did not decline" as they were not ELs and not disabled. However, the EL sample did range from a high on Picture Concepts (SS=98) to a low on Vocabulary (SS=85), largely in accordance with prior research and the C-LIM classifications.

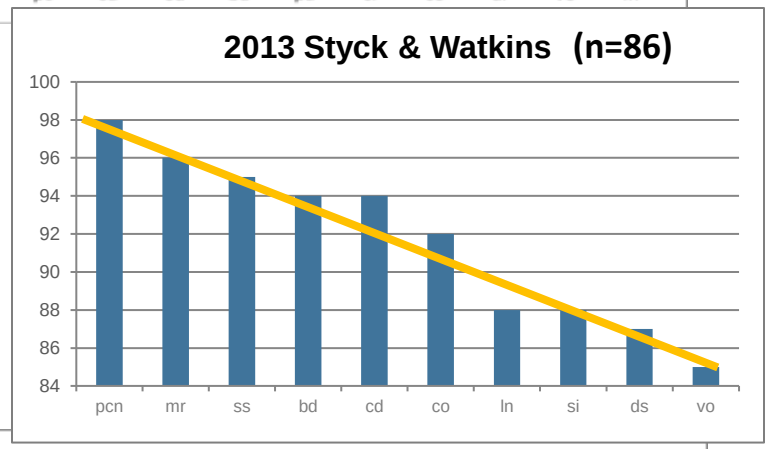
**Source: Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.*

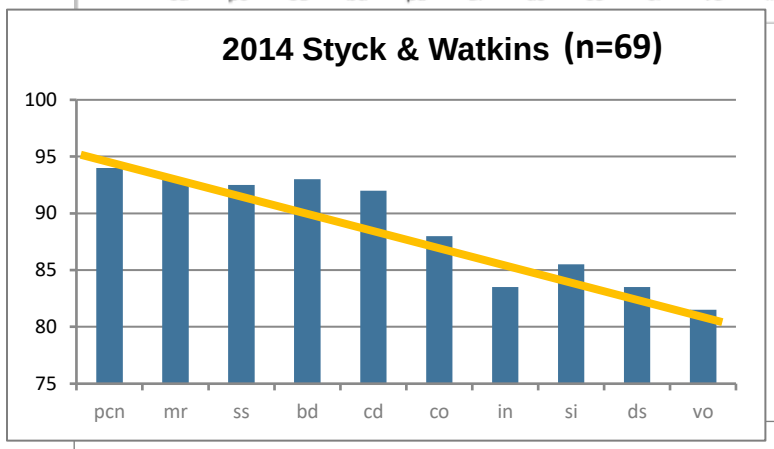
Interpretive Errors in C-LIM Studies: Styck & Watkins

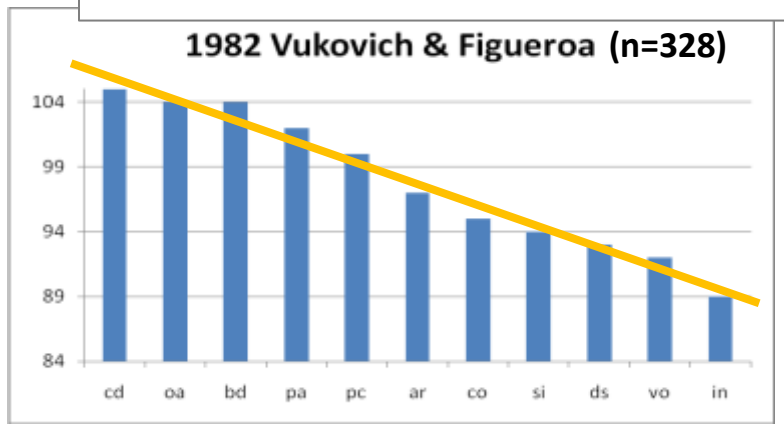
Comparison of overall "average" test performance at the subtest level: EL to ES

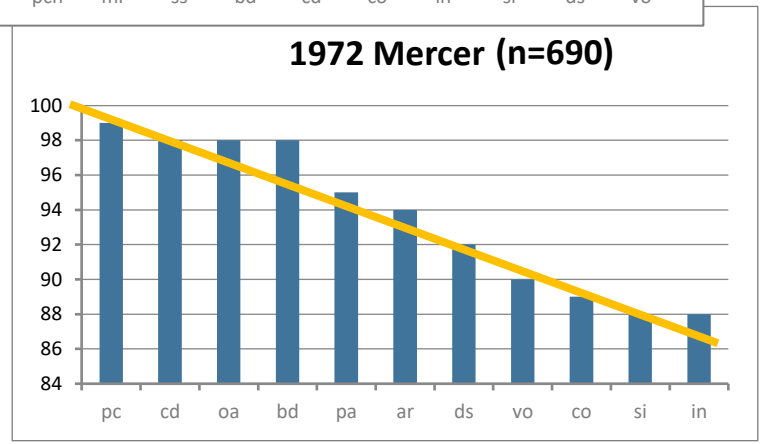












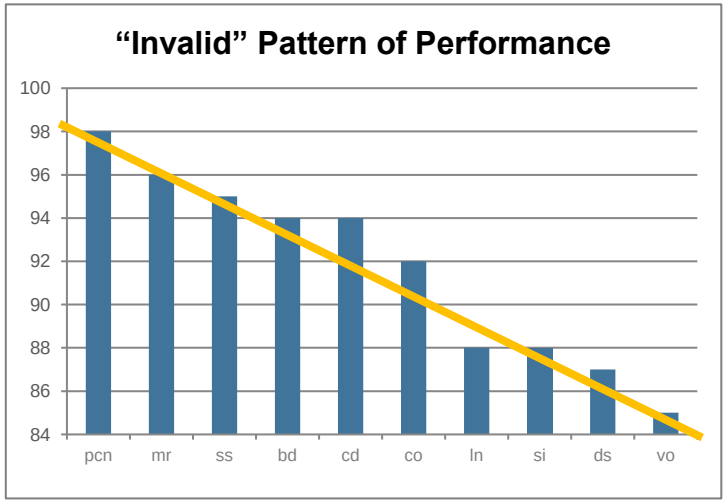
Interpretive Errors in C-LIM Studies: Styck & Watkins

The most egregious error in the Styck & Watkins studies is found in the examination of *individual* patterns of performance within the C-LIM.

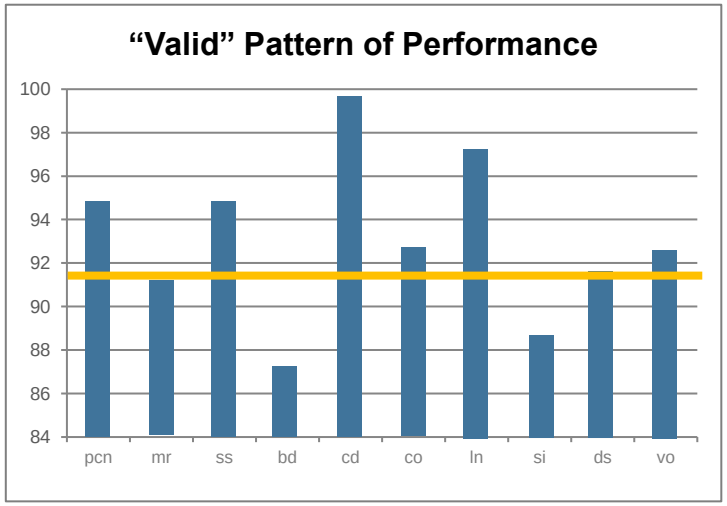
In this regard, the researchers incorrectly expected to find patterns of decline (does not support disability) in each individual case which is precisely opposite of what they should have expected to find, no decline (support for a disability) because their sample was comprised of ELs who had already been identified as having a disability.

The authors noted that *"roughly 97% of (n = 83) of participants were identified as meeting criteria for an educational disability (86% as SLD)" (p. 371).* Yet, only 9 ELL cases (10.5%) resulted in invalid scores (no disability). Thus, the C-LIM suggested invalid scores in 9 cases, 3 of which were likely correct (those without disabilities) so that the C-LIM was consistent with and supported the placement decision of the child by the district in 93% of the cases (89.5% + 3.5%). Moreover, the results of analyses with the WISC-IV normative sample show that declines relative to language are unusual, perhaps even indications of potential SLI in monolingual, native English speakers as described by Cormier et al. (2014).

Interpretive Errors in C-LIM Studies: Styck & Watkins



Invalid Pattern	Expected N = 3 (out of 86)
No evidence of disability	**Correct (N = 3, 3.5%)** Incorrect (N = 6, 7.0%)



Correct C-LIM pattern found in 89.5% + 3.5% = 93% of all cases

Valid Pattern	Expected N = 83 (out of 86)
Evidence of disability	**Correct (N = 77, 89.5%)**

Far from undermining the validity of the C-LIM, the Styck & Watkins studies provide powerful support for the clinical utility and validity of the C-LIM when evaluating EL test performance using current research and an evidence-base.

Interpretive Errors in C-LIM Studies by Styck & Watkins

Overall decline *and* within expected range = no disability

No decline *or* below expected range = possible disability

		EL Sample (with disability)	Norm Sample (no disability)
WISC-IV C-LIM Analysis	Invalid Scores (decline)	N=9 (N=6, 7.0%) (N=3, 3.5%)	N = 100 (4.9%)
	Valid Scores (no decline)	N = 77 (89.5%)	N = 1,933 (95.1%)

The authors noted that *"roughly 97% of (n = 83) of participants were identified as meeting criteria for an educational disability (86% as SLD)" (p. 371).* Yet, only 9 ELL cases (10.5%) resulted in invalid scores (no disability). Thus, the C-LIM suggested invalid scores in 9 cases, 3 of which were likely correct (those without disabilities) indicating that the C-LIM was consistent with and supported the placement decision of the child by the district in 93% of the cases (89.5% + 3.5%).

Moreover, the results of analyses with the WISC-IV normative sample show that declines relative to language are unusual, perhaps even indications of potential SLI in monolingual, native English speakers as described by Cormier et al. (2014).

**Table adapted from: Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.*

Summary of Research Foundations for EL Evaluation

1. COMPARED TO ENGLISH SPEAKERS (EL to ES): *Test performance of ELs is moderated by the degree to which a given index or subtest relies on or requires age- or grade-expected English language development and the acquisition of incidental acculturative knowledge.*

2. COMPARED TO ENGLISH LEARNERS (EL to EL): *Test performance of ELs is further moderated by the degree to which an EL varies in terms of their own developmental English language proficiency and acculturative knowledge acquisition.*

Proper interpretation of EL test performance thus requires a true peer group of other ELs that is based not on the language spoken by the individual but on comparison to other ELs with the same degree of English exposure and development.

With two exceptions, current test norm samples lack control for developmental differences in English language exposure. This means that interpretation of test scores at any level must be made within the context of research which provides the only empirically-derived, albeit very rough, true peer standard or "norm group".

Use of research on the relative test performance of ELs based on language exposure (as reflected by the degree of "difference" the student displays relative to the norm samples of the tests being used) is the very foundation and sole purpose of the C-LIM.

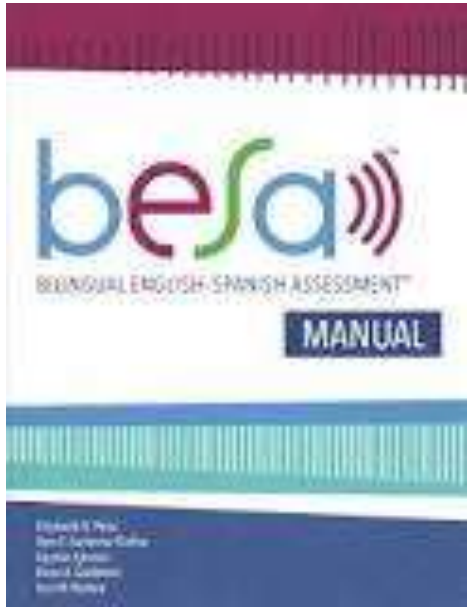
Translating Research into Practice

Evaluation Issues and Methods	Norm sample representative of bilingual development	Measures a wider range of school-related abilities	Does not require the evaluator to be bilingual	Adheres to the test's standardized protocol	Substantial research base on bilingual performance	Sufficient to identify or diagnosis disability	Accounts for variation in bilingual development	Most likely to yield reliable and valid data and information	Provides extensive data regarding development
Modified or Altered Assessment	✗	✓	✓	✗	✗	✗	✗	✗	✗
Reduced-language Assessment	✗	✗	✓	✓	✗	✗	✗	✗	✗
Dominant Monolingual Assessment in L1: native only	✗	✓	✗	✓	✗	✗	✗	✗	✗
Dominant Monolingual Assessment in L2: English only	✗	✓	✓	✓	✓	✗	✗	✗	✗
Integrated Approach (L2+L1)	✓	✓	✓	✓	✓	✓	✓	✓	✓

An evidence-based approach to evaluation of ELs must consider issues beyond test score validity. An integrated approach can resolve relevant validity issues by applying research on EL test performance to establish a "true peer" reference group for disability-based evaluations that does not require the evaluator to have bilingual competency.

The Bilingual English-Spanish Assessment:

Sampling bilinguals—categorical (3 levels of exposure)



Authors: Elizabeth D. Pena, Vera F. Gutierrez-Clellen, Aquiles Iglesias, Brian A. Goldstein, Lisa M. Bedore.

Performance is based on comparison to peers grouped by three categories based on language development.

Table 5.2 Sample Distribution by Age and Language Exposure

	Language Group					
Age	**Functional Monolingual: English**	**Bilingual Dominant: English**	**Balanced Bilingual**	**Bilingual Dominant: Spanish**	**Functional Monolingual: Spanish**	**Total**
4	7.80%	2.90%	3.20%	3.80%	12.80%	31%
5	9.20%	3.90%	6.70%	6.40%	12.00%	38%
6	6.70%	3.80%	7.00%	4.90%	8.90%	31%
Total	24%	11%	17%	15%	34%	

The Ortiz Picture Vocabulary Acquisition Test

Sampling bilinguals—continuous (99 levels of exposure: 1%-99%)

Author: Samuel O. Ortiz

Performance is based on comparison of exact amount of language development determined by percentage of lifetime exposure—not by category.

Table 5. Length of Exposure to English: Ortiz PVAT English Learner Normative Sample

Length of Time Exposed to English	English Learner Normative Sample (*N*)	English Learner Normative Sample (%)
0–6 months	128	10.8
7–11 months	131	11.0
1–2 years	168	14.1
3–4 years	165	13.9
5 years	119	10.0
6–7 years	118	9.9
8–9 years	113	9.5
10–11 years	90	7.6
12–13 years	70	5.9
14–15 years	51	4.3
16 years or more	37	3.1
Total	**1,190**	**100.0**

Table 6. Percentage of Life Exposed to English: Ortiz PVAT English Learner Normative Sample

Percentage of Life Exposed to English (%)	English Learner Normative Sample (*N*)	English Learner Normative Sample (%)
0–20	280	23.5
21–40	196	16.5
41–60	196	16.5
61–80	209	17.6
81–100	309	26.0
Total	**1,190**	**100.0**

Practical Considerations in Addressing Score Validity When Testing ELs

The usual purpose of testing is to identify deficits in ability (i.e., low scores) which means that validity is more of a concern for low scores than average/higher scores because:

- *Test performances in the average range are NOT likely a chance finding and strongly suggests average ability (i.e., no deficits in ability)*
- *Test performances that are below average MAY be a chance finding because of experiential or developmental differences and thus do not automatically confirm below average ability (i.e., possible deficits in ability)*

Therefore, testing in one language only (English or native language) means that:

- *It can be determined that a student DOES NOT have a disability (i.e., if all scores are average or higher, they are very likely to be valid)*
- *It CANNOT be determined if the student has a disability (i.e., low scores must be validated as true indicators of deficit ability)*

Thus, testing in both languages (English and native language) is necessary to determine disability because:

- *Testing requires confirmation that deficits are not language-specific and exist in both languages (although low performance in both can result from other factors)*

No matter the order, all low test scores, whether in English or the native language, must be validated:

- *Low scores from testing in English can be validated via research underlying the C-LIM*
- *Low scores from testing in the native language cannot be validated with research*

Multilingual Testing - L1 then L2 or L2 then L1?

- Testing cannot be conducted in two languages simultaneously. In addition, testing must necessarily begin in one language or another. The considerations presented previously and the limitations and problems regarding establishing test score validity in the native language tend to favor initial evaluation in English.
- However, the language used for initial evaluation also depends in large part on the purpose and the area of suspected disability. If the evaluation is for SLD, then students should have been given sufficient opportunity to learn in which time they will have gained enough proficiency to easily allow testing in English and testing should begin in English (L2). Immediate referral for LD evaluation is discriminatory.
- For other disabilities, particularly those that are more developmentally based and have physical manifestations, there may be a need to evaluate immediately without having to be concerned with allowing a lengthy period of time to observe learning. In these cases, evaluation in the heritage language (L1) may well be preferable and appropriate, particularly since such disabilities require much more than just valid test scores.
- Remember that even scores from tests that are not administered in English must be examined for validity and there is no formal way at present of doing so. This is less of a problem when evaluating disabilities that have clear physiological indicators or where additional data are necessary for determination. Because SLD has no such obvious markers and because it is a disability determined by exclusion with heavy reliance on test scores, evaluation should likely proceed in English (L2) first to permit evaluation of test score validity and then any weaknesses followed up in the heritage language (L1).
- An L2+L1 approach is the most feasible and logical sequence for testing in the vast majority of evaluations.

A Best Practice Framework for Nondiscriminatory Evaluation of SLD:

Pre-referral Activities

Assess and evaluate factors that affect opportunity to learn and age/grade-expected development (baseline functioning)

- *Include assessment of first and second language acquisition, type and length of formal schooling, opportunity for learning via systematic exposure to linguistic and acculturative experiences, parental level of education, literacy, and socio-economic status.*

Monitor and evaluate academic skills growth relative to true peers including native/heritage language (pre-referral evaluation)

- *Formally monitor and systematically evaluate progress in academic skills in English (or native/heritage language, as appropriate) using true peer comparison. Directly examine the effectiveness of interventions and academic growth. Methods may include authentic and informal data (e.g., work samples, portfolios, etc.) or more formal data collected within an MTSS/RtI framework (e.g., CBM, progress monitoring charts, standardized test data). Goal is to evaluate progress and growth, not determine disability.*

RTI/MTSS addresses concerns regarding fairness and equity in the assessment process

Post-referral Testing

Assess and evaluate construct validity in all areas in English first* (exclusion of cultural/linguistic factors)

- *Evaluate in English first (when possible and appropriate) using true peer comparison and standards for expected performance. For formal testing, the C-LIM can be used for this purpose. If all data indicate average performance, a disability is unlikely and further evaluation unnecessary. If some data suggest performance is below true peers, continue evaluation.*

Re-assess and re-evaluate construct validity in areas of poor performance in the native language (cross-linguistic evidence)

- *If performance in some areas evaluated in English is lower than expected compared to true peers, re-assess the same areas in the native/heritage language (when possible and appropriate) to support them as areas of true weakness.*

Multilingual Testing addresses possible bias in use of test scores

Decision Making

Cross-validate all data with contextual factors and pre-referral information (ecological validity for disability)

- *Use all other case data and information to serve as the context by which to evaluate the L1 and L2 data and ensure ecological validity for any conclusions that have been made.*

**This procedure assumes that an EL has been given sufficient opportunity to learn before being referred for evaluation as not doing so is discriminatory. Because this period of time should not generally be shorter than one year, students will have had enough time to also learn enough English to permit testing in English.*

Multilingual Testing (L2+L1): Step 1

Step 1. Test first in English (L2) and examine test scores

- If all composites indicate normative strengths (SS > ≈ 90), scores are valid (to the extent that a disability is not likely), no further testing is necessary and poor academic performance is due to cultural/linguistic factors.
- If one or more composites indicate a normative weakness (SS < ≈ 90), enter and evaluate subtest score validity in the C-LIM.

Step 2. Evaluate impact of cultural/linguistic factors on test scores

- If all criteria are met, then cultural/linguistic factors are primary (cannot be excluded), scores are likely to be invalid BUT do indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.
- If C-LIM indicates only a contributory or minimal influence of cultural/linguistic factors, then test scores are likely to be valid and evaluation should continue.

Step 3. Re-evaluate areas of weakness only in native language (L1) cross-linguistic confirmation of validity

- If L1 data indicate an area is a strength (i.e., average), then original L2 (English) score is invalid, use the L1 score.
- If L1 data indicate an area is still a weakness, then original L2 (English) score is valid, use the L2 (English) score.

Step 4. Cross-validate L1 and L2 scores with contextual factors and pre-referral concerns for ecological validity

- Use information from detailed history of language development, formal education, parental SES, family experiences, as the context by which to evaluate consistency, explanatory logic, and ecological validity regarding data from pre-referral interventions, classroom work, achievement data, and pre-referral concerns.

Nondiscriminatory Assessment of Specific Learning Disability with an English Learner

Evaluation of Maria Ayala
Tests Used: WISC-V, WIAT-III, and WJ IV
DOE: 6/22/2016
DOB: 10/4/2006
Grade: 4

Multilingual Testing (L2+L1): Step 1

Step 1. Test first in English (L2) and examine test scores

- If all composites indicate normative strengths (SS > ≈ 90), scores are valid (to the extent that a disability is not likely), no further testing is necessary and poor academic performance is due to cultural/linguistic factors.
- If one or more composites indicate a normative weakness (SS < ≈ 90), enter and evaluate subtest score validity in the C-LIM.

Step 2. Evaluate impact of cultural/linguistic factors on test scores

- If all criteria are met, then cultural/linguistic factors are primary (cannot be excluded), scores are likely to be invalid BUT do indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.
- If C-LIM indicates only a contributory or minimal influence of cultural/linguistic factors, then test scores are likely to be valid and evaluation should continue.

Step 3. Re-evaluate areas of weakness only in native language (L1) cross-linguistic confirmation of validity

- If L1 data indicate an area is a strength (i.e., average), then original L2 (English) score is invalid, use the L1 score.
- If L1 data indicate an area is still a weakness, then original L2 (English) score is valid, use the L2 (English) score.

Step 4. Cross-validate L1 and L2 scores with contextual factors and pre-referral concerns for ecological validity

- Use information from detailed history of language development, formal education, parental SES, family experiences, as the context by which to evaluate consistency, explanatory logic, and ecological validity regarding data from pre-referral interventions, classroom work, achievement data, and pre-referral concerns.

Multilingual Testing (L2+L1): Step 1

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Visual-Spatial Index	*95*
Block Design	9
Visual Puzzles	9

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*94*
Coding	9
Symbol Search	8

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*
Word Reading	92
Pseudoword Decoding	98

Reading Comprehension	*76*
Reading Comprehension	76
Oral Reading Fluency	80

Written Expression	*92*
Spelling	100
Sentence Composition	86
Essay Composition	93

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

Not all composites are average and there are four areas that may suggest possible cognitive weakness. Because none of these tests are valid for English learners, however, test performance must be evaluated with the C-LIM to evaluate the impact of cultural/linguistic factors.

Multilingual Testing (L2+L1): Step 2

Step 1. Test first in English (L2) and examine test scores

- If all composites indicate normative strengths (SS > ≈ 90), scores are valid (to the extent that a disability is not likely), no further testing is necessary and poor academic performance is due to cultural/linguistic factors.
- If one or more composites indicate a normative weakness (SS < ≈ 90), enter and evaluate subtest score validity in the C-LIM.

Step 2. Evaluate impact of cultural/linguistic factors on test scores

- If all criteria are met, then cultural/linguistic factors are primary (cannot be excluded), scores are likely to be invalid BUT do indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.
- If C-LIM indicates only a contributory or minimal influence of cultural/linguistic factors, then test scores are likely to be valid and evaluation should continue.

Step 3. Re-evaluate areas of weakness only in native language (L1) cross-linguistic confirmation of validity

- If L1 data indicate an area is a strength (i.e., average), then original L2 (English) score is invalid, use the L1 score.
- If L1 data indicate an area is still a weakness, then original L2 (English) score is valid, use the L2 (English) score.

Step 4. Cross-validate L1 and L2 scores with contextual factors and pre-referral concerns for ecological validity

- Use information from detailed history of language development, formal education, parental SES, family experiences, as the context by which to evaluate consistency, explanatory logic, and ecological validity regarding data from pre-referral interventions, classroom work, achievement data, and pre-referral concerns.

Clear ALL Data in Matrix

Culture-Language Interpretive Matrix - Basic Version 4.0

Clear Unused Subtests

WISC-V | WAIS-IV | WJ IV COG | DAS-II | CAS2 | RIAS-2 | LEITER-3 | CELF-5 | TAPS-3 | WRAML2 | NEPSY-II | Bateria III*
WPPSI-IV | KABC-II | WJ IV OL | SB5 | WNV | UNIT-2 | CELF-Pre2 | CASL-2 | CTOPP-2 | WMS-IV | D-KEFS | WISC Spanish*

Culture-Language Interpretive Matrix (Basic v4.0) - Analyzer and Data Entry

Print C-LIM Matrix

Name: _____ Age: _____ Grade: _____ Date: _____

DEGREE OF LINGUISTIC DEMAND

LOW	Score		MODERATE	Score		HIGH	Score	
WISC-V Matrix Reasoning	7	85	WISC-V Block Design	9	95	WISC-V Digit Span	5	75
WISC-V Visual Puzzles	9	95	WISC-V Coding	9	95			
			WISC-V Picture Span	7	85			
			WISC-V Symbol Search	8	90			
Cell Average =	90		Cell Average =	91		Cell Average =	75	
	Score			Score			Score	
			WISC-V Figure Weights	7	85	WJ IV COG Phonological Processing	99	99
			WJ IV COG Nonword Repetition	84	84			
			WJ IV COG Visual-Auditory Learning	75	75			
Cell Average =			Cell Average =	81		Cell Average =	99	
	Score			Score			Score	
						WISC-V Similarities	5	75
						WISC-V Vocabulary	6	80
						WJ IV COG Story Recall	79	79
Cell Average =			Cell Average =			Cell Average =	78	

Click to select the core battery (WISC-V) from the drop down menu list and the C-LIM automatically populates the subtests according to their classifications.

Click to select the supplemental battery (WJ-IV) from the drop down menu list and enter the subtest scores.

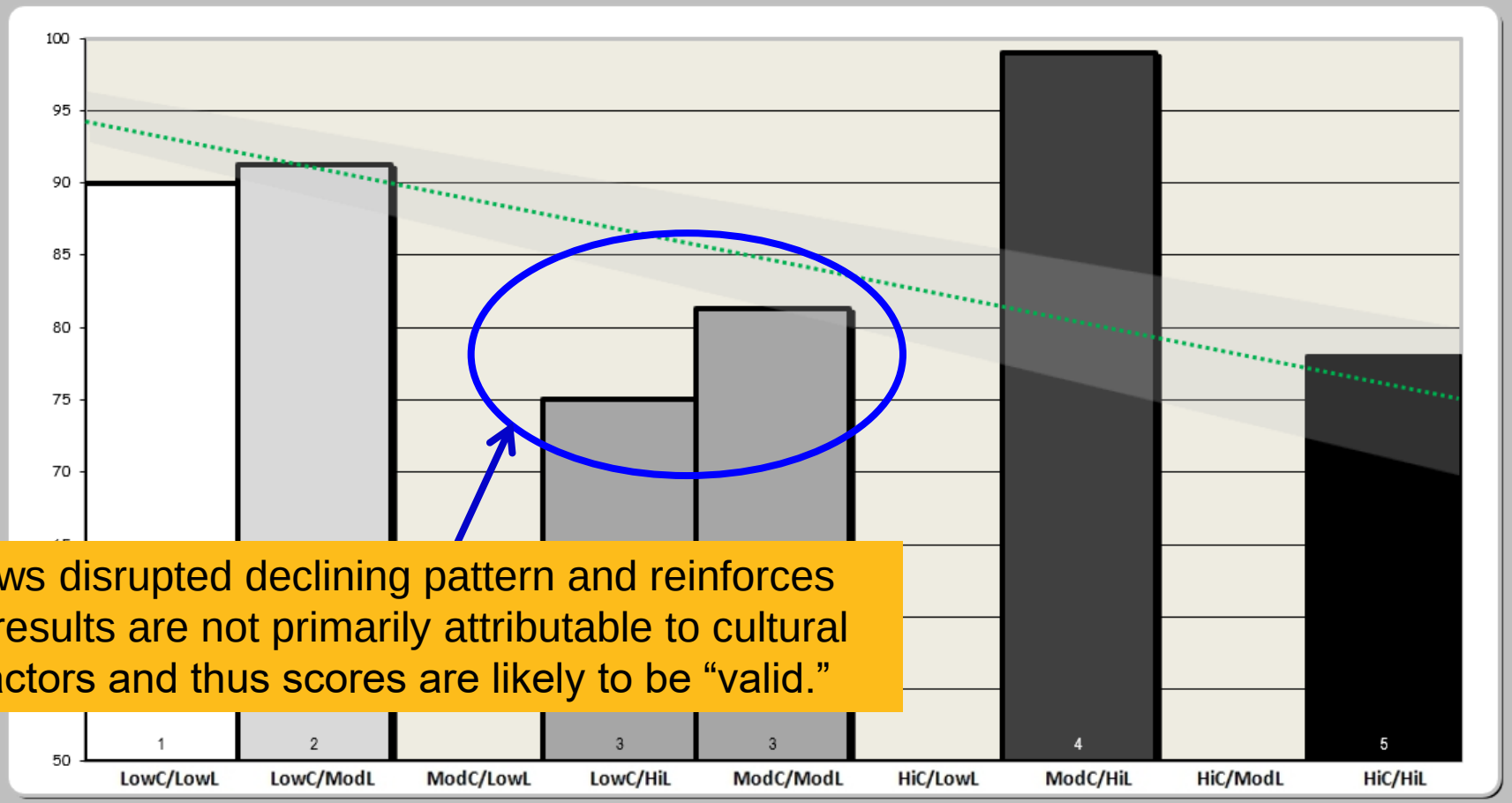
Click to delete subtests for which no scores have been entered.

Multilingual Testing (L2+L1): Step 2

Name: _____ Age: _____ Grade: _____ Date: _____

C-LIM Summary Graph: Evaluation of Cultural/Linguistic Influences

DEGREE OF DIFFERENCE FOR EVALUATION: ○ Slightly Different ◉ Moderately Different ○ Markedly Different



Graph also shows disrupted declining pattern and reinforces conclusion that results are not primarily attributable to cultural and linguistic factors and thus scores are likely to be “valid.”

Multilingual Testing (L2+L1): Step 3

Step 1. Test first in English (L2) and examine test scores

- If all composites indicate normative strengths (SS > ≈ 90), scores are valid (to the extent that a disability is not likely), no further testing is necessary and poor academic performance is due to cultural/linguistic factors.
- If one or more composites indicate a normative weakness (SS < ≈ 90), enter and evaluate subtest score validity in the C-LIM.

Step 2. Evaluate impact of cultural/linguistic factors on test scores

- If all criteria are met, then cultural/linguistic factors are primary (cannot be excluded), scores are likely to be invalid BUT do indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.
- If C-LIM indicates only a contributory or minimal influence of cultural/linguistic factors, then test scores are likely to be valid and evaluation should continue.

Step 3. Re-evaluate areas of weakness only in native language (L1) cross-linguistic confirmation of validity

- If L1 data indicate an area is a strength (i.e., average), then original L2 (English) score is invalid, use the L1 score.
- If L1 data indicate an area is still a weakness, then original L2 (English) score is valid, use the L2 (English) score.

Step 4. Cross-validate L1 and L2 scores with contextual factors and pre-referral concerns for ecological validity

- Use information from detailed history of language development, formal education, parental SES, family experiences, as the context by which to evaluate consistency, explanatory logic, and ecological validity regarding data from pre-referral interventions, classroom work, achievement data, and pre-referral concerns.

Multilingual Testing (L2+L1): Step 3

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Visual-Spatial Index	*95*
Block Design	9
Visual Puzzles	9

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*94*
Coding	9
Symbol Search	8

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*
Word Reading	92
Pseudoword Decoding	98

Reading Comprehension	*76*
Reading Comprehension	76
Oral Reading Fluency	80

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

There are four possible areas of cognitive weakness that may suggest deficits related to the reported academic difficulties as well as three areas of strength. However, because these tests are not designed for English learners, for the areas of suspected weakness it is necessary to generate additional information and data to cross-linguistically confirm that they are true deficits. Strengths do not support disability identification and therefore do not require any further validation.

Multilingual Testing (L2+L1): Step 3

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Visual-Spatial Index	*95*
Block Design	9
Visual Puzzles	9

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*84*
Coding	9
Symbol Search	8

In addition, because Gc itself is "language," it cannot be compared fairly to native English speaker norms to determine whether it is a strength or weakness even when scores are deemed "valid" using the C-LIM. Thus, in the case, additional procedures must be employed to determine whether Gc is actually a true weakness or not and whether it does or does not require re-evaluation.

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*
Word Reading	92
Pseudoword Decoding	98

Reading Comprehension	*76*
Reading Comprehension	76
Oral Reading Fluency	80

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

Basic Disability Evaluation with an English Learner: A Case Study – Step 2

Determining if and when to re-test Gc via the C-LIM

Re-evaluation of suspected areas of weakness is necessary to provide cross-linguistic confirmation of potential deficits in functioning. A disability cannot be identified in an English learner if the observed difficulties occur only in one language. Even then, deficits that are identified in both languages are not definitive evidence of dysfunction and evaluation of expectations for native language performance is as relevant for native language evaluation as it is for evaluation in English.

Because of the nature of Gc, it should be treated slightly differently when it comes to re-evaluation as compared to other cognitive abilities. The following guidelines from the best practice recommendations apply specifically to Gc:

- *Review results from testing in English and identify domains of suspected weakness or difficulty:
    a. For Gc only, evaluate weakness according to high/high cell in C-LIM or in context of other data and information
- *For Gc only:
    a. If high/high cell in C-LIM is within/above expected range, consider Gc a strength and assume it is at least average (re-testing is not necessary)
    b. If high/high cell in C-LIM is below expected range, re-testing of Gc in the native language is recommended
- For Gc only, scores obtained in the native language should only be interpreted relative to developmental and educational experiences of the examinee in the native language and only as compared to others with similar developmental experiences in the native language.

It is important that the actual, obtained Gc score, regardless of magnitude, be reported when required, albeit with appropriate nondiscriminatory assignment of meaning, and that it be used for the purposes of instructional planning and educational intervention.

If Gc is evaluated with the Ortiz PVAT, use the actual score obtained from the English Learner norms (NOT the English Speaker norms) to determine if it is an area of weakness. If the score indicates a weakness, it should then be further re-evaluated in the native language.

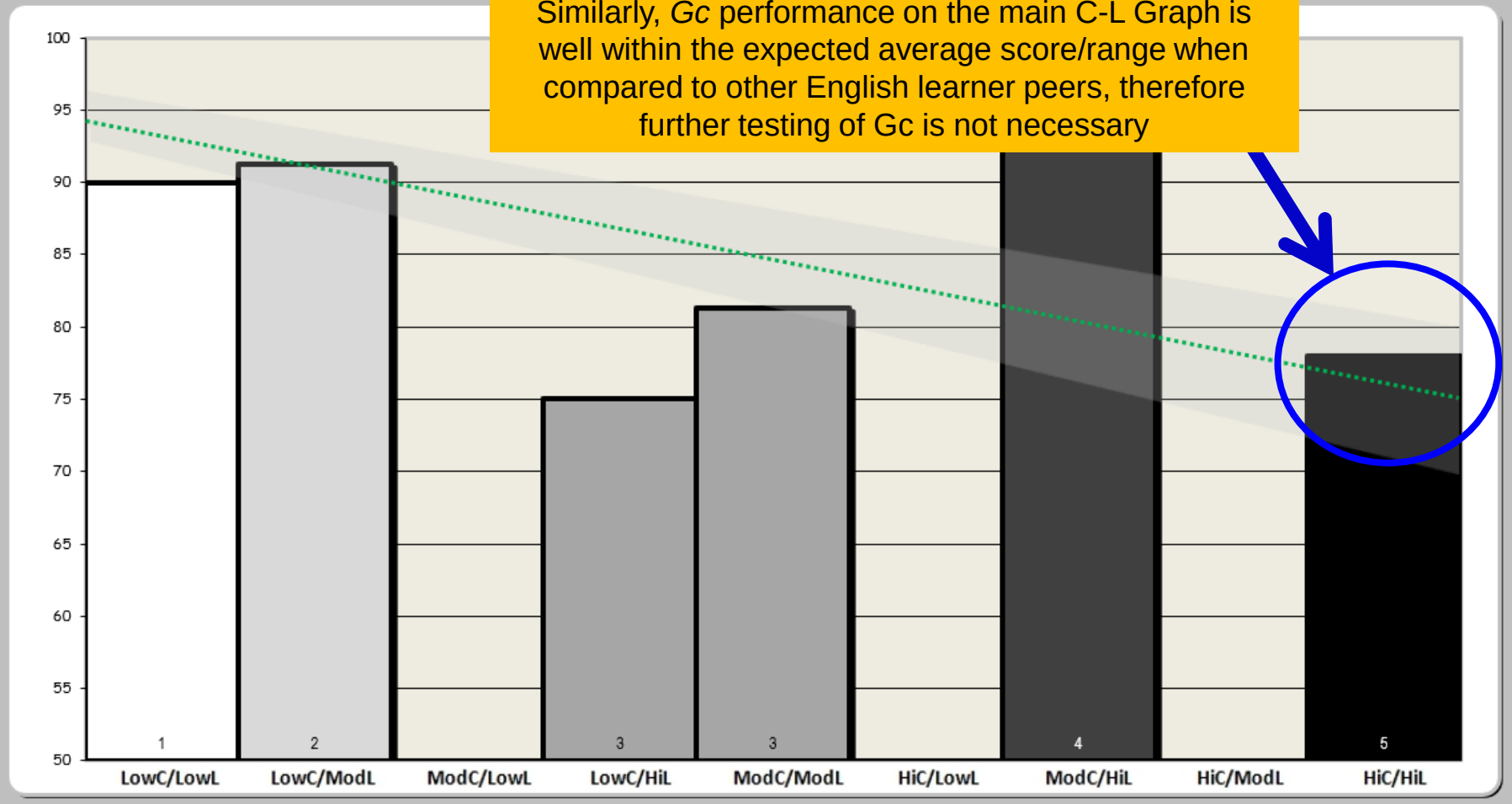
Basic Disability Evaluation with an English Learner: A Case Study

Name: _____ Age: _____ Grade: _____ Date: _____

C-LIM Summary Graph: Evaluation of Cultural/Linguistic Influences

DEGREE OF DIFFERENCE FOR EVALUATION: ○ Slightly Different ◉ Moderately Different ○ Markedly Different

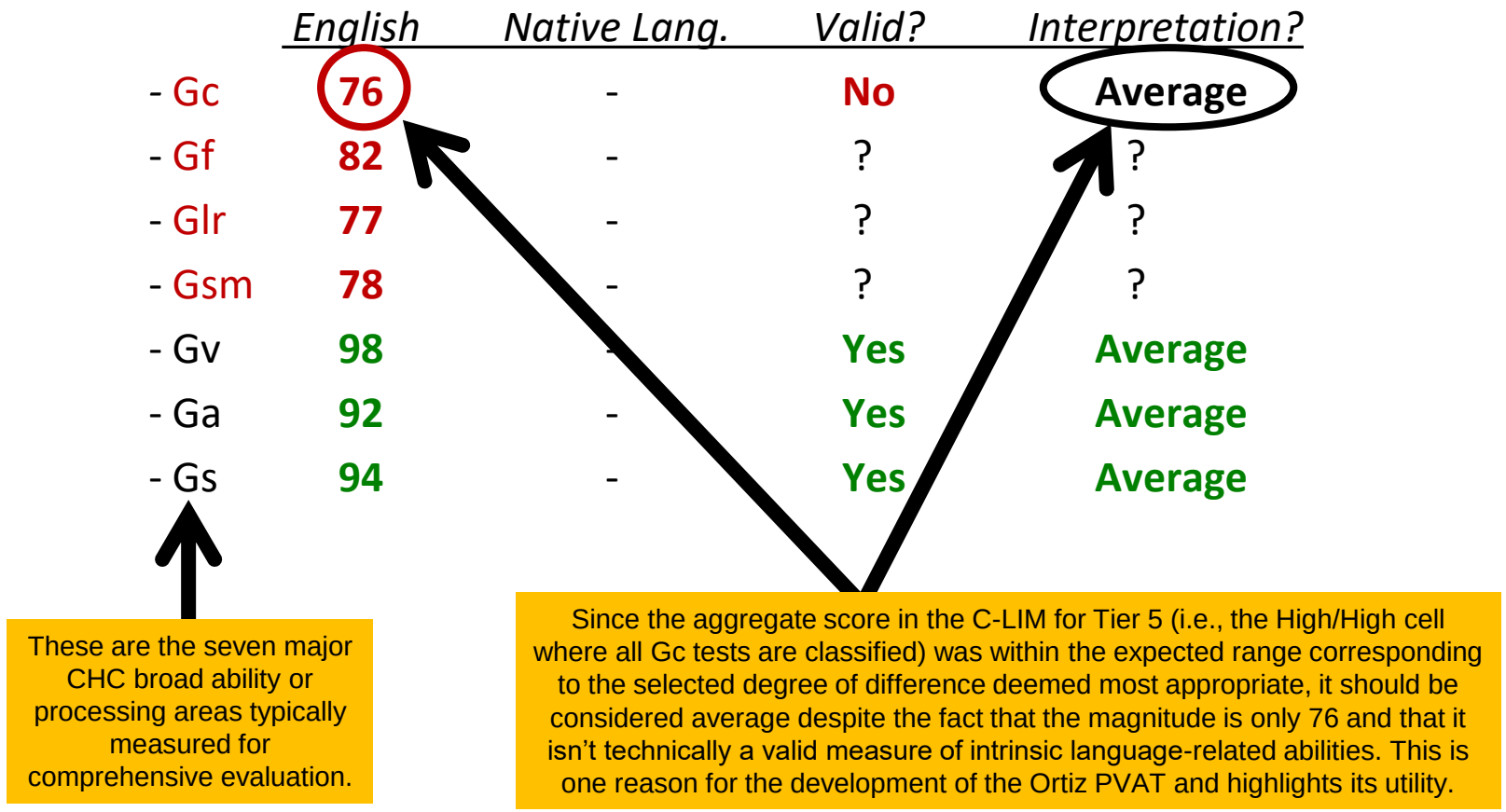
Similarly, *Gc* performance on the main C-L Graph is well within the expected average score/range when compared to other English learner peers, therefore further testing of Gc is not necessary



Basic Disability Evaluation with an English Learner: A Case Study – Step 2

Interpretive Problems with Gc Scores with English Learners

Although the C-LIM helped determine that Gc is NOT an area of weakness, further evaluation and interpretation is complicated because of the low magnitude of the score (i.e., SS=76). Other corrections are necessary to prevent discriminatory decisions, particularly in evaluation of SLD or SLI. However, use of the Ortiz PVAT provides a simple and more direct solution to all of these problems.

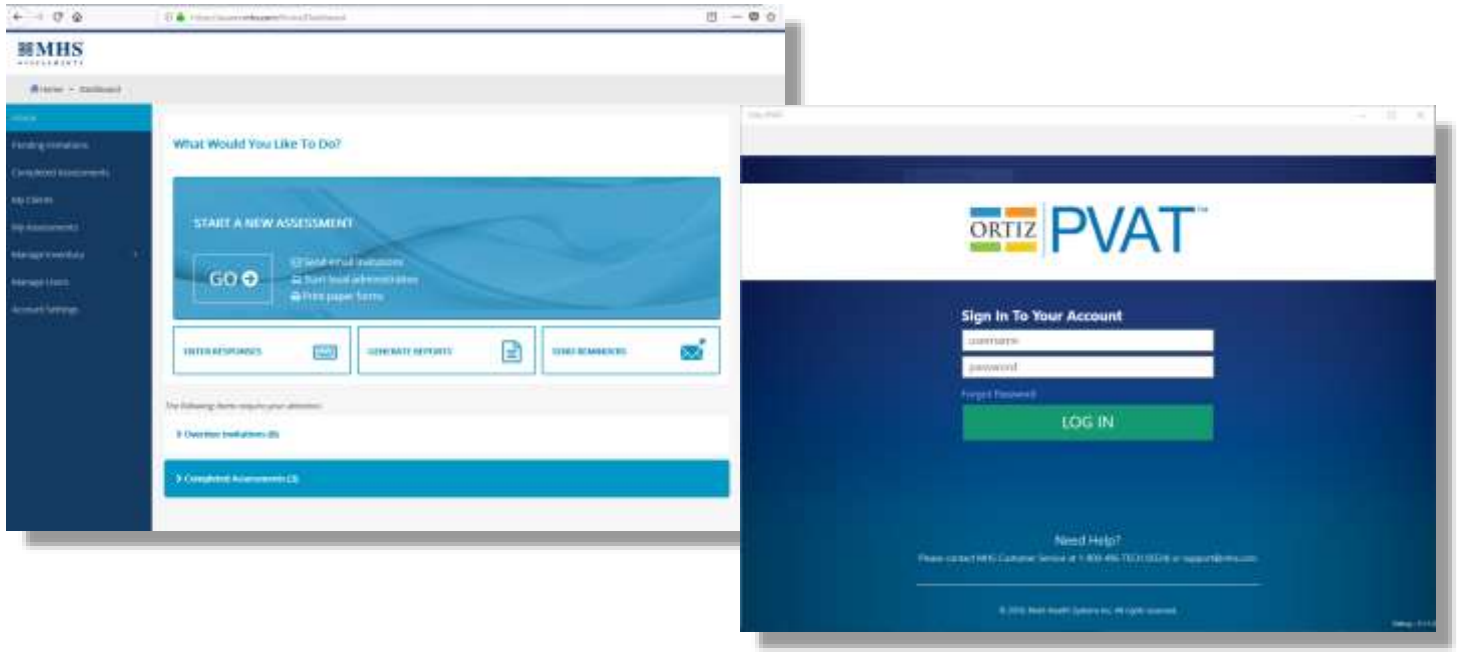


Basic Disability Evaluation with an English Learner: A Case Study – Step 2

An alternative to resolving problems with Gc scores for ELs

Use of the C-LIM to resolve for addressing test score validity with Gc is necessary due to the lack of tests that provide valid measures of language via the use of norms that control for differences among ELs relative to the language being evaluated. This leaves practitioners in the unenviable position of having to defend a low score (e.g., SS=76) as being technically invalid but which is interpreted as an area of processing "strength."

Partly in response to the difficulties posed by current limitations, a new test has been developed that yields valid Gc scores for any individual who is learning English, including as a second language and regardless of the native language or amount of time/exposure to English. That test is the **Ortiz PVAT.**



Fairness and English Learners:

Ensuring True Peer Comparability

Stratification Variables in Dual Standardization Norm Samples of the Ortiz PVAT

English Speakers (*N* = 1,530)

- Ages 2:6 to 22:11
- Gender: equal split
- Stratification:
  - Geographic region
  - Parental education level (PEL)
  - Race/ethnicity

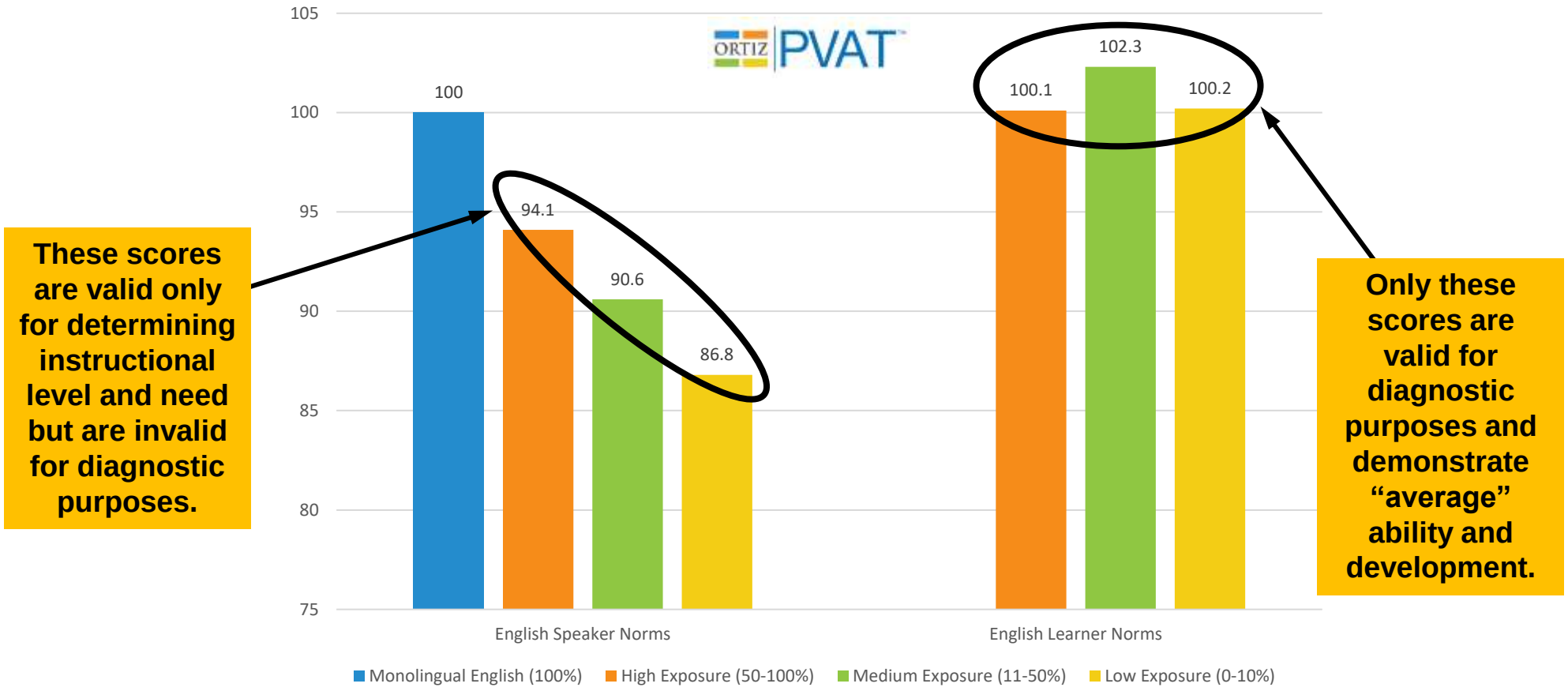
English Learners (*N* = 1,190)

- Ages 2:6 to 22:11
- Gender: equal split
- Stratification:
  - Geographic region
  - Parental education level (PEL)
  - Language spoken at home (53 different languages)
  - Proportion of lifetime exposure to English (i.e., opportunity to learn English):
    - 11 categories for length of exposure to English
    - 0-6 months up to 16+ years

Inclusion of these variables in the stratification of the EL Norm Sample is a completely unique feature of the Ortiz PVAT not found in any other test.

The Ortiz PVAT – Advances in fairness and testing

Developmental Language/Exposure-based Comparison Provides Validity and Fairness for ELs



The Ortiz PVAT – Fairness for ALL Learners

Removal of all variance due to language results in no influence of race or ethnicity

Norm sample for native English speakers demonstrates negligible effect of race/ethnicity.

Form	Racial/Ethnic Group	*N*	*M*	*SD*	*F (df)*	*p*	Pairwise Comparisons ($p < .01$)	Partial η^2
Form A	Black	280	99.4	15.2	2.60 (3, 1523)	.051	*ns*	.005
	Hispanic	126	99.5	15.4				
	White	1,018	100.5	15.3				
	Other	106	96.3	15.3				
Form B	Black	280	99.6	15.1	2.47 (3, 1523)	.060	*ns*	.005
	Hispanic	126	99.7	15.3				
	White	1,018	100.6	15.2				
	Other	106	96.4	15.2				

The Ortiz PVAT – Fairness for ALL English Learners

First language learned (L1) does not alter the sequence of learning English (L2)

English language acquisition is an invariant process, irrespective of the native language

Form	Language Spoken	*N*	*M*	*SD*	*F* (*df*)	*p*	Pairwise Comparisons ($p < .01$)	Partial η^2
Form A	Spanish & Spanish Creole	872	101.5	15.5	1.63 (3, 1183)	.181	*ns*	.004
	Indo-European Languages	161	99.4	15.7				
	Asian & Pacific Islander Languages	129	98.8	15.4				
	All Other Languages	28	99.9	15.4				
Form B	Spanish & Spanish Creole	872	101.7	15.5	1.52 (3, 1183)	.208	*ns*	.004
	Indo-European Languages	161	99.8	15.7				
	Asian & Pacific Islander Languages	129	99.0	15.4				
	All Other Languages	28	99.9	15.4				

The Ortiz PVAT – Applications

Pre-school Screening and Evaluation – dual norms permit evaluation of basic language development (receptive vocabulary) in very young children (minimum age: 2 years, 6 months) in both native English speakers and English learners prior to the beginning of formal instruction.

Progress Monitoring of English Language Proficiency – many tests, for example those used to monitor compliance with Title III ELA requirements are not well designed for that purpose and give misleading results regarding progress and growth and no information relative to the acquisition of BICS vs. CALP.

Determination of Instructional Level – the Assessment Report indicates the linguistically appropriate level of instruction and the degree of intensity required to assist the student in making progress toward grade-level standards and expectations. Specific instructional strategies are also provided.

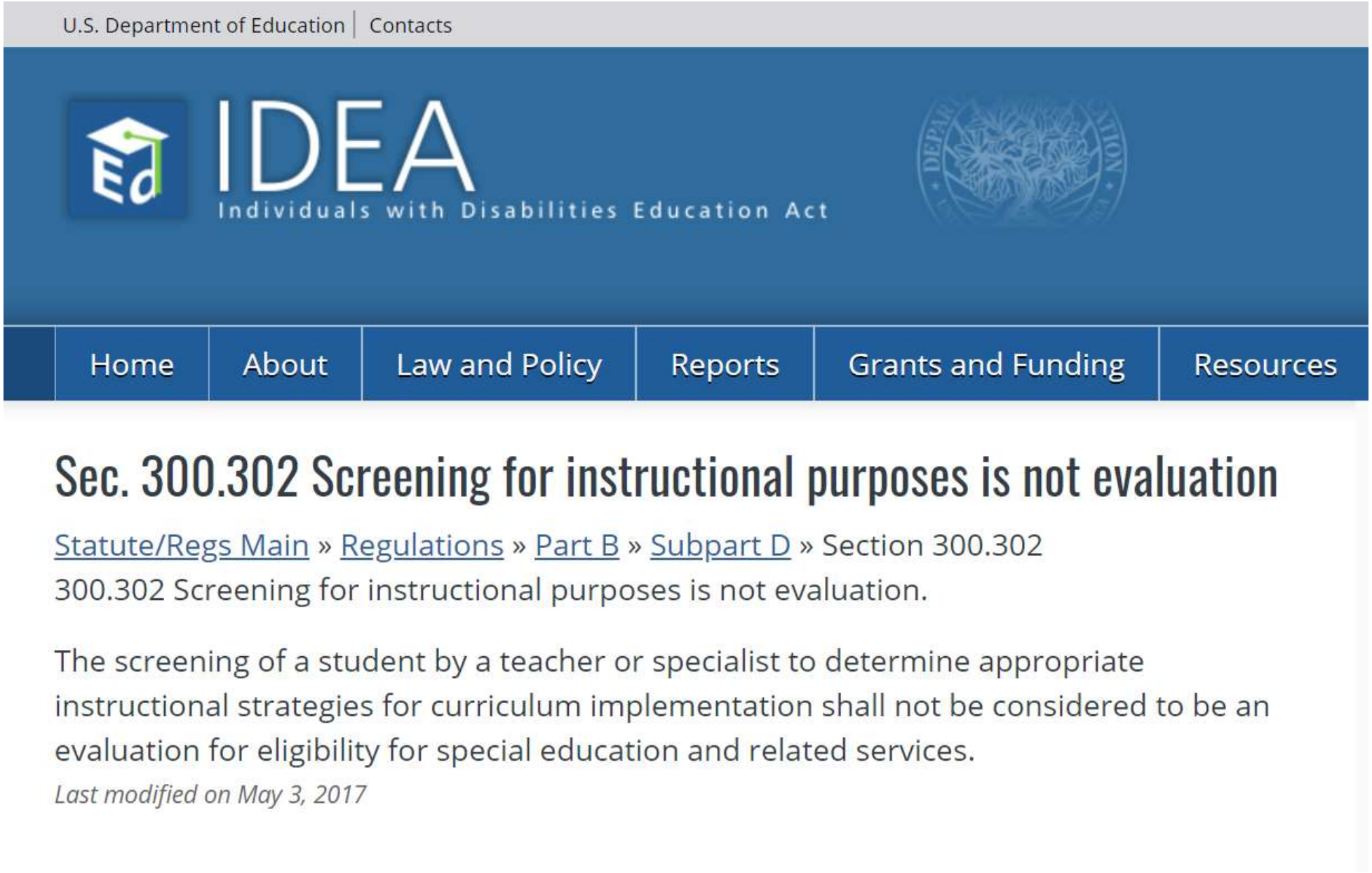
Progress monitoring of Reading and Writing Vocabulary – the Progress Report provides data for evaluating increases in receptive vocabulary that may reflect relative progress in response to specific interventions that are being employed.

Evaluation of Growth in General Language Ability – unlike tests that do not allow measurement of growth, a specific index documenting actual growth in English vocabulary/language acquisition across short and long intervals is provided.

Development of Intervention/Treatment Strategies – performance is linked directly to specific and customized recommendations for language-based intervention and treatment strategies relative to true peers.

Diagnostic and Disability Evaluation – provides the only norm-referenced “true peer” comparison necessary for evaluating “difference vs. disorder” in general language-related disabilities/disorders related to vocabulary acquisition.

The Ortiz PVAT – Pre-referral Applications



The Ortiz PVAT – Diagnostic Applications

ORTIZ PVAT™

ORTIZ PICTURE VOCABULARY ACQUISITION TEST™

Samuel O. Ortiz, Ph.D.

Assessment Report

Examinee Information

Name/ID:	**Maria Ayala / Case Study**
Age:	9 years 8 months
Gender:	Female
Date of Birth:	Unspecified
Language(s) Spoken at Home:	Spanish and English
Age at First Exposure to English:	5 years
Exposure to English:	44% of life
Primary Language of Instruction:	English
School Grade:	4

Assessment Information

Administration Date:	April 14, 2021
Examiner Name:	Maria's Examiner
Form Administered:	A
Norms Used:	**English Learner Norms (accounting for exposure to English)**
Number of Items Presented:	63
Number of Items Omitted:	0

This computerized report provides quantitative information about the performance of the examinee. Additional interpretive information can be found in the *Ortiz PVAT Technical Manual*. **This Assessment Report is intended for use by qualified evaluators only, and is not to be used as the sole basis for clinical diagnosis or intervention.**

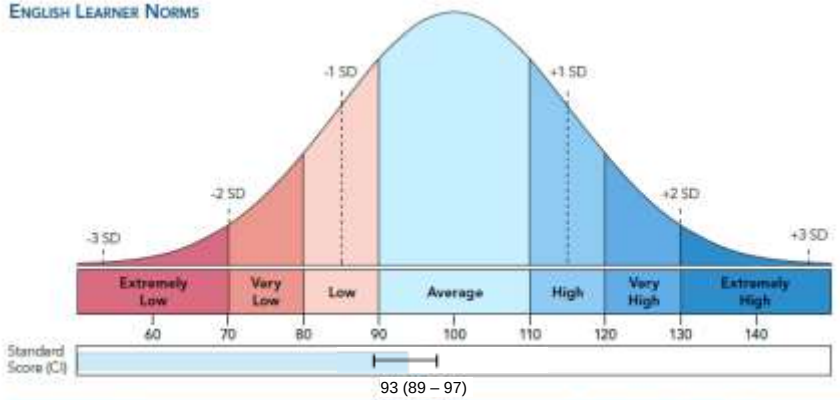
About the Ortiz PVAT™

The Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT) is a test that assesses the ability of a child, youth, or young adult (aged 2 years 6 months to 22 years 11 months) to comprehend the meaning of spoken English words (i.e., receptive vocabulary). It is appropriate for both native English speakers and English learners. In addition, it can be used to measure and track growth and development in English vocabulary, investigate possible speech-language difficulties, and guide instruction and educational intervention.

Vocabulary Acquisition and Development

This section of the report compares the examinee's scores against the **English Learner norms** to assess vocabulary acquisition in English relative to other English learners of the same age who have similar exposure to English. This comparison assists in differentiating the normal process of learning another language from an underlying language disorder. Please see the *Ortiz PVAT Technical Manual* for more information on the importance of using English Learner norms that account for exposure to English.

ENGLISH LEARNER NORMS



Ortiz PVAT Scores	English Learner Norms*
Raw Score	75
Standard Score (95% Confidence Interval)	93 (89 – 97)
Percentile	32nd
Stanine	4
Age Equivalent (Years:Months)	9:9
Classification	Average

*Compared to other English learners of the same age who have similar exposure to English.

Interpretation

- Compared to same-age peers who have been exposed to English for 44% of their lives (**English Learner norms**), Maria's ability to recognize spoken English words is **average**.
- Her performance is comparable to that of other English learners with similar exposure to English, which may indicate a normal process of learning another language and **does not suggest the presence of any underlying difficulties in language acquisition**.

The Ortiz PVAT – Diagnostic Applications

Instructional Level

This section of the report compares the examinee's scores against the **English Speaker norms** to assess instructional needs. A comparison to native English speakers provides a baseline of current functioning relative to peers from the same age group and can be used to inform instructional level or services required. Please see the *Ortiz PVAT Technical Manual* for more information on the use of English Speaker norms for English learners.

Instructional level recommendations:

- With respect to the level of instruction required for continued academic growth and success in English, Maria's vocabulary acquisition is **below** the level typically associated with same-age native English-speaking peers (**English Speaker norms**).
- **Individualized classroom instruction and strategies are needed** to support her level of English comprehension. **Interventions are also needed** to assist in making further progress toward grade-level standards in English.

Important Note: English-language experiences should not be viewed as a replacement or substitute for continued native-language development which may offer better educational outcomes for Maria, in both English and her native language.

Recommendations For Further Vocabulary Growth (English and Native Language)

Instructional strategies for English language development:

- Provide significant opportunities to use and express content vocabulary in the English language:
  - Continue to facilitate language learning through speech production and interaction so that the learning process remains active, not passive. Support and encourage active participation rather than just presenting information.
  - Create interactive educational settings where there is continued exposure to age/grade level-appropriate English language models for advanced social and academic conversations. Such interactive environments should focus on providing increased opportunities for using language along with corrective feedback that is appropriate for the student's current vocabulary level.
- Increase contextualization of information:
  - Use clear, consistent relational language (i.e., descriptions and analogies that illustrate similarities) for objects, key vocabulary words, and ideas, especially when introducing new or more complex ways of using social and academic language in English and the native language (if the student speaks their native language).
  - Provide continued opportunities for scaffolding, focusing on both social and academic language acquisition in English via the use of rich, visual imagery and varying the amount of context (e.g., hand gestures, pointing to surrounding objects) in order to aid comprehension. This technique may include requiring the student to access information that they were taught previously.
  - Provide continued opportunities for drawing (especially for younger students), writing, and expressions in the English language in order to connect the student's own ideas to both social and academic settings.
- Use visual aids and graphic organizers (e.g., picture dictionaries, icons, or flowcharts) during instruction to tag and connect vocabulary and ideas.
- Allow the student to incorporate their own experiences into learning situations.
- Probe the student to summarize and synthesize information by using their own words in English.
- Provide continued opportunities for the student to connect the English language with ideas or concepts within the context of academic and social settings.
- Provide continued opportunities for the student to read aloud in English in order to practice effective language use and appropriate expression.
- Encourage the student to create picture dictionaries using illustrations and images to support semantic language development and acquisition of English vocabulary.

ORTIZ PVAT
Assessment Report – Form A for Maria Ayala / Case Study
Admin Date: 04/14/2021

Recommendations For Further Vocabulary Growth (English and Native Language) *continued...*

- Categorize words by concept or by similar features to develop a network of connections.
- Teach the student to monitor their understanding and ask questions while reading.
- Present the student with illustrations of people engaged in various activities. Ask the student to dictate a story in English about the people and watch while they transcribe it in front of you.
- Ask both literal and inferential questions (e.g., who, what, where, when, why, and how) about a picture or idea to collaboratively build a structured definition or mental concept. Have the student respond in English as much as possible.
- Create stories and promote both social and academic language by encouraging the student's use of story maps and illustrations as well as written/spoken English.
- Employ the Whole-Part-Whole instructional model (Roseberry-McKibbin, 2008):
  - Provide the student with the "big picture" and support semantic relationships or overlaps with the parts.
  - Opt for small group instruction; encourage students to interact with one another during learning experiences instead of passive listening.
- Use thematic instructional units focusing on topics relevant to the student's interests that are aligned with the classroom curriculum.
- Provide the student with a moderate- to high-level scope of a concept and help them identify the parts and relationships within the larger concept.
- Provide interactive settings with significant exposure to social and academic English vocabulary and opportunity for feedback.
- Introduce intermediate- to advanced-level English synonyms and antonyms to deepen understanding of new social and academic vocabulary words.
- Encourage curiosity to discover and explore new English words and definitions.

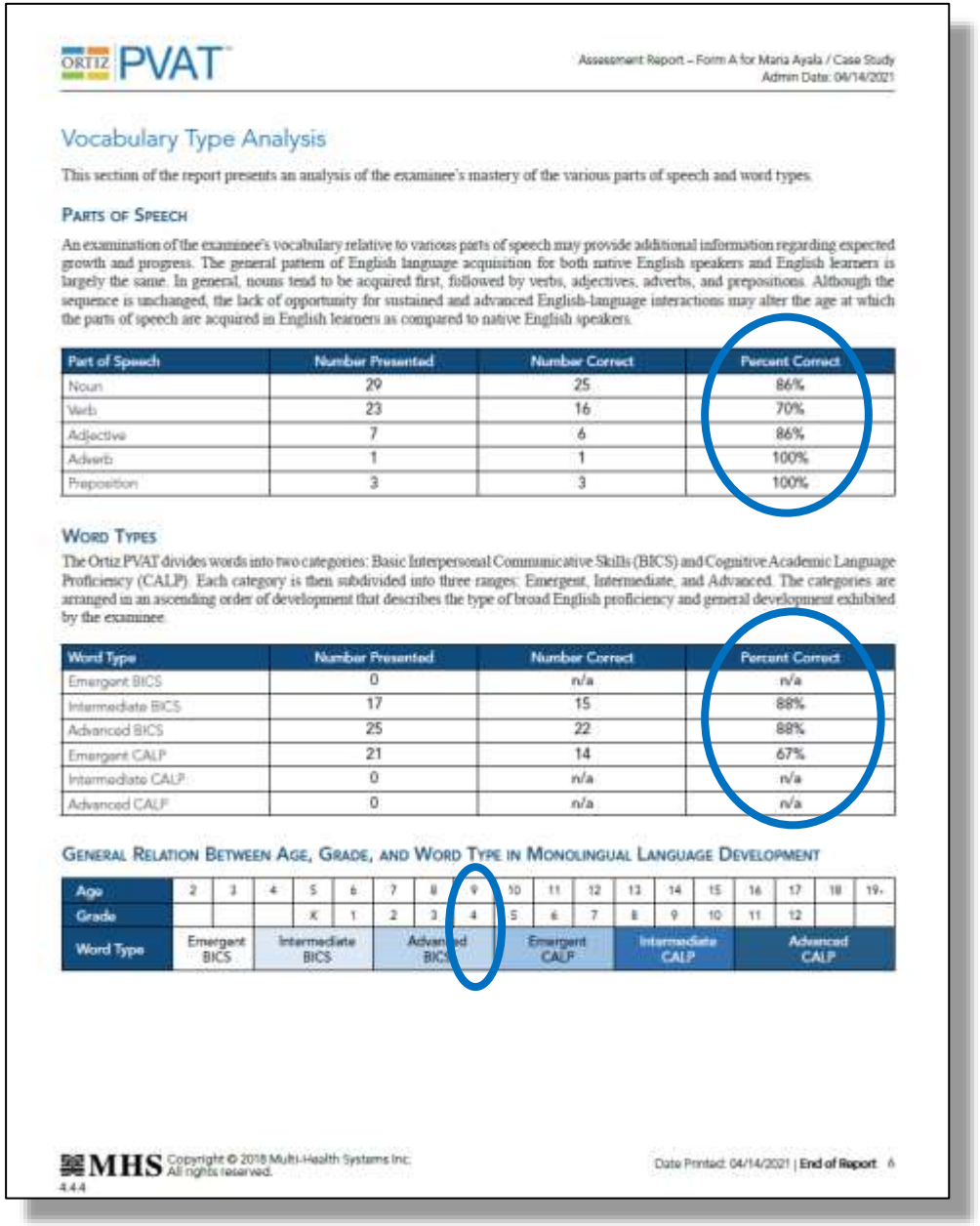
Practical strategies for intervention (if desired):

- Focus on teaching concepts or strategies that help the student learn new academic English words.
- Expand the student's academic English vocabulary within familiar and naturalistic contexts. In other words, create authentic communication situations that focus on meaning rather than just the structure of language.
- Provide continued English-speaking models for social and academic language communication.
- Encourage active peer-to-peer communication in English (e.g., discussions, acting out a concept or scenario) to balance passive listening to oral information (e.g., rote lesson plans, videos).
- Frequently repeat and review newly acquired English vocabulary words.
- Relate new words to previously learned words to accelerate acquisition. Help students relate or connect new information to what they already know.
- Carefully listen and respond to the student's communication attempts.

Instructional strategies in the native language to assist in English language development:

- Read aloud to the student developmentally appropriate passages from bilingual books, if they understand their native language and if there are native language models available at home.
- Create opportunities to support cross-linguistic learning from the student's native language to English. For example, "mama" sounds fairly similar across many languages. For languages that share roots with English, this can be accomplished through cognates (i.e., words from different languages that share both form and meaning [e.g., rose/rosa in Spanish; lamp/lampa in Polish; alligator/alligatore in Italian; eat/essen in German]).
- Foster exposure to the student's native language through media (e.g., books, songs, television, or movies) and conversations with native language models (e.g., family, friends, and relatives).

The Ortiz PVAT – Diagnostic Applications



ORTIZ PVAT

Assessment Report – Form A for Maria Ayala / Case Study
Admin Date: 04/14/2021

Vocabulary Type Analysis

This section of the report presents an analysis of the examinee's mastery of the various parts of speech and word types.

Parts of Speech

An examination of the examinee's vocabulary relative to various parts of speech may provide additional information regarding expected growth and progress. The general pattern of English language acquisition for both native English speakers and English learners is largely the same. In general, nouns tend to be acquired first, followed by verbs, adjectives, adverbs, and prepositions. Although the sequence is unchanged, the lack of opportunity for sustained and advanced English-language interactions may alter the age at which the parts of speech are acquired in English learners as compared to native English speakers.

Part of Speech	Number Presented	Number Correct	Percent Correct
Noun	29	25	86%
Verb	23	16	70%
Adjective	7	6	86%
Adverb	1	1	100%
Preposition	3	3	100%

Word Types

The Ortiz PVAT divides words into two categories: Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP). Each category is then subdivided into three ranges: Emergent, Intermediate, and Advanced. The categories are arranged in an ascending order of development that describes the type of broad English proficiency and general development exhibited by the examinee.

Word Type	Number Presented	Number Correct	Percent Correct
Emergent BICS	0	n/a	n/a
Intermediate BICS	17	15	88%
Advanced BICS	25	22	88%
Emergent CALP	21	14	67%
Intermediate CALP	0	n/a	n/a
Advanced CALP	0	n/a	n/a

General Relation Between Age, Grade, and Word Type in Monolingual Language Development

Age	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19+
Grade				K	1	2	3	4	5	6	7	8	9	10	11	12		
Word Type	Emergent BICS		Intermediate BICS			Advanced BICS			Emergent CALP			Intermediate CALP			Advanced CALP			

A true peer norm sample helps establish current baseline functioning and identifies areas of possible need.

For example, by using the "Parts of Speech" information, specific **intervention goals** can be designed to help an EL improve vocabulary acquisition with respect to various parts of speech as the list is arranged by order of acquisition.

By using the "Word Types" information, additional **intervention goals** can be designed to help an EL improve vocabulary acquisition with respect to social/conversational language as well as content/subject matter words.

The Ortiz PVAT – Diagnostic Applications

ORTIZ PVAT™

Ortiz Picture Vocabulary Acquisition Test™

Samuel O. Ortiz, Ph.D.

Progress Report

Examinee Information (Based on the most recent administration)

Name/ID:	Maria Ayala / Case Study
Gender:	Female
Date of Birth:	Unspecified
Language(s) Spoken at Home:	Spanish and English[1]
Age at First Exposure to English:	5 years[1]
Primary Language of Instruction:	English[1]
Norms Used:	English Learner Norms (accounting for exposure to English)

Assessment Information

	Admin 1	Admin 2	Admin 3
Administration Date:	Apr 14, 2021	Apr 14, 2021	Apr 14, 2021
Age at Testing (Years:Months):	9:8	9:8	9:8
Grade:	4	4	4
Exposure to English:	44% of life	44% of life	44% of life
Form Administered:	A	A	A
Number of Items Presented:	35	46	63
Number of Items Omitted:	0	0	0
Examiner Name:	Maria's Examiner	Case Study	Maria's Examiner

[1]See Assessment Report for details.

About the Ortiz PVAT™

The Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT) is a test that assesses the ability of a child, youth, or young adult (aged 2 years 6 months to 22 years 11 months) to comprehend the meaning of spoken English words (i.e., receptive vocabulary). It is appropriate for both native English speakers and English learners. In addition, it can be used to measure and track growth and development in vocabulary, investigate possible speech-language difficulties, and guide instruction and educational intervention. This computerized report compares the

After 2 administrations, a Progress Report can be generated, and The Growth Index provides an indication of actual change or true growth across two or more administrations.

ORTIZ PVAT™

Vocabulary Acquisition and Development Across Administrations

Vocabulary is expected to improve with age and with increased exposure to English. The Growth Index provides an indication of the amount of growth of an examinee's receptive vocabulary ability from one administration to another. When interpreting results from repeated administrations, all reported Ortiz PVAT scores should be considered in conjunction with the Growth Index. Since the Growth Index only estimates the *amount of change* in vocabulary skills over time, an examination of the examinee's standard scores at various administrations is also required to determine their relative standing at any given point in time (i.e., performance compared to that of their same-aged peers in the **English Learner** normative sample with the same exposure to English). See the *Ortiz PVAT Technical Manual* for more information.

Ortiz PVAT Scores	Administration: Admin 1 04/14/21	Admin 2 04/14/21	Admin 3 04/14/21	Growth Index: Admin 1 to 2	Admin 2 to 3	Overall (1 to 3)
Raw Score	52	60	75			
Standard Score (95% CI)	85 (81–89)	89 (85–93)	96 (92–100)			
Age (Age Equivalent)	9:8 (7:0)	9:8 (7:9)	9:8 (9:9)			
Classification	Low	Low	Average			
Rate of Growth				As Expected	More than Expected	More than Expected

Note. CI = Confidence Interval. Age and age equivalent scores are denoted in terms of years and months (e.g., 7:6 = 7 years 6 months).
Growth Index: Much Less than Expected = -4.00 to -3.01; Less than Expected = -3.00 to -2.01; As Expected = -2.00 to 2.00; More than Expected = 2.01 to 4.00.

Instructional Level Recommendations Across Administrations

This section presents the examinee's vocabulary level compared to same-aged native English-speaking peers (using the **English Speaker** norms) at each administration to assess instructional needs or services required for academic growth and success in English.

Administration	Admin 1 04/14/21	Admin 2 04/14/21	Admin 3 04/14/21
Vocabulary Level	Well Below Average	Below Average	Below Average
Instructional Level Recommendation	Classroom instruction requires substantial modifications and intensive interventions are needed.	Classroom instruction requires modifications and interventions are needed.	Classroom instruction requires modifications and interventions are needed.

The Ortiz PVAT – Diagnostic Applications

How much of a difference does "true language peer" comparison make for diagnostic decisions?

		EL vs. EL	EL vs. ES	EL vs. SS
Grade	Age	Ortiz PVAT	WMLS-III English	WMLS-III Spanish
4	9	97	64	40
3	8	87	69	43
4	10	105	63	40
2	7	84	58	42
1	6	98	45	104
5	10	92	42	88
K	5	71	45	40
4	9	97	61	41
4	9	95	55	42
4	9	94	40	61
2	7	92	65	48
1	6	104	68	55
5	9	84	40	73
1	7	89	43	59
	Average=	92	54	56
	Percentile Rank =	30th	0.1st	0.1st
	Potential False Positive Rate =	7-21%	100%	86%

EL = English Learner
ES = English speaker
SS = Spanish speaker

WMLS-III Oral Language

Oral Comprhension

Picture Vocabulary

L1 dominance approach = 12/14 with language impairment

L2 dominance approach = 14/14 with language impairment

True peer comparison = 3/14 with language impairment*

**Without true peer comparison, false positive error rates for misidentification of ELs could be exceptionally high.*

**Of the 3 scores in the true peer comparison, two are very close to being WNL (SEM=2) and may not actually represent a disability.*

Basic Disability Evaluation with an English Learner: A Case Study – Step 2

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 6/22/2016 DOB: 10/4/2006 Grade: 4

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*94*
Coding	9
Symbol Search	8

Although we are adding the Ortiz PVAT at this point in the evaluation, it would have been easiest to simply include it as a standard part of any battery particularly because it can be administered to any individual to generate a valid Gc score, and in the case of ELs, it will also address the Gc problem that will always exist and provide that information in an interpretive summary report.

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*
Word Reading	92
Pseudoword Decoding	98

Reading Comprehension	*76*
Reading Comprehension	76
Oral Reading Fluency	80

Written Expression	*92*
Spelling	100
Sentence Composition	86
Essay Composition	93

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

Ortiz PVAT (EL Norms)	*93*

Basic Disability Evaluation with an English Learner: A Case Study – Step 2

Avoiding Interpretive Problems by Use of the Ortiz PVAT

Derivation of an Ortiz PVAT score using the English learner norms eliminates the Gc problem completely. The Ortiz PVAT score simply replaces any Gc/language-related/verbal ability score because it was derived precisely on "true peers" and therefore inherently valid in terms of both meaning/classification and actual magnitude (e.g., 90 - 109 = average).

	English	*Spanish*	*Valid?*	*Interpretation?*
- Gc	**76**	-	**No**	?
- Gf	**82**	-	?	?
- Glr	**77**	-	?	?
- Gsm	**78**	-	?	?
- Gv	**98**	-	**Yes**	**Average**
- Ga	**92**	-	**Yes**	**Average**
- Gs	**94**	-	**Yes**	**Average**
- Gc (Ortiz PVAT)	**93**	-	**Yes**	**Average**

Use of the Ortiz PVAT requires no native language confirmation since the score is derived from norms that control for amount of exposure to English and is based on a true peer comparison group for both English speakers and English learners. Therefore, it is valid and may be interpreted directly as a strength or weakness without requiring any further cross-linguistic validation. It also eliminates the potential confusion and difficulty in having to explain why a low score (e.g. 76) is a strength, not a weakness.

Multilingual Testing (L2+L1): Step 3

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Visual-Spatial Index	*95*
Block Design	9
Visual Puzzles	9

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*94*
Coding	9
Symbol Search	8

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

Gf, Gsm, and Glr need to be re-tested in the native language to provide additional confirmation that they are true weaknesses. The same or similar tests can be used and scores may be generated but the main purpose is to observe performance qualitatively in the domain to provide cross-linguistic validation of suspected difficulties.

Multilingual Testing (L2+L1): Step 3

Determining if and when to re-evaluate all other (non-Gc) abilities

Because cultural knowledge and language ability are not the primary focus in measurement of other abilities, the influence of cultural/linguistic factors can be determined via the C-LIM and scores below the expected range of performance may well be deemed to be the result of factors other than cultural knowledge or language ability. Thus, there is no limitation requiring comparison of performance to a true ELL peer group as there is with Gc. Thus, use of a test's norms and the attendant standard classification scheme is appropriate for determining areas of suspected weakness using tests administered in English for abilities other than Gc.

However, to establish validity for a low score obtained from testing in English with an ELL, native language evaluation is required. The following guidelines from the best practice recommendations apply to all abilities, including Gc—when Gc has been determined to be a weakness because it falls below the expected range of difference in the C-LIM:*

- Review results from testing in English and identify domains of suspected weakness or difficulty:
    a. For all abilities, except Gc, evaluate weakness using standard classifications (e.g., SS < 90)
- Re-test all domains of suspected weakness, including Gc when it is not within the expected range of difference in the C-LIM* using native language tests
- Administer tests in manner necessary to ensure full comprehension including use of any modifications and alterations necessary to reduce barriers to performance, while documenting approach to tasks, errors in responding, and behavior during testing, and analyze scores both quantitatively and qualitatively to confirm and validate areas as true weaknesses

**Or, if Gc was evaluated with the Ortiz PVAT, the actual score when compared to the English Learner norms (NOT the English Speaker norms) indicates that it is likely an area of weakness.*

Multilingual Testing (L2+L1): Step 3

Procedures for Follow-up Evaluation in the Native Language

When providing cross-linguistic confirmation of areas of weakness that were found via scores derived from testing in English, it is helpful (but not actually necessary) to generate scores. Qualitative information and data (e.g., process or error analysis, dynamic assessment, task observations, etc.) are equally helpful and useful with respect to confirming areas of weakness.

It is also reasonable to use the exact same tests for follow up evaluation in the native language as were initially used in English language evaluation because, in this case, practice effects are diagnostically helpful in terms of discerning "learning ability" from "learning disability."

Evaluation in the native language can be accomplished in several different ways and will likely depend on the competency of the evaluator and the available resources. Completion of the task may include one or more of the following procedures:

More defensible

1. Use of native language tests (if available) administered by a bilingual evaluator

2. Use of native language tests (if available) administered by a trained translator

---

In the absence of parallel or similar native language tests with which to evaluate the necessary domains, follow up evaluation will need to resort to other procedures for task completion, including:

3. Use of English language tests translated directly by a bilingual evaluator

4. Use of English language tests administered via assistance of trained translator

5. Use of developmental or dynamic assessment, informal tasks accompanied by careful observation, error analysis, and other probing with the assistance of a translator for communication.

Less defensible

Multilingual Testing (L2+L1): Step 3

WECHSLER INTELLIEGENCE SCALE ~~FOR CHILDREN V~~

Verbal Comprehension Index	*76*
Similarities	5
Vocabulary	6

Fluid Reasoning Index	*82*
Matrix Reasoning	7
Figure Weights	7

Visual-Spatial Index	*95*
Block Design	9
Visual Puzzles	9

Working Memory Index	*79*
Digit Span	5
Picture Span	7

Processing Speed Index	*94*
Coding	9
Symbol Search	8

WISC V Spanish	91
Matrix Reasoning	8
Figure Weights	9

WISC V Spanish WMI	*72*
Digit Span	5
Picture Span	4

WOODCOCK JOHNSON-IV TESTS ~~OF COGNITIVE ABILITY~~

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	79
Visual-Auditory Learning	75

Ortiz PVAT	*93*

Bateria IV LT Retrieval	*79*
Story Recall	78
Visual-Auditory Learning	81

Results of native language testing for Gf, Gsm, and Glr

Multilingual Testing (L2+L1): Step 3

Determining which scores are valid and interpretable

Average* or higher scores in testing are unlikely to be due to chance. Thus, when a score obtained from native language testing is found to be in the average range or higher, it serves to effectively invalidate the original low score from testing in English since deficits must exist in both languages. Conversely, if another low score in the same domain is obtained from native language evaluation, it may serve to bolster the validity of the original score obtained in English.

Based on these premises, the following guidelines from the best practice recommendations offer guidance regarding selection and use of the most appropriate and valid score for the purposes of PSW analysis (or any other situation in which the validity of test scores is central or relevant):

- For all domains, including Gc, if a score obtained in the native language suggests a domain is a strength (SS $\geq$ 90), it serves to invalidate/disconfirm the corresponding weakness score obtained in English—thus, report, use, and interpret the domain score obtained in the native language
- For all domains, except Gc, if a score obtained in the native language also suggests weakness in the same domain (SS < 90), it serves to validate/confirm the corresponding weakness score obtained in English—thus, report, use, and interpret the original domain score obtained in English
- For Gc only, if a score obtained in the native language also suggests weakness in Gc (SS < 90), it may serve to validate/confirm the corresponding weakness score obtained in English but only if low performance in Gc cannot be attributed to factors related to a lack or interruption of native language instruction and education, low family SES, or other lack of opportunity to learn—thus, in the absence of such mitigating factors, report, use, and interpret the domain score obtained in English

*Although "average or higher" (e.g., SS$\geq$90) is used as a recommended cutoff for supporting the validity of test scores, use of a lower standard (e.g., SS$\geq$85) may also represent a reasonable standard for practice since it is based on performance that can be categorized as being within normal limits.

Multilingual Testing (L2+L1): Step 3

DETERMINING STRENGTHS AND WEAKNESSES IN MULTILINGUAL EVALUATION

	Original score when tested in English	Follow up score when tested in native language	Most appropriate and valid score for use in PSW analysis		Rationale for Use as Strength or Weakness in PSW Analysis
			Original Score (in English)	Follow Up Score (in native lang)	
For ALL domains*	S	n/a	✓		Strength—scores in or above the average range (or even WNL) are unlikely to occur by chance and very likely to be valid thus re-evaluation in the native language is unnecessary
For ALL domains (and when Gc is below expected range in C-LIM)	W	S		✓	Strength—because a deficit cannot exist in one language only, the original score from testing in English is invalidated and should be replaced by the follow up average score which is likely to be valid
For ALL domains (and when Gc is below expected range in C-LIM)	W	W	✓		Weakness—low scores in both languages suggest a true deficit but additional, convergent and consistent ecological evidence is required to substantiate scores as deficits
For Gc Only (and when Gc is within the expected range in C-LIM)	S	n/a	✓		Strength—Gc can only be compared fairly to other ELLs, thus its position within the expected range in the C-LIM should be considered to be average and native language testing may not be necessary unless there is reason to believe it may be informative

*Although this table uses "average or higher" (e.g., SS≥90) as a recommended cutoff for supporting the validity of test scores, use of a lower standard (e.g., SS≥85) may also represent a reasonable standard for practice since it is based on performance that can be categorized as being within normal limits.

Multilingual Testing (L2+L1): Step 3

Determining which scores are valid and interpretable

Derivation of an Ortiz PVAT score using the English learner norms eliminates the Gc problem completely. The Ortiz PVAT score simply replaces any Gc/language-related/verbal ability score because it was derived precisely on EL "true peers" and therefore inherently valid in terms of both meaning/classification and actual magnitude (e.g., 90 - 109 = average).

	English	*Spanish*	*Valid?*	*Interpretation?*
- Gc	**76**	-	76 - **No**	-
- Gf	**(82)**	**91**	**91 - Yes**	**Average**
- Glr	**77**	**(79)**	**77 - Yes**	**Not Average**
- Gsm	**78**	**(72)**	**78 - Yes**	**Not Average**
- Gv	**98**	-	**Yes**	**Average**
- Ga	**92**	-	**Yes**	**Average**
- Gs	**94**	-	**Yes**	**Average**
- Gc (Ortiz PVAT)	**93**	-	**Yes**	**Average**

Additional native language investigation of areas of weakness noted in scores derived from testing in English (with the exception of the score from the Ortiz PVAT), resulted in an average Gf score that invalidated the original Gf score, and two below average scores that simply cross-linguistically confirmed Glr and Gsm as areas of weakness as indicated by the test scores in English.

Multilingual Testing (L2+L1): Step 4

Step 1. Test first in English (L2) and examine test scores

- If all composites indicate normative strengths (SS > ≈ 90), scores are valid (to the extent that a disability is not likely), no further testing is necessary and poor academic performance is due to cultural/linguistic factors.
- If one or more composites indicate a normative weakness (SS < ≈ 90), enter and evaluate subtest score validity in the C-LIM.

Step 2. Evaluate impact of cultural/linguistic factors on test scores

- If all criteria are met, then cultural/linguistic factors are primary (cannot be excluded), scores are likely to be invalid BUT do indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.
- If C-LIM indicates only a contributory or minimal influence of cultural/linguistic factors, then test scores are likely to be valid and evaluation should continue.

Step 3. Re-evaluate areas of weakness only in native language (L1) cross-linguistic confirmation of validity

- If L1 data indicate an area is a strength (i.e., average), then original L2 (English) score is invalid, use the L1 score.
- If L1 data indicate an area is still a weakness, then original L2 (English) score is valid, use the L2 (English) score.

Step 4. Cross-validate L1 and L2 scores with contextual factors and pre-referral concerns for ecological validity

- Use information from detailed history of language development, formal education, parental SES, family experiences, as the context by which to evaluate consistency, explanatory logic, and ecological validity regarding data from pre-referral interventions, classroom work, achievement data, and pre-referral concerns.

Multilingual Testing (L2+L1): Step 4

The Importance of Converging Evidence in Establishing Validity

Validity is based on an accumulation of evidence. The evaluation approach described herein is designed to assist in generating test scores that may be interpreted as valid indicators of an individual's abilities. Embedded in the broader framework are two basic forms of evidence that bolster the validity of obtained test scores by using expectations of test performance that are grounded in research on individuals of comparable cultural and linguistic backgrounds and the extent to which their development differs from the individuals on whom the tests were normed. Validity is thus inferred by:

> *1.Test scores from evaluation in English that have been subjected to systematic analysis of the influence of cultural and linguistic variables where such factors have been found to be either minimal or contributory but not primary factors in test performance;*
>
> *2. Test scores or qualitative data regarding evaluation of weak areas in the native language that either further confirm suspected areas of deficit as being true or dis-confirm suspected areas of deficit due to evidence of average or higher performance.*

To these, a third form of evidence needs to be added to fully support conclusions and interpretation of the test scores:

> *3. Ecological and contextual evidence regarding consistency of the test scores with ecological data and information on developmental influences (e.g., L1 and L2 exposure, language of instruction, socio-economic status, parental education level, etc.) and convergence of patterns of performance with other case data (e.g., progress monitoring data, pre-referral concerns, work samples, observations, school records, teacher/parent reports, grades, interviews, observations, etc.).*

Only when all three forms of evidence are seen to converge can there be sufficient confidence in the use and interpretation of test scores obtained in an evaluation of English learners.

Multilingual Testing (L2+L1): Step 4

Context provides an additional layer of validity

Examination and due consideration of exclusionary factors related to linguistic/cultural differences can only be established via written documentation that includes consideration the following ecological variables which collectively, comprise the context of the developmental experiences of the student:

- generational history
- language proficiency
- socio-economic status
- opportunity to learn
- academic history
- familial history
- developmental data
- work samples
- curriculum based data
- intervention results, etc.

This information is neither incidental nor merely "background" as it forms the context within which all collected data, including both qualitative and quantitative information, must be viewed so as to provide nondiscriminatory evaluation that meets the standards for fairness and supports the validity of any decisions that may be made, particularly those related to the fundamental question which is whether:

"the student's observed learning problems due primarily
to cultural or linguistic differences or disorder?"

Multilingual Testing (L2+L1): Step 4

The Importance of the Context of Difference

Subtests	Standard Score	Confidence Interval (95% Band)	Descriptions
Verbal Comprehension	64	56 – 72	Very Low
Visual-Auditory Learning	88	76 – 100	Low Average
Spatial Relations	98	91 – 107	Average
Sound Blending	75	64 – 87	Low
Concept Formation	70	62 – 78	Low
Visual Matching	86	76 – 97	Low Average
Numbers Reversed	80	67 – 93	Low
Incomplete Words	78	65 – 91	Low
Auditory Working Memory	85	76 – 94	Low Average
Analysis-Synthesis	78	66 – 90	Low
Auditory Attention	81	67 – 95	Low
Decision Speed	72	63 – 81	Low
Retrieval Fluency	82	69 – 95	Low
General Information	69	60 – 78	Very Low

Multilingual Testing (L2+L1): Step 4

The Importance of the Context of Difference

XBA Culture-Language Interpretive Matrix (XBA C-LIM v2.0) for WJ III NU COG

Name: _____ Age: _____ Grade: _____ CLEAR DATA SAVE DATA

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING

	LOW			MODERATE			HIGH		
LOW		Score			Score			Score	
	WJ III Spatial Relations	93	93	WJ III Numbers Reversed	80	80	WJ III Analysis-Synthesis	78	78
				WJ III Visual Matching	86	86	WJ III Auditory Working Memory	85	85
							WJ III Concept Formation	70	70
	Cell Average =	93		Cell Average =	83		Cell Average =	78	
MODERATE		Score			Score			Score	
	WJ III Pair Cancellation			WJ III Delayed Recall: Visual Auditory Learning			WJ III Auditory Attention	81	81
	WJ III Picture Recognition			WJ III Rapid Picture Naming			WJ III Decision Speed	72	72
	WJ III Planning			WJ III Retrieval Fluency	82	82	WJ III Incomplete Words	78	78
				WJ III Visual Auditory Learning	88	88	WJ III Memory for Words		
							WJ III Sound Blending	75	75
	Cell Average =			Cell Average =	85		Cell Average =	77	
HIGH		Score			Score			Score	
							WJ III General Information	69	69
							WJ III Verbal Comprehension	64	64
	Cell Average =			Cell Average =			Cell Average =	67	

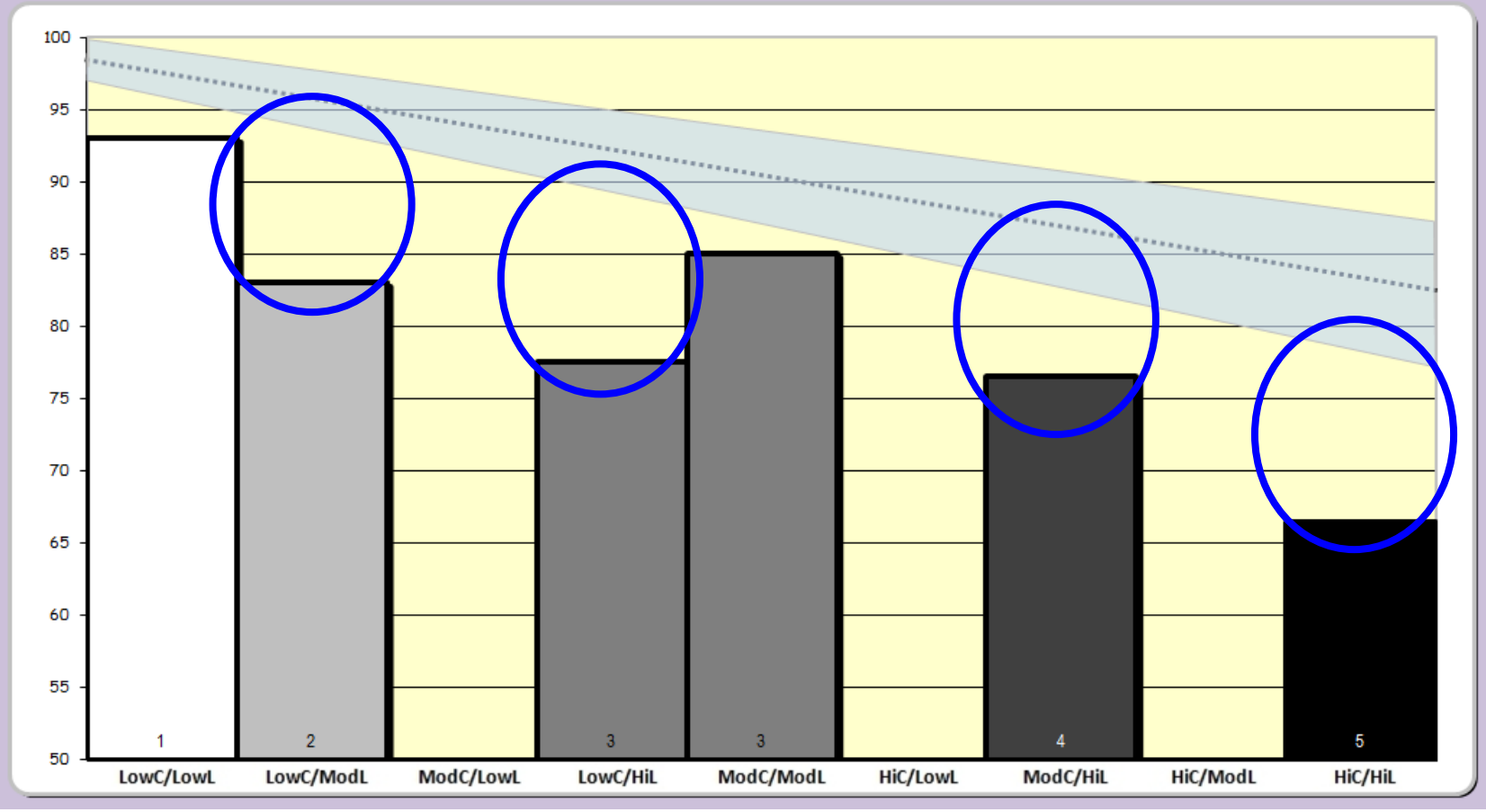
Multilingual Testing (L2+L1): Step 4

The Importance of the Context of Difference

Name: _____ Age: _____ Grade: _____

DIFFERENCE LEVEL FOR EVALUATION: (•) Slightly Different ( ) Moderately Different ( ) Markedly Different

XBA C-LIM Graph for WJ III NU COG: Primary Evaluation of Cultural and Linguistic Influences



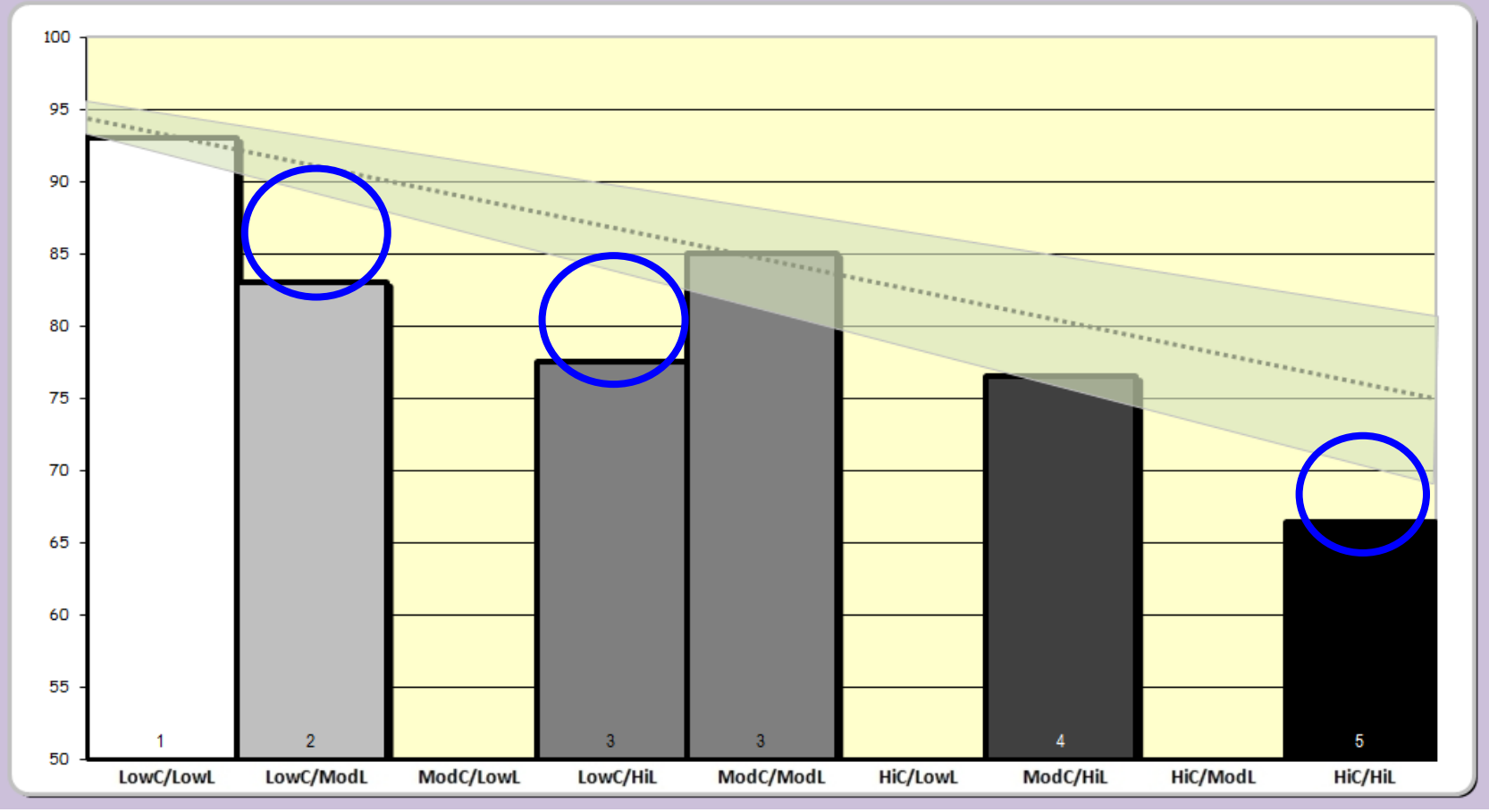
Multilingual Testing (L2+L1): Step 4

The Importance of the Context of Difference

Name: _____ Age: _____ Grade: _____

DIFFERENCE LEVEL FOR EVALUATION: ○ Slightly Different ◉ Moderately Different ○ Markedly Different

XBA C-LIM Graph for WJ III NU COG: Primary Evaluation of Cultural and Linguistic Influences



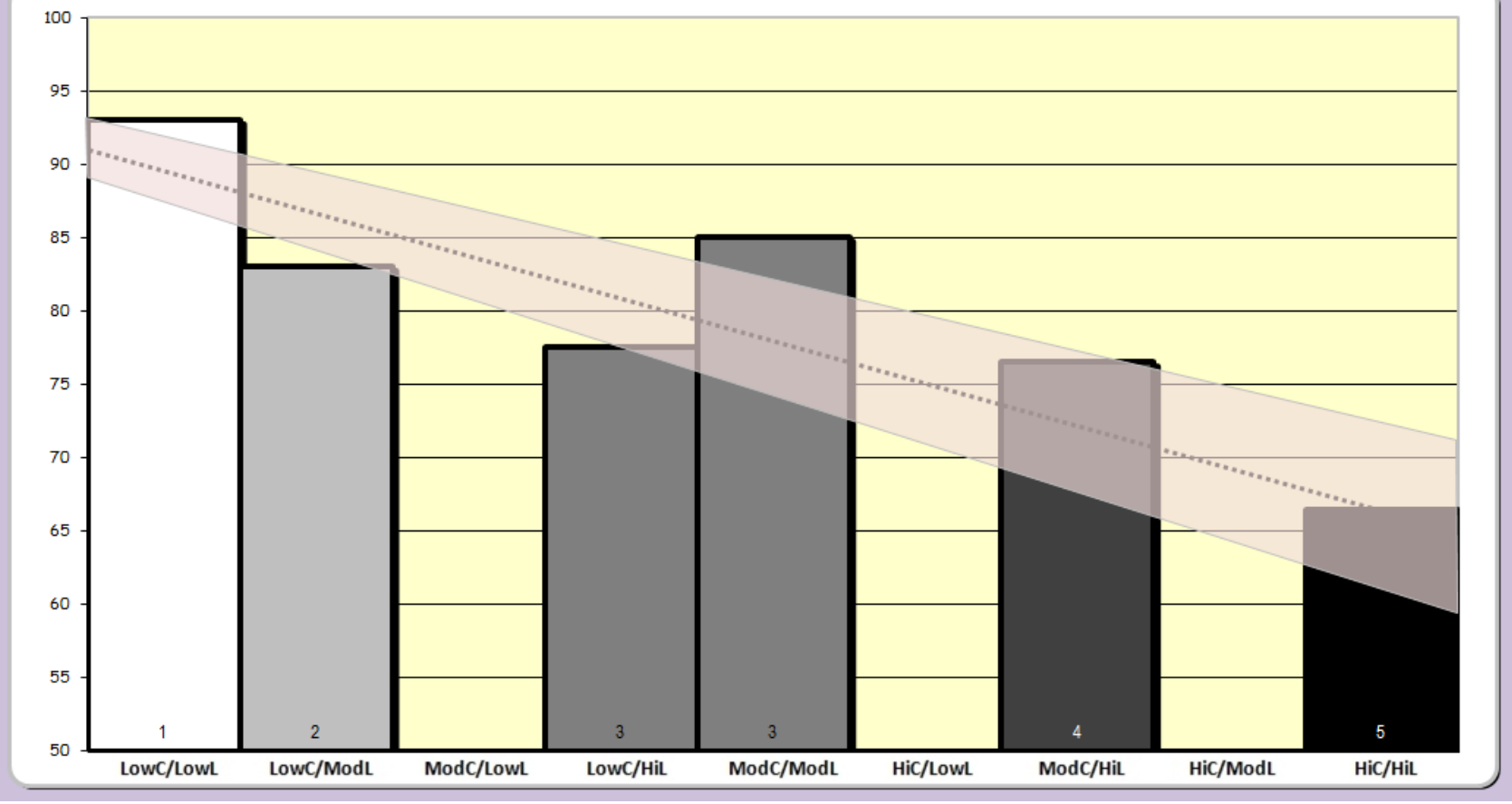
Multilingual Testing (L2+L1): Step 4

The Importance of the Context of Difference

Name: _____ Age: _____ Grade: _____

DIFFERENCE LEVEL FOR EVALUATION: ○ Slightly Different ○ Moderately Different ◉ Markedly Different

XBA C-LIM Graph for WJ III NU COG: Primary Evaluation of Cultural and Linguistic Influences



Multilingual Testing (L2+L1): Step 4

The Importance of Converging Evidence in Establishing Validity

The student's developmental history relative to culture, language, and education provide the context by which test scores acquire sufficient validity for diagnosing any condition. When test scores are consistent with the referral concerns and the student's experiences, the necessary ecological validity is established for conclusions that suggest the presence of a disability.

	English	*Spanish*	*Valid?*	*Interpretation?*
- Gc	**76**	-	76 - **No**	-
- Gf	**(82)**	**91**	**91 - Yes**	**Average**
- Glr	**77**	**(79)**	**77 - Yes**	**Not Average**
- Gsm	**78**	**(72)**	**78 - Yes**	**Not Average**
- Gv	**98**	-	**Yes**	**Average**
- Ga	**92**	-	**Yes**	**Average**
- Gs	**94**	-	**Yes**	**Average**
- Gc (Ortiz PVAT)	**93**	-	**Yes**	**Average**

To support disability identification on the basis of this pattern of test scores requires consideration of various factors including federal/state regulations and policies, the criteria for whatever approach or model is employed to establish the disability, and especially integration with other data and information that provide a valid, defensible, and consistent picture of the final determination and conclusions. ***Test scores will bolster interpretation only to the extent with which they are consistent with what else is known about the student. Failure to consider contextual factors may still lead to discriminatory interpretation and use of scores that lack sufficient validity with which to identify a disability.***

SLD Identification with an English Learner via PSW

Determination of Strengths and Weaknesses

Indicate whether the CHC domains (highlighted in blue) and neuropsychological domains (highlighted in beige) represent strengths or weaknesses for the individual. Determination of strengths and weaknesses is a judgment that is made by the evaluator based on what is known about the examinee. In general, ability and processing strengths facilitate learning and academic performance, whereas weaknesses inhibit learning and academic performance. Typically, scores that fall in the average range or higher likely facilitate learning and scores that fall below average or lower likely inhibit learning. Also, indicate whether the academic areas (highlighted in purple) represent strengths or weaknesses for the individual. Achievement standard scores that are about 90 or higher are considered strengths and scores that fall below 90 are considered weaknesses.

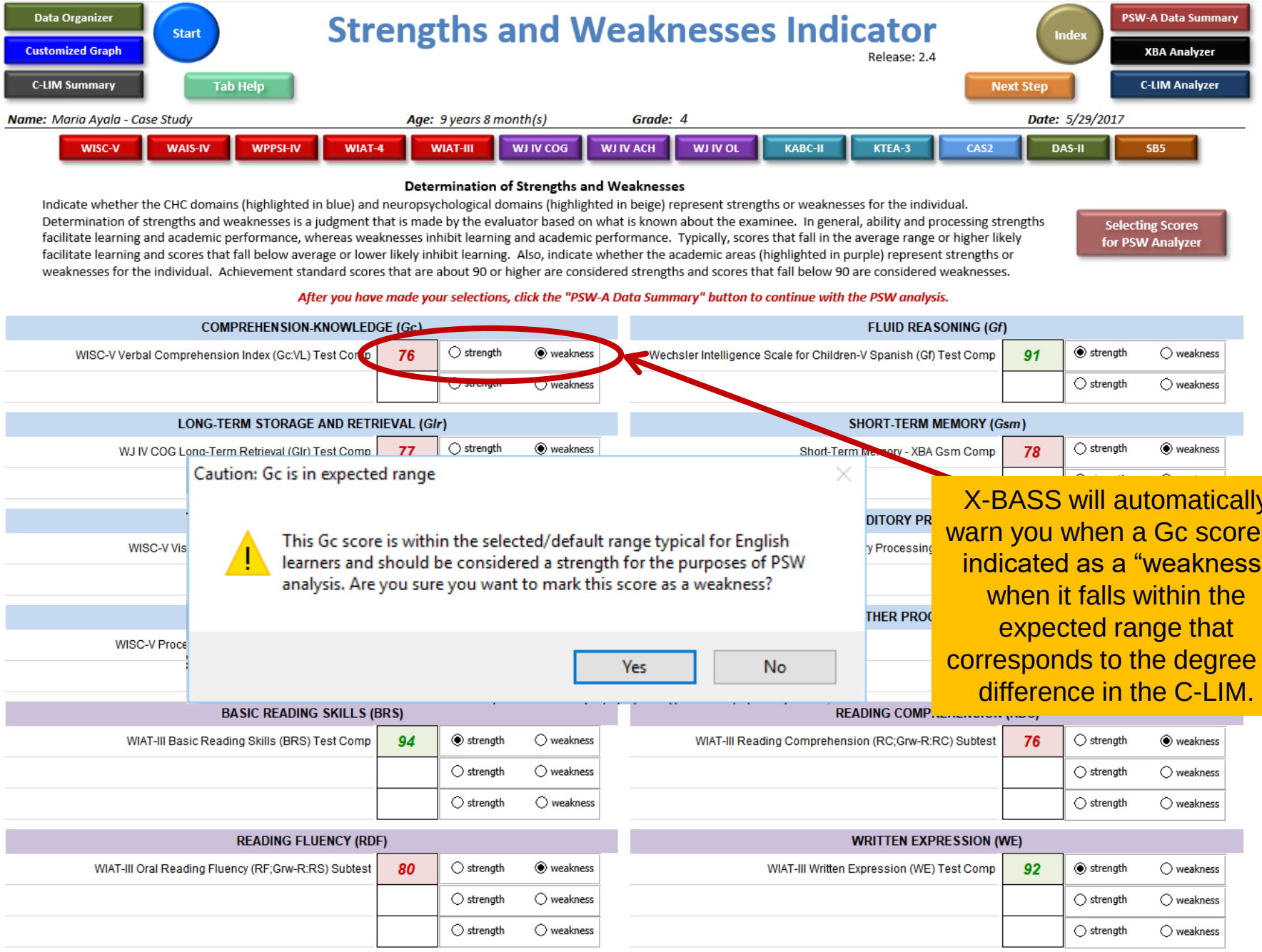
Selecting Scores for PSW Analyzer

After you have made your selections, click the "PSW-A Data Summary" button to continue with the PSW analysis.

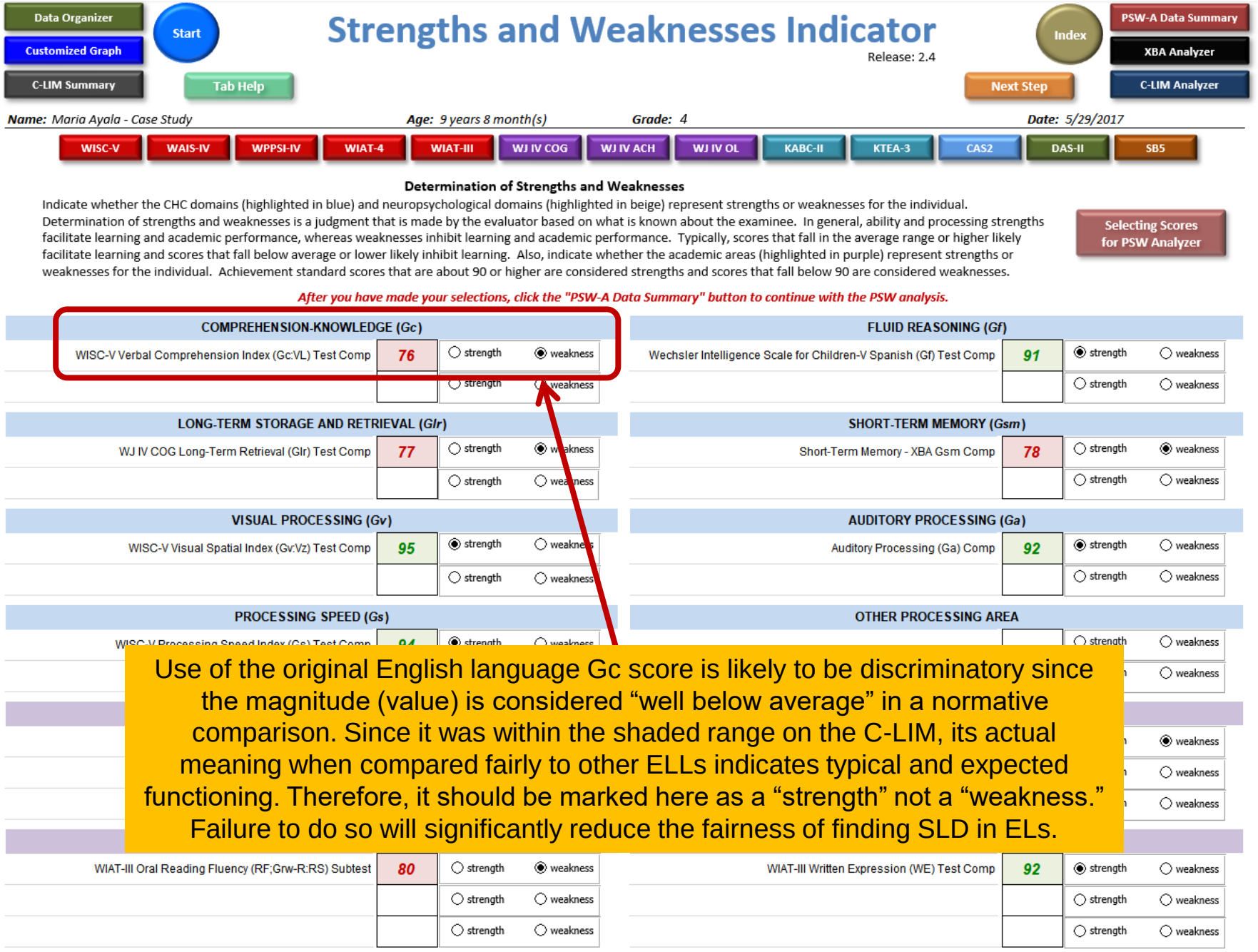
COMPREHENSION-KNOWLEDGE (*Gc*)				FLUID REASONING (*Gf*)			
WISC-V Verbal Comprehension Index (Gc:VL) Test Comp	76	○ strength	○ weakness	Wechsler Intelligence Scale for Children-V Spanish (Gf) Test Comp	91	○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
LONG-TERM STORAGE AND RETRIEVAL (*Glr*)				**SHORT-TERM MEMORY (*Gsm*)**			
V COG Long-Term Retrieval (Glr) Test Comp	77	○ strength	○ weakness	Short-Term Memory - XBA Gsm Comp	78	○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
VISUAL PROCESSING (*Gv*)				**AUDITORY PROCESSING (*Ga*)**			
SC-V Visual Spatial Index (Gv:Vz) Test Comp	95	○ strength	○ weakness	Auditory Processing (Ga) Comp	92	○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
PROCESSING SPEED (*Gs*)				**OTHER PROCESSING AREA**			
-V Processing Speed Index (Gs) Test Comp	94	○ strength	○ weakness			○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
BASIC READING SKILLS (BRS)				**READING COMPREHENSION (RDC)**			
AT-III Basic Reading Skills (BRS) Test Comp	94	○ strength	○ weakness	WIAT-III Reading Comprehension (RC;Grw-R:RC) Subtest	76	○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
READING FLUENCY (RDF)				**WRITTEN EXPRESSION (WE)**			
WIAT-III Oral Reading Fluency (RF;Grw-R:RS) Subtest	80	○ strength	○ weakness	WIAT-III Written Expression (WE) Test Comp	92	○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness
		○ strength	○ weakness			○ strength	○ weakness

Strengths and weaknesses MUST be designated by the user. X-BASS does NOT make this determination as the meaning of any given score requires more information than just its magnitude.

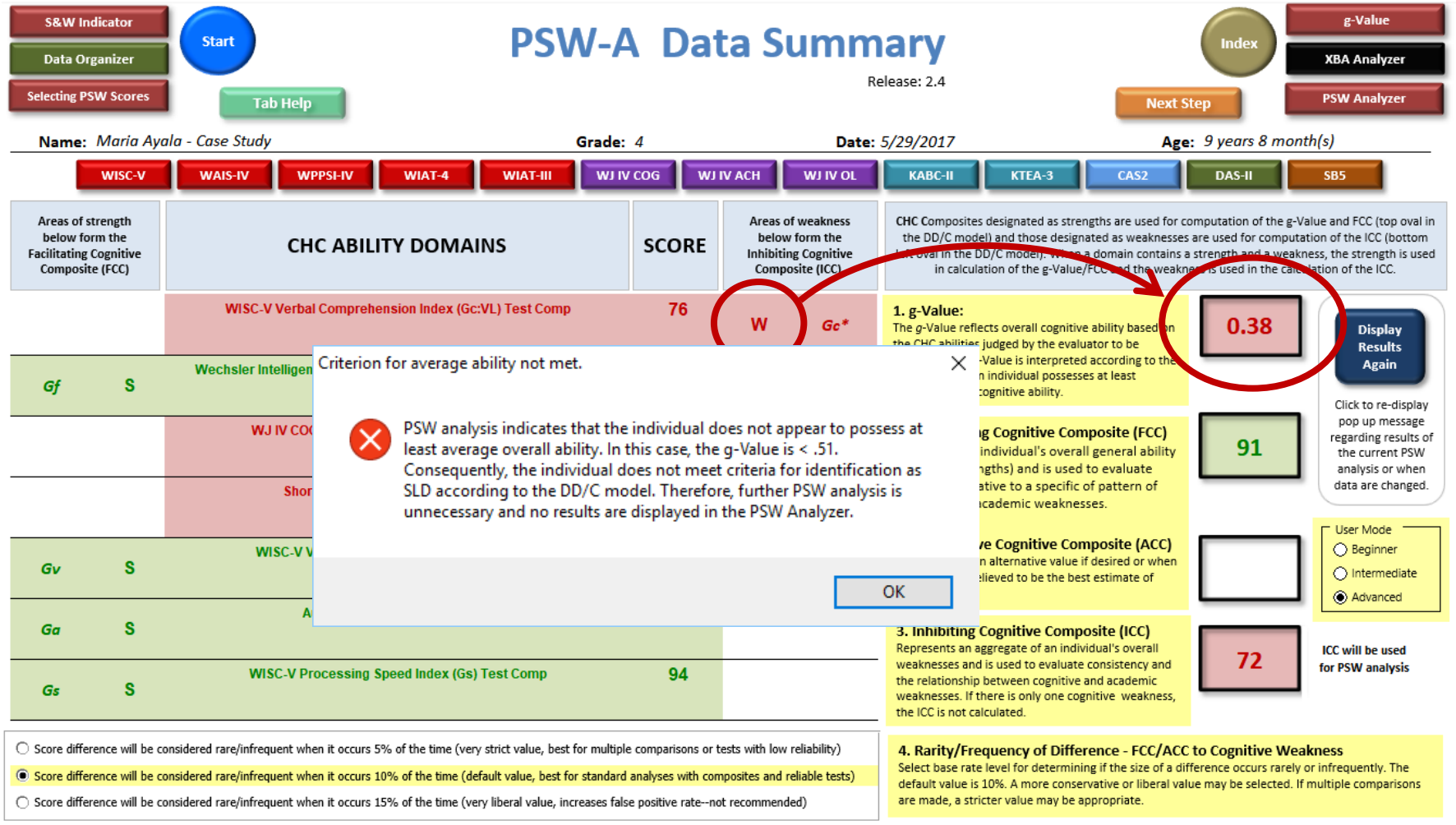
SLD Identification with an English Learner via PSW



SLD Identification with an English Learner via PSW

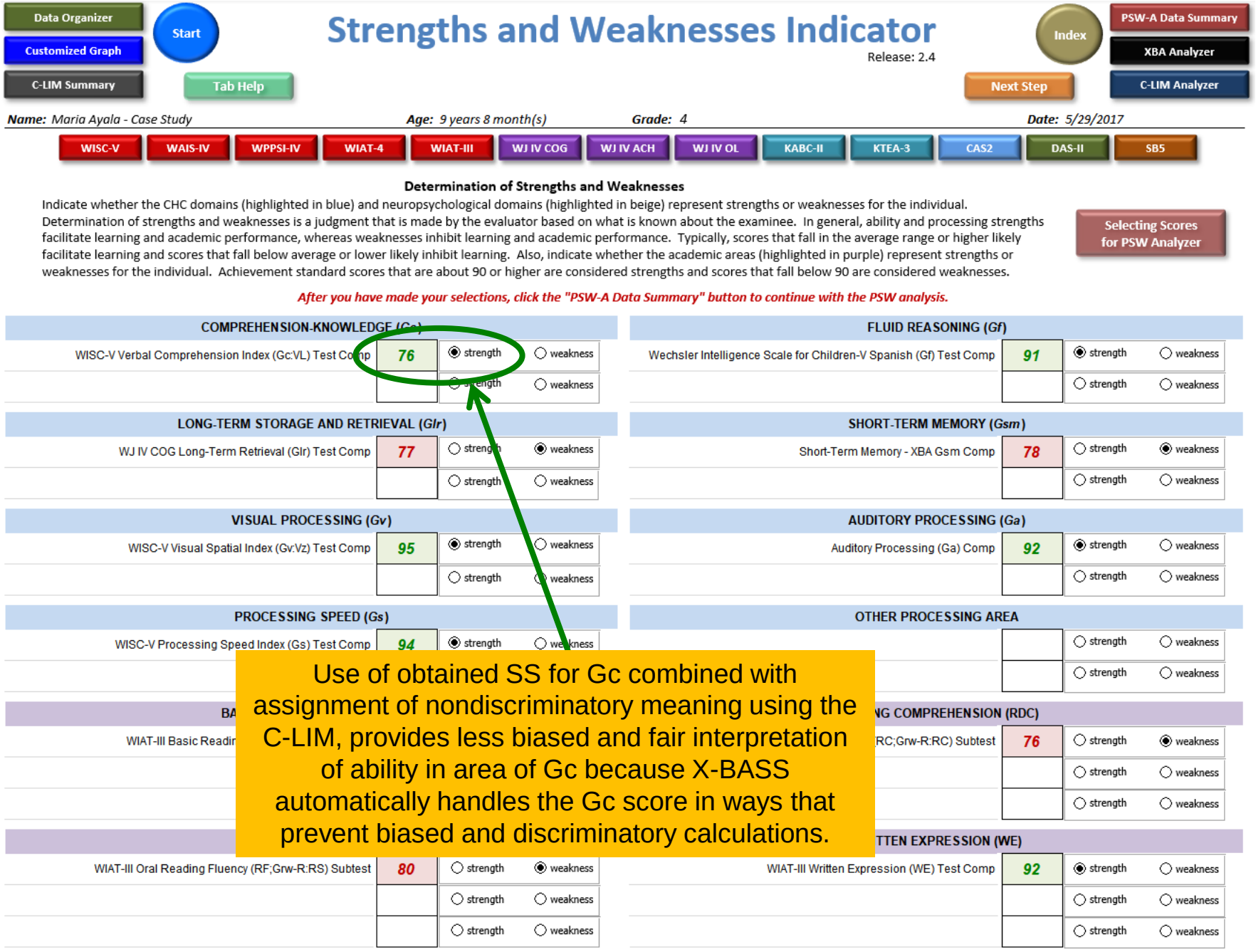


SLD Identification with an English Learner via PSW

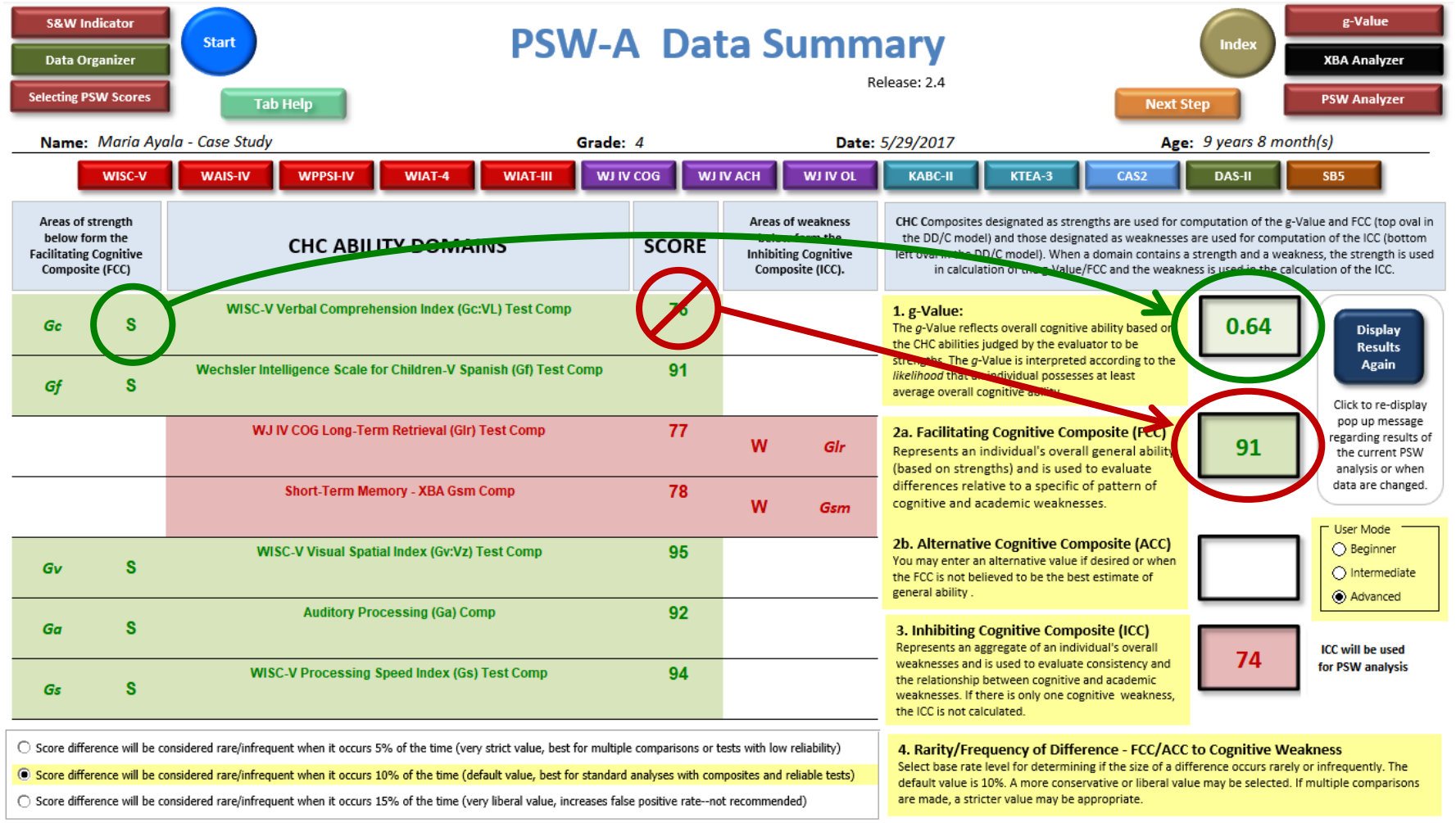


Because Gc is the most important ability related to academic success and accounts for the majority of variance in overall general ability, failure to properly evaluate it against other ELs with comparable backgrounds may result in highly attenuated g-Values that suggest low ability and mask possible SLD. In this case, the Gc score was within the expected range and should be indicated as a “strength” not “weakness.”

SLD Identification with an English Learner via PSW

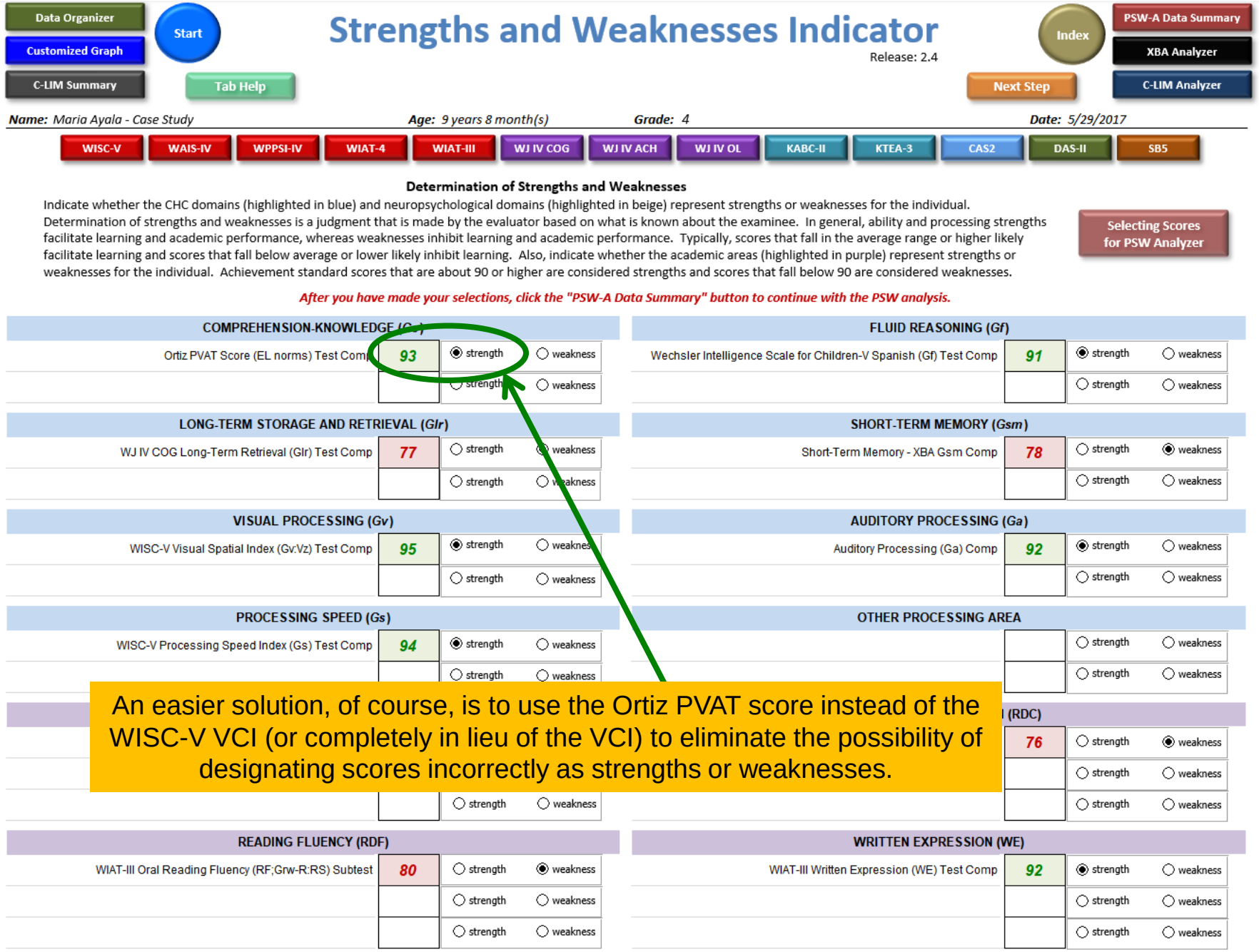


SLD Identification with an English Learner via PSW



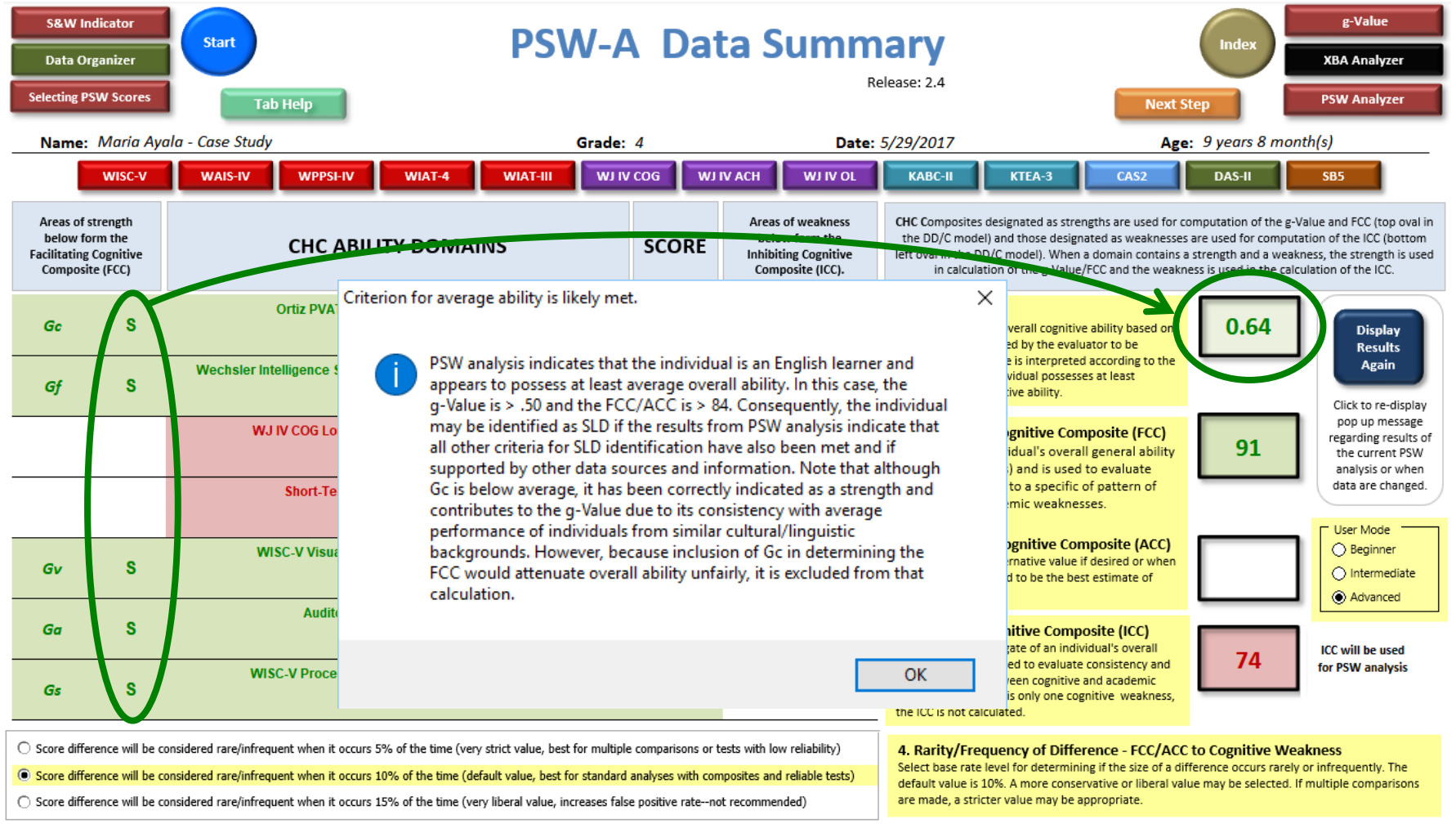
To prevent discriminatory attenuation of overall general cognitive ability in the case of ELs, if the Gc score is designated as a strength, and it is SS < 90 but within or above the expected range in the C-LIM, X-BASS will automatically exclude it from the calculations for the FCC.

SLD Identification with an English Learner via PSW



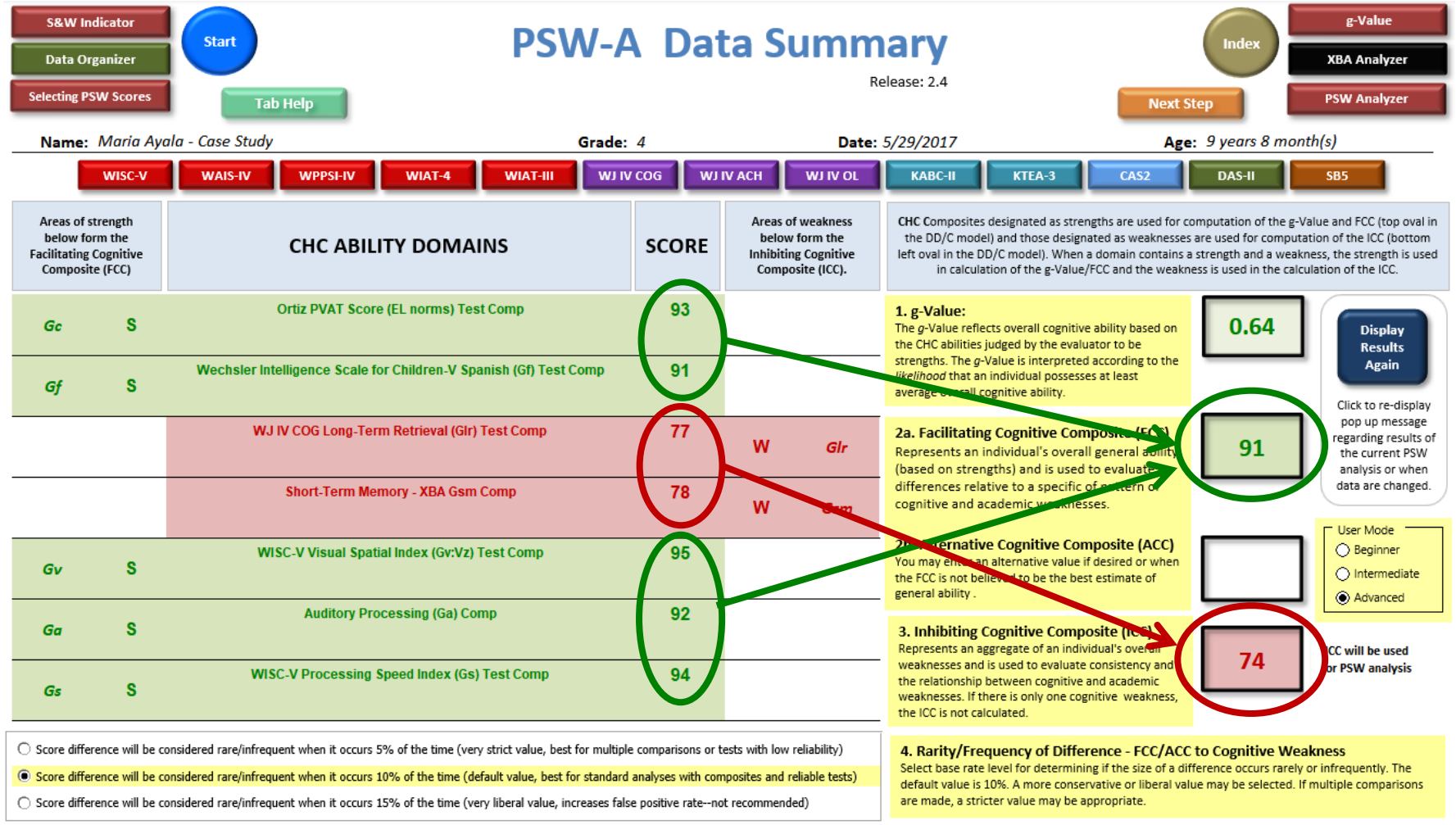
An easier solution, of course, is to use the Ortiz PVAT score instead of the WISC-V VCI (or completely in lieu of the VCI) to eliminate the possibility of designating scores incorrectly as strengths or weaknesses.

SLD Identification with an English Learner via PSW



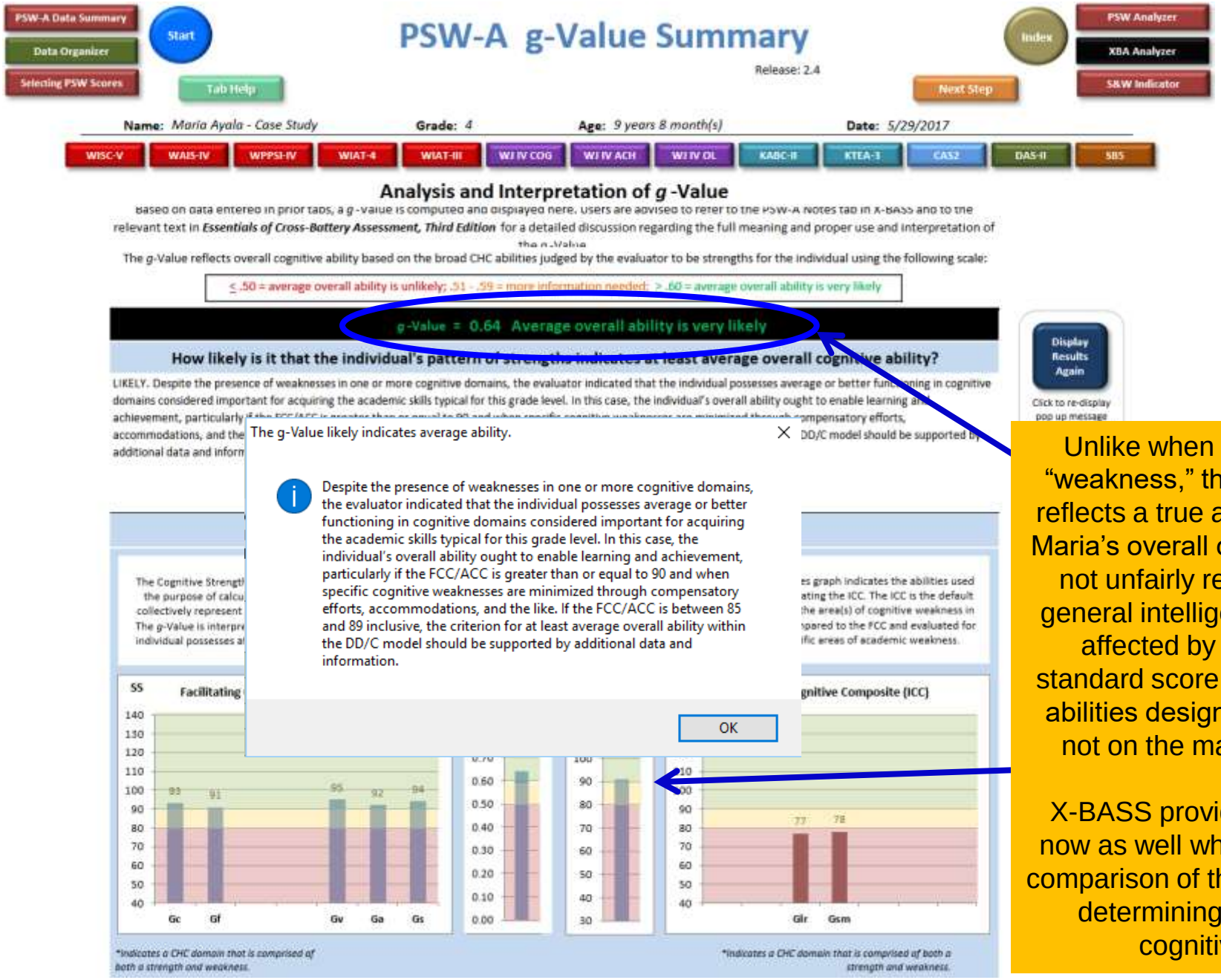
When a Gc (and any other CHC ability domain score), whether a broad or narrow composite, is marked as a "strength," it is included in calculations for determining the g-Value. Likewise, any score marked as a "weakness" are not used in deriving the g-Value. This keeps the g-Value free from the influence of the magnitude of the scores and thus complements the FCC which is based directly on the magnitude of the "strength" scores.

SLD Identification with an English Learner via PSW



When a Gc (and any other CHC ability domain score), whether a broad or narrow composite, is marked as a “strength,” (typically SS ≥ 90), X-BASS will always include its value in calculation of the FCC. Likewise, any scores marked as “weakness” are always factored into calculation of the ICC.

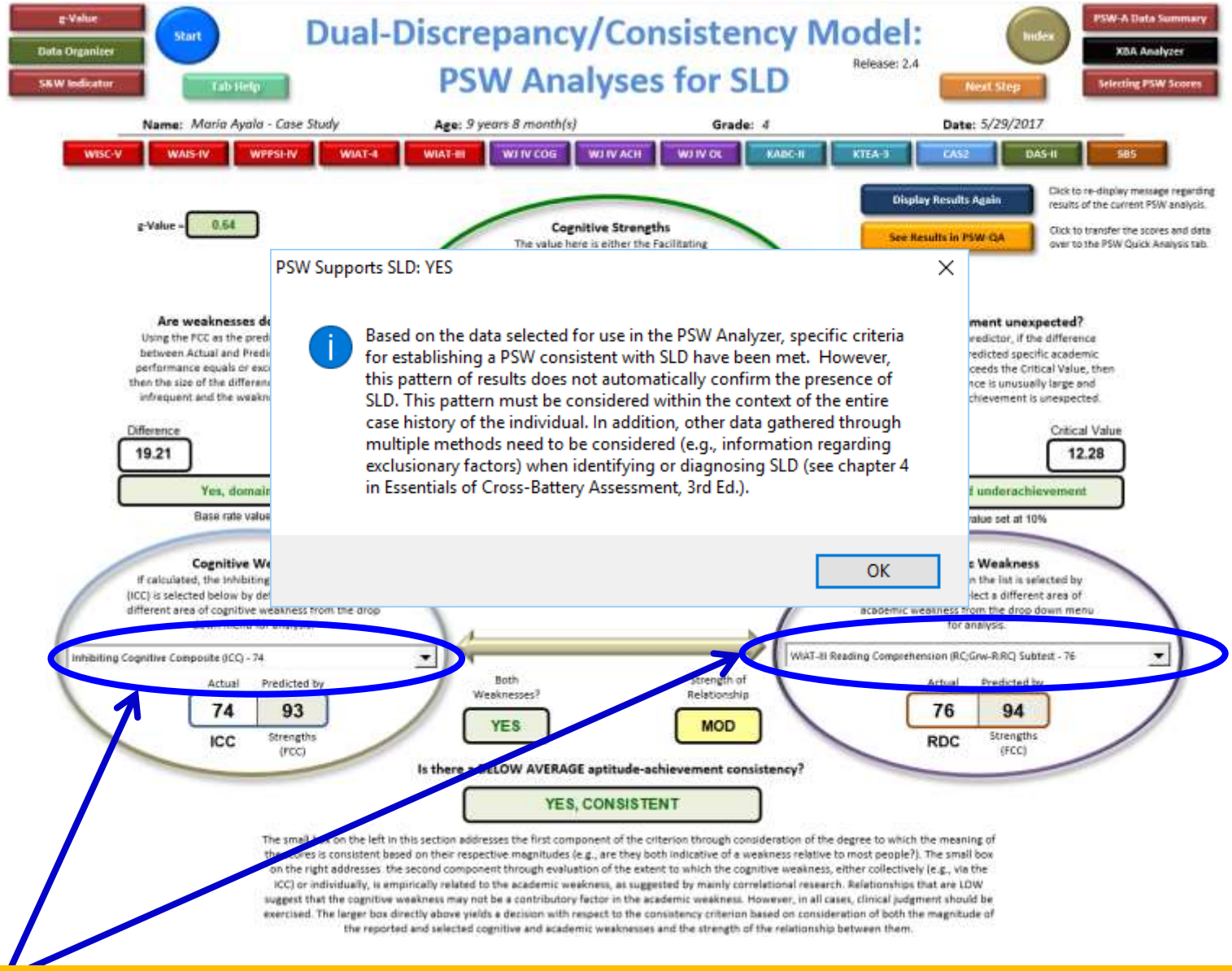
SLD Identification with an English Learner via PSW



Unlike when Gc was indicated as a "weakness," the g-Value now correctly reflects a true and equitable estimate of Maria's overall cognitive ability and does not unfairly represent her as lacking general intelligence. The g-Value is not affected by the magnitude of the standard score since it is based only on abilities designated as "strengths" and not on the magnitude of the scores.

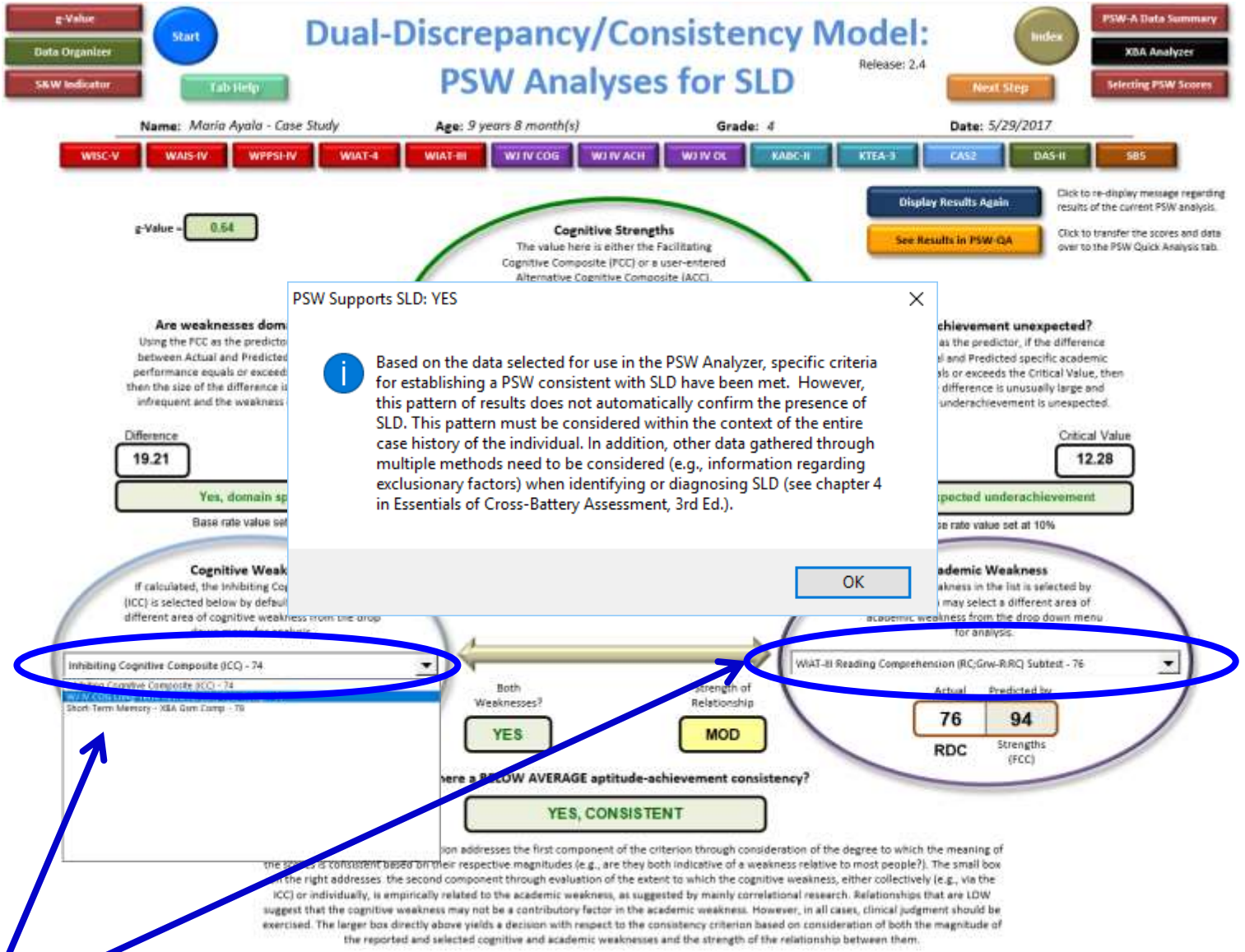
X-BASS provides a graph of the FCC now as well which allows simultaneous comparison of the two values as a part of determining an otherwise normal cognitive ability profile.

SLD Identification with an English Learner via PSW



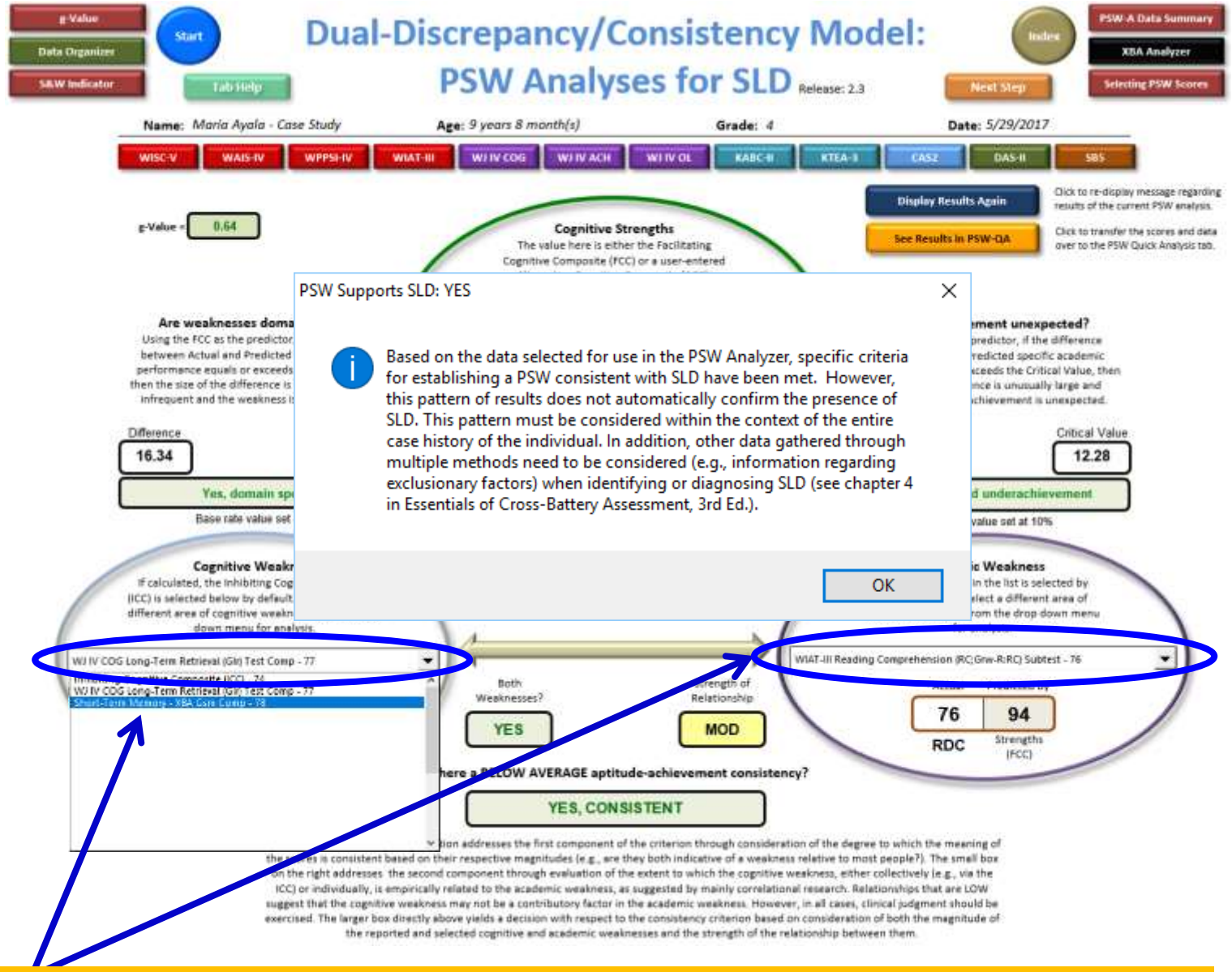
Using the ICC, data are consistent overall with SLD. But because the ICC is a trans-domain composite with greater reliability than a domain specific composite, it is more likely to reveal a significant difference. In addition, the ICC does not provide specific information regarding the nature of the cognitive deficit or inform intervention and instruction. As such, it may be beneficial to also explore SLD via specific areas of cognitive weakness that may be related to the areas of academic weakness.

SLD Identification with an English Learner via PSW



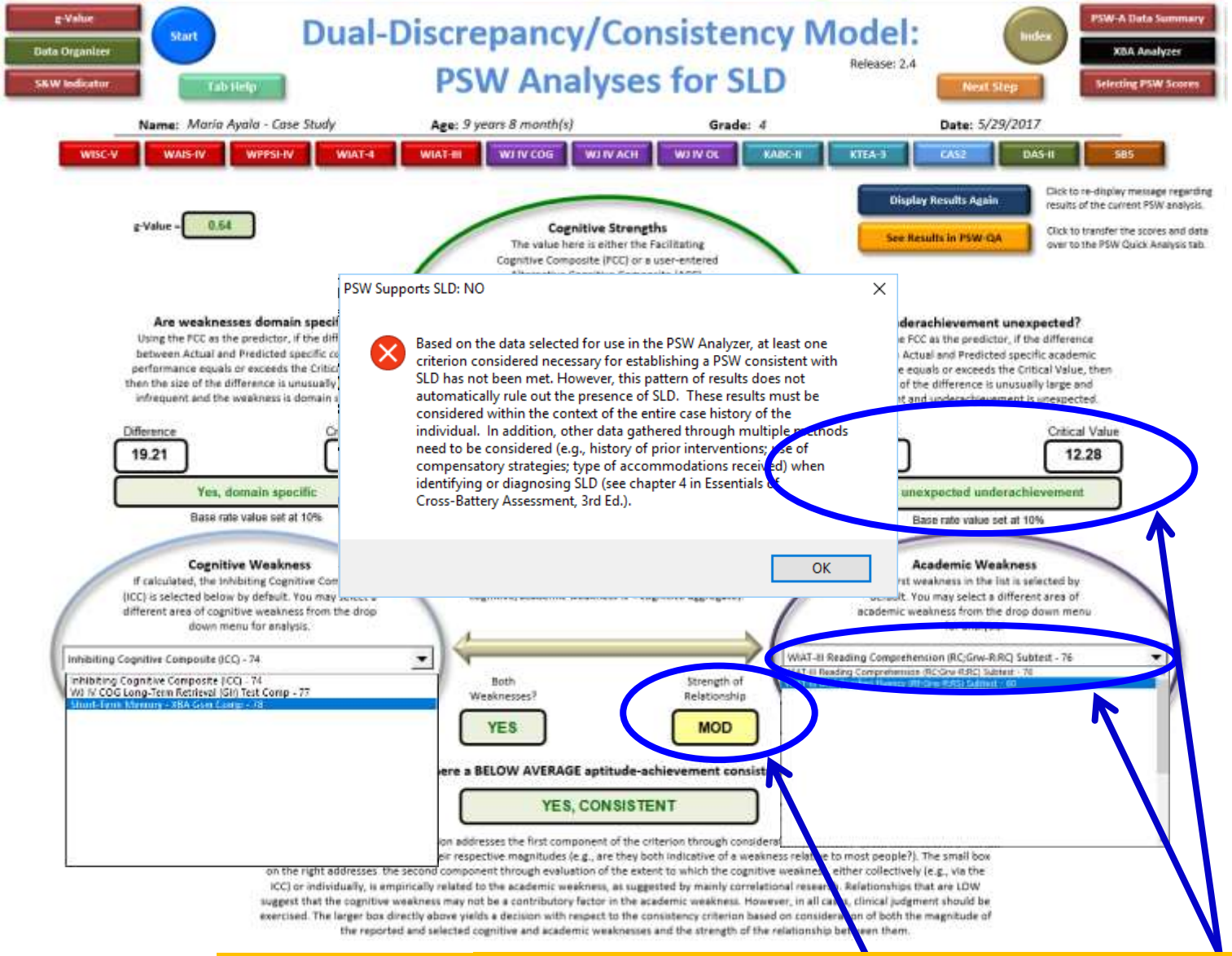
Changing the cognitive weakness to Glr also reveals a PSW consistent with SLD. It also provides a better indication that the academic problems are likely the result of deficits in Glr.

SLD Identification with an English Learner via PSW



In this case, changing the cognitive weakness to Gsm also results in a PSW consistent with SLD and provides additional information regarding the likely cause of the academic problems as having a basis, at least in part, to deficits in Gsm.

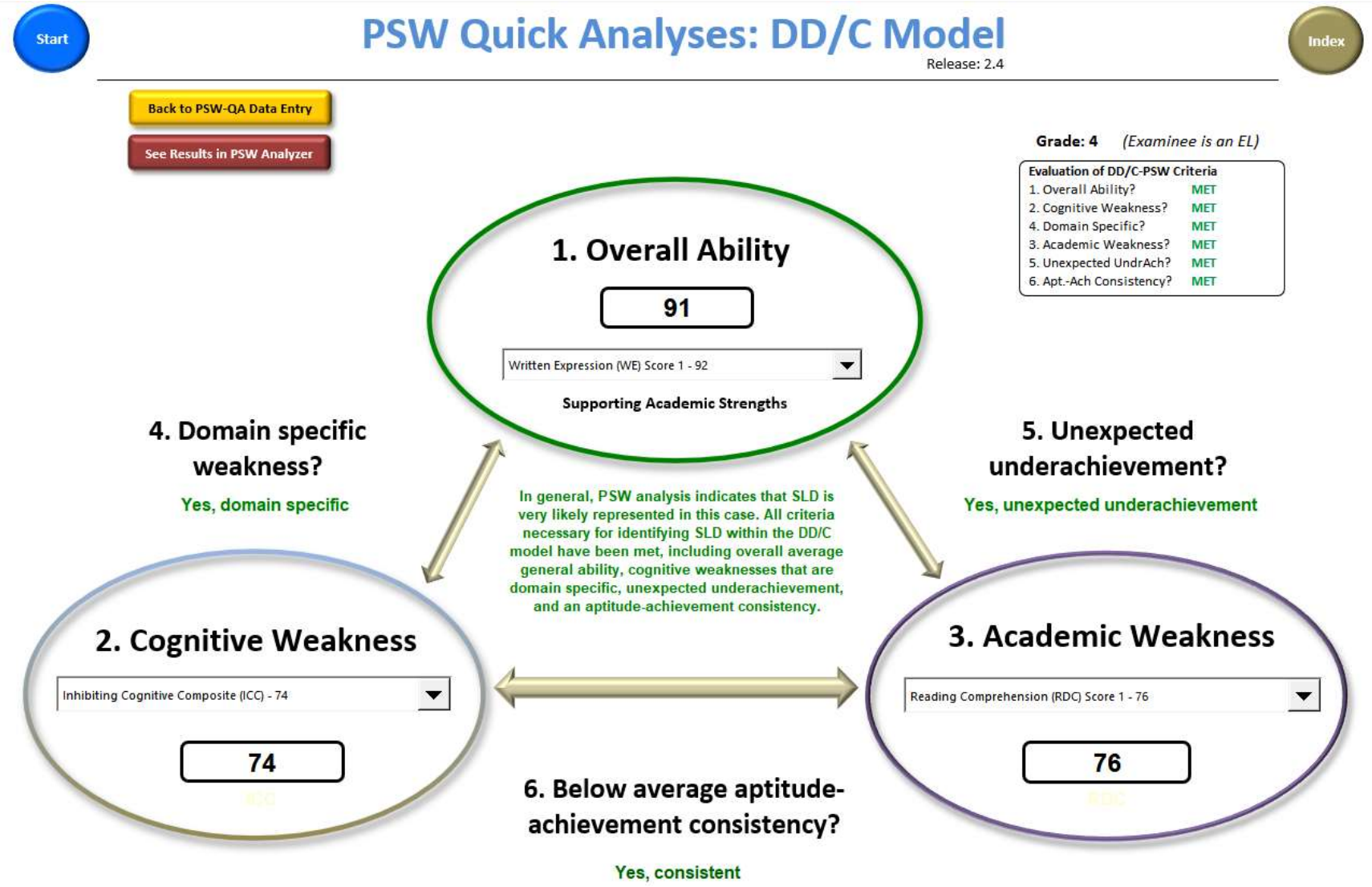
SLD Identification with an English Learner via PSW



When compared against short-term memory, reading fluency shows a poor relationship to reading fluency and further argues against SLD in this area.

consistent with SLD because pected underachievement.

SLD Identification with an English Learner via PSW



Transferring the scores into the PSW-QA provides a more simplified view of the results and is far more suitable for explaining results to others and including in typical psychoeducational reports.

Meeting Standards for Fairness in Evaluation of ELs

Sample Validity Statement for EL Evaluations

Statement 2. Evaluations of Suspected Learning Disability - Valid Results

The following sample validity statement may be used in cases where a clear declining pattern is NOT evident, that is, there is no primary effect of culture and language thus the results ARE valid and there may be a disability.

Because the student is not a native English speaker, it is necessary to establish the validity of the results obtained from testing to ensure that they are accurate estimates of ability or knowledge and not the manifestation of cultural or linguistic differences. To this end, a systematic evaluation of the possible effects of a relative lack of opportunity for the acquisition of acculturative knowledge and English proficiency was carried out via use of the Culture-Language Interpretive Matrix (C-LIM).

A careful review of the student's test data, as entered into the C-LIM, revealed either no overall pattern of decline or a partial pattern of decline combined with performance in one or more area that was below the range that would be expected of other individuals with similar cultural and linguistic backgrounds. This pattern of test performance suggests that cultural and linguistic factors were either ***minimal*** (no evident decline) or ***contributory*** (some decline) influences on the measured test performance but can not account for the entirety of the results. Accordingly, the test results were not considered to be due primarily to the influence of cultural and linguistic factors but still required additional information to fully establish their validity. Evidence to further support the validity of the obtained results was provided by converging sources of information including results from native language evaluation, progress-monitoring data, qualitative analysis, and authentic assessment methods. In addition, other extraneous factors that might account for the observed pattern (for example, lack of motivation, fatigue, incorrect administration/scoring, emotional/behavioral problems) were also evaluated and excluded. Taken together, the reported test results were deemed likely to be valid, interpretable, and to be reliable estimates of the student's actual ability or knowledge. However, equitable interpretation of Gc (cultural knowledge and language development), required comparison relative to other English learners with comparable linguistic development and educational experiences which was accomplished via examination of the magnitude of the high culture/high language cell in the C-LIM and whether it was within the selected range of difference. Consequently, the academic difficulties observed in classroom performance and which prompted this evaluation are not likely to attributable primarily to the process of normal second language and acculturative knowledge acquisition.

In summary, the observed pattern of the student's test results is not consistent with performance that is typical of non-disabled, culturally and linguistically diverse individuals who are of average ability or higher. Therefore, it can be reasonably concluded that the test data evaluated with the C-LIM are likely to be valid, are supported by additional converging data, and suggest that that the student's test performance can be used to support the presence of a learning disability or other cognitive-based disorder.

The statement above is the one most appropriate for this case where a) the evaluation focused on suspected SLD; and b) where it was determined that the obtained test results were NOT influenced primarily by cultural and linguistic factors, albeit they remained contributory. Thus, the test results (except for Gc) could be considered valid estimates of the abilities that were measured. In addition, native language testing was conducted to further support cognitive test score validity. This statement (and three others contained in X-BASS) have been placed in the public domain and may be freely copied, modified, and distributed for non-profit purposes without the need to secure permission.

Meeting Standards for Fairness in Evaluation of ELs

Sample Validity Statement for EL Evaluations

Simplified Validity Statement for LIKELY disability and Determination of VALID Results

Because XXXX is not a native English speaker, it is necessary to establish the validity of test scores to ensure that they are true estimates of their ability and not the result of limited English proficiency.

XXXX's test data were entered into the Culture-Language Interpretive Matrix which permitted evaluation of the extent to which the scores were primarily affected by cultural or linguistic factors. A review of the pattern of test scores indicated that performance *was not consistent* with what would be expected of other individuals with similar cultural and linguistic backgrounds. This means that the scores may be interpreted as fair estimates of XXXX's abilities, with the exception of language which can only be determined to be an area of strength or weakness via comparison to other English learners which was accomplished by further use of the C-LIM.

The statement above is most appropriate for this case where a) the evaluation focused on identification of a suspected cognitive/academic-based disability; and b) where it was determined that the obtained test results were ***not*** influenced primarily by cultural and linguistic factors, albeit these factors may have remained contributory. Thus, the test results (except for Gc) could be considered valid estimates of the abilities that were measured. Native language testing should also have been conducted to further support cognitive test score validity.

Meeting Standards for Fairness in Evaluation of ELs

Sample Validity Statement for EL Evaluations

Simplified Validity Statement for UNLIKELY disability and Determination of INVALID Results

Because XXXX is not a native English speaker, it is necessary to establish the validity of test scores to ensure that they are true estimates of their ability and not the result of limited English proficiency.

XXXX's test data were entered into the Culture-Language Interpretive Matrix which permitted evaluation of the extent to which the scores were primarily affected by cultural or linguistic factors. A review of the pattern of test scores indicated that performance *was consistent* with what would be expected of other individuals with similar cultural and linguistic backgrounds. This means that the scores cannot be interpreted as fair estimates of XXXX's abilities.

However, because the scores were compared to other individuals from research studies who were of average ability and who had not been identified as having a disability, it suggests that XXXX's performance is also average (possibly higher) and that it is not likely that a disability is present in this case. This means that although XXXX is having difficulties in the classroom, the problems are most likely to attributable to, and primarily the result of, the normal process of second language and acculturative knowledge acquisition.

Meeting Standards for Fairness in Evaluation of ELs

Although there are no professional or legal standards that specify actual procedures for evaluation of English learners or determining the impact of exclusionary factors related to linguistic/cultural differences, there are consensus recommendations that provide some guidance in being able to document and establish that a given evaluation has been conducted in compliance with standards necessary to demonstrate and establish such consideration and fairness. The following are standards that may be used to bolster conclusions regarding exclusionary factors and fairness.

1. **TOOLS AND PROCEDURES:** The report contains a section detailing the deliberate selection of tools, methods, and procedures with respect to the cultural and linguistic factors in the examinee's background—simply listing tests, even native language ones, is not sufficient. Explanations are provided for any modification or alteration to the administration or scoring of any standardized instrument, including use of a translator or translated test.

2. **DEVELOPMENTAL LANGUAGE HISTORY:** The report contains a specific and distinct section on language development which contains a detailed history and sufficient information with which to formulate appropriate expectations of current proficiency. Information should include, at a minimum, age of first exposure to all languages, parental/home language use, parental levels of proficiency in all languages, parental education and socio-economic status, individual's experiences with all languages, current proficiency in all languages, amount of formal education in all languages, and type of educational programming.

3. **VALIDITY:** The report contains a section that provides a discussion regarding the validity of the obtained assessment data and any collected test scores including specification regarding how the impact of cultural/linguistic differences were considered and excluded as factors that might have compromised validity of the information—simply stating that scores or data are valid is insufficient.

4. **INTERPRETATION OF RESULTS:** Discussion of results, whether cognitive, linguistic, or academic, are always presented in terms of the extent to which cultural or linguistic factors may have compromised performance and affected interpretive validity and the extent to which they are consistent with or not consistent with what would be reasonably expected of the examinee, given their unique cultural and linguistic background.

5. **DIAGNOSTIC IMPRESSIONS:** The report contains conclusions and interpretations that are supported by integration of data and includes discussion regarding how cultural/linguistic factors are not the primary reasons for any claimed deficits and that such deficits are above and beyond what would be expected given the examinee's unique cultural/linguistic background.

Multilingual Testing (L2+L1)

Combined with the comprehensive framework for nondiscriminatory assessment, the Multilingual Assessment (L2+L1) testing approach provides an efficient, research-based, and IDEA-compliant process that makes best use of available resources for evaluation consistent with current standards, and it permits ANY evaluator to begin (and in some cases, complete) testing without being bilingual or requiring outside assistance. The approach does require *knowledge of research on Els and other issues that are not generally taught in training programs and must be applied in a systematic way to establish defensibility and bolster decisions regarding exclusionary factors and their bearing on any disability determination*.

These activities do not preclude or prevent monolingual, English-speaking or bilingual school psychologists from being able to engage in “bilingual evaluations” with speakers of any language.

Assessment and Related Resources

RESOURCES:

C-LIM Resources - free
http://facpub.stjohns.edu/~ortizs/CLIM/index.html

Ortiz, S. O. (2019). On the Measurement of Cognitive Abilities in English Learners. Contemporary School Psychology, Vol. 23(1) 68-86. https://doi.org/10.1007/s40688-018-0208-8

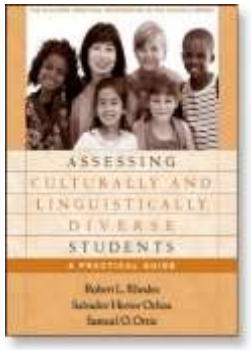
Ortiz, S. O. (2017). Evaluation of English Learners: Issues in measurement, interpretation and reporting. The Score, APA Division 5 (Quantitative and Qualitative Methods) Newsletter, January 2017. Available at http://www.apadivisions.org/division-5/publications/score/2017/01/english-learners.aspx

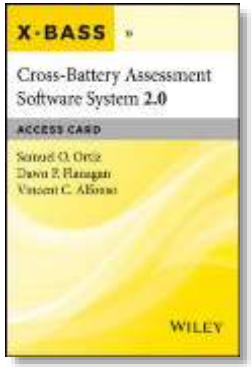
Kovaleski, J. F., Lichtenstein, R. Naglieri, J., Ortiz, S. O., Klotz, M. B. & Rossen, E. (2015). Current Perspectives in the Identification of Specific Learning Disabilities. Communiqué, 44(4).

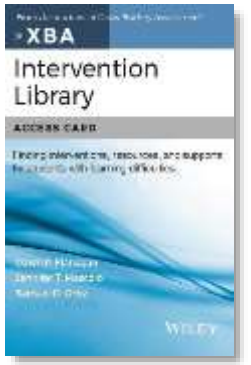
Whittaker, M. & Ortiz, S. O. (2019). Exclusionary Factors—What a Specific Learning Disability is Not: Examining exclusionary factors. National Center for Learning Disabilities, Washington DC. Available at https://www.ncld.org/wp-content/uploads/2019/09/What-a-Specific-Learning-Disability-Is-Not-Examining-Exclusionary-Factors.pdf

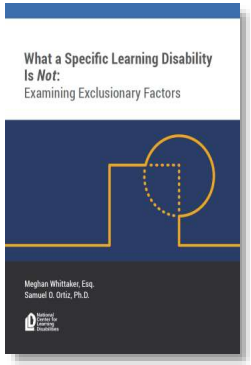
Ortiz, S. O., Flanagan, D. P. & Alfonso, V. C. (2015). Cross-Battery Assessment Software System (X-BASS v2.X). New York: Wiley & Sons, Inc.

Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT)
https://www.mhs.com/ortizpvat









Additional Readings and Related References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing. Washington DC: American Educational Research Association.
- Bialystok, E. (2001). Bilingualism in development: Language, literacy and cognition. Cambridge: Cambridge University Press.
- Cormier, D. C., McGrew, K. S., & Ysseldyke, J. E. (2014). The influences of linguistic demand and cultural loading on cognitive test scores. Journal of Psychoeducational Assessment, 32(7), 610–623.
- Cummins, J. C. (1984). Bilingual and special education: Issues in assessment and pedagogy. Austin, TX: PRO-ED.
- Dynda, A. M., (2008). The relation between language proficiency and IQ test performance. Unpublished manuscript, St. John's University, Jamaica, NY.
- Grosjean, F. (1989). Neurolinguists beware!: The bilingual is not two monolinguals in one person. Brain and Language, 36, 3-15.
- Krashen, S.D. (l982). Principles and Practice in second language acquisition. New York: Pergamon Press.
- Ortiz, S. O. (2019). On the Measurement of Cognitive Abilities in English Learners. Contemporary School Psychology, Vol. 23(1) 68-86. https://doi.org/10.1007/s40688-018-0208-8
- Ortiz, S. O. (2018). Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT). Toronto, Canada: Multi-Health Systems.
- Ortiz, S. O., Piazza, N., Ochoa, H. S. & Dynda, A. M. (2018). Testing with Culturally and Linguistically Diverse Populations: Moving beyond the verbal-performance dichotomy into evidence-based practice. In D. P. Flanagan and E. McDonough (Eds.), Contemporary Intellectual Assessment, Fourth Edition (pp. 684-712). New York: Guilford Press.
- Rhodes, R., Ochoa, S. H., & Ortiz, S. O. (2005). Assessment of culturally and linguistically diverse students: A practical guide. New York: Guilford Press. (UNDER REVISION – DUE OUT Spring 2021)
- Sotelo-Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. (2013). English language proficiency and test performance: Evaluation of bilinguals with the Woodcock-Johnson III Tests of Cognitive Ability. Psychology in the Schools, 50(8), 781–797.
- Valdés, G., & Figueroa, R. A. (1994). Bilingualism and testing: A special case of bias. Norwood, NJ: Ablex.