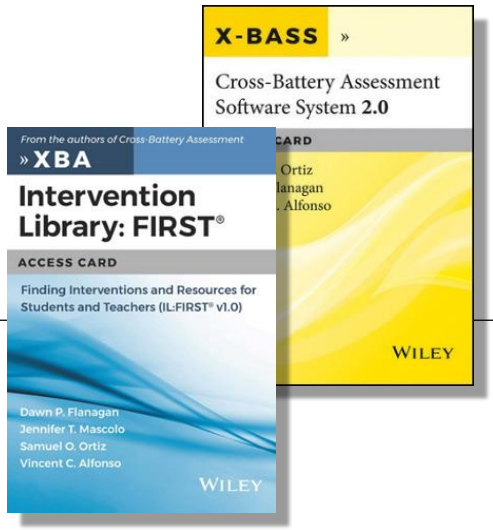
Evidence-based Assessment of English Learners:
Contemporary PSW evaluation for SLD with X-BASS, C-LIM, and the Ortiz PVAT.

Texas Educational Diagnostician's Association

Waco, TX
December 9, 2019

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



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."

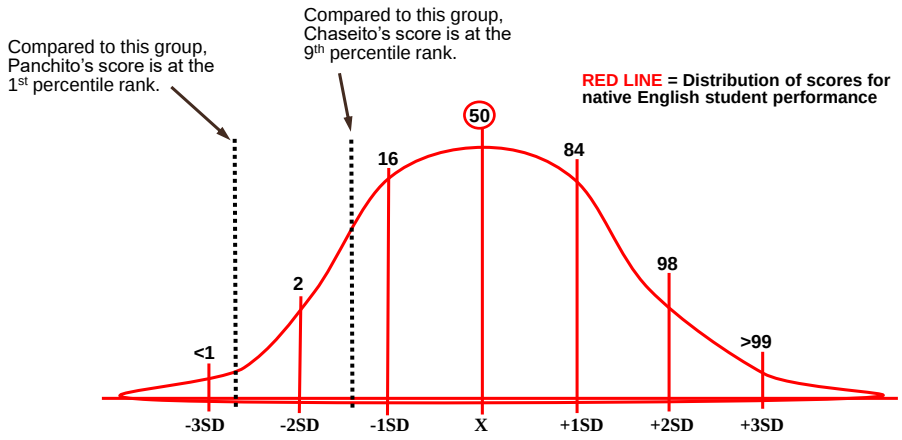
A test designed to measure visual processing (Gv) in ELs must avoid over-reliance on language ability (Gc) or else measurement of visual processing may be confounded with language ability.

Example of Potential Interpretive Invalidity:

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

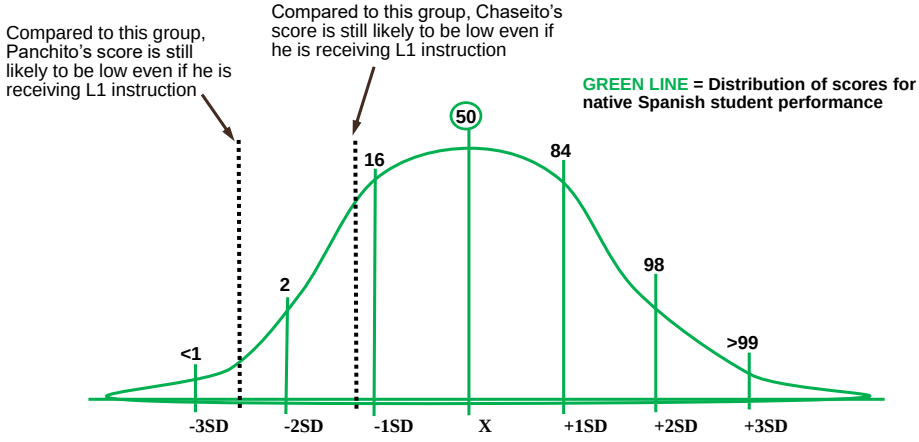
A test designed to measure English language ability (Gc) is valid for EL's ability ***in English***, but poor performance cannot be ascribed to a potential disability unless developmental differences in English have been controlled.

Diagnostic Question: Does Chaseito's or Panchito's rate of progress suggest cultural/linguistic difference or possible disorder?



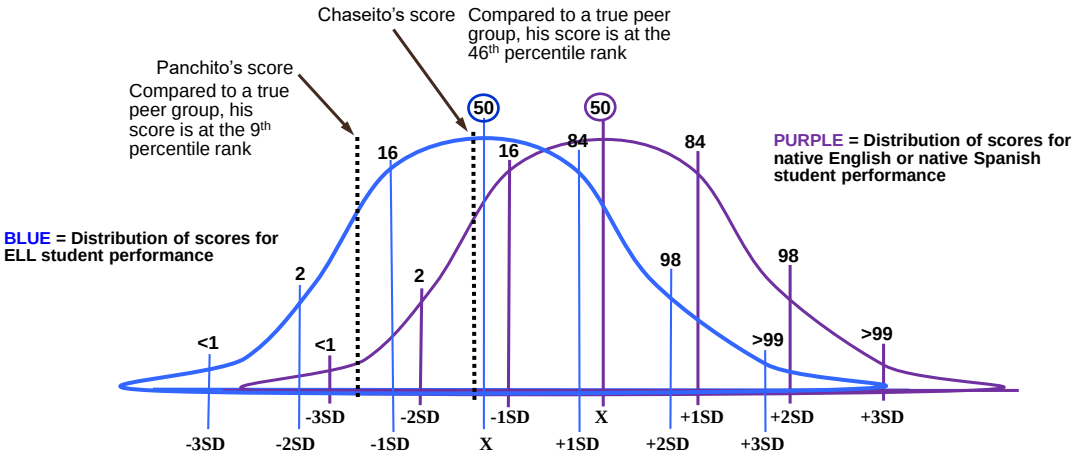
For the purposes of determining whether a disability exists, use of a monolingual English speaking comparison group is discriminatory and makes it appear incorrectly that both students might have some type of disability.

Diagnostic Question: Does Chaseito's or Panchito's rate of progress suggest cultural/linguistic difference or possible disorder?



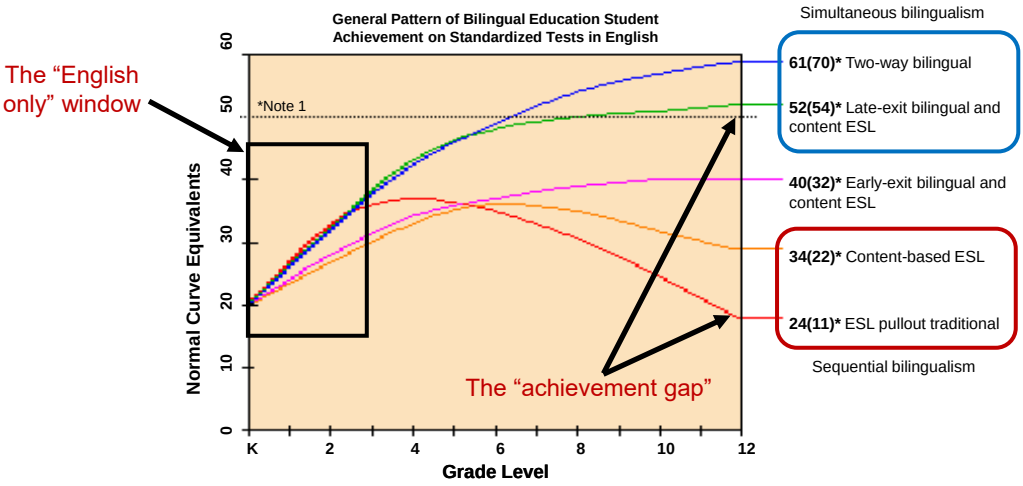
Similarly, use of a monolingual, native-language speaking group remains discriminatory because neither student is monolingual anymore (even when receiving native language instruction) and it continues to make it appear incorrectly that both Chaseito and Panchito have some type of disability.

Diagnostic Question: Does Chaseito's or Panchito's rate of progress suggest cultural/linguistic difference or possible 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.

Academic Test Score Validity Requires "True Peer" Comparison



*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.

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

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.

Development Varies by Exposure to English – Not relative dominance

"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

Processes and Procedures for Addressing Test Score Validity

In what manner exactly, is evidence-based, nondiscriminatory assessment conducted and to what extent is there any research to support the use of any of the following methods as being capable of establishing sufficient 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*

Current Approaches Fail to Establish Test Score Validity

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	✗	✓	✓	✓	✓	✗	✗	✗	✗

All approaches are limited in some manner when addressing test score validity and none are sufficient to diagnosis a disability, account for variation in bilingual development, represent a form or manner that automatically yields reliable and valid results, and do not provide extensive data regarding cognitive and school-based learning and development.

The validity of an interpretation regarding disability requires an unbiased standard for comparison.

Whatever method or approach may be employed in evaluation of EL's, the fundamental obstacle to nondiscriminatory interpretation rests on the degree to which the examiner is able to defend claims of test score (construct and interpretive) validity that is being used to support diagnostic conclusions. This idea is captured by and commonly referred to as a question of:

"DIFFERENCE vs. DISORDER?"

Simply absolving oneself from responsibility of establishing test score validity, for example via wording such as, "all scores should be interpreted with extreme caution" does not in any way provide a defensible argument regarding the validity of obtained test results and does not permit valid diagnostic inferences or conclusions to be drawn from them.

Test score validity must be evaluated or established via use of a "true peer" comparison standard and the only manner in which to accomplish this task is with evidence and data.

The Culture-Language Interpretive Matrix (C-LIM)

Addressing test score validity for ELLs

Translation of Research into Practice

1. The use of various traditional methods for evaluating ELLs, including testing in the dominant language, modified testing, nonverbal testing, or testing in the native language do not ensure valid results and provide no mechanism for determining whether results are valid, let alone what they might mean or signify.
2. The pattern of ELL test performance, when tests are administered in English, has been established by research and is predictable and based on the examinee's degree of English language proficiency and acculturative experiences/opportunities as compared to native English speakers.
3. The use of research on ELL test performance, when tests are administered in English, provides the only current method for applying evidence to determine 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)**.
4. The principles of ELL test performance as established by research are the foundations upon which the C-LIM is based and serve as a de facto norm sample for the purposes of comparing test results of individual ELLs to the performance of a group of average ELLs with a specific focus on the attenuating influence of cultural and linguistic factors.

The Culture-Language Interpretive Matrix (C-LIM)

GENERAL RULES AND GUIDANCE FOR EVALUATION OF TEST SCORE VALIDITY

There are two basic criteria that, when both 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. 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. 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.

Results are **INVALID** only if both conditions are met.

When both criteria are observed, it may be concluded that the test scores are likely to have been influenced primarily by the presence of cultural/linguistic variables and therefore are not likely to be valid and should not be interpreted. If either criterion is not met, the results can be assumed to be **VALID**.

Research Foundations of the C-LIM
Additional Issues in Evaluation of Test Score Patterns

Evaluation of test score validity, particularly in cases where results are "possibly valid," includes considerations such as:

1. Is the Tiered graph consistent with the main Culture-Language graph or the other secondary (language-only/culture-only) graphs?

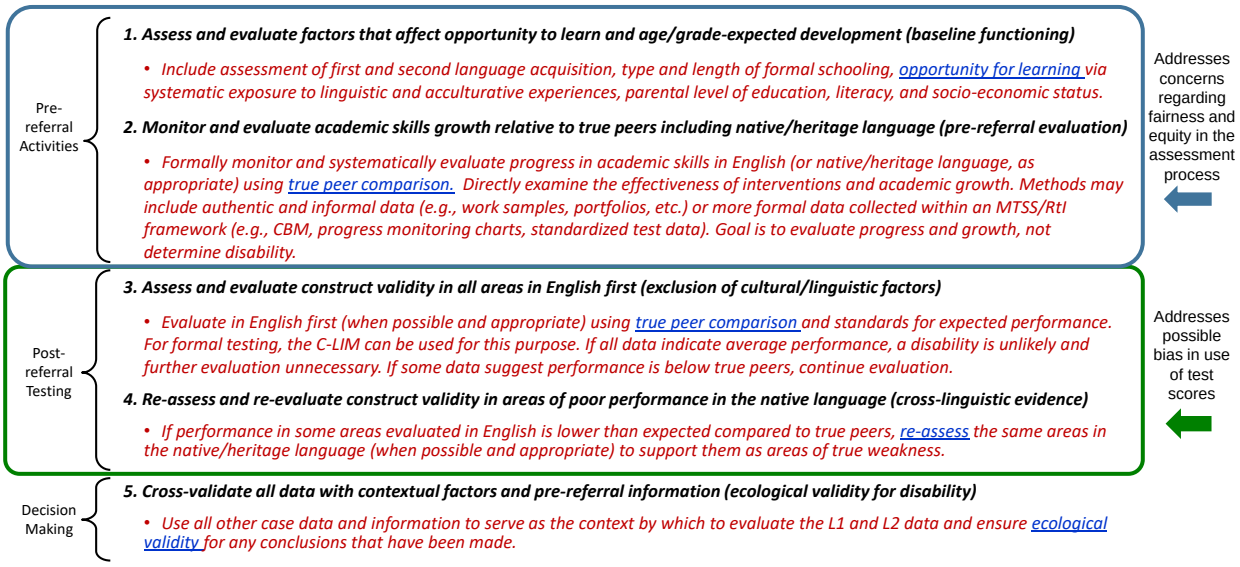
2. Is there any variability in the scores that form the aggregate in a particular cell that may be masking low performance?

3. Is the pattern of scores consistent with a developmental explanation of the examinee's educational program and experiences?

4. Is the pattern of scores consistent with a developmental explanation of the examinee's linguistic/acculturative learning experiences?

Evaluation of results using all graphs, including secondary ones, identification of score variability in relation to CHC domains or task characteristics, and evaluation of educational, cultural, and linguistic developmental experiences assists in determining the most likely cause of score patterns and overall test score validity.

A Best Practice Framework for Comprehensive Evaluation of ELs



Practical Considerations for Addressing Test Score Validity in Evaluation of ELs

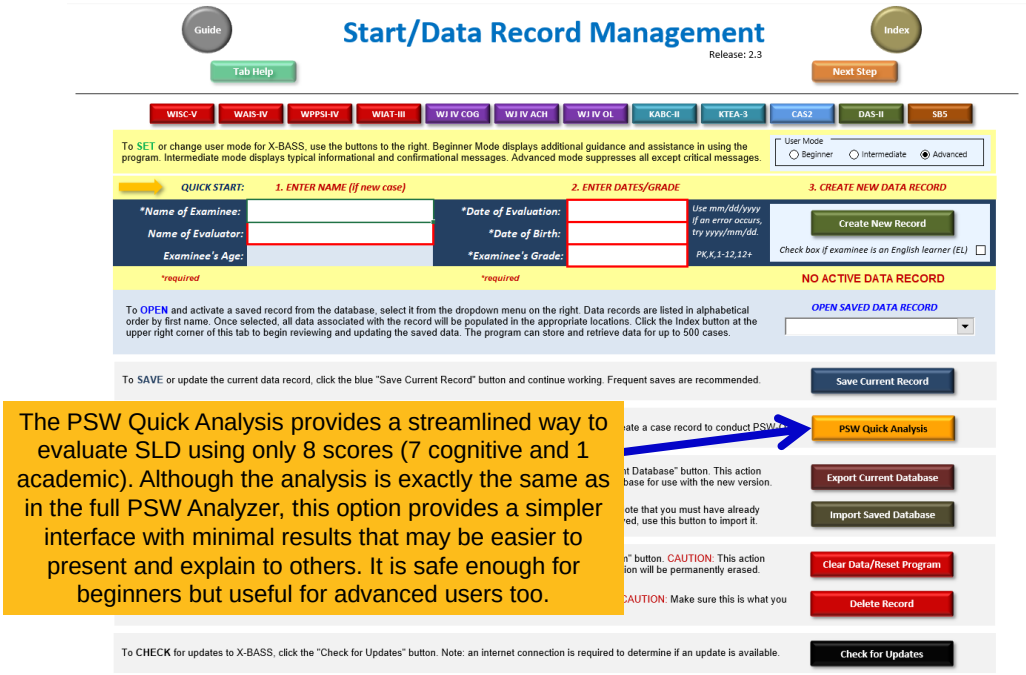
1. *The usual purpose of testing is to identify deficits in ability (i.e., low scores)*
2. *Validity is more of a concern for low scores than average/higher scores because:*
   - *Test performances in the average range are NOT likely a chance finding and strongly suggests average ability (i.e., no deficits 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)*
3. *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)*
4. *Testing in both languages (English and native language) is necessary to determine disability*
   - *Testing requires confirmation that deficits are not language-specific and exist in both languages (although low performance in both can result from other factors)*
5. *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*

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	✗	✓	✓	✓	✓	✗	✗	✗	✗
Multilingual Assessment in L1 + L2	✓	✓	✓	✓	✓	✓	✓	✓	✓

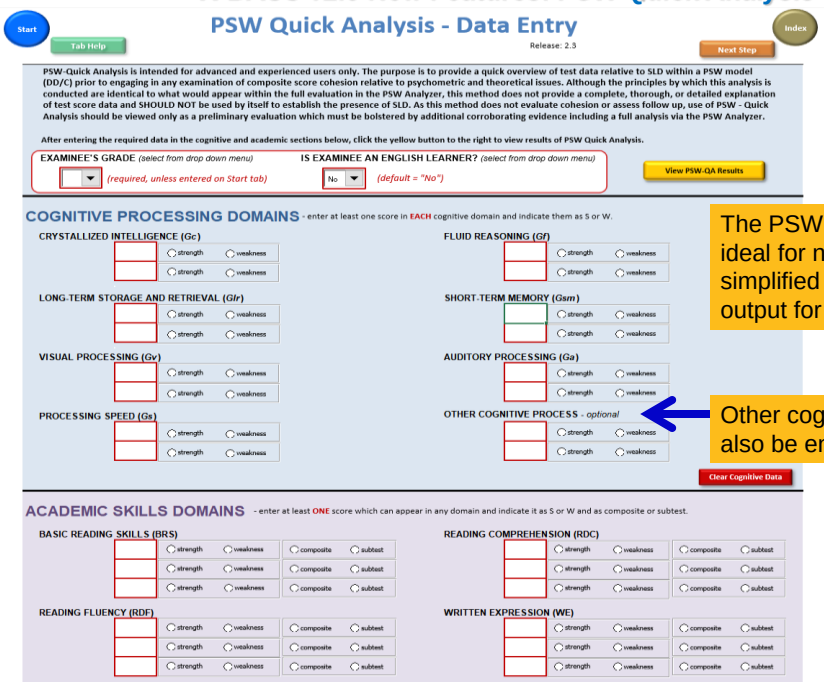
Multilingual Assessment combined with the C-LIM resolves all validity issues, and by applying research on EL test performance, they can be used to define and establish a "true peer" reference group for disability-based evaluations.

X-BASS v2.3 New Features: PSW Quick Analysis



The PSW Quick Analysis provides a streamlined way to evaluate SLD using only 8 scores (7 cognitive and 1 academic). Although the analysis is exactly the same as in the full PSW Analyzer, this option provides a simpler interface with minimal results that may be easier to present and explain to others. It is safe enough for beginners but useful for advanced users too.

X-BASS v2.3 New Features: PSW Quick Analysis

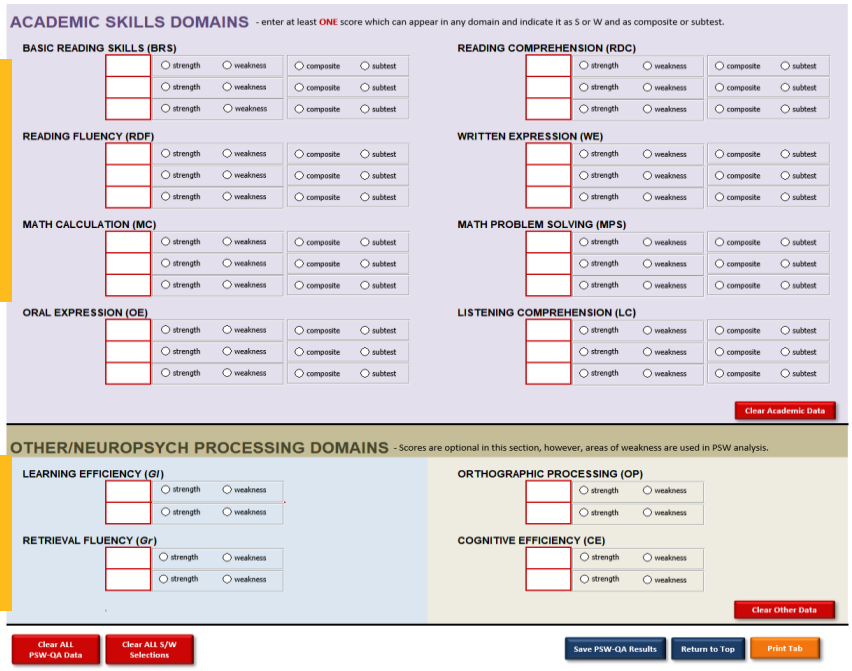


The PSW Quick Analysis is ideal for new users and offers a simplified interface and results output for easy interpretation.

Other cognitive processes may also be entered for analysis.

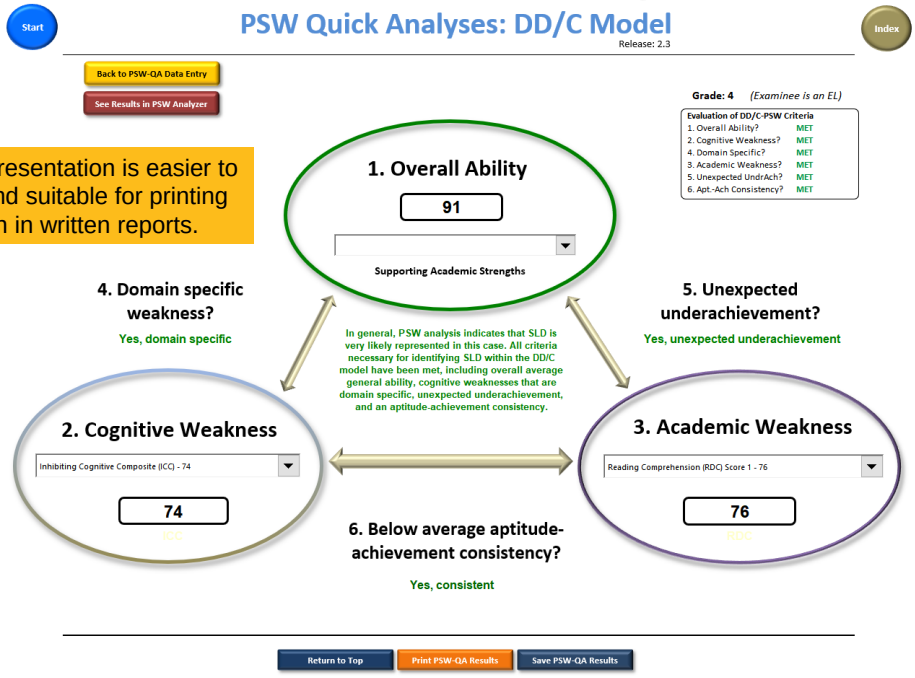
X-BASS v2.3 New Features: PSW Quick Analysis

Only one score each of the seven cognitive areas and one score in any of the academic areas (8 scores total) is sufficient to conduct PSW Quick Analysis.



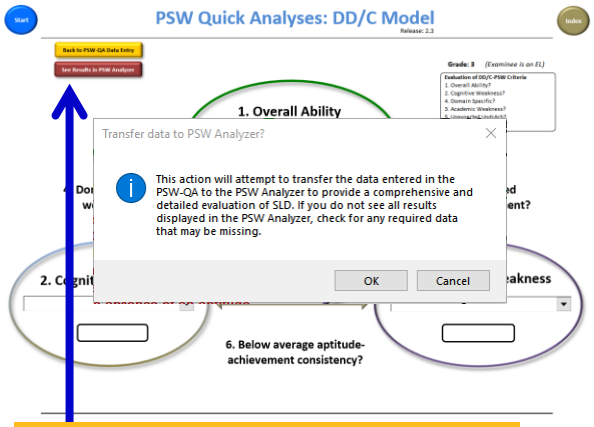
PSW Quick Analysis can include "other cognitive" and neuropsych processing domains.

X-BASS v2.3 New Features: PSW Quick Analysis



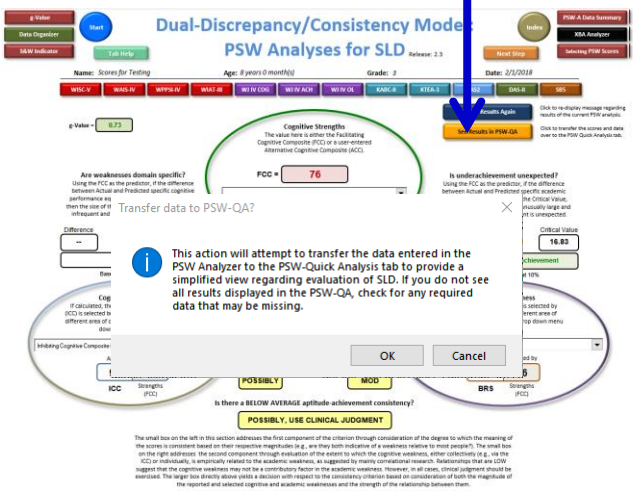
The simplified presentation is easier to comprehend and suitable for printing and inclusion in written reports.

X-BASS v2.3 New Features: Two-way PSW data/results transfer

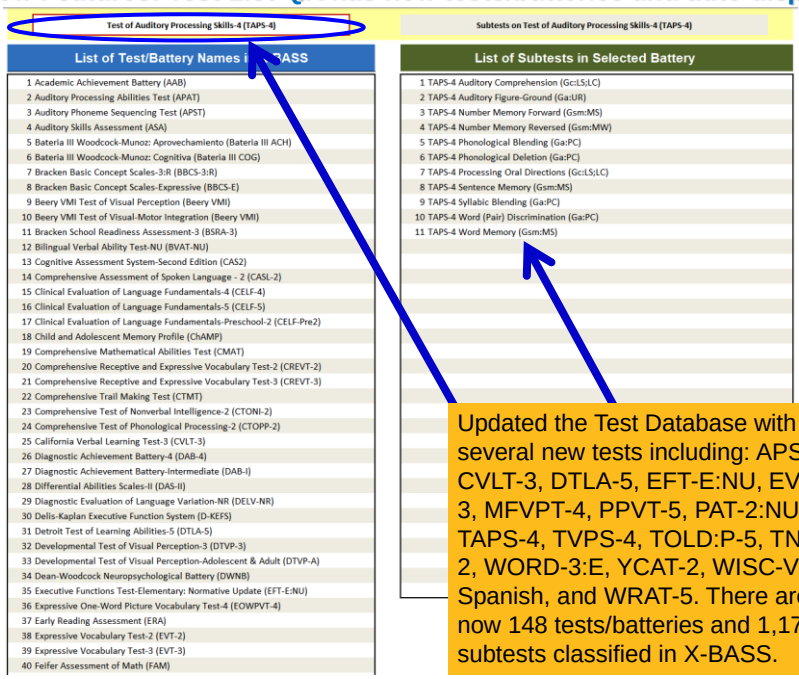


If you use the PSW-QA first, you can click the brown button and have the results transferred to and displayed in the full PSW Analyzer.

If you use the PSW Analyzer first, you can click the yellow button and have the results transferred to and displayed in the PSW Quick Analyzer.

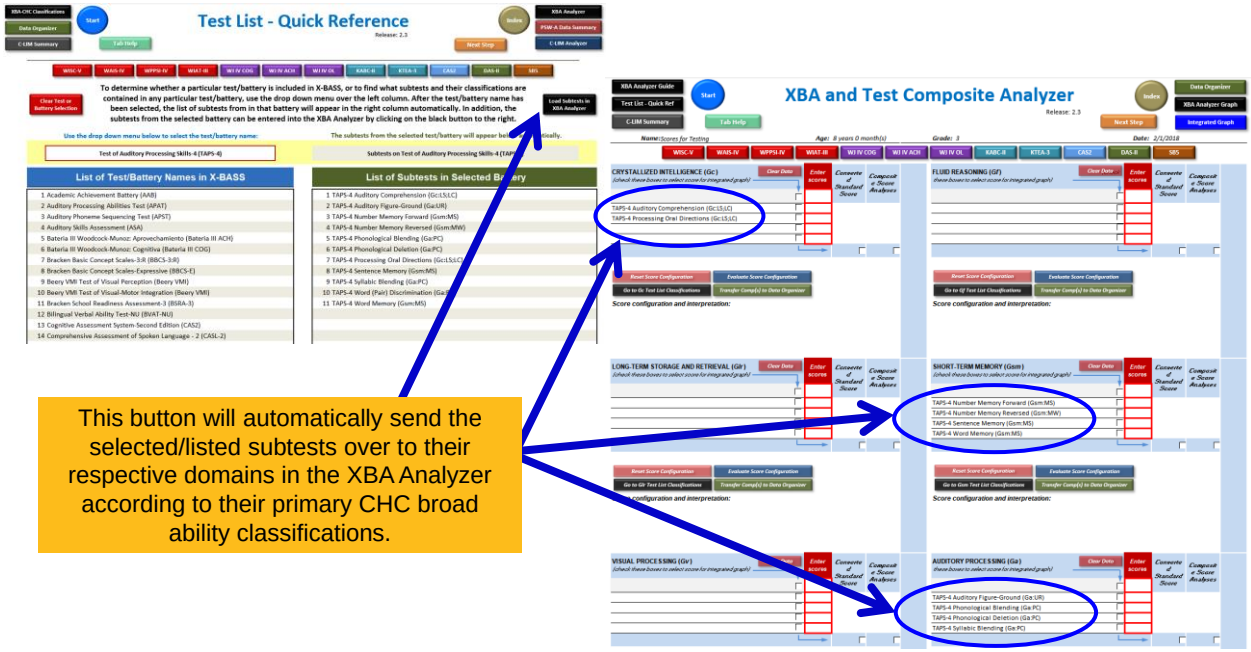


X-BASS v2.3 New Features: Test List-QR has new tests/batteries and auto display of subtests

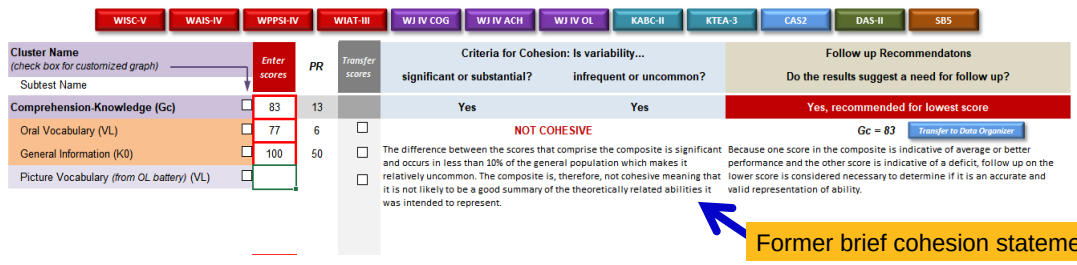


Updated the Test Database with several new tests including: APST, CVLT-3, DTLA-5, EFT-E:NU, EVT-3, MFVPT-4, PPVT-5, PAT-2:NU, TAPS-4, TVPS-4, TOLD:P-5, TNL-2, WORD-3:E, YCAT-2, WISC-V Spanish, and WRAT-5. There are now 148 tests/batteries and 1,175 subtests classified in X-BASS.

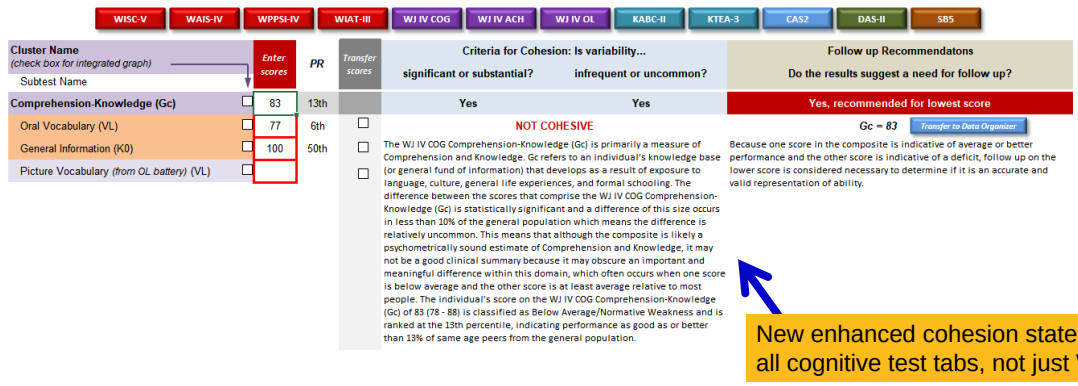
X-BASS v2.3 New Features: Auto loading of subtests from Test List-QR to XBA Analyzer



X-BASS v2.3 New Features: Enhanced Cohesion Statements



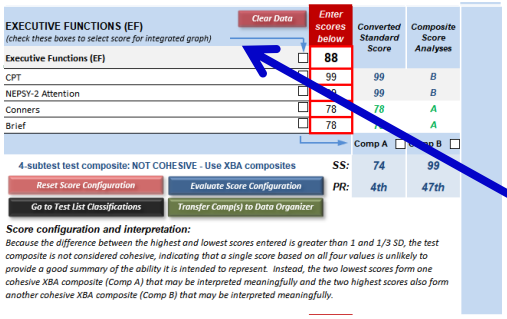
Former brief cohesion statements.



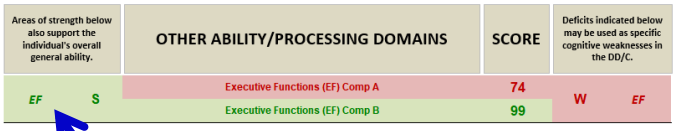
New enhanced cohesion statements on all cognitive test tabs, not just WISC-V.

X-BASS v2.3 New Features: Selectable/modifiable "other ability" domain

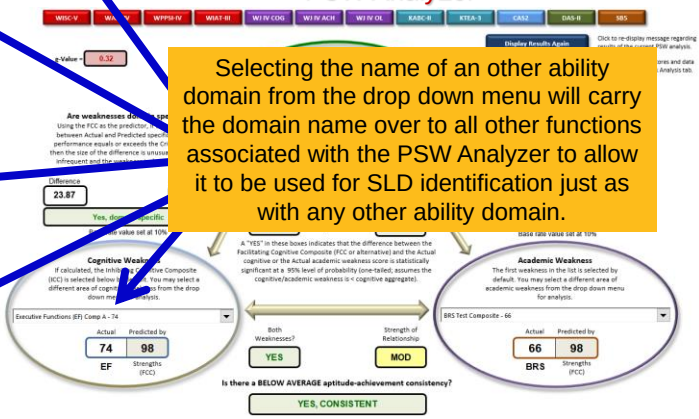
XBA Analyzer



PSW-A Data Summary

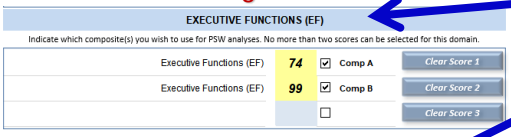


PSW Analyzer

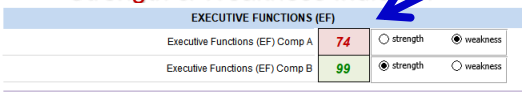


Selecting the name of an other ability domain from the drop down menu will carry the domain name over to all other functions associated with the PSW Analyzer to allow it to be used for SLD identification just as with any other ability domain.

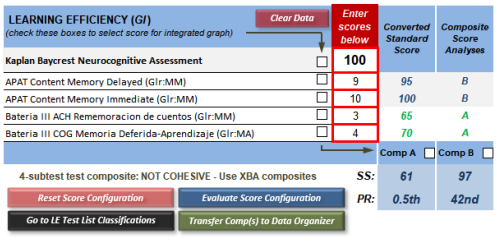
Data Organizer

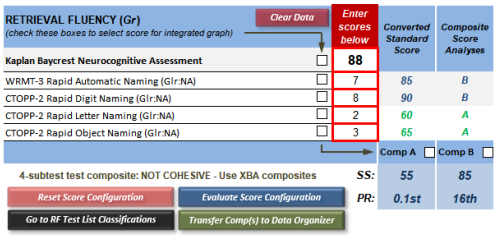


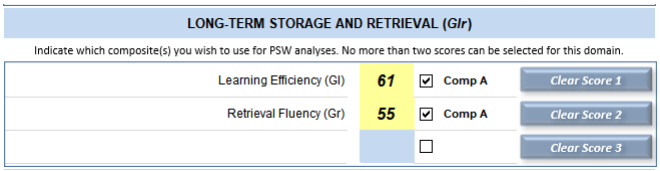
Strength & Weakness Indicator



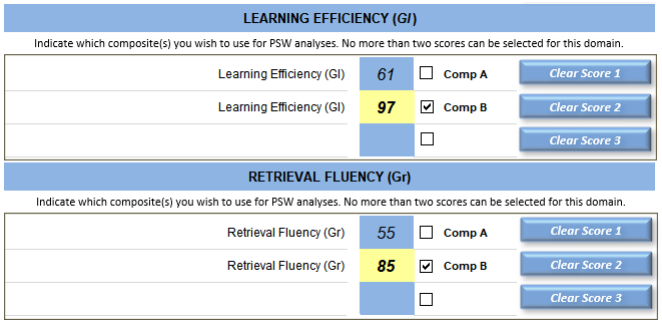
X-BASS v2.3 New Features: Separation of Glr into Gl and Gr



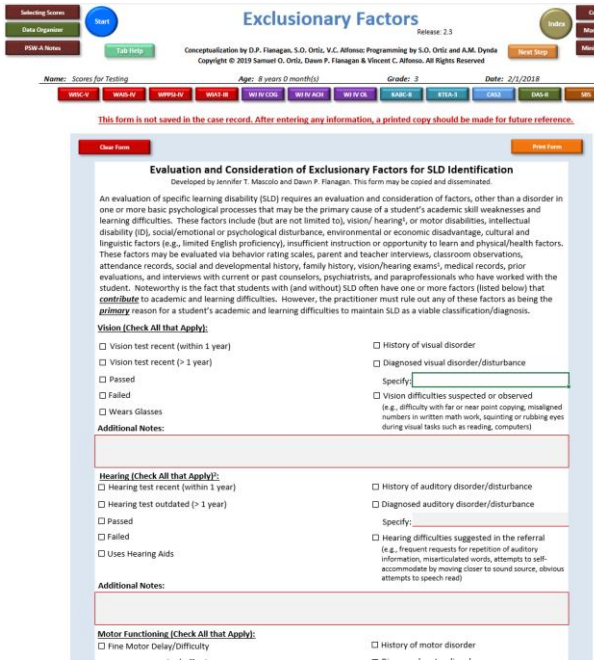




Gl (learning efficiency) and Gr (retrieval fluency) scores can be transferred to either the Gl and Gr domains in the "neuropsych/other cognitive" section or into the broad Glr domain, or both.

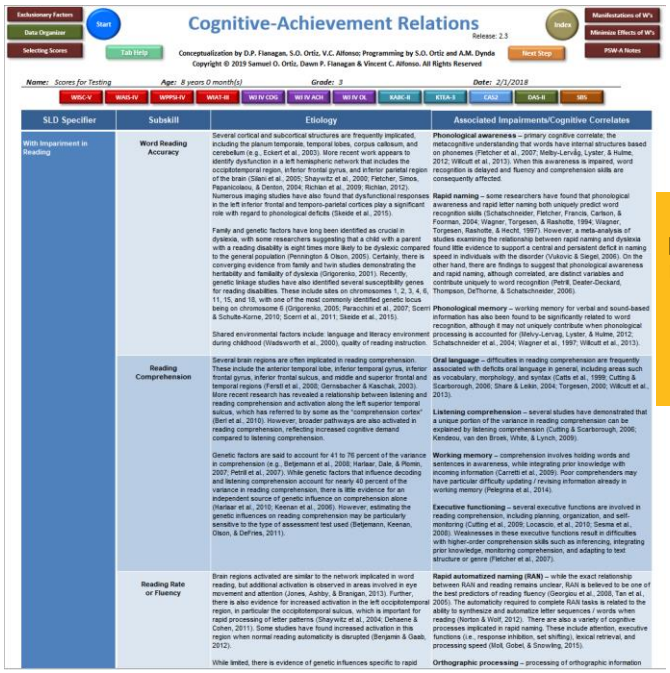


X-BASS v2.3 New Features: Exclusionary Factors Form Tab



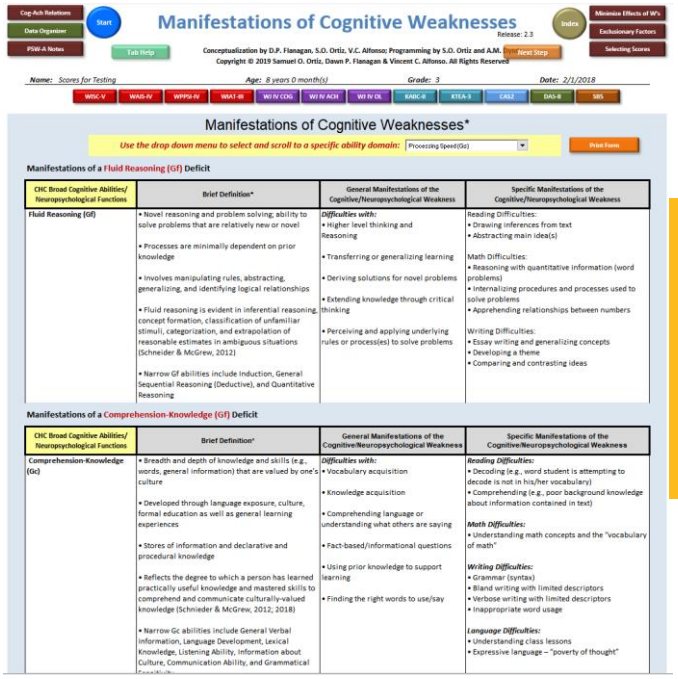
Simply check off the appropriate boxes, enter any additional information, including notes, and click the Print Form button to print out a completed form that examines and considers all possible exclusionary factors that must be ruled out to diagnose SLD

X-BASS v2.3 New Features: Cognitive-Achievement Relations Tab



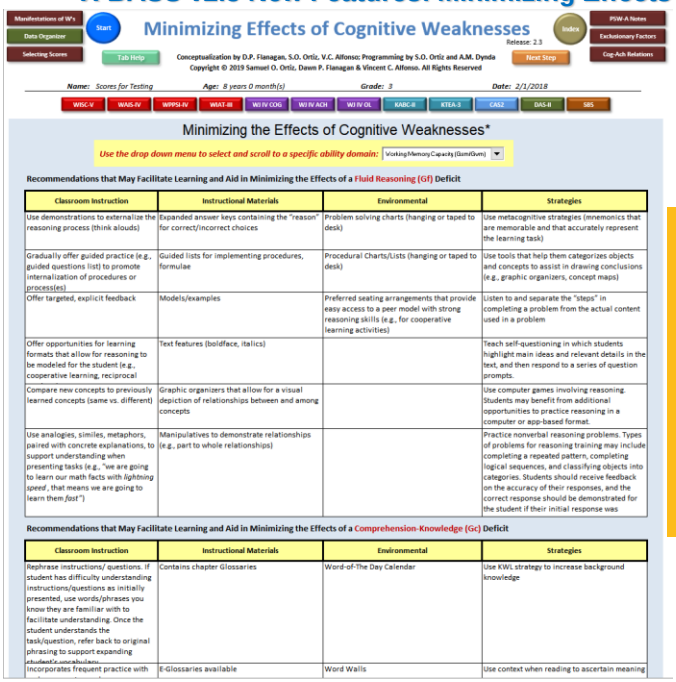
This new tab contains a table that provides information regarding the relationship between an academic area (and subskill) to specific areas of cognitive functioning. An explanation of the possible etiology is also provided.

X-BASS v2.3 New Features: Manifestations of Cognitive Weaknesses Tab



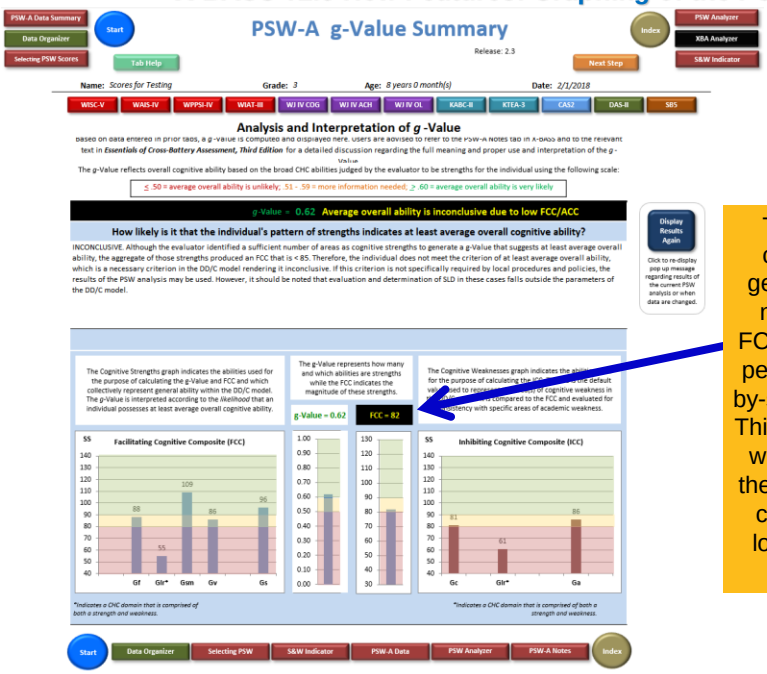
This new tab contains a table that provides a definition of academically-related cognitive abilities as well as their general and specific manifestations in terms of academic functioning and skills acquisition. The table can be quickly navigated by selecting the cognitive domain from the drop down menu at the top.

X-BASS v2.3 New Features: Minimizing Effects of Cognitive Weaknesses Tab



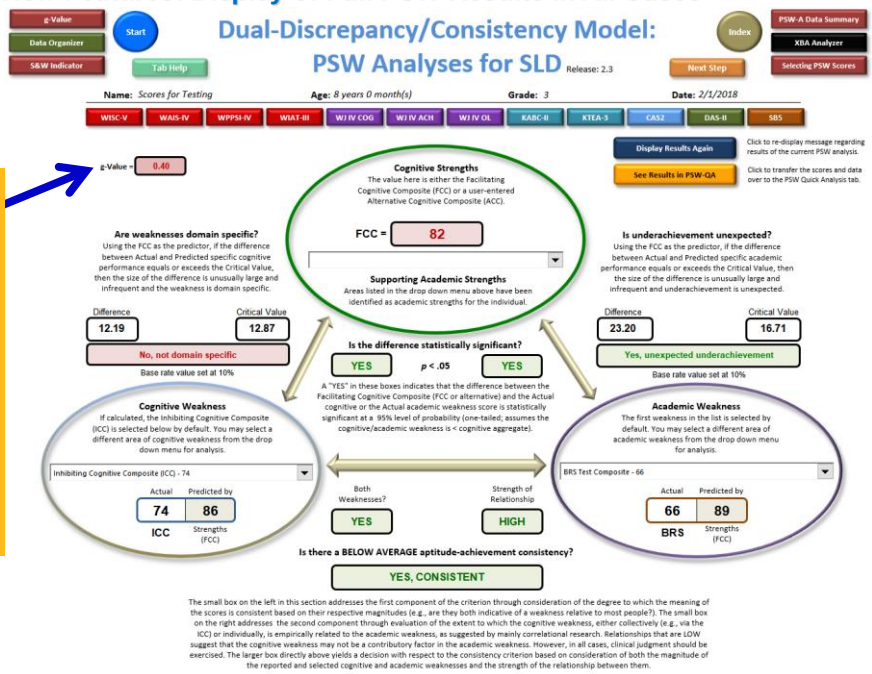
This new tab contains a table that provides information regarding instructional, environmental, and other strategies for minimizing the effects of cognitive weaknesses which may be helpful in determining appropriate avenues for intervention. The table can be quickly navigated by selecting the cognitive domain from the drop down menu at the top.

X-BASS v2.3 New Features: Graphing of the FCC on the g-Value Tab



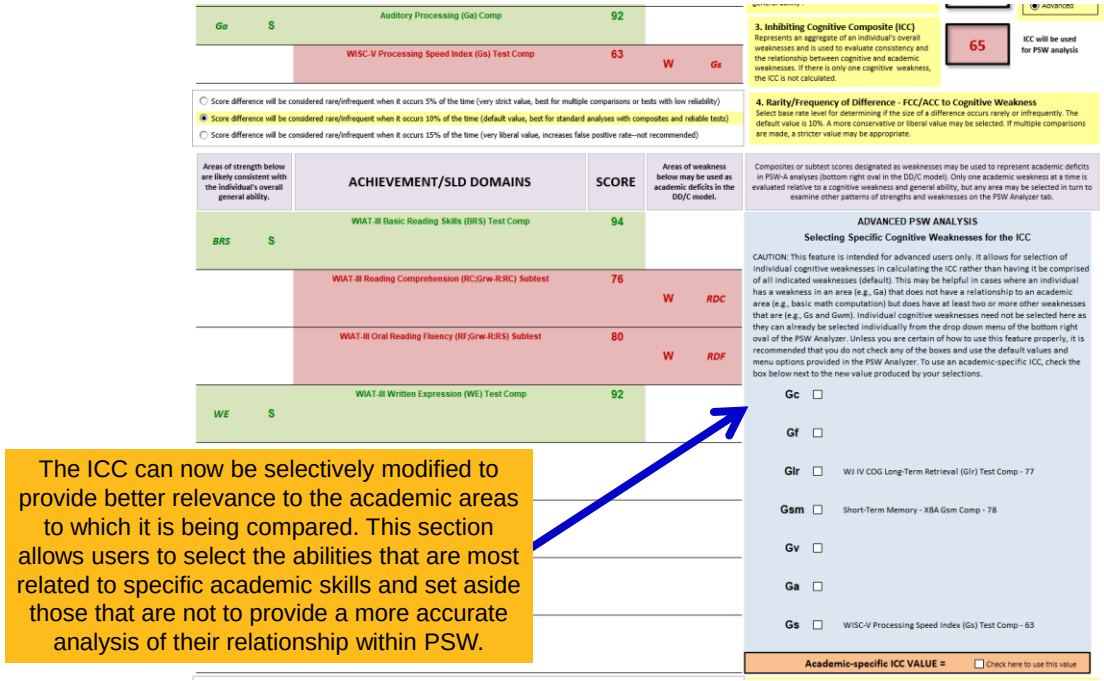
To assist in determining the criterion for overall average general ability, the g-Value tab now provides a graph of the FCC or ACC value in a way that permits consideration in a side-by-side manner with the g-Value. This is especially useful in cases where the g-Value is good but the FCC may be less than 85 or conversely, when the FCC is lower than .51 but the FCC is greater than 85.

X-BASS v2.3 New Features: Display of Full PSW Results in All Cases



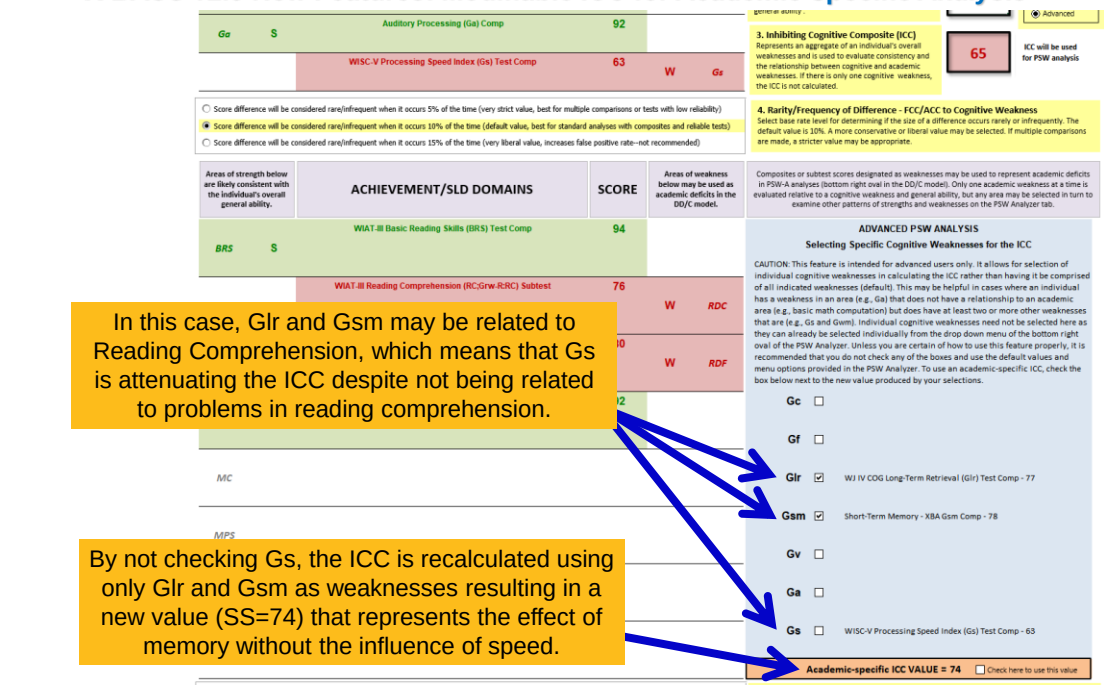
In previous versions, no PSW results were displayed if the g-Value was below .51 (except when Gf and Gc were indicated as the only two strengths). Now, the PSW results are displayed regardless of the g-Value as a way of helping practitioners determine where problems in PSW analyses exist.

X-BASS v2.3 New Features: Modifiable ICC for Academic Specific Analysis



The ICC can now be selectively modified to provide better relevance to the academic areas to which it is being compared. This section allows users to select the abilities that are most related to specific academic skills and set aside those that are not to provide a more accurate analysis of their relationship within PSW.

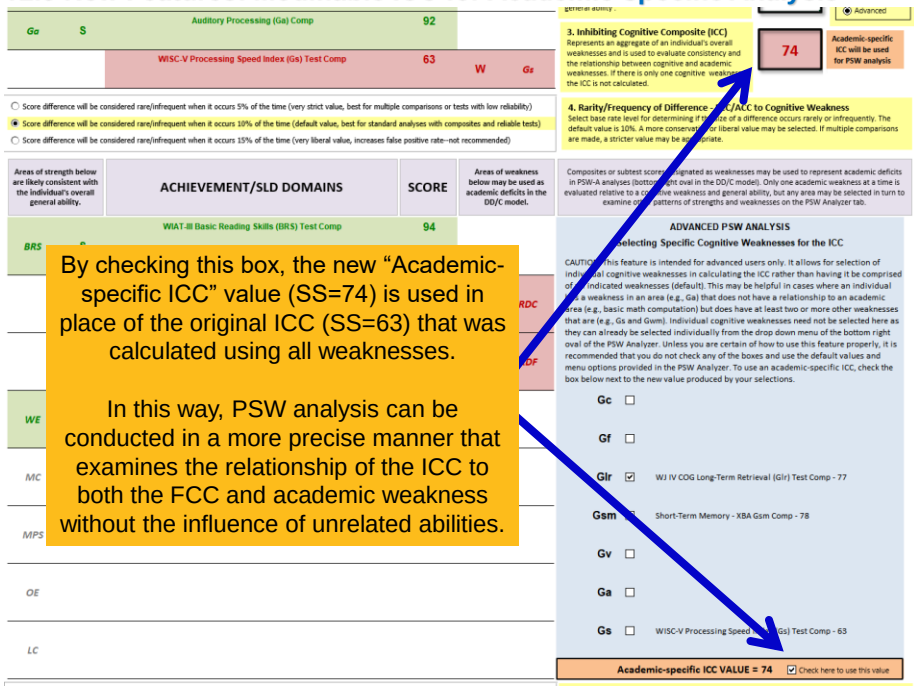
X-BASS v2.3 New Features: Modifiable ICC for Academic Specific Analysis



In this case, Glr and Gsm may be related to Reading Comprehension, which means that Gs is attenuating the ICC despite not being related to problems in reading comprehension.

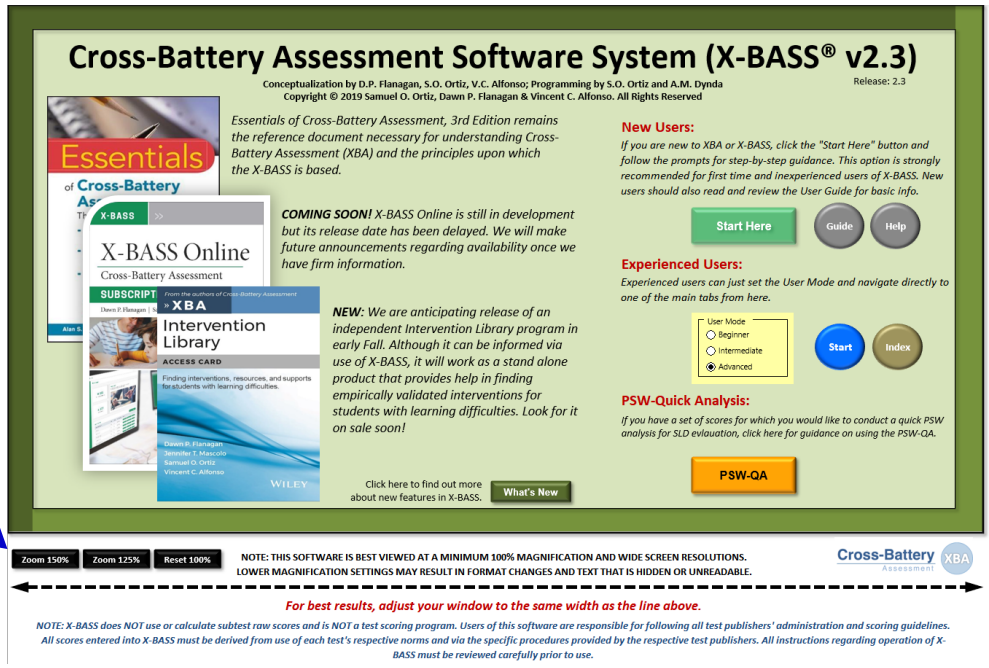
By not checking Gs, the ICC is recalculated using only Glr and Gsm as weaknesses resulting in a new value (SS=74) that represents the effect of memory without the influence of speed.

X-BASS v2.3 New Features: Modifiable ICC for Academic Specific Analysis

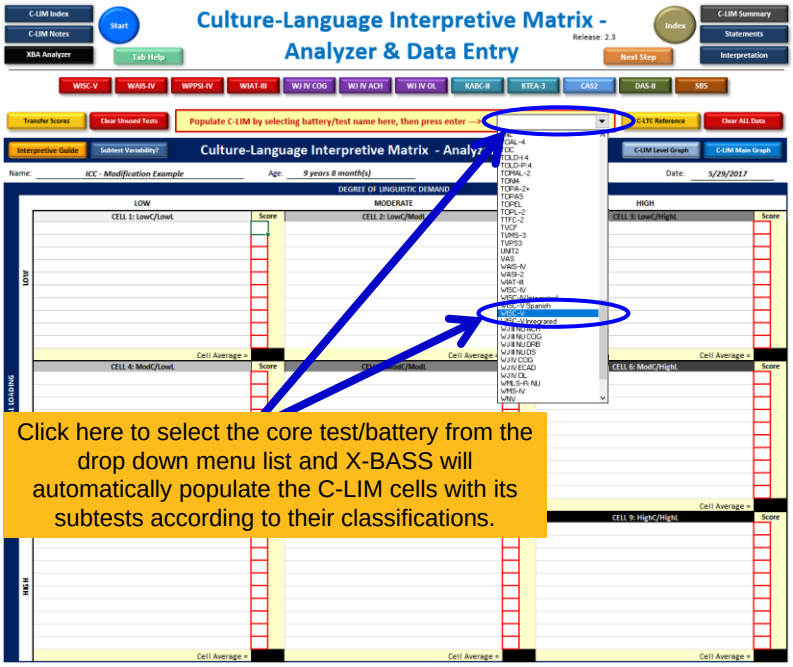


X-BASS v2.3 New Features: Buttons to Auto-Zoom (enlarge and reset) Display

These buttons will zoom all tabs in X-BASS making it easier to read. The reset button will return all tabs to 100%, which is the default and standard view.

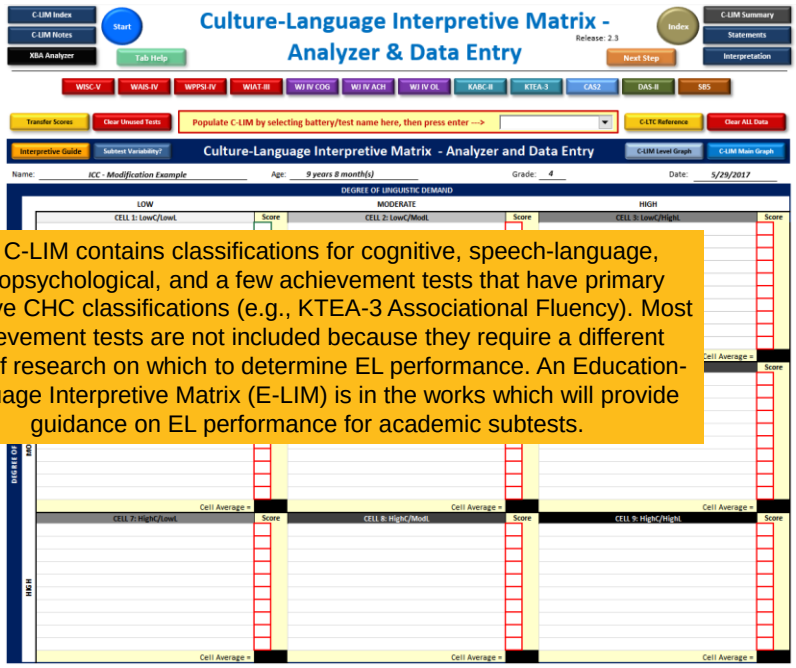


X-BASS v2.2 and v2.3: C-LIM Updates



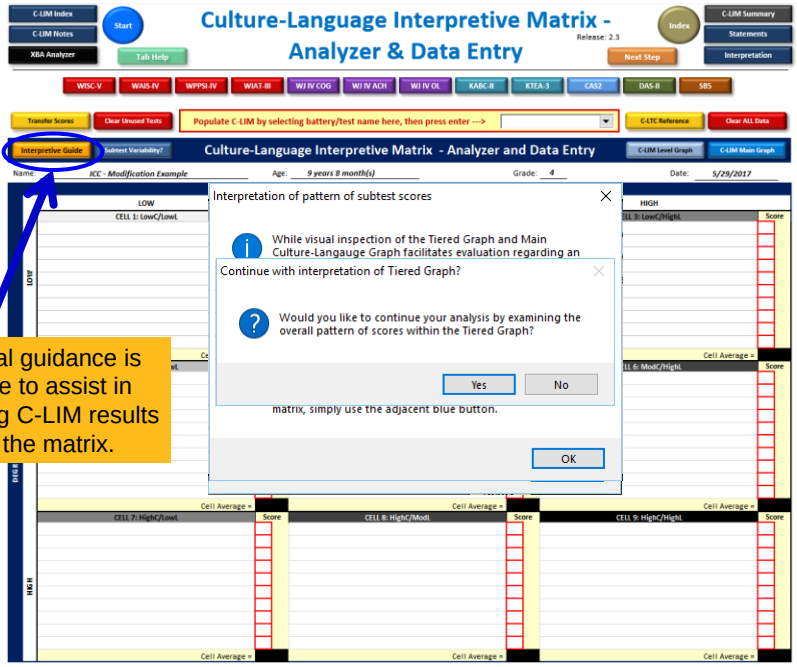
Click here to select the core test/battery from the drop down menu list and X-BASS will automatically populate the C-LIM cells with its subtests according to their classifications.

X-BASS v2.2 and v2.3: C-LIM Updates



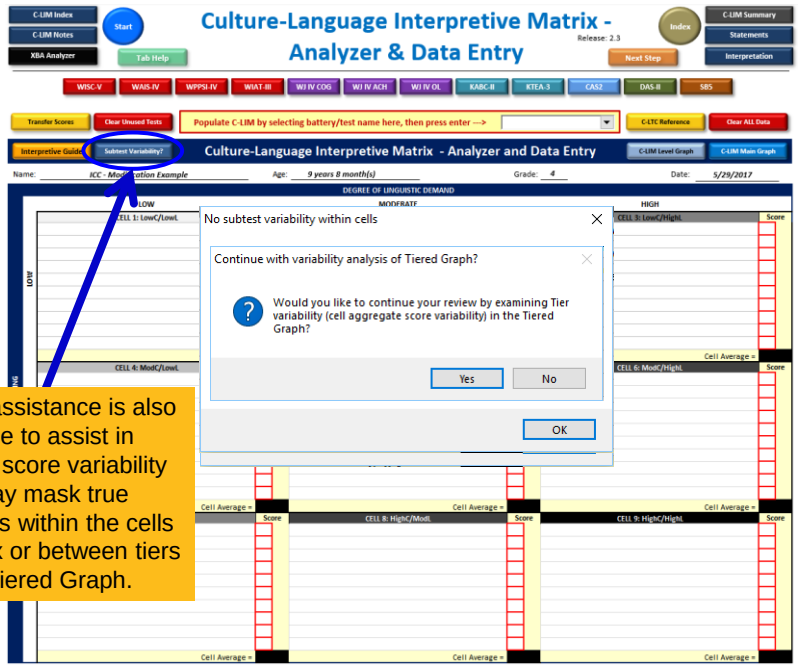
The C-LIM contains classifications for cognitive, speech-language, neuropsychological, and a few achievement tests that have primary cognitive CHC classifications (e.g., KTEA-3 Associational Fluency). Most achievement tests are not included because they require a different body of research on which to determine EL performance. An Education-Language Interpretive Matrix (E-LIM) is in the works which will provide guidance on EL performance for academic subtests.

X-BASS v2.2 and v2.3: C-LIM Updates



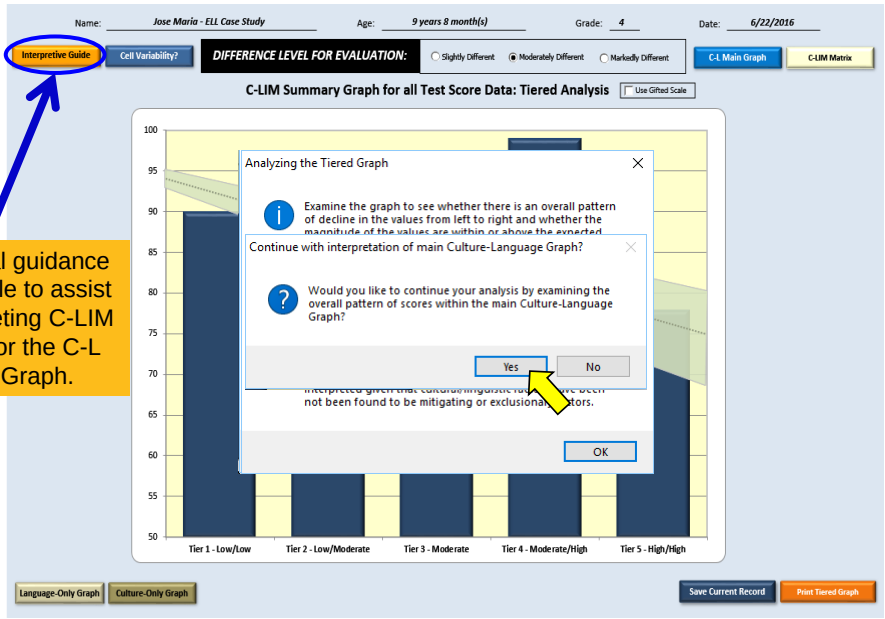
Additional guidance is available to assist in interpreting C-LIM results within the matrix.

X-BASS v2.2 and v2.3: C-LIM Updates



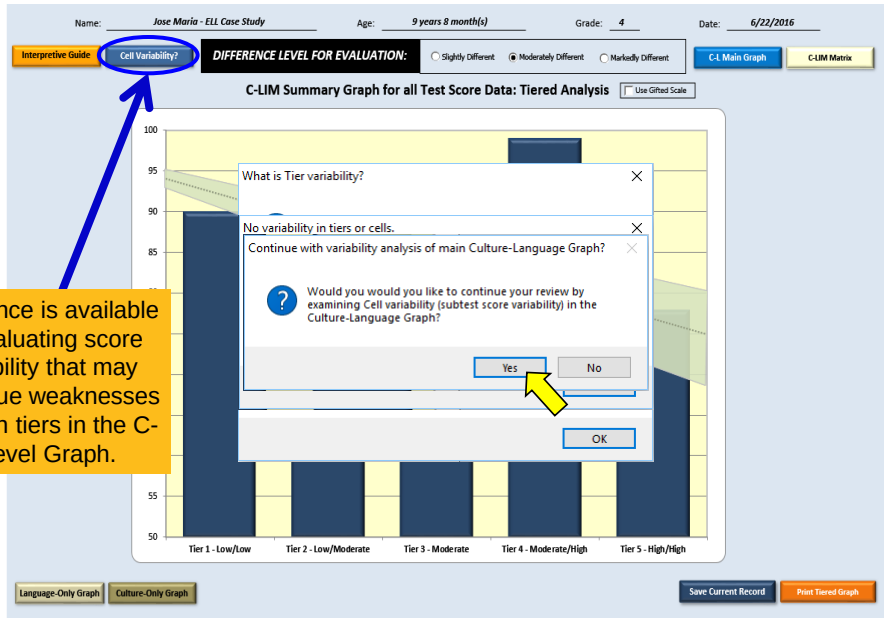
Additional assistance is also available to assist in evaluating score variability that may mask true weaknesses within the cells in the matrix or between tiers in the Tiered Graph.

X-BASS v2.2 and v2.3: C-LIM Updates



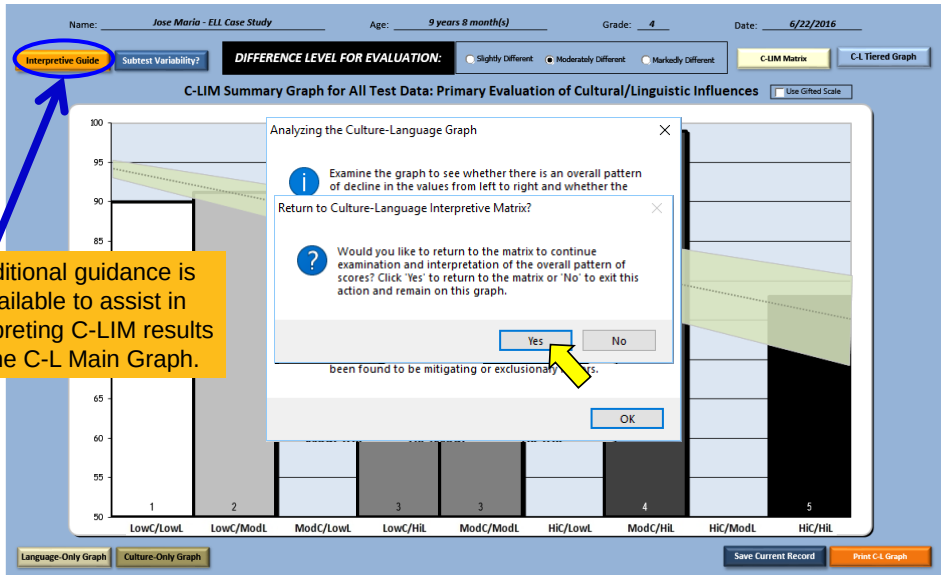
Additional guidance is available to assist in interpreting C-LIM results for the C-L Level Graph.

X-BASS v2.2 and v2.3: C-LIM Updates



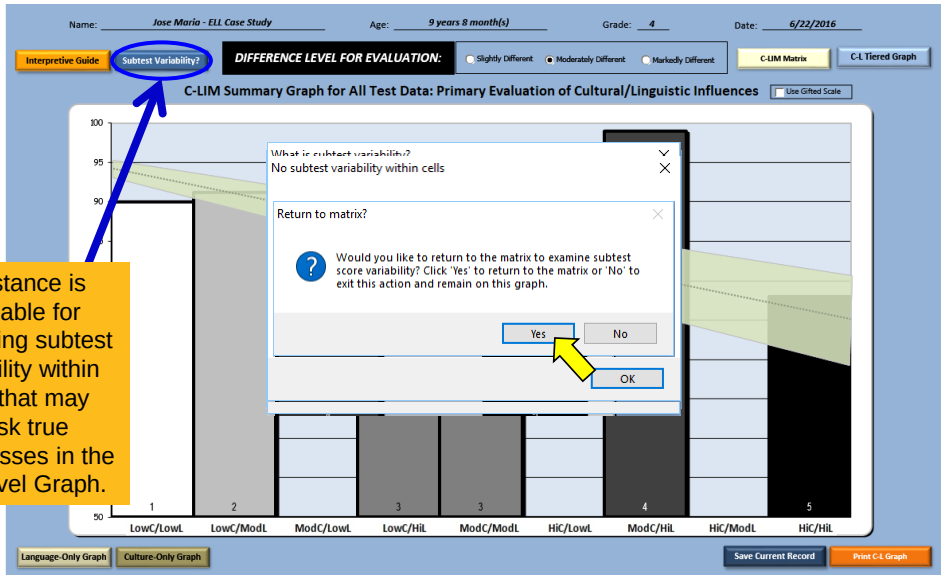
Assistance is available for evaluating score variability that may mask true weaknesses between tiers in the C-L Level Graph.

X-BASS v2.2 and v2.3: C-LIM Updates



Additional guidance is available to assist in interpreting C-LIM results in the C-L Main Graph.

X-BASS v2.2 and v2.3: C-LIM Updates



Assistance is available for evaluating subtest variability within cells that may mask true weaknesses in the C-L Level Graph.

X-BASS v2.2 and v2.3: C-LIM Updates

Statement 1. Evaluations of Suspected Learning Disability - Invalid Results

The following sample validity statement is appropriate for cases where there is an overall declining pattern and the magnitude of the scores are generally within the selected range of difference. In such cases, the effect of culture and language is primary, the results are NOT likely to be valid, and performance suggests average functioning.

Simplified Statement:

Because the student 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.

The student'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 the student'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 the student's performance is also average (possibly higher) and that it is not likely that a learning disability is present in this case. This means that although the student 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.

Detailed Statement:

Because the student is not a native English speaker, it is nec[illegible]sary 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 [illegible]is e[illegible]ative knowledge and English language proficiency was carried out via use [illegible]

A careful review of the student's test data, as entered into the C-LI[illegible]viduals with similar cultural and linguistic backgrounds. This overall, declining patte[illegible]nguistic factors rather than lack of actual ability. Accordingly, the test results evaluated[illegible]ities that were the focus of the evaluation. However, given that the observed pattern [illegible]abled individuals with comparable linguistic development and educational e[illegible]e also at least within the average range of performance (or possibly higher) and stron[illegible]l that the academic difficulties observed in classroom performance that prompted this evaluation are most likely to attributable primarily to the normal process of second language and acculturative knowledge acquisition.

In summary, the observed pattern of the student's test results is consistent with performance that is typical of culturally and linguistically diverse individuals of similar backgrounds who are not disabled and possess average general ability or higher. Therefore, it can be reasonably concluded that the test data evaluated with the C-LIM are likely to be invalid due to the presence of overarching cultural and linguistic influences and suggest that the student's test performance can not be used to support the presence of any type of learning disability.

New, simplified validity statements for use with the C-LIM are provided alongside the previous detailed statements. These may be more helpful in explaining procedures, results, and interpretation within written reports in comparison to the more detailed and technical versions.

A Guided Case Study Example of Evaluation of an English Learner for Specific Learning Disability

Evaluation of Maria Ayala
Tests Used: WISC-V, WIAT-III, and WJ IV
DOE: 5/29/2017
DOB: 9/6/2007
Grade: 4

Multilingual Assessment of ELs: Step by Step

Step 1. Test first in English (L2) and evaluate construct validity in all areas in English (exclusion of cultural/linguistic factors)

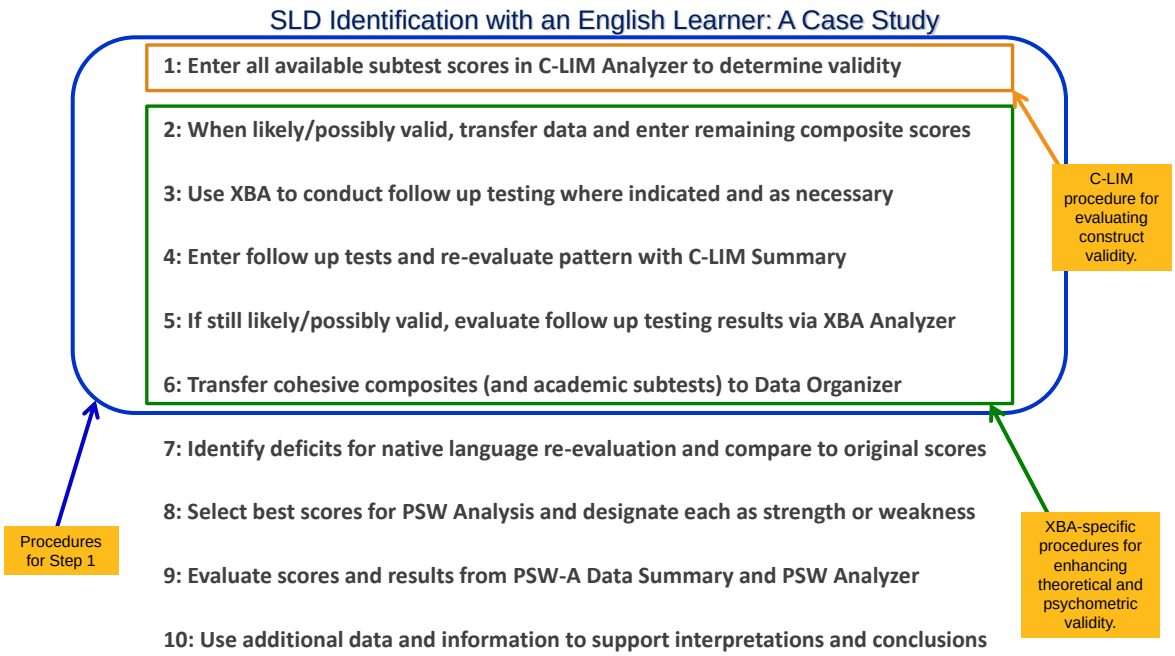
- *If all scores indicate normative strengths (SS ≈ 90 or higher) when tested in English (L2), scores are valid to the extent that a disability is not likely, thus no further testing is necessary.*
- *If some scores are normative weaknesses (SS < ≈ 90) evaluate test score validity in a research-based manner, e.g., via the C-LIM.*
- *If C-LIM indicates primary influence of language/culture, test scores are likely invalid and indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.*
- *If C-LIM indicates contributory or minimal influence of language/culture, test scores are likely to be valid and the evaluation should continue.*

Step 2. Re-evaluate areas of weakness in native language (L2) to provide additional supporting evidence of validity (cross-linguistic confirmation)

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

Step 3. Further cross-validate L1 and L2 test scores with contextual factors and pre-referral data and academic concerns (ecological validity for disability)

- *Use all other case data and information to serve as the context by which to evaluate the test scores and ensure ecological validity to conclusions*



SLD Identification with an English Learner: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 5/29/2017 DOB: 9/6/2007 Grade: 4

WECHSLER INTELLIGENCE 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

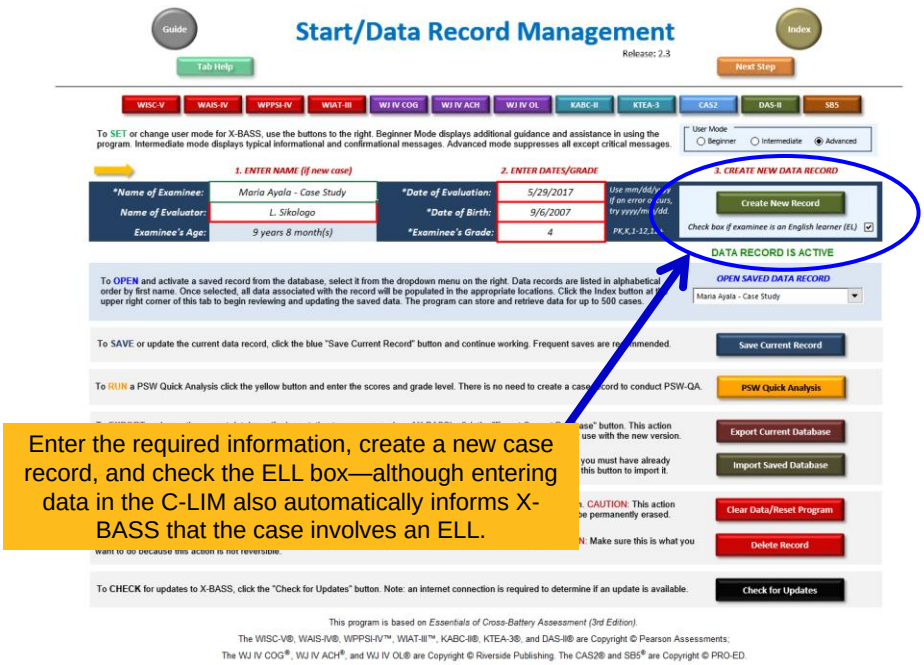
Reading Comprehension	*76*
Reading Comprehension	76
Oral Reading Fluency	80

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

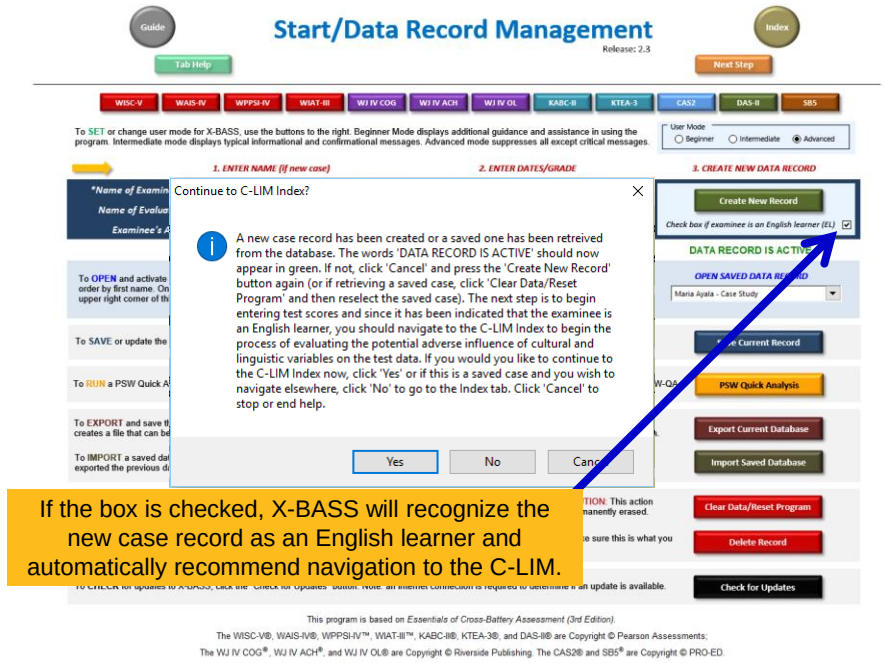
WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

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

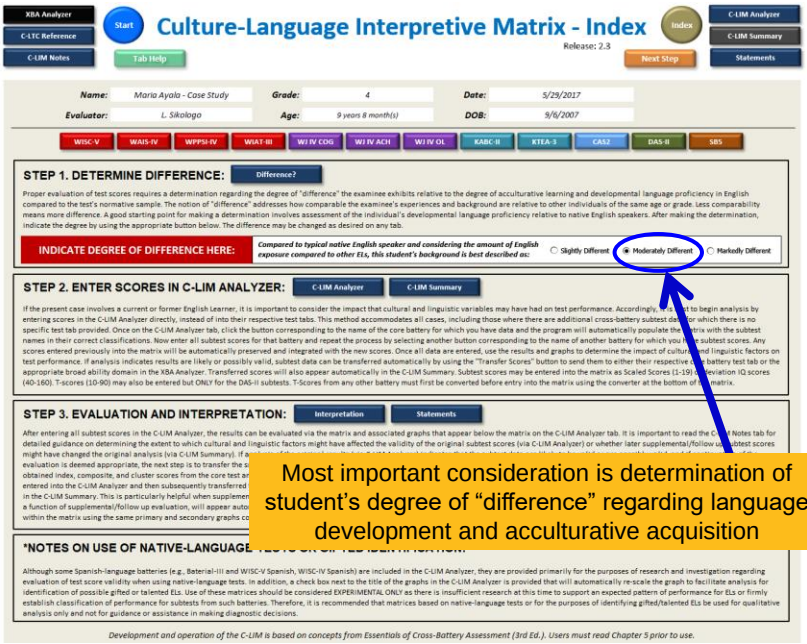


Enter the required information, create a new case record, and check the ELL box—although entering data in the C-LIM also automatically informs X-BASS that the case involves an ELL.



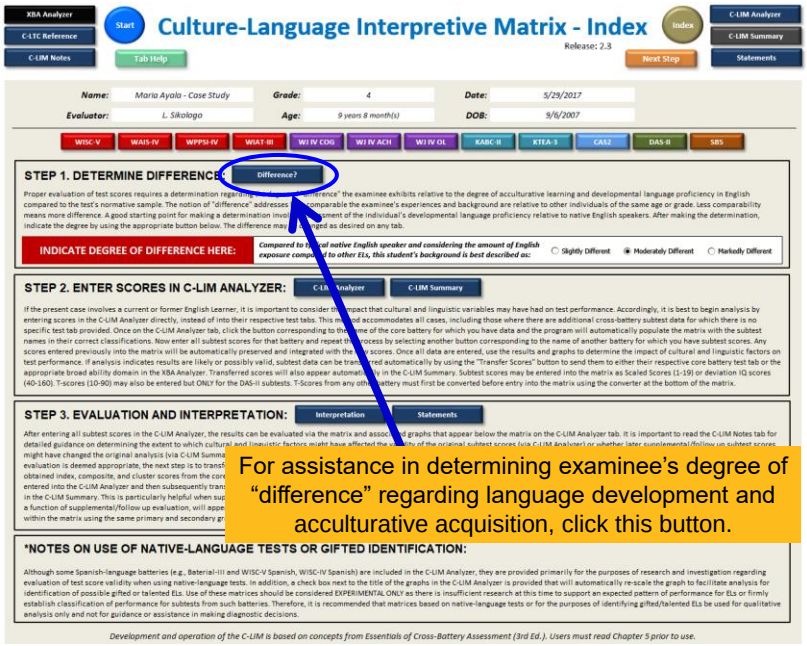
If the box is checked, X-BASS will recognize the new case record as an English learner and automatically recommend navigation to the C-LIM.

SLD Identification with an English Learner: A Case Study



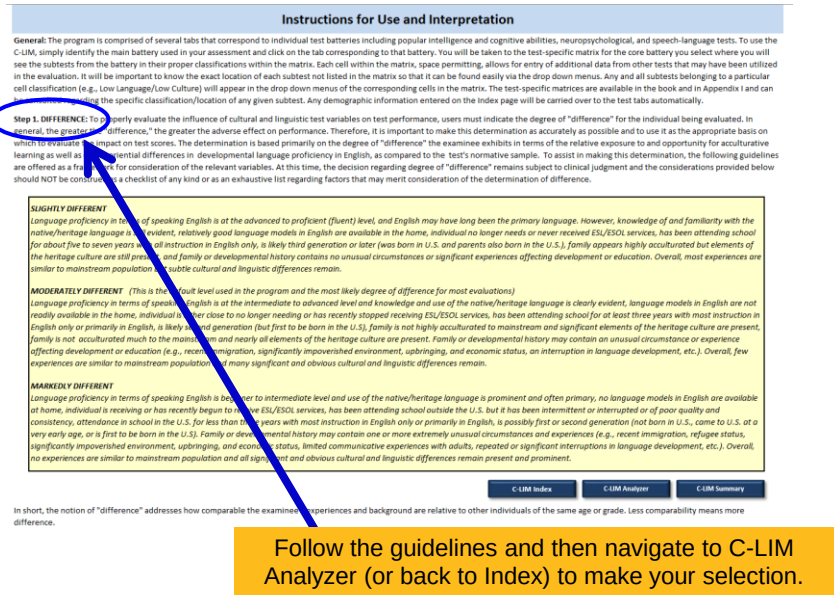
Most important consideration is determination of student's degree of "difference" regarding language development and acculturative acquisition

SLD Identification with an English Learner: A Case Study



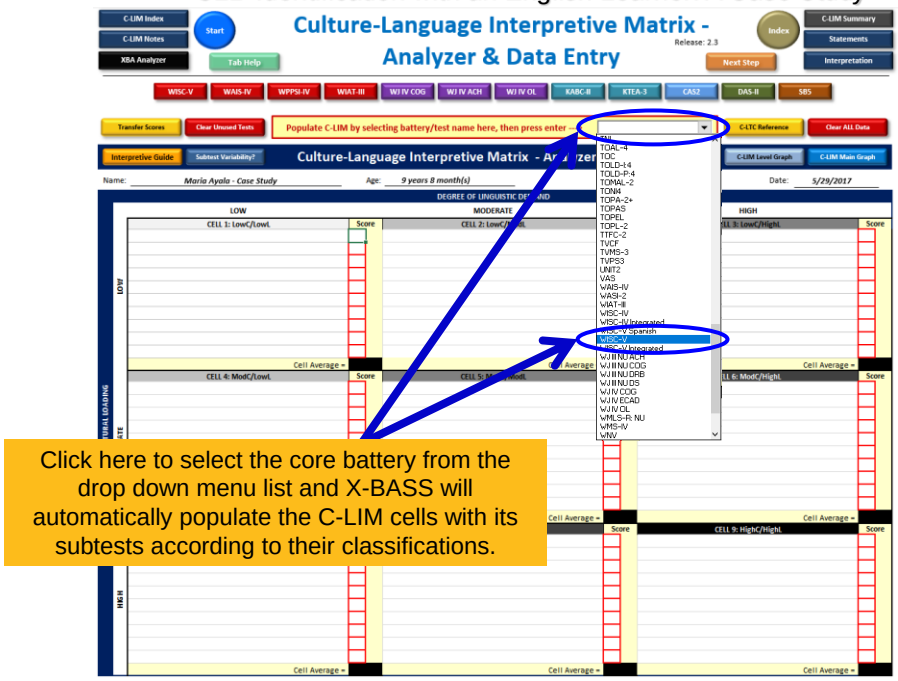
For assistance in determining examinee's degree of "difference" regarding language development and acculturative acquisition, click this button.

SLD Identification with an English Learner: A Case Study



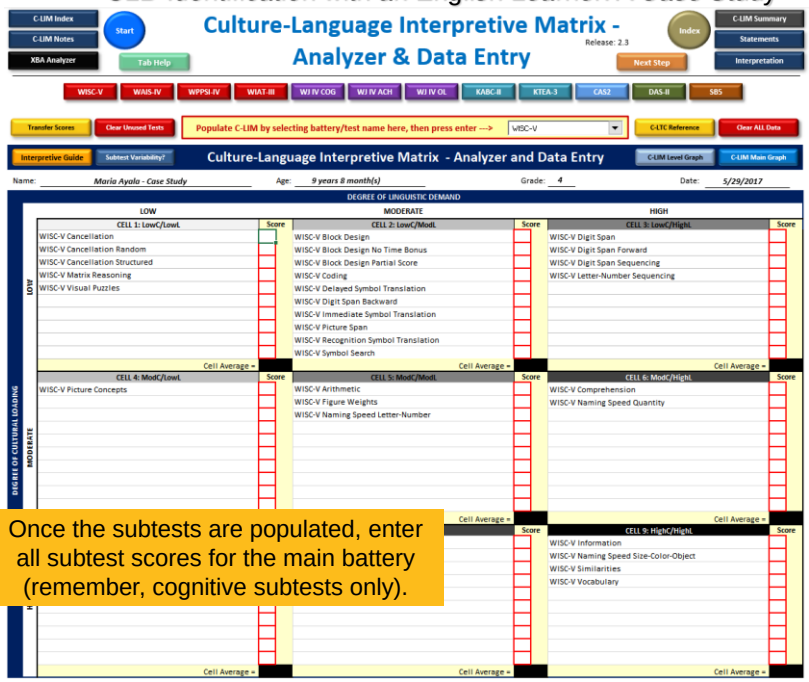
Follow the guidelines and then navigate to C-LIM Analyzer (or back to Index) to make your selection.

SLD Identification with an English Learner: A Case Study



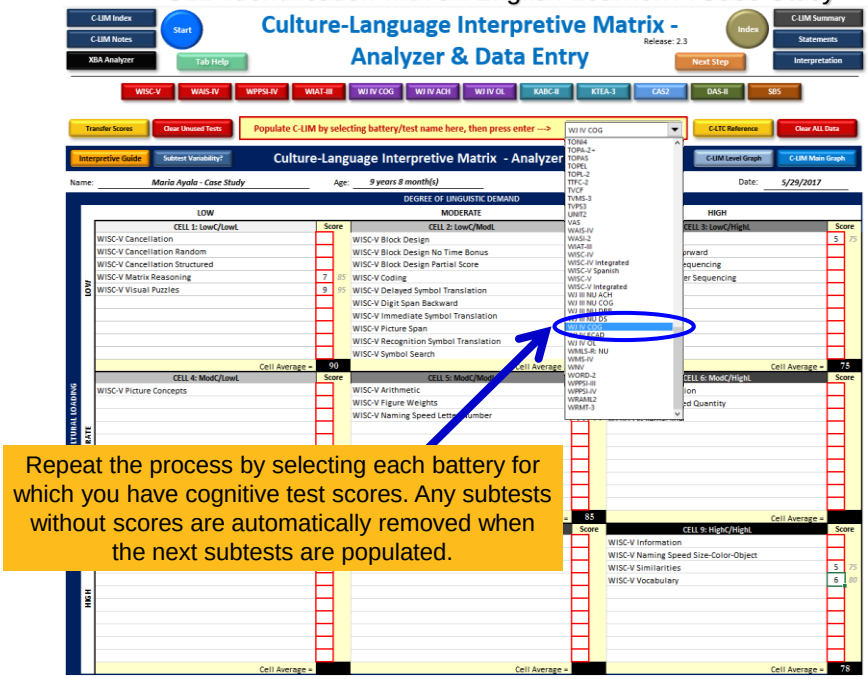
Click here to select the core battery from the drop down menu list and X-BASS will automatically populate the C-LIM cells with its subtests according to their classifications.

SLD Identification with an English Learner: A Case Study



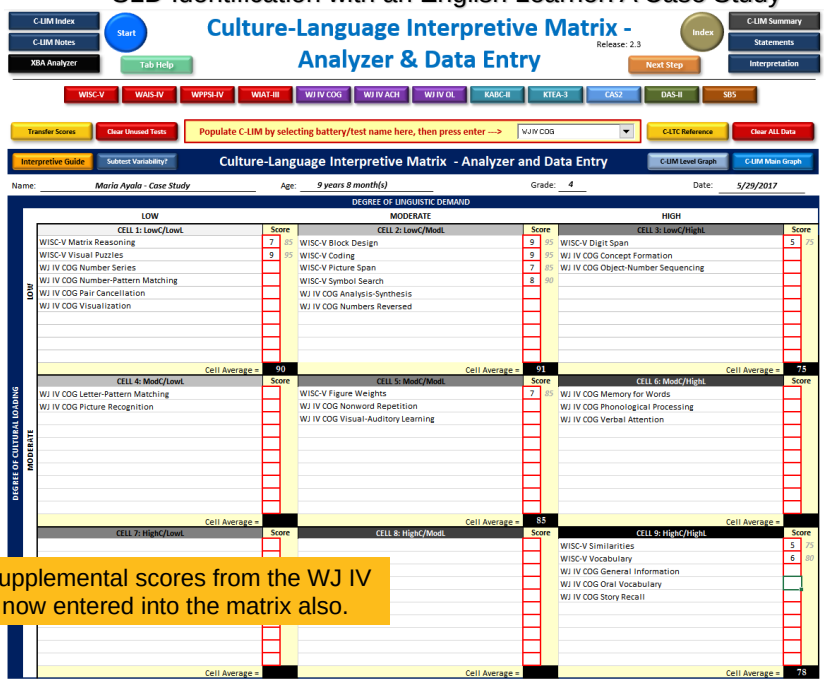
Once the subtests are populated, enter all subtest scores for the main battery (remember, cognitive subtests only).

SLD Identification with an English Learner: A Case Study



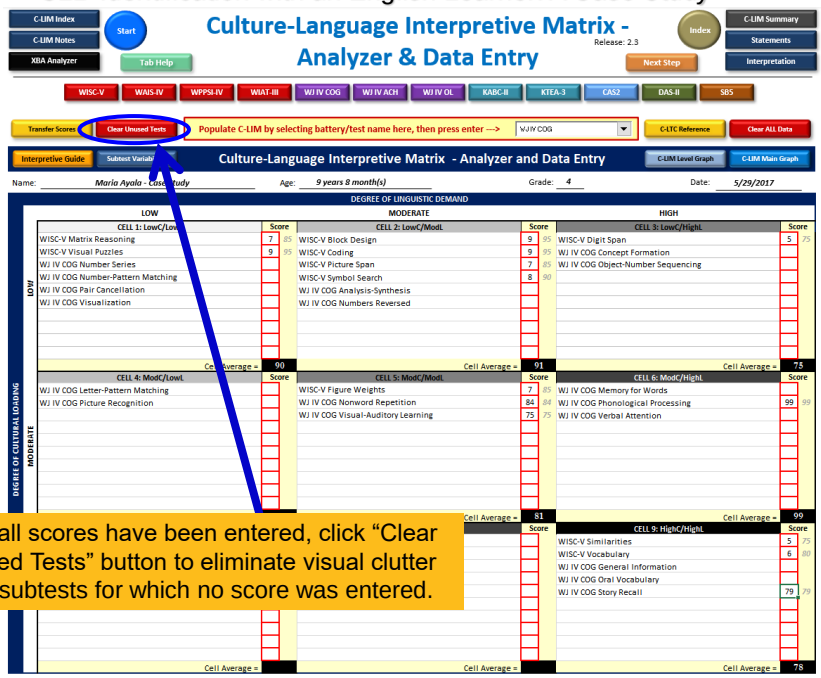
Repeat the process by selecting each battery for which you have cognitive test scores. Any subtests without scores are automatically removed when the next subtests are populated.

SLD Identification with an English Learner: A Case Study



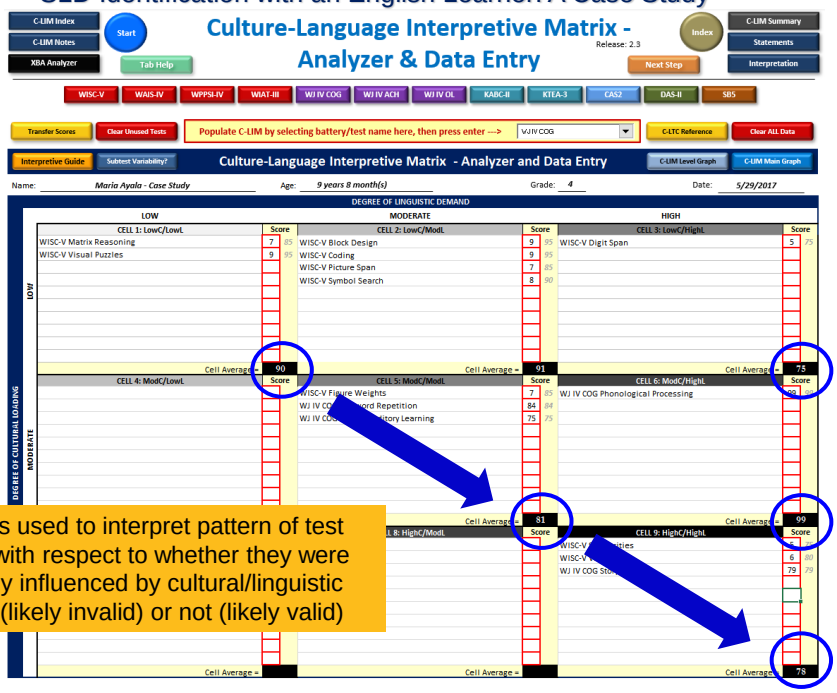
The supplemental scores from the WJ IV are now entered into the matrix also.

SLD Identification with an English Learner: A Case Study



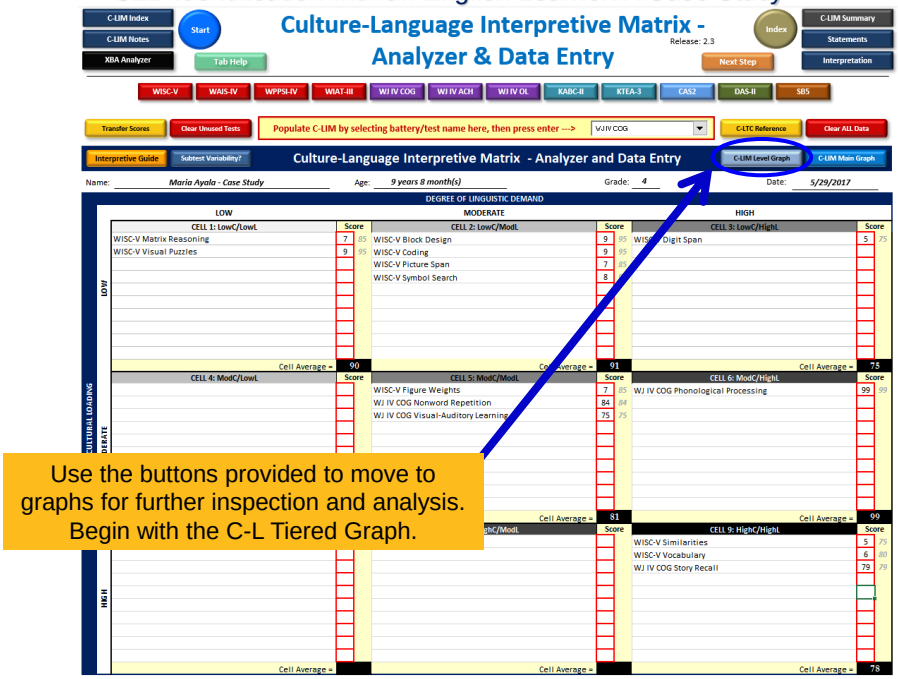
After all scores have been entered, click “Clear Unused Tests” button to eliminate visual clutter from subtests for which no score was entered.

SLD Identification with an English Learner: A Case Study



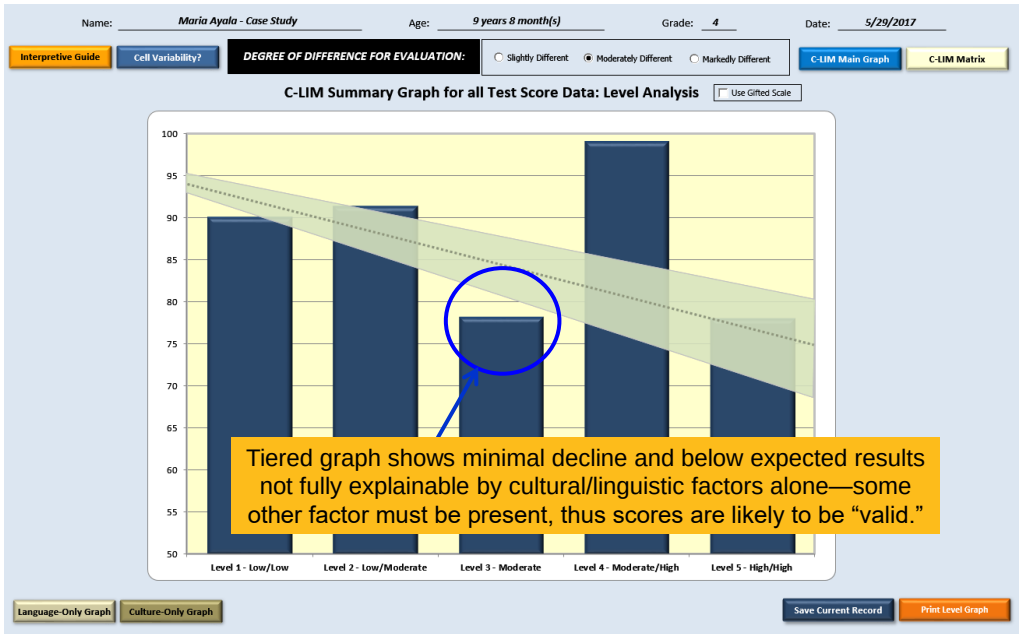
C-LIM is used to interpret pattern of test scores with respect to whether they were primarily influenced by cultural/linguistic factors (likely invalid) or not (likely valid)

SLD Identification with an English Learner: A Case Study



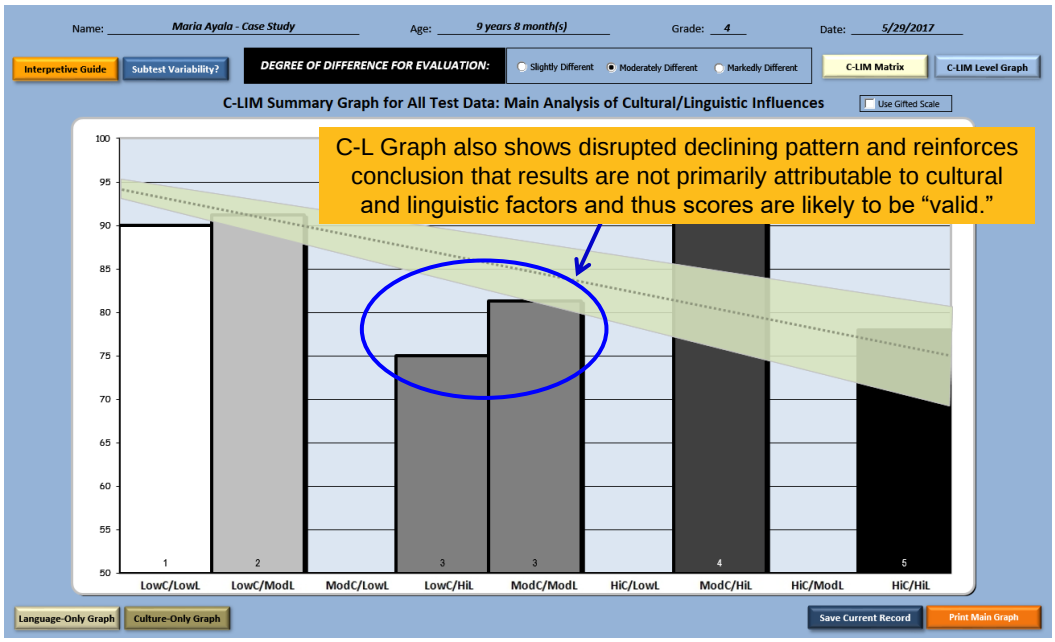
Use the buttons provided to move to graphs for further inspection and analysis. Begin with the C-L Tiered Graph.

SLD Identification with an English Learner: A Case Study

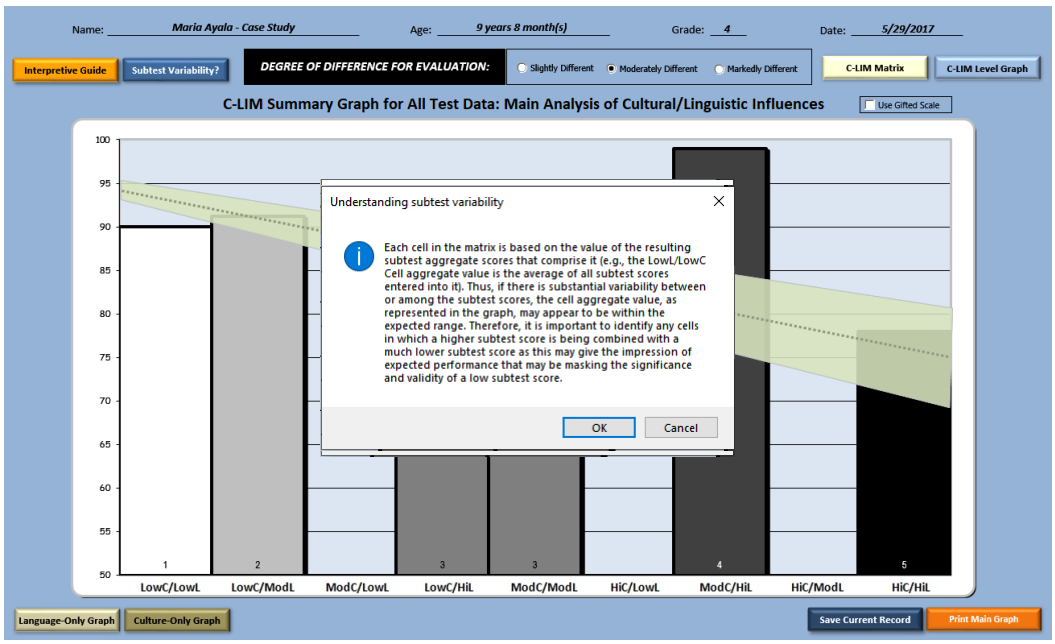


Tiered graph shows minimal decline and below expected results not fully explainable by cultural/linguistic factors alone—some other factor must be present, thus scores are likely to be "valid."

SLD Identification with an English Learner: A Case Study

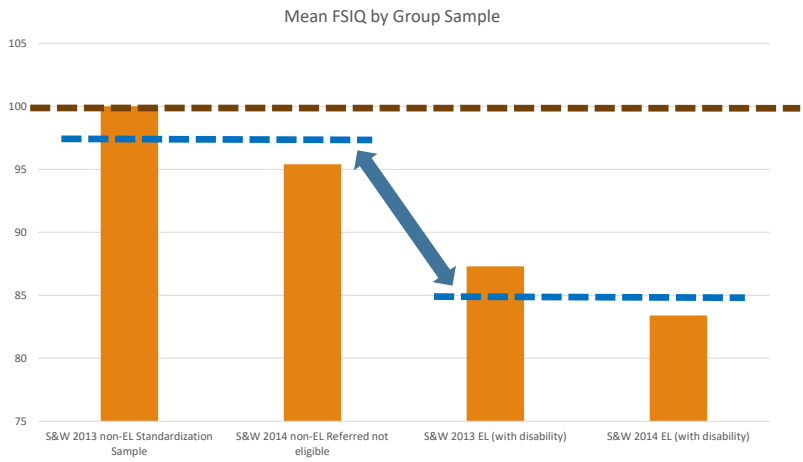


SLD Identification with an English Learner: A Case Study



Research Foundations for EL Evaluation

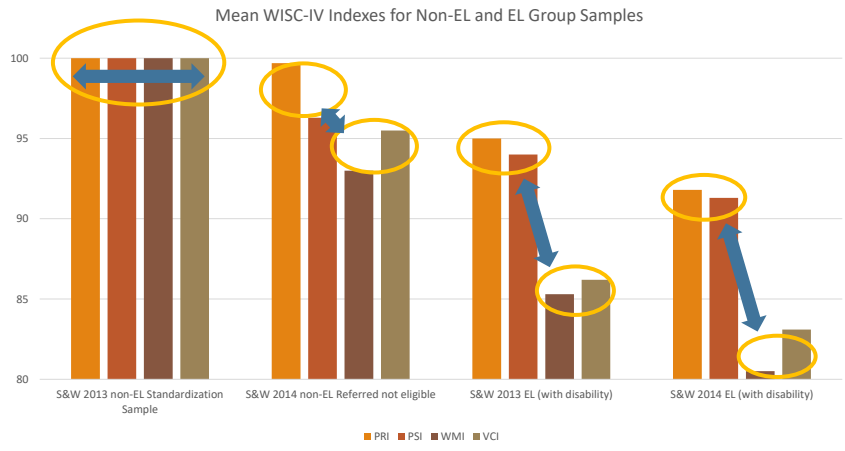
ELs and non-EL's perform differently: Broad ability level



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

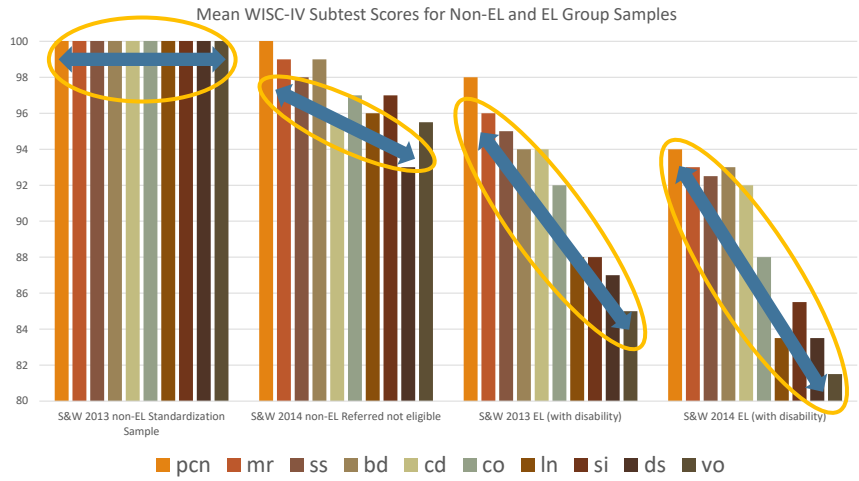
ELs and non-EL's perform differently: Index level



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

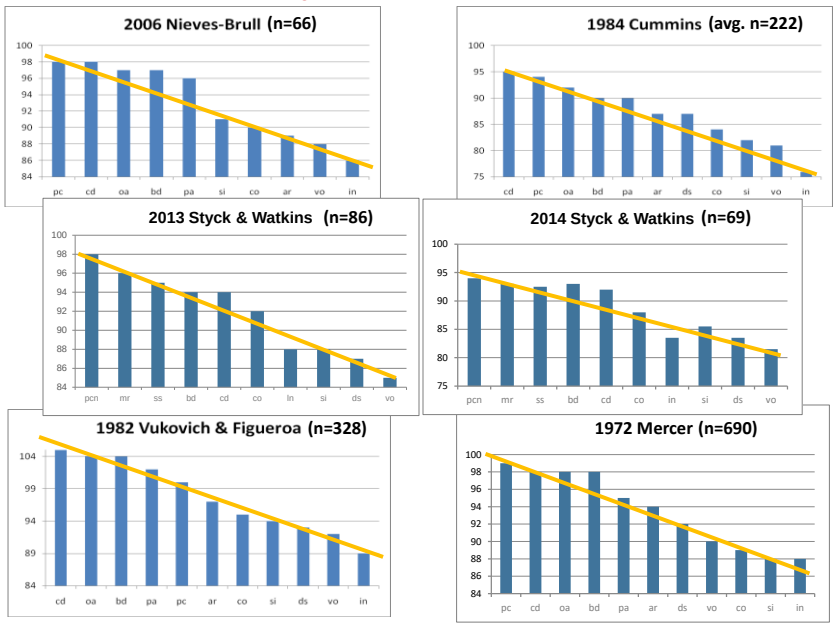
ELs and non-EL's perform differently: Subtest level



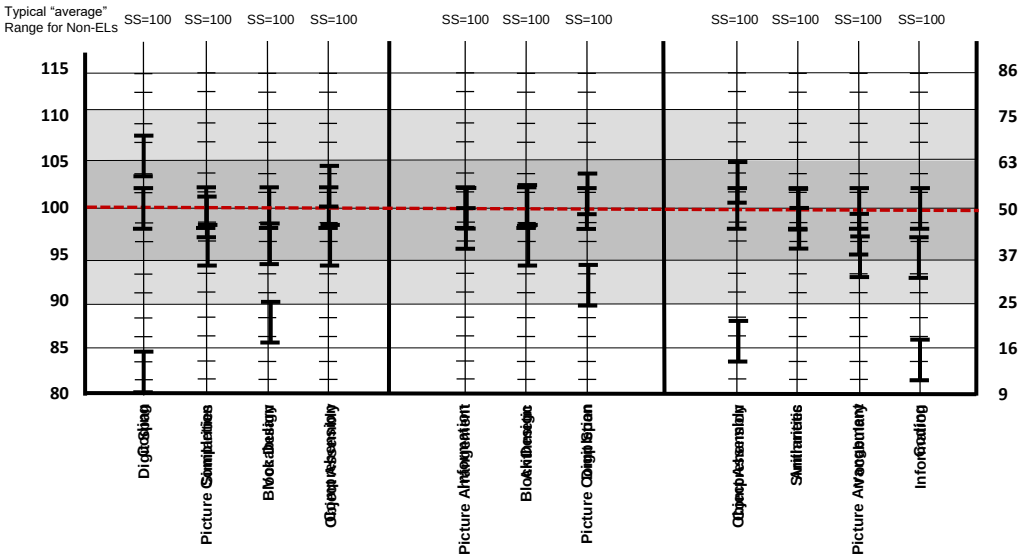
Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.

Research Foundations for EL Evaluation: EL to ES

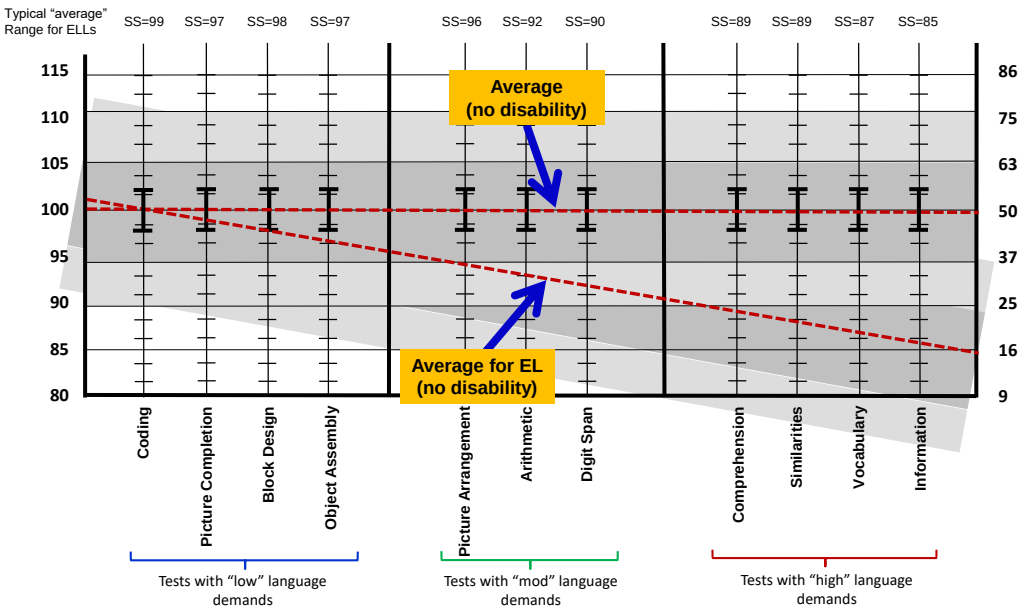
Comparison of overall "average" test performance at the subtest level: EL to ES



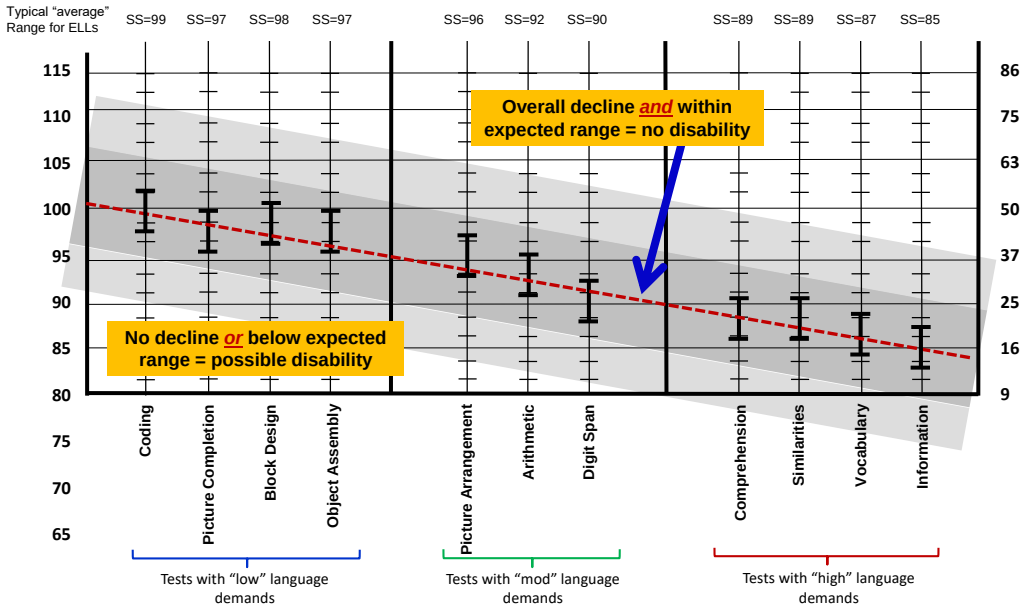
Fairness in Determining "Average" Performance: ES to ES



Fairness in Determining "Average" Performance: EL to ES



Fairness in Determining "Average" Performance: EL to ES



Interpretive Errors in C-LIM Studies: 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) 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).

To summarize, far from undermining the validity of the C-LIM, the Styck & Watkins studies provide strong and powerful support for the clinical utility and validity of the C-LIM when evaluating EL test performance.

**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.*

Research Foundations for EL Evaluation: EL to ES

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

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]	

Highest Language Demands ↓ Lowest Language Demands

*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.

Research Foundations for EL Evaluation: EL to ES

EL performance is moderated by level of English proficiency as compared to ES

Subtest Name	Mercer 1972 Scale Score	Vukovich & Figueroa, 1982 Scale Score	Cummins 1982 Scale Score	Nieves-Brull 2006 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

Tests with "high" language demands

Tests with "mod" language demands

Tests with "low" language demands

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

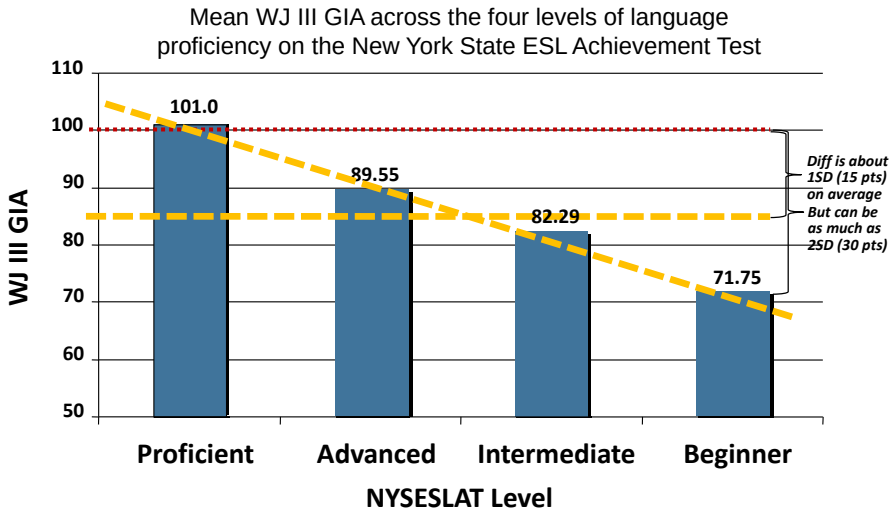
Fairness in Determining "Average" Performance: EL to ES

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

DEGREE OF CULTURAL LOADING	DEGREE OF LINGUISTIC DEMAND: 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

Research Foundations for EL Evaluation: EL to EL

General ability level performance as compared to other English learners

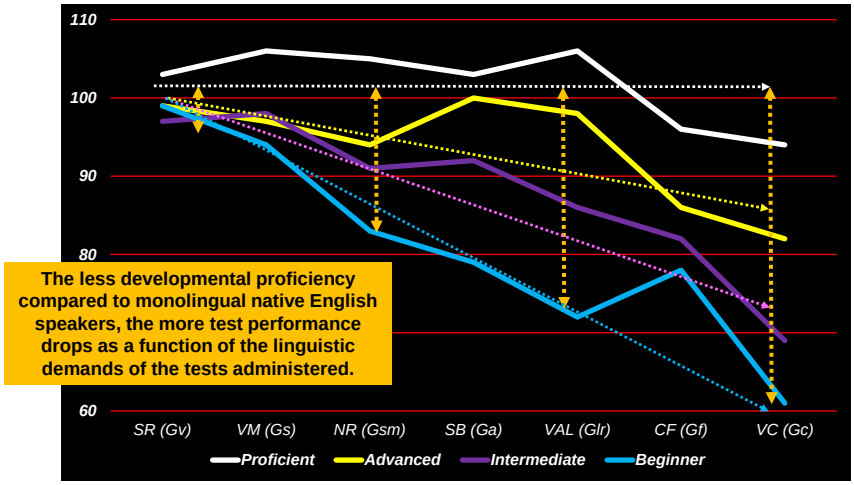


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013).

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

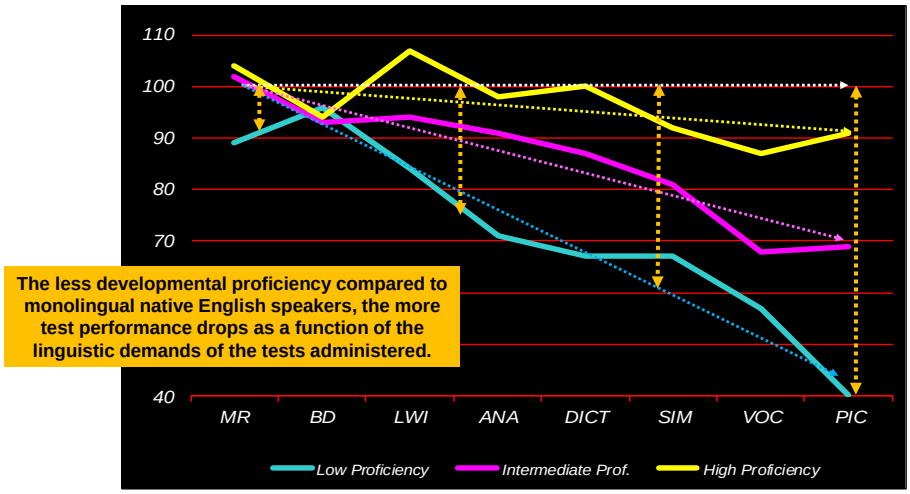


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013). English Language Proficiency and Test Performance: Evaluation of bilinguals with the Woodcock-Johnson III Tests of Cognitive Ability. Psychology in the Schools, Vol 50(8), pp. 781-797.

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



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

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 one exception, 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.

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

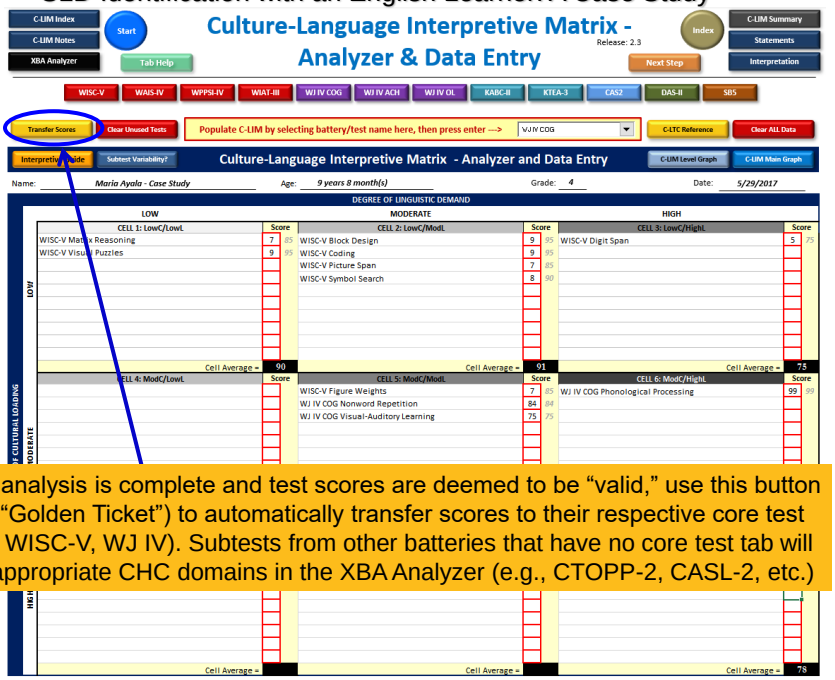
8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

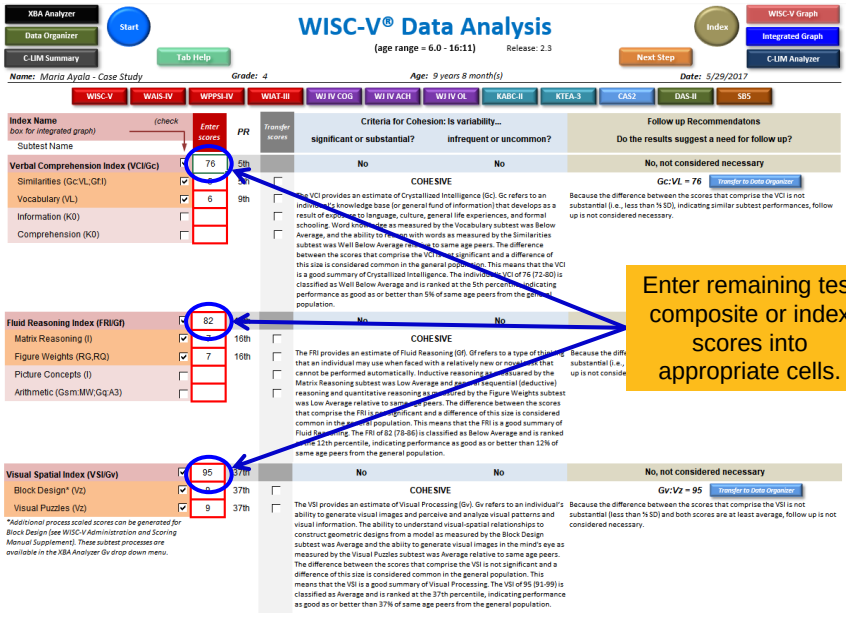
Procedures for Step 1

SLD Identification with an English Learner: A Case Study



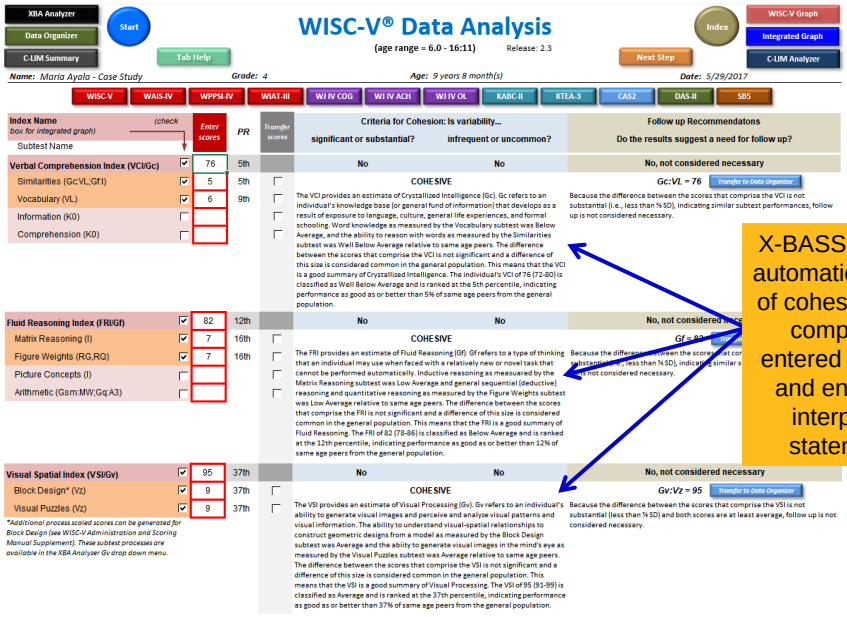
Once your analysis is complete and test scores are deemed to be "valid," use this button (aka, the "Golden Ticket") to automatically transfer scores to their respective core test tabs (e.g., WISC-V, WJ IV). Subtests from other batteries that have no core test tab will go to the appropriate CHC domains in the XBA Analyzer (e.g., CTOPP-2, CASL-2, etc.)

SLD Identification with an English Learner: A Case Study

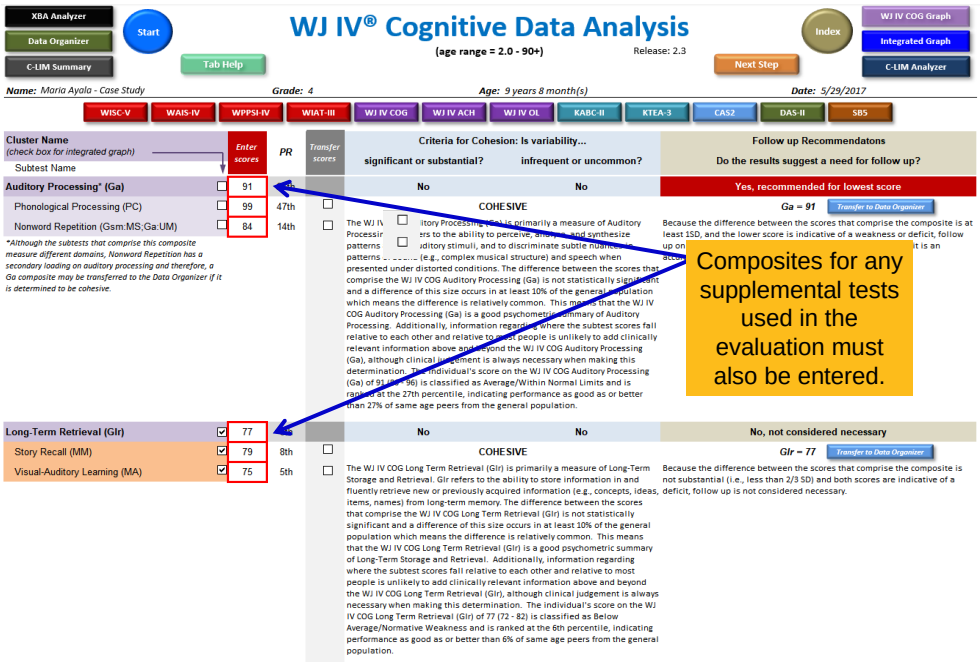


Enter remaining test composite or index scores into appropriate cells.

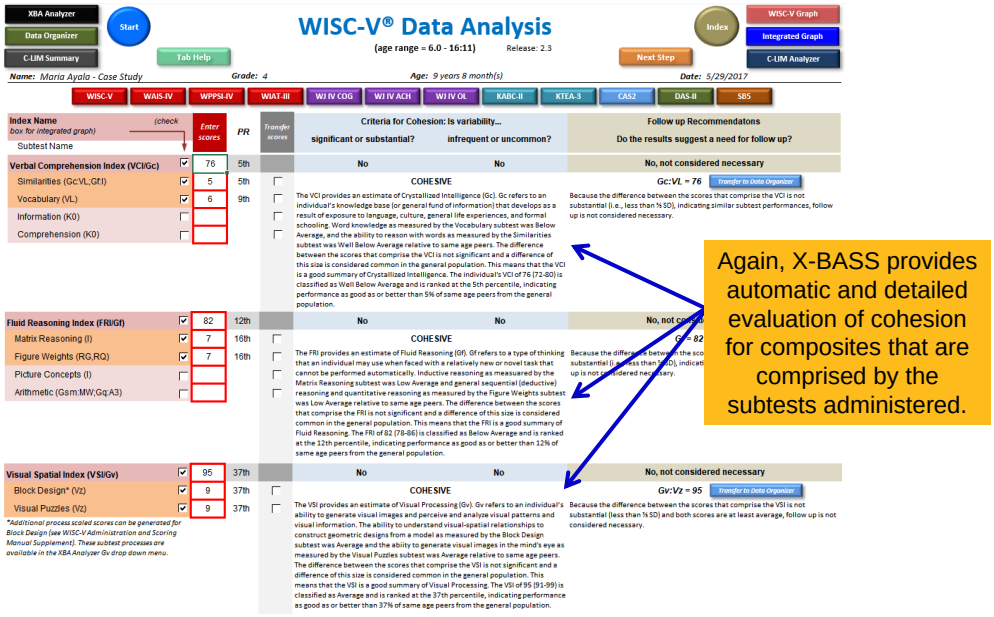
SLD Identification with an English Learner: A Case Study



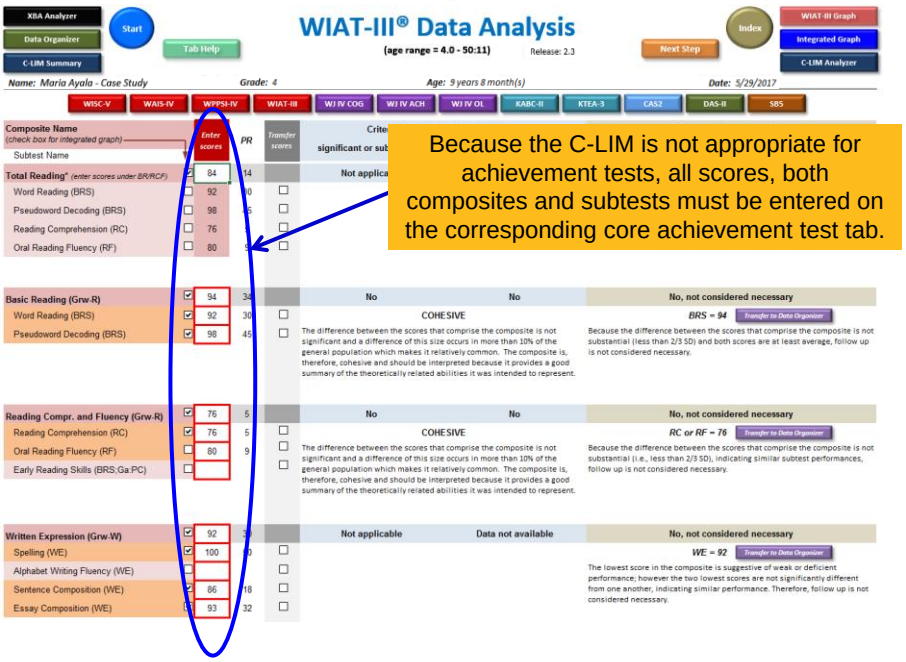
SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

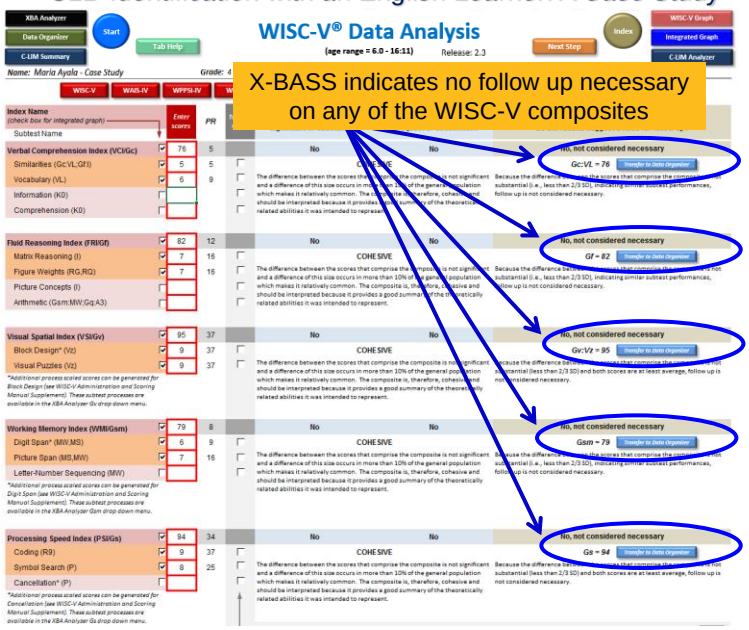
8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

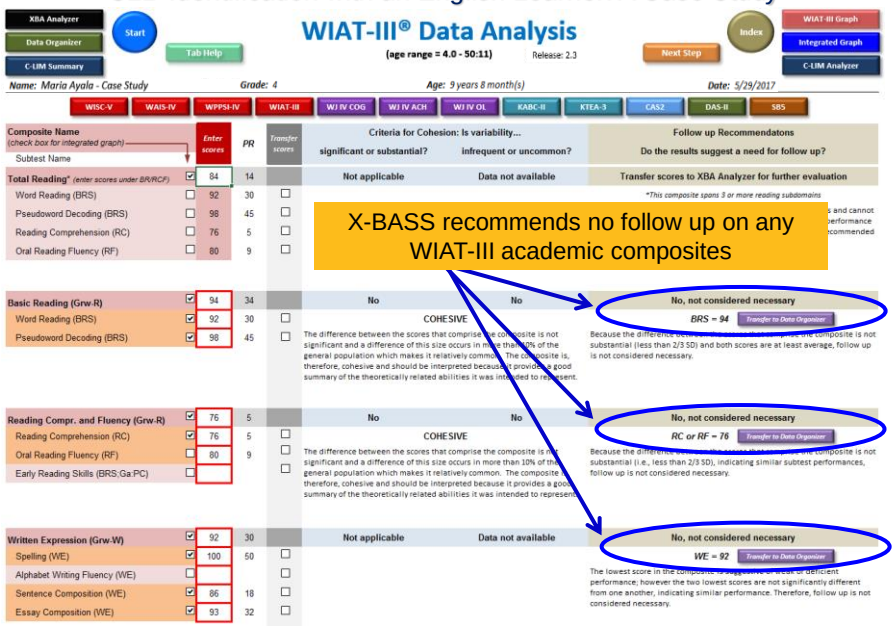
10: Use additional data and information to support interpretations and conclusions

Procedures for Step 1

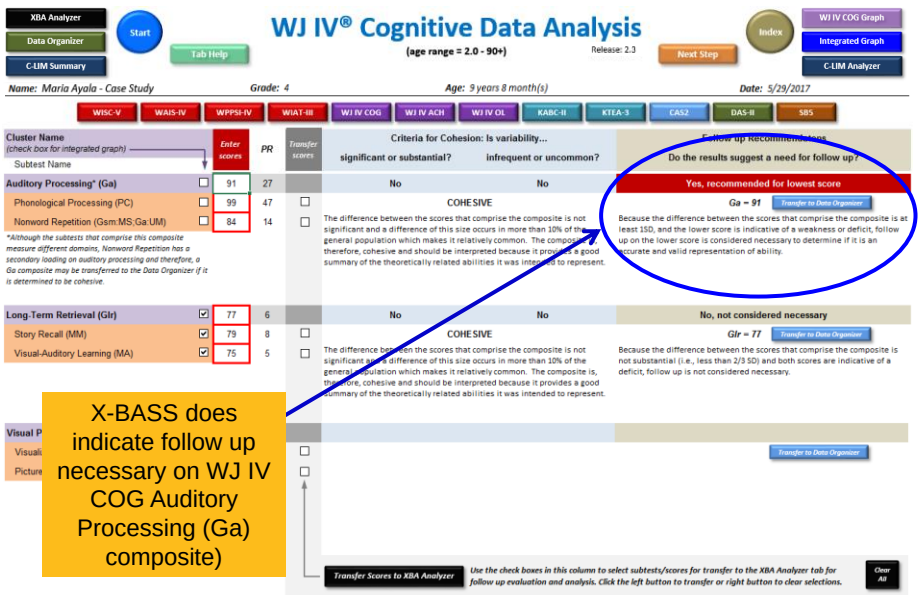
SLD Identification with an English Learner: A Case Study



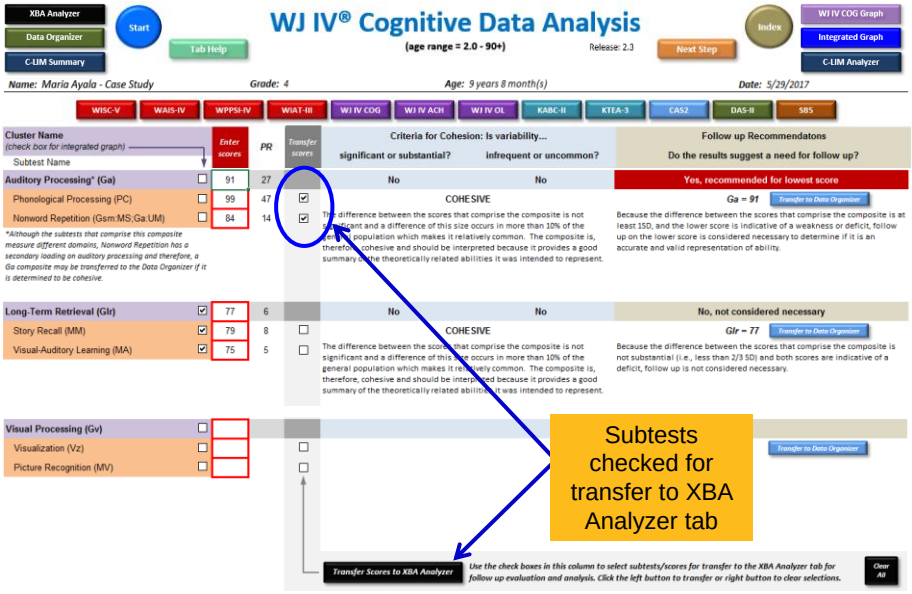
SLD Identification with an English Learner: A Case Study



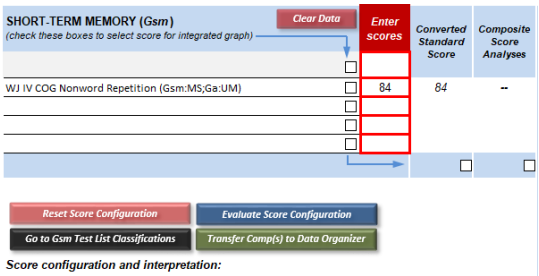
SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study

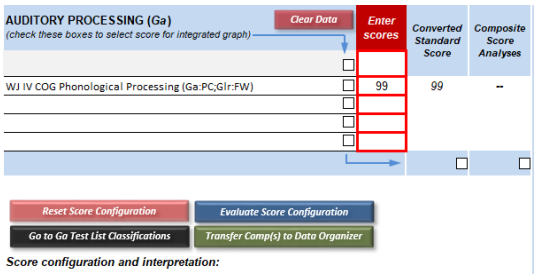


SLD Identification with an English Learner: A Case Study



The WJ IV COG Nonword Repetition subtest loads primarily on Gsm, not Ga. It can be combined with other WISC-V Gsm subtests to form an XBA composite or the WISC-V WMI can be used by itself if it has been determined to be cohesive.

The WJ IV COG Phonological Processing subtest loads primarily on Ga. Thus, it needs to be supplemented with another Ga subtest (e.g., WJ IV OL Sound Blending) to form a useable composite since the original composite was not cohesive.



SLD Identification with an English Learner: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 5/29/2017 DOB: 9/6/2007 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

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

Follow Up Testing	
WJ IV OL Sound Blending	88

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

8: Select best scores for PSW Analysis and designate each as strength or weakness

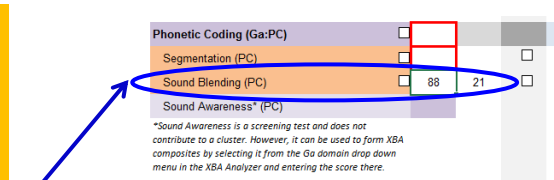
9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

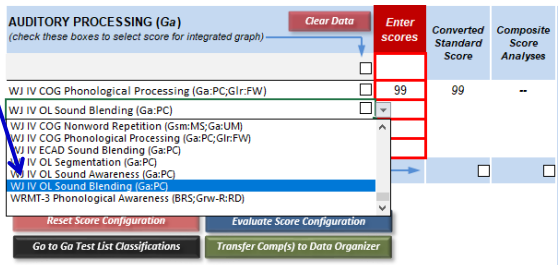
10: Use additional data and information to support interpretations and conclusions

Procedures for Step 1

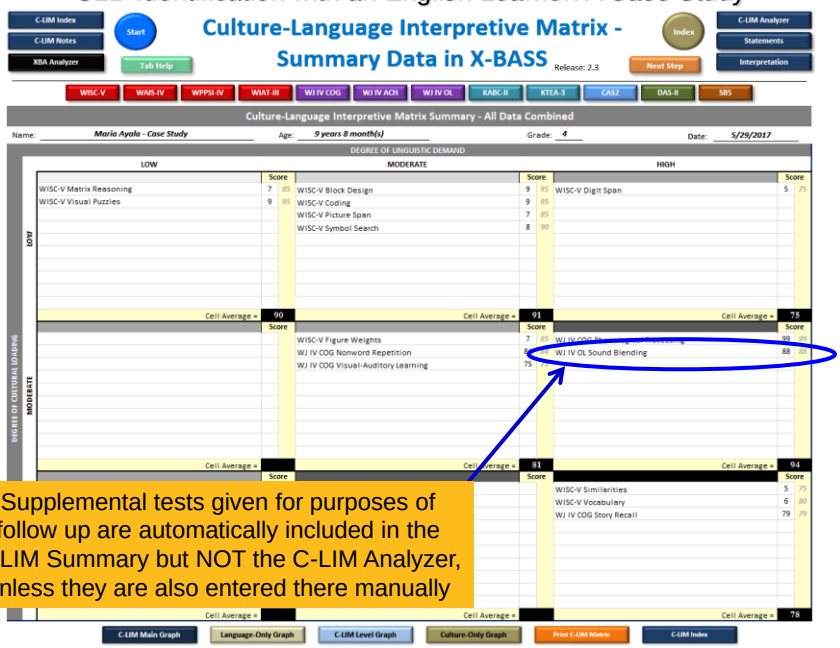
SLD Identification with an English Learner: A Case Study

Although supplemental tests can be entered on their respective core test tabs (if one is available for them), it is easier and quicker to simply enter them directly into the XBA Analyzer by selecting them from the appropriate drop down menus. In either case, they will automatically appear in the C-LIM Summary which permits re-examination of test score validity that now includes the additional scores.



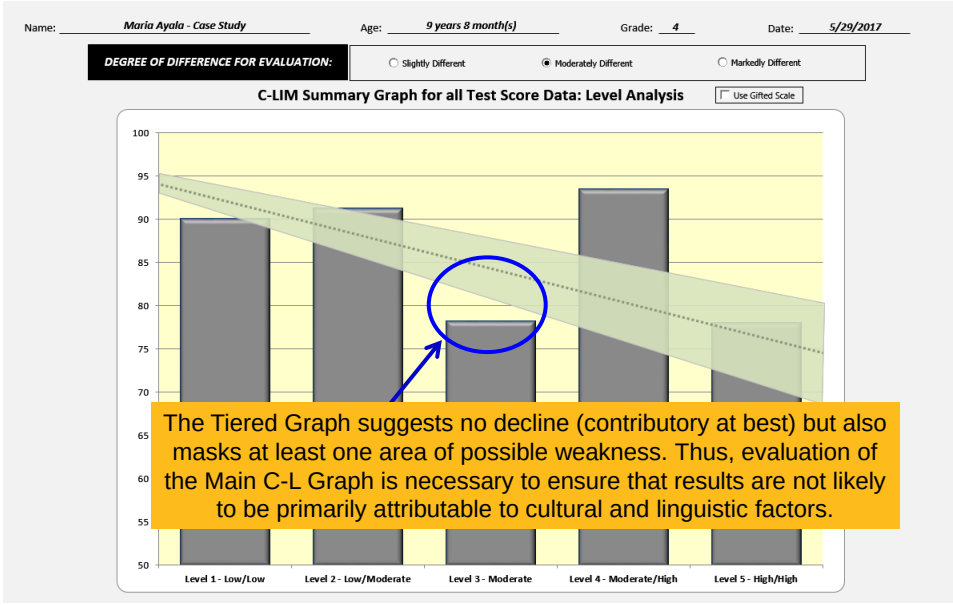


SLD Identification with an English Learner: A Case Study

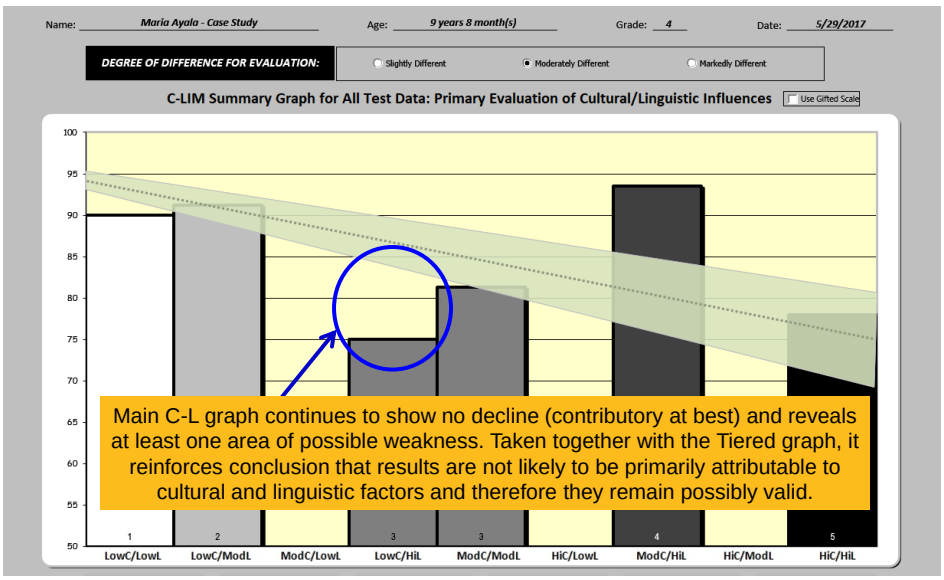


Supplemental tests given for purposes of follow up are automatically included in the C-LIM Summary but NOT the C-LIM Analyzer, unless they are also entered there manually

SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

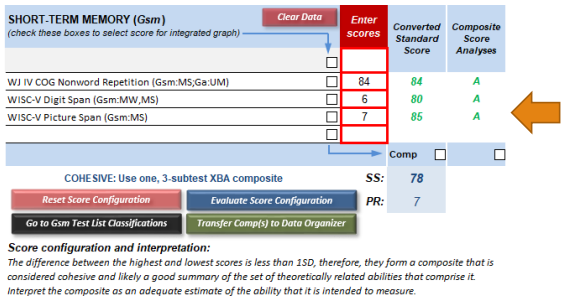
8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

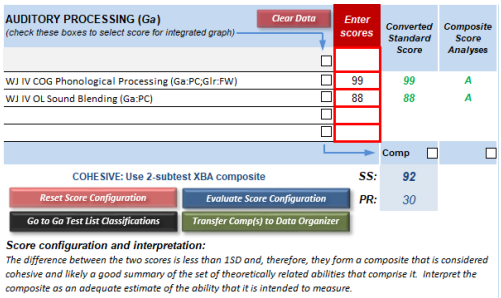
Procedures for Step 1

SLD Identification with an English Learner: A Case Study



Combining WISC-V subtests from WMI creates a cohesive 3-subtest XBA composite (SS=78). Although it's ok to use existing WMI, a 3-subtest composite is more reliable than a 2-subtest test composite so the XBA composite is preferable and will be transferred to the Data Organizer.

Follow up for Ga indicates that scores do form a cohesive 2-subtest XBA composite (SS=92). Thus, performance in auditory processing domain is within average range and the XBA composite will be transferred to Data Organizer.



SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

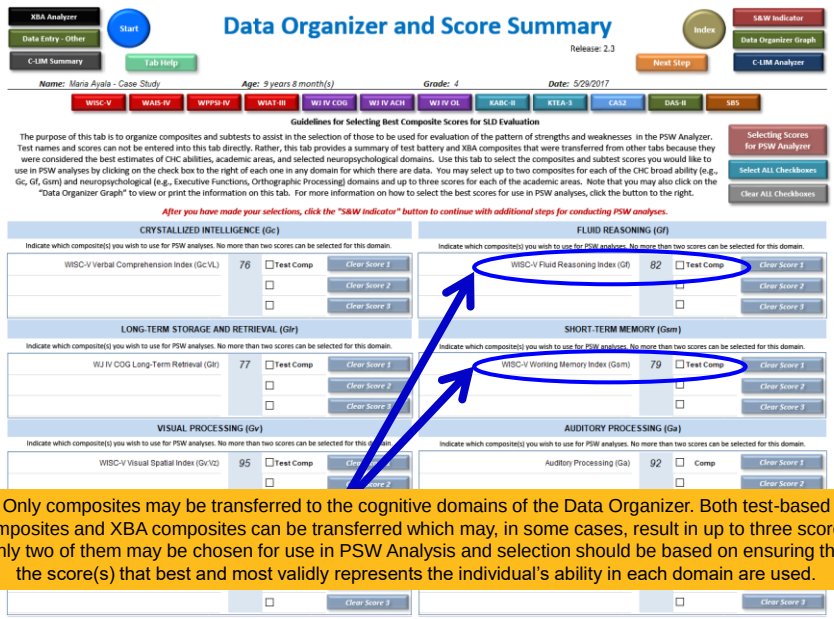
8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

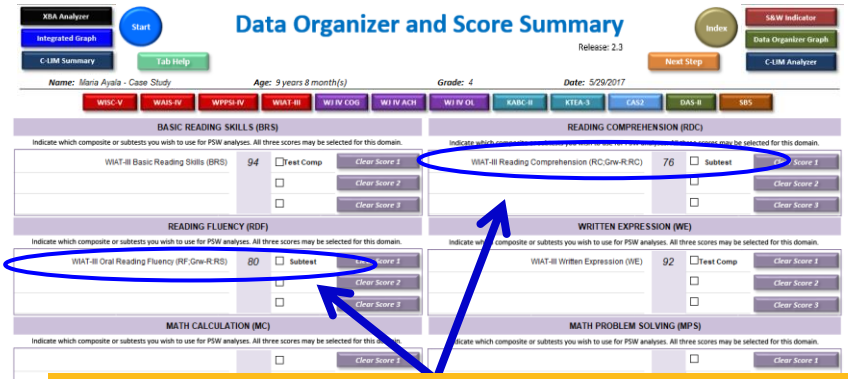
Procedures for Step 1

SLD Identification with an English Learner: A Case Study



Only composites may be transferred to the cognitive domains of the Data Organizer. Both test-based composites and XBA composites can be transferred which may, in some cases, result in up to three scores. Only two of them may be chosen for use in PSW Analysis and selection should be based on ensuring that the score(s) that best and most validly represents the individual's ability in each domain are used.

SLD Identification with an English Learner: A Case Study



Although both achievement composite and subtest scores may be transferred to the Data Organizer, use of individual achievement subtests rather than composites is often useful for specifying areas of academic difficulty and pinpointing skills for targeted intervention. This also helps avoid having to decide where a composite should be used, for example, the WIAT-III Reading Comprehension and Fluency composite can be used for RC or RF which is ambiguous. Using the subtests that make up this composite clarifies the domains for each score.

Data Organizer provides a summary of test-based composites, any derived XBA composites, and any specific achievement subtests from a test tab or the XBA Analyzer.

Multilingual Assessment of ELs: Step by Step

Step 1. Test first in English (L2) and evaluate construct validity in all areas in English (exclusion of cultural/linguistic factors)

- *If all scores indicate normative strengths (SS ≈ 90 or higher) when tested in English (L2), scores are valid to the extent that a disability is not likely, thus no further testing is necessary.*
- *If some scores are normative weaknesses (SS < ≈ 90) evaluate test score validity in a research-based manner, e.g., via the C-LIM.*
- *If C-LIM indicates primary influence of language/culture, test scores are likely invalid and indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.*
- *If C-LIM indicates contributory or minimal influence of language/culture, test scores are likely to be valid and the evaluation should continue.*

Step 2. Re-evaluate areas of weakness in native language (L2) to provide additional supporting evidence of validity (cross-linguistic confirmation)

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

Step 3. Further cross-validate L1 and L2 test scores with contextual factors and pre-referral data and academic concerns (ecological validity for disability)

- *Use all other case data and information to serve as the context by which to evaluate the test scores and ensure ecological validity to conclusions*

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

Procedures for Step 2

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

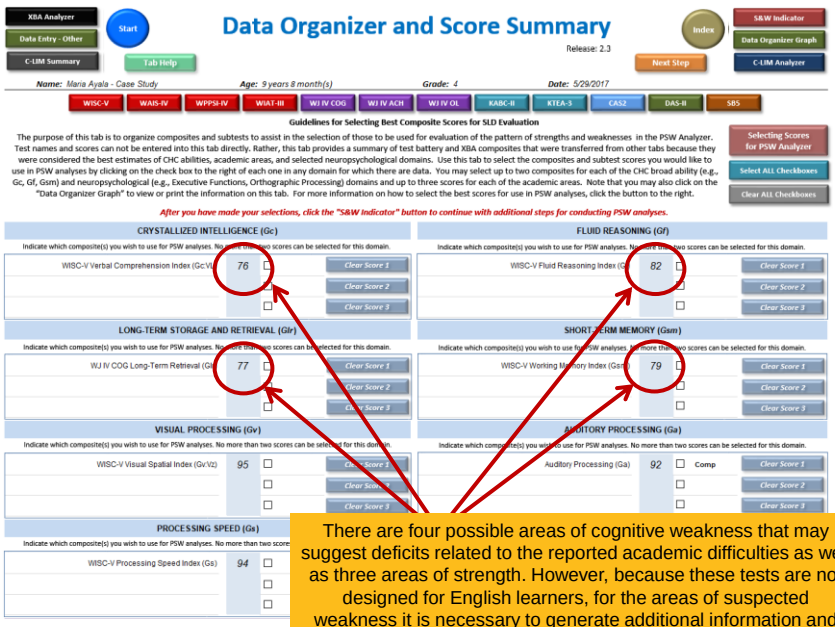
7: Identify deficits for native language re-evaluation and compare to original scores

8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

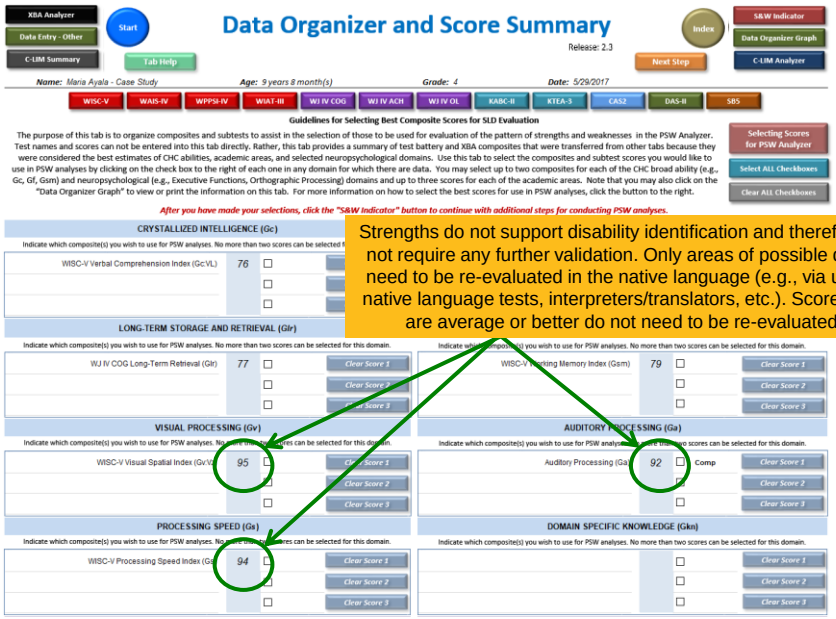
10: Use additional data and information to support interpretations and conclusions

SLD Identification with an English Learner: A Case Study



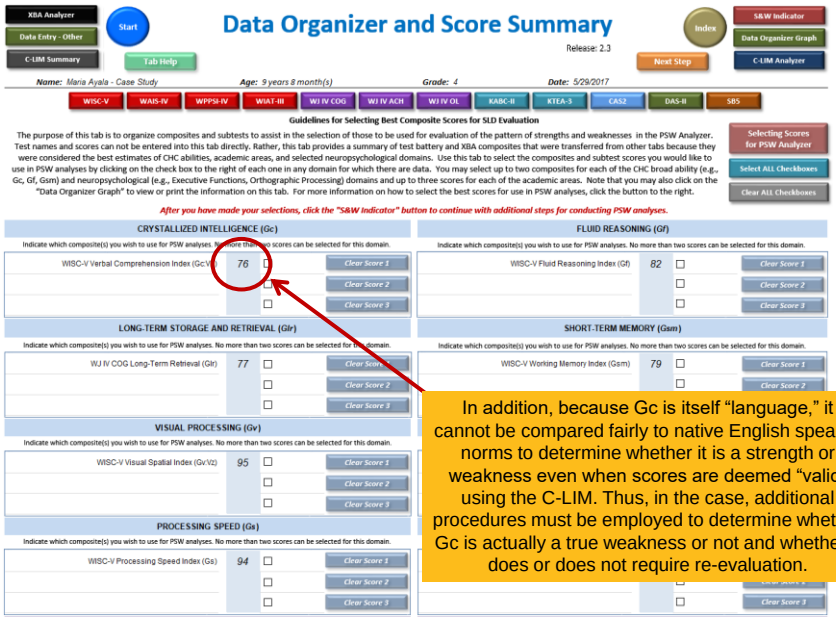
There are four possible areas of cognitive weakness that may suggest deficits related to the reported academic difficulties as well as three areas of strength. However, because these tests are not designed for English learners, for the areas of suspected weakness it is necessary to generate additional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

SLD Identification with an English Learner: A Case Study



Strengths do not support disability identification and therefore do not require any further validation. Only areas of possible deficit need to be re-evaluated in the native language (e.g., via use of native language tests, interpreters/translators, etc.). Scores that are average or better do not need to be re-evaluated.

SLD Identification with an English Learner: A Case Study



In addition, because Gc is itself "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.

SLD Identification with an English Learner: A Case Study

Interpretive Problems with Gc Scores with English Learners

Because Gc is, by definition, comprised of cultural knowledge and language development, the influence of these factors cannot be separated from tasks designed to measure them. Thus, unless exposure to English is a controlled variable in a test's norm sample and the sample includes many different languages, *Gc scores for ELLs always remain at risk for inequitable interpretation even when the overall pattern of scores within the C-LIM is determined to be valid.*

For example, a Gc score of 76 would be viewed as "deficient" relative to a norm sample comprised primarily of native English speakers. Moreover, testing in the native language doesn't solve this problem because current native-language tests treat ELs as being all the same (they aren't), as if being behind in English is only temporary (it isn't), as if the country they come from is important (it's not), and as if five years of English learning makes them native English speakers (it doesn't).

Therefore, practitioners must find and rely on a "true peer" comparison group such as that which is formed within the High Culture/High Language cell of the C-LIM to help *ensure that ELLs are not unfairly regarded as having either deficient Gc ability or significantly lower overall cognitive ability*—conditions that may simultaneously decrease identification of SLD and increase suspicion of ID and speech impairment.

SLD Identification with an English Learner: A Case Study

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.

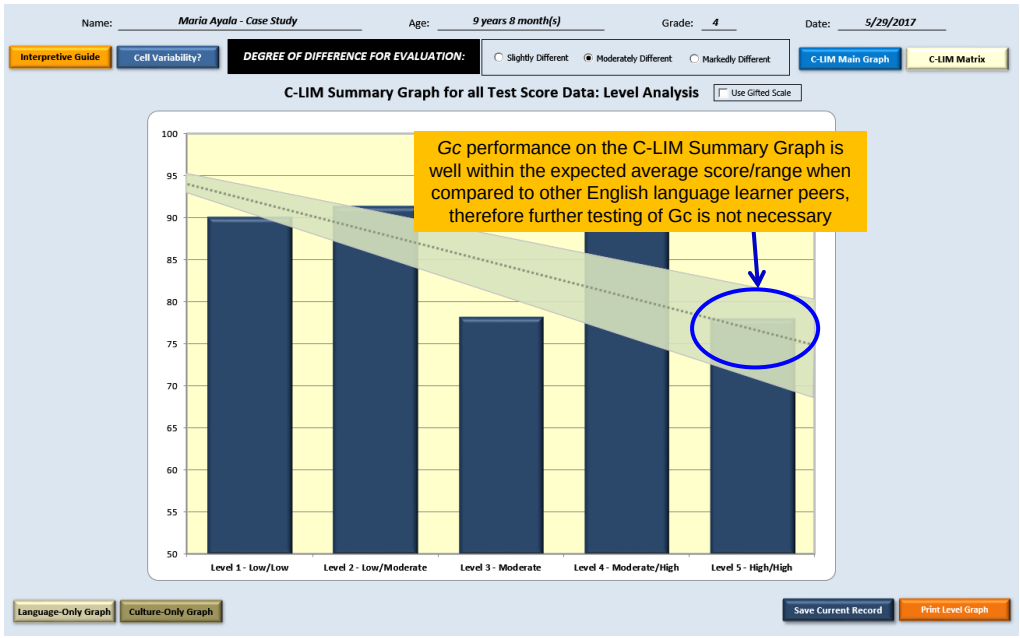
Because of the nature of Gc, it should be treated slightly differently when it comes to re-evaluation as compared to other cognitive abilities. The following guidelines from the best practice recommendations apply specifically to Gc:

- **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.*

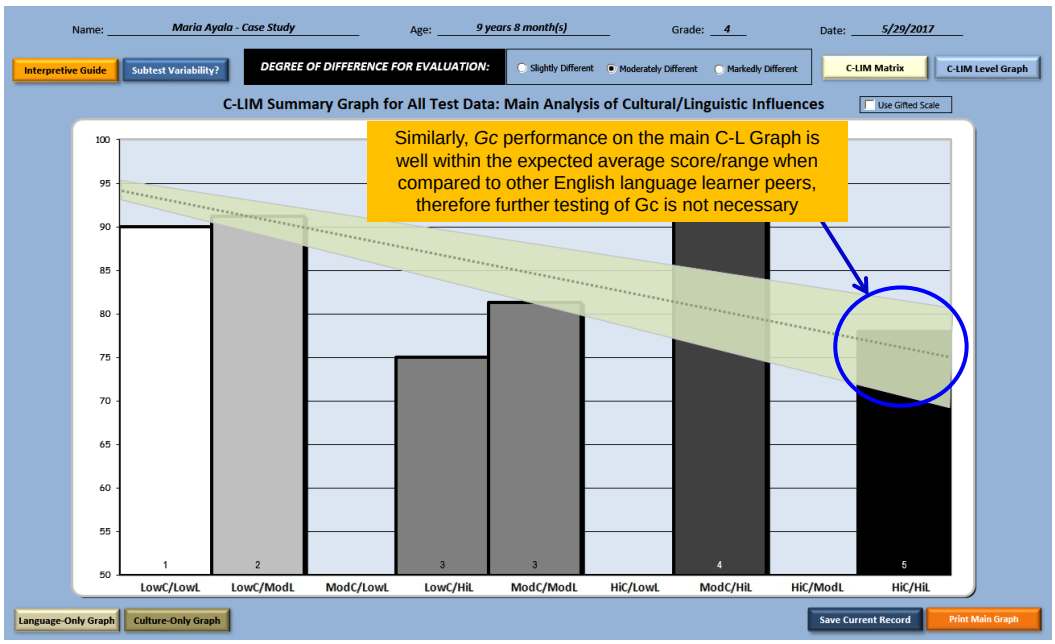
It is important that the actual, obtained Gc score, regardless of magnitude, be reported when required, albeit with appropriate nondiscriminatory assignment of meaning, and that it be used for the purposes of instructional planning and educational intervention.

****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.***

SLD Identification with an English Learner: A Case Study



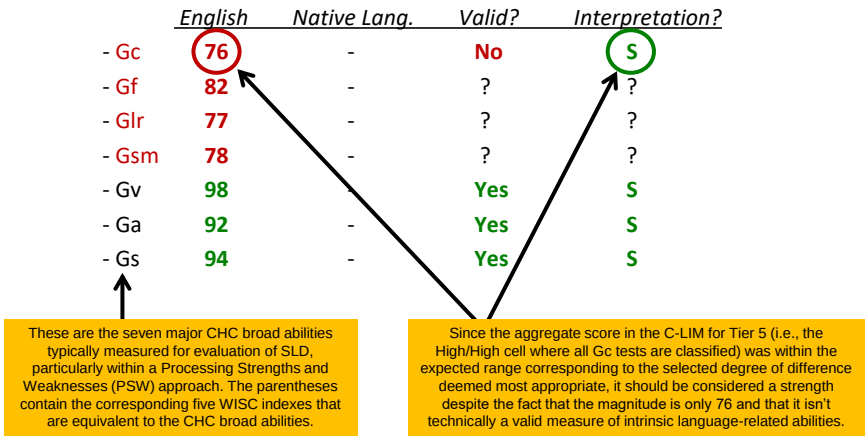
SLD Identification with an English Learner: A Case Study



SLD Identification with an English Learner: A Case Study

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.



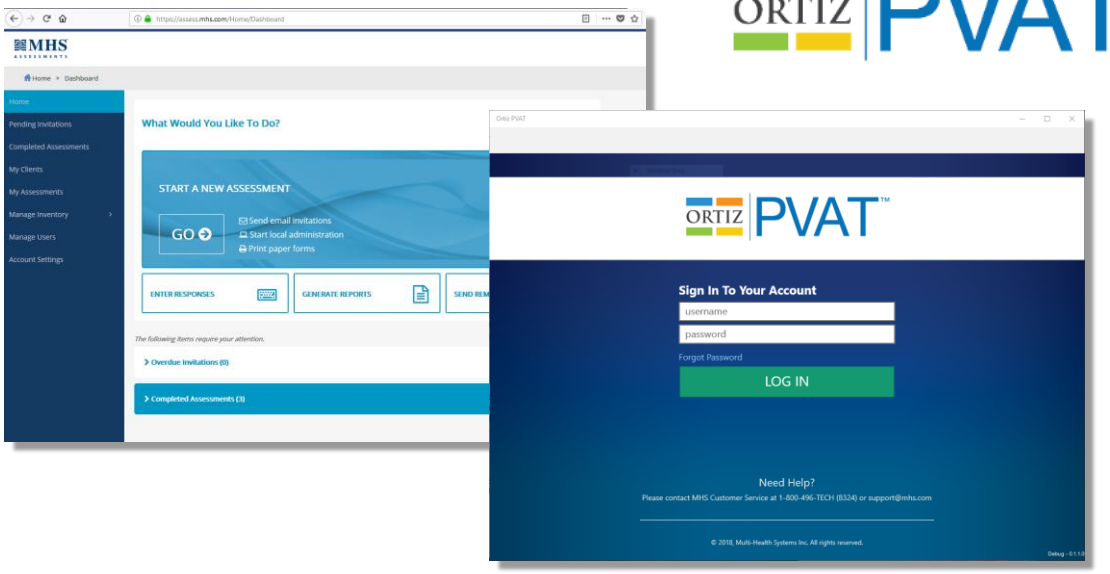
SLD Identification with an English Learner: A Case Study

Resolving Problems with Gc Scores for ELs: The Ortiz PVAT

Clearly, the preceding procedures necessary to address validity issues related to the measurement of Gc and language/culture-related abilities are complicated, somewhat cumbersome, and not very efficient. It may also leave the practitioner in the unenviable position of having to defend a very low score (SS=76) as being technically invalid, but still considered to be an area of processing "strength."

This one issue, more than any other, best highlights the shortcomings of today's tests relative to their failure to provide a true peer comparison group for English learners that would alleviate all of the extra work and potential confusion. There simply is no substitute for being able to make fair and equitable interpretations than comparison to peers with similar developmental experiences.

That said, there is in fact an easier way to do all of this. In response to the many difficulties posed by these issues, *a new test has been developed with dual-norm samples, including one specifically for English learners that yields valid Gc scores for English learners of any language background and level of English exposure—and 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

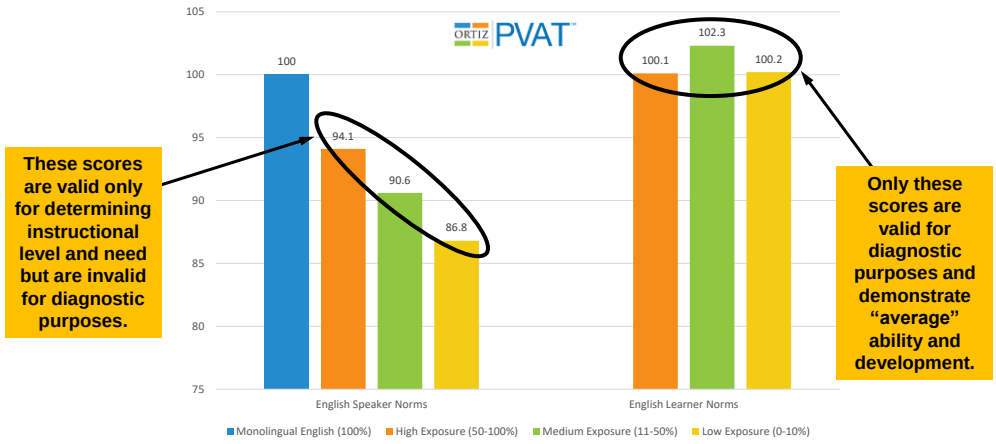
English Learners (*N* = 1,190)

- 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



This graph is reproduced from the Technical Manual of the Ortiz PVAT and is Copyright © 2017 Multi-Health Systems Inc. All rights reserved.

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				

This table is reproduced from the Technical Manual of the Ortiz PVAT and is Copyright © 2017 Multi-Health Systems Inc. All rights reserved.

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				

This table is reproduced from the Technical Manual of the Ortiz PVAT and is Copyright © 2017 Multi-Health Systems Inc. All rights reserved.

The Ortiz PVAT – Recommended 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.

Assessment Report from the Ortiz PVAT

ORTIZ PVAT™

Ortiz Picture Vocabulary Acquisition Test™

Samuel O. Ortiz, Ph.D.

Assessment Report

Examinee Information

Name/ID:	**Josephine Cruz**
Age:	15 years 6 months
Gender:	Female
Date of Birth:	July 18, 2003
Language(s) Spoken at Home:	Unspecified
Age at First Exposure to English:	14 years
Exposure to English:	3% of life
Primary Language of Instruction:	Unspecified
School Grade:	10

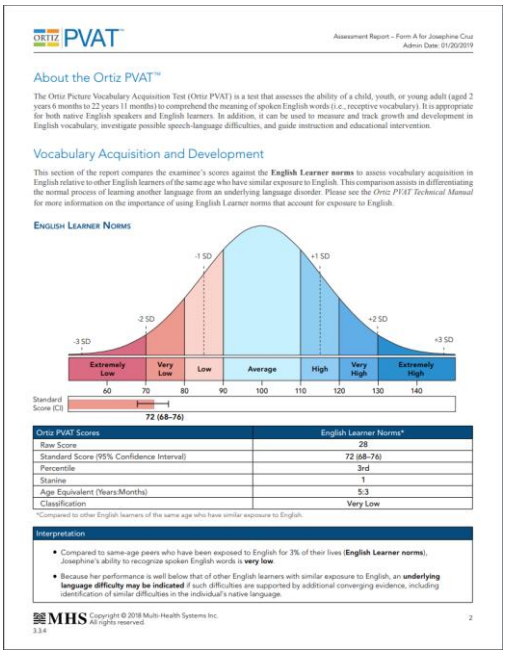
Assessment Information

Administration Date:	January 20, 2019
Examiner Name:	
Form Administered:	A
Norms Used:	**English Learner Norms (accounting for exposure to English)**
Number of Items Presented:	24
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.**

MHS ASSESSMENTS Copyright © 2018 Multi-Health Systems Inc. All rights reserved.
P.O. Box 950, North Tonawanda, NY 14120-0950
3770 Victoria Park Ave., Toronto, ON M2H 3M6

Assessment Report from the Ortiz PVAT



ORTIZ PVAT™

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	28
Standard Score (95% Confidence Interval)	72 (68–76)
Percentile	3rd
Stanine	1
Age Equivalent (Years:Months)	5:3
Classification	Very Low

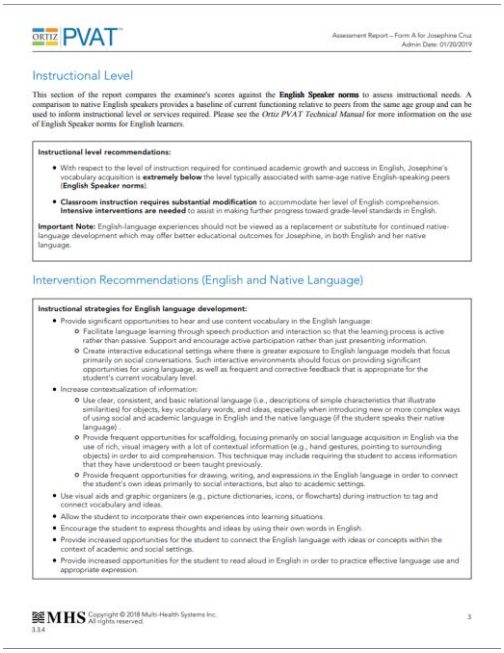
*Compared to other English learners of the same age who have similar exposure to English.

Interpretation

- Compared to same-age peers who have been exposed to English for 3% of their lives (**English Learner norms**), Josephine's ability to recognize spoken English words is **very low**.
- Because her performance is well below that of other English learners with similar exposure to English, an **underlying language difficulty may be indicated** if such difficulties are supported by additional converging evidence, including identification of similar difficulties in the individual's native language.

MHS Copyright © 2018 Multi-Health Systems Inc. All rights reserved.

Assessment Report from the Ortiz PVAT



ORTIZ PVAT

Assessment Report – Form A for Josephine Cruz
Admin Date: 01/20/2019

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, Josephine's vocabulary acquisition is **extremely below** the level typically associated with same-age native English-speaking peers (**English Speaker norms**).
- **Classroom instruction requires substantial modification** to accommodate her level of English comprehension. **Intensive interventions are 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 Josephine, in both English and her native language.

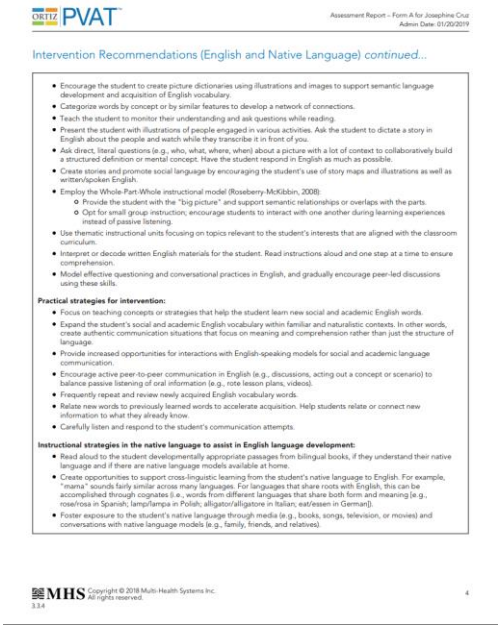
Intervention Recommendations (English and Native Language)

Instructional strategies for English language development:

- Provide significant opportunities to hear and use content vocabulary in the English language:
  - Facilitate language learning through speech production and interaction so that the learning process is active rather than passive. Support and encourage active participation rather than just presenting information.
  - Create interactive educational settings where there is greater exposure to English language models that focus primarily on social conversations. Such interactive environments should focus on providing significant opportunities for using language, as well as frequent and corrective feedback that is appropriate for the student's current vocabulary level.
- Increase contextualization of information:
  - Use clear, consistent, and basic relational language (i.e., descriptions of simple characteristics 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 frequent opportunities for scaffolding, focusing primarily on social language acquisition in English via the use of rich, visual imagery with a lot of contextual information (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 have understood or been taught previously.
  - Provide frequent opportunities for drawing, writing, and expressions in the English language in order to connect the student's own ideas primarily to social interactions, but also to 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.
- Encourage the student to express thoughts and ideas by using their own words in English.
- Provide increased opportunities for the student to connect the English language with ideas or concepts within the context of academic and social settings.
- Provide increased opportunities for the student to read aloud in English in order to practice effective language use and appropriate expression.

MHS Copyright © 2018 Multi-Health Systems Inc.
All rights reserved.

Assessment Report from the Ortiz PVAT



ORTIZ PVAT

Assessment Report – Form A for Josephine Cruz
Admin Date: 01/20/2019

Intervention Recommendations (English and Native Language) *continued...*

- Encourage the student to create picture dictionaries using illustrations and images to support semantic language development and acquisition of English vocabulary.
- 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 direct, literal questions (e.g., who, what, where, when) about a picture with a lot of context to collaboratively build a structured definition or mental concept. Have the student respond in English as much as possible.
- Create stories and promote social 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.
- Interpret or decode written English materials for the student. Read instructions aloud and one step at a time to ensure comprehension.
- Model effective questioning and conversational practices in English, and gradually encourage peer-led discussions using these skills.

Practical strategies for intervention:

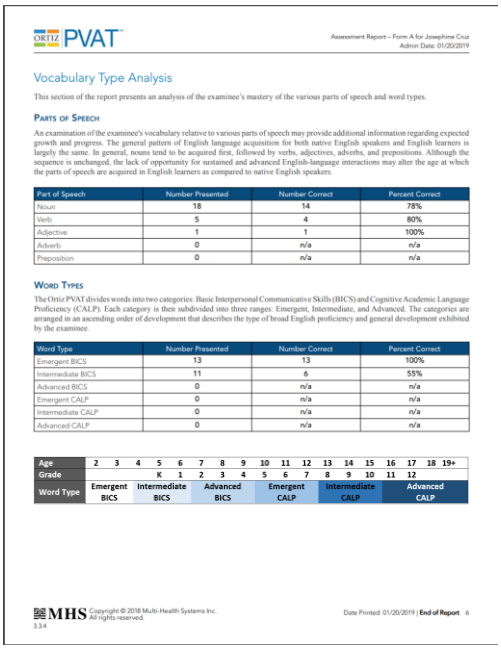
- Focus on teaching concepts or strategies that help the student learn new social and academic English words.
- Expand the student's social and academic English vocabulary within familiar and naturalistic contexts. In other words, create authentic communication situations that focus on meaning and comprehension rather than just the structure of language.
- Provide increased opportunities for interactions with 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 of 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).

MHS Copyright © 2018 Multi-Health Systems Inc.
All rights reserved.

Assessment Report from the Ortiz PVAT



ORTIZ PVAT

Assessment Report – Form A for Josephine Cruz
Admin Date: 01/20/2019

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	18	14	78%
Verb	5	4	80%
Adjective	1	1	100%
Adverb	0	n/a	n/a
Preposition	0	n/a	n/a

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.

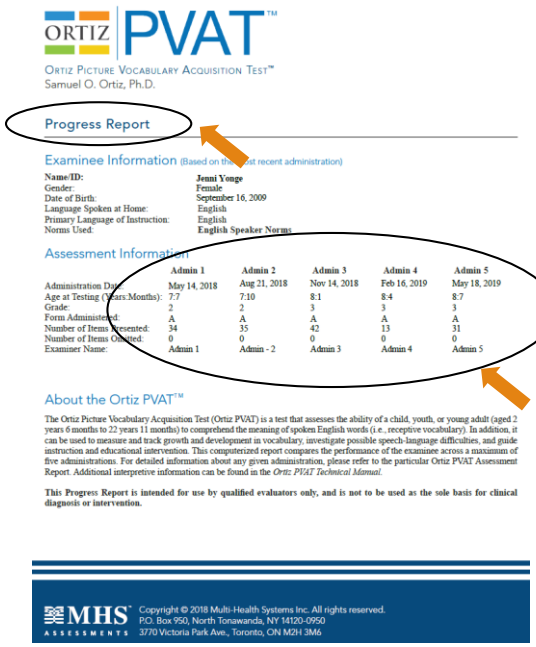
Word Type	Number Presented	Number Correct	Percent Correct
Emergent BICS	13	13	100%
Intermediate BICS	11	6	55%
Advanced BICS	0	n/a	n/a
Emergent CALP	0	n/a	n/a
Intermediate CALP	0	n/a	n/a
Advanced CALP	0	n/a	n/a

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			

MHS Copyright © 2018 Multi-Health Systems Inc. All rights reserved.
3.3.4

Date Printed: 01/20/2019 | End of Report 6

Progress Report from the Ortiz PVAT



ORTIZ PVAT™

Ortiz Picture Vocabulary Acquisition Test™

Samuel O. Ortiz, Ph.D.

Progress Report

Examinee Information (Based on the most recent administration)

Name/ID:	Jenni Yonge
Gender:	Female
Date of Birth:	September 16, 2009
Language Spoken at Home:	English
Primary Language of Instruction:	English
Norms Used:	English Speaker Norms

Assessment Information

	Admin 1	Admin 2	Admin 3	Admin 4	Admin 5
Administration Date:	May 14, 2018	Aug 21, 2018	Nov 14, 2018	Feb 16, 2019	May 18, 2019
Age at Testing (Years:Months):	7:7	7:10	8:1	8:4	8:7
Grade:	2	2	3	3	3
Form Administered:	A	A	A	A	A
Number of Items Presented:	34	35	42	13	31
Number of Items Omitted:	0	0	0	0	0
Examiner Name:	Admin 1	Admin - 2	Admin 3	Admin 4	Admin 5

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). 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 performance of the examinee across a maximum of five administrations. For detailed information about any given administration, please refer to the particular Ortiz PVAT Assessment Report. Additional interpretive information can be found in the *Ortiz PVAT Technical Manual*.

This Progress Report is intended for use by qualified evaluators only, and is not to be used as the sole basis for clinical diagnosis or intervention.

MHS ASSESSMENTS Copyright © 2018 Multi-Health Systems Inc. All rights reserved.
P.O. Box 950, North Tonawanda, NY 14120-0950
3770 Victoria Park Ave., Toronto, ON M2H 3M6

Alternate forms of the test (Form A and B) are fully parallel and allow for repeated testing in cases where progress monitoring or evaluation of growth is desired.

In such cases, a Progress Report can be generated that permits comparison of up to 5 different administration of the same examinee.

Progress Report from the Ortiz PVAT

ORTIZ PVAT™

Progress Report for Benni Ayala
Admin Dates: May 14, 2018 to May 18, 2019

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.

	Administration					Growth Index				
Ortiz PVAT Scores	Admin 1 05/14/18	Admin 2 08/21/18	Admin 3 11/14/18	Admin 4 02/16/19	Admin 5 05/18/19	Admin 1 to 2	Admin 2 to 3	Admin 3 to 4	Admin 4 to 5	Overall (1 to 5)
Raw Score	53	58	64	68	66					
Standard Score (95% CI)	95 (91–99)	96 (92–100)	97 (93–101)	98 (94–102)	95 (91–99)	0.45	0.45	0.45	0.00	
Age (Age Equivalent)	7:7 (7:6)	7:10 (7:9)	8:1 (8:0)	8:4 (8:3)	8:7 (8:6)				-1.36	
Classification	Average	Average	Average	Average	Average					
Rate of Growth						As Expected	As Expected	As Expected	As Expected	As Expected

More than Expected
As Expected
Less than Expected
Much Less than Expected

Note. CI = Confidence Interval. Age and age equivalent scores are denoted in terms of years and months (e.g., 2:6 = 2 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.

The Growth Index provides an indication of actual change or true growth across two or more administrations.

It is useful for both progress monitoring purposes as well as for determining whether an individual's language acquisition is typical or not as compared to other English learners of the same age.

Performance Across Different Norm Sample Comparisons

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*

**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.*

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

Data in this table are provided courtesy of an urban school district and may not be copied or reproduced. Used here with permission of the owner.

SLD Identification with an English Learner: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 5/29/2017 DOB: 9/6/2007 Grade: 4

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 INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*	*Fluid Reasoning Index*	*82*
Similarities	5	Matrix Reasoning	7
Vocabulary	6	Figure Weights	7
Working Memory Index	*79*	*Processing Speed Index*	*94*
Digit Span	5	Coding	9
Picture Span	7	Symbol Search	8

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*	*Reading Comprehension*	*76*	*Written Expression*	*92*
Word Reading	92	Reading Comprehension	76	Spelling	100
Pseudoword Decoding	98	Oral Reading Fluency	80	Sentence Composition	86
				Essay Composition	93

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*	*LT Storage/Retrieval*	*77*	*Ortiz PVAT (EL Norms)*	*93*
Phonological Processing	99	Story Recall	79		
Nonword Repetition	84	Visual-Auditory Learning	75		

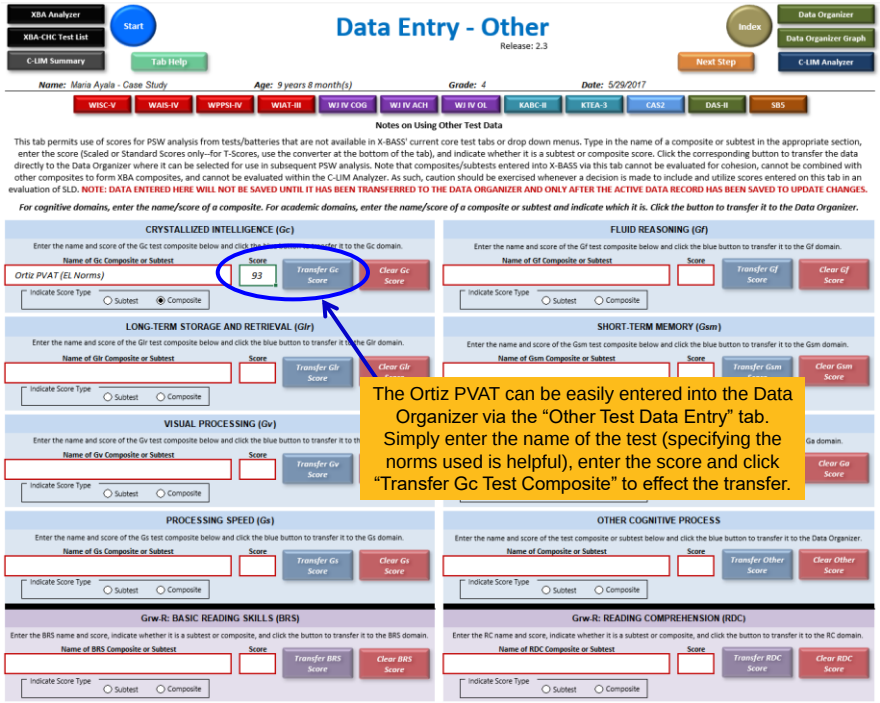
SLD Identification with an English Learner: A Case Study

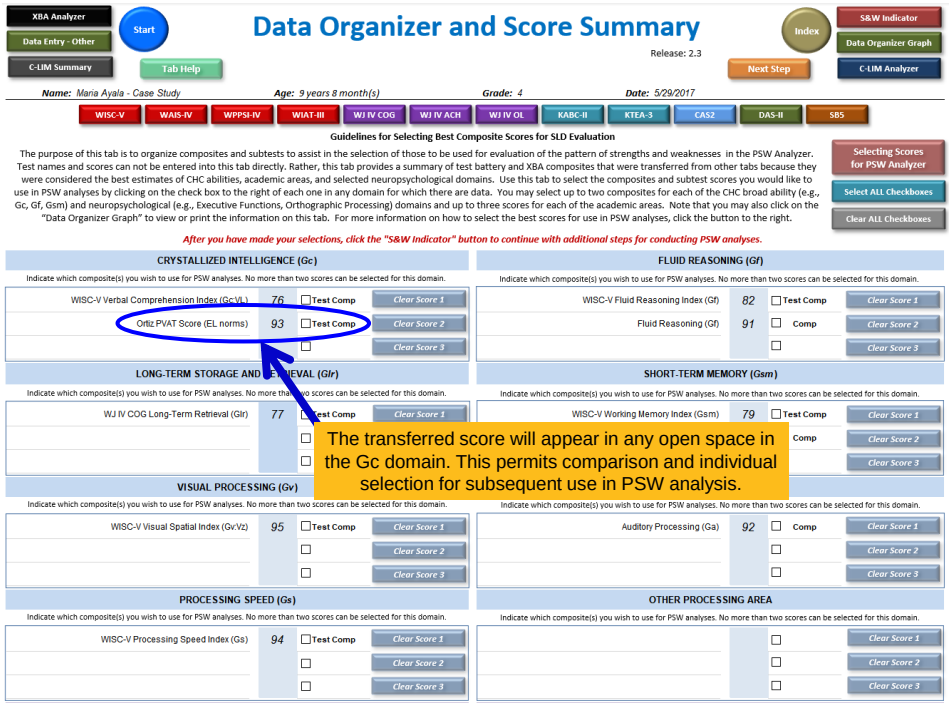
Avoiding Interpretive Problems by Use of the Ortiz PVAT

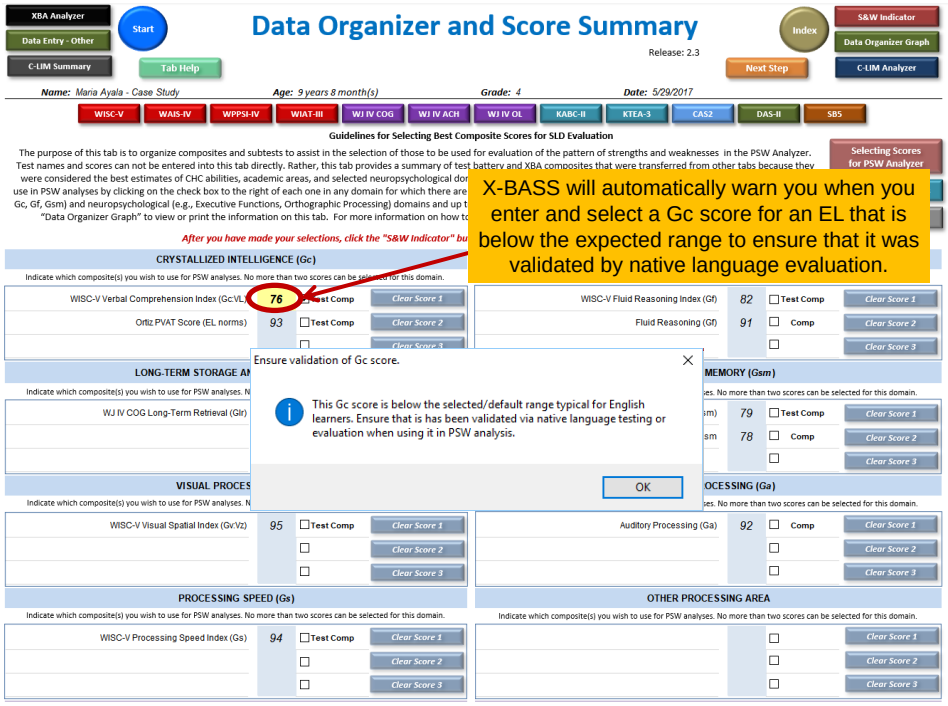
Derivation of an Ortiz PVAT score using the English learner norms eliminates the Gc problem completely. The Ortiz PVAT score simply replaces any Gc/language-related/verbal ability score because it was derived precisely on "true peers" and therefore inherently valid in terms of both meaning/classification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

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

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.


Data Entry - Other
Release: 2.3
Name: Maria Ayala - Case Study
Age: 9 years 8 month(s)
Grade: 4
Date: 5/29/2017
Notes on Using Other Test Data
CRYSTALLIZED INTELLIGENCE (Gc)
Ortiz PVAT (EL Norms)
93
FLUID REASONING (Gf)
LONG-TERM STORAGE AND RETRIEVAL (Glr)
SHORT-TERM MEMORY (Gsm)
VISUAL PROCESSING (Gv)
PROCESSING SPEED (Gs)
OTHER COGNITIVE PROCESS
Grw-R: BASIC READING SKILLS (BRS)
Grw-R: READING COMPREHENSION (RDC)
The Ortiz PVAT can be easily entered into the Data Organizer via the "Other Test Data Entry" tab. Simply enter the name of the test (specifying the norms used is helpful), enter the score and click "Transfer Gc Test Composite" to effect the transfer.


Data Organizer and Score Summary
Release: 2.3
Name: Maria Ayala - Case Study
Age: 9 years 8 month(s)
Grade: 4
Date: 5/29/2017
Guidelines for Selecting Best Composite Scores for SLD Evaluation
CRYSTALLIZED INTELLIGENCE (Gc)
WISC-V Verbal Comprehension Index (Gc:VL) 76
Ortiz PVAT Score (EL norms) 93
FLUID REASONING (Gf)
WISC-V Fluid Reasoning Index (Gf) 82
Fluid Reasoning (Gf) 91
LONG-TERM STORAGE AND RETRIEVAL (Glr)
WJ IV COG Long-Term Retrieval (Glr) 77
SHORT-TERM MEMORY (Gsm)
WISC-V Working Memory Index (Gsm) 79
VISUAL PROCESSING (Gv)
WISC-V Visual Spatial Index (Gv:Vz) 95
Auditory Processing (Ga) 92
PROCESSING SPEED (Gs)
WISC-V Processing Speed Index (Gs) 94
OTHER PROCESSING AREA
The transferred score will appear in any open space in the Gc domain. This permits comparison and individual selection for subsequent use in PSW analysis.



Nondiscriminatory Interpretation of Test Scores: A Case Study

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 native language tests or conduct re-testing using one of the following methods:*
  - *a. Native language test administered in the native language (e.g., WJ III/Bateria III or WISC-IV/WISC-IV Spanish)*
  - *b. Native language test administered via assistance of a trained interpreter*
  - *c. English language test translated and administered via assistance of a trained interpreter*
- *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.

SLD Identification with an English Learner: A Case Study

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

SLD Identification with an English Learner: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 5/29/2017 DOB: 9/6/2007 Grade: 4

WECHSLER INTELLIEGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*	*Fluid Reasoning Index*	*82*	*Visual-Spatial Index*	*95*
Similarities	5	Matrix Reasoning	7	Block Design	9
Vocabulary	6	Figure Weights	7	Visual Puzzles	9
Working Memory Index	*79*	*Processing Speed Index*	*94*	WISC IV Spanish **(Gf subtests)**	91
Digit Span	5	Coding	9	Matrix Reasoning	8
Picture Span	7	Symbol Search	8	Picture Concepts	9
WISC IV Spanish WMI	*72*				
Digit Span	5				
Letter-Number Sequencing	4				

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*	*LT Storage/Retrieval*	*77*	*Ortiz PVAT*	*93*
Phonological Processing	99	Story Recall	79		
Nonword Repetition	84	Visual-Auditory Learning	75		
		Bateria III LT Retrieval	*79*		
		Visual-Auditory Learning	81		
		Retrieval Fluency	78		

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.

SLD Identification with an English Learner: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria Ayala
DOE: 5/29/2017 DOB: 9/6/2007 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

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 IV Spanish **(Gf subtests)**	91
Matrix Reasoning	8
Picture Concepts	9

WISC IV Spanish WMI	*72*
Digit Span	5
Letter-Number Sequencing	4

WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*
Phonological Processing	99
Nonword Repetition	84

LT Storage/Retrieval	*77*
Story Recall	78
Visual-Auditory Learning	75

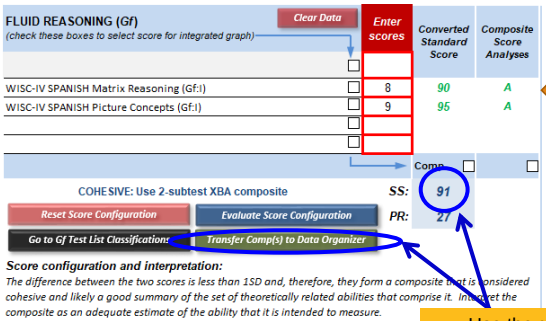
Ortiz PVAT	*93*

Bateria III LT Retrieval	*79*
Visual-Auditory Learning	81
Retrieval Fluency	78

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

SLD Identification with an English Learner: A Case Study

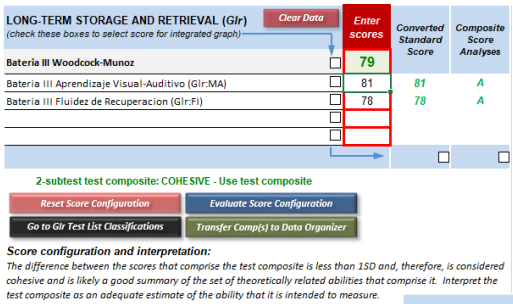
The original WISC-V Gf-based score (FRI) was cohesive and suggested a deficit (SS=82). Because the corresponding domain (PRI) of the older WISC-IV Spanish was based on three subtests (Matrix Reasoning, Block Design, and Picture Concepts) and because Block Design is now a part of the new Visual Spatial Index of the WISC-V, it should not be re-tested or used again as a part of the Gf domain. It is, however, appropriate to use the two Gf subtests to form a composite via the XBA Analyzer shown below.



The original score (WISC-V FRI=82) suggested a deficit. However, follow up native language testing resulted in a higher and cohesive XBA composite score (SS=91) indicating likely average ability. Thus, the original score is invalidated and should be replaced by the native language score for the purposes of analysis and interpretation.

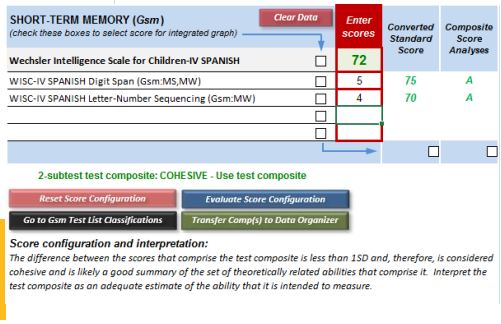
Use the green button to transfer the native language XBA Gf composite to the Data Organizer

SLD Identification with an English Learner: A Case Study



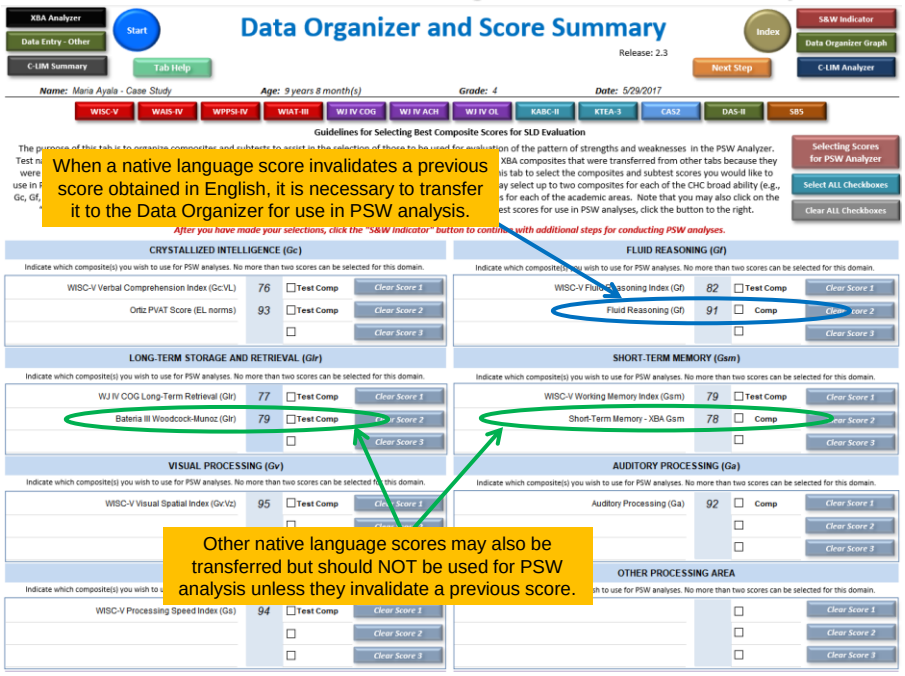
The original Glr score on the WJ IV COG was cohesive and suggested a deficit (SS=77). Follow up native language testing resulted in a similar score (SS=79) that also indicated possible deficit. Thus, the original score is validated, but additional converging evidence is necessary (work samples, observations, progress monitoring info, etc.).

The original Gsm score from the WISC-IV (WMI) was cohesive and suggested a deficit (SS=78). Follow up native language testing provided a norm-based test composite for Gsm similar in value (SS=72) that also indicates a possible deficit. Thus, the original score is validated but additional converging evidence is necessary (e.g., work samples, observations, progress monitoring info, etc.).



In these two cases, the native language scores do not need to be transferred to the Data Organizer as they merely provide cross-linguistic confirmation of the original scores obtained in English which will be used instead.

SLD Identification with an English Learner: A Case Study



When a native language score invalidates a previous score obtained in English, it is necessary to transfer it to the Data Organizer for use in PSW analysis.

Other native language scores may also be transferred but should NOT be used for PSW analysis unless they invalidate a previous score.

Nondiscriminatory Interpretation of Test Scores: A Case Study

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 ≥ 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≥90) is used 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.*

A Recommended Best Practice Approach for Using Tests with ELs

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.*

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

Procedures for Step 2

SLD Identification with an English Learner: A Case Study

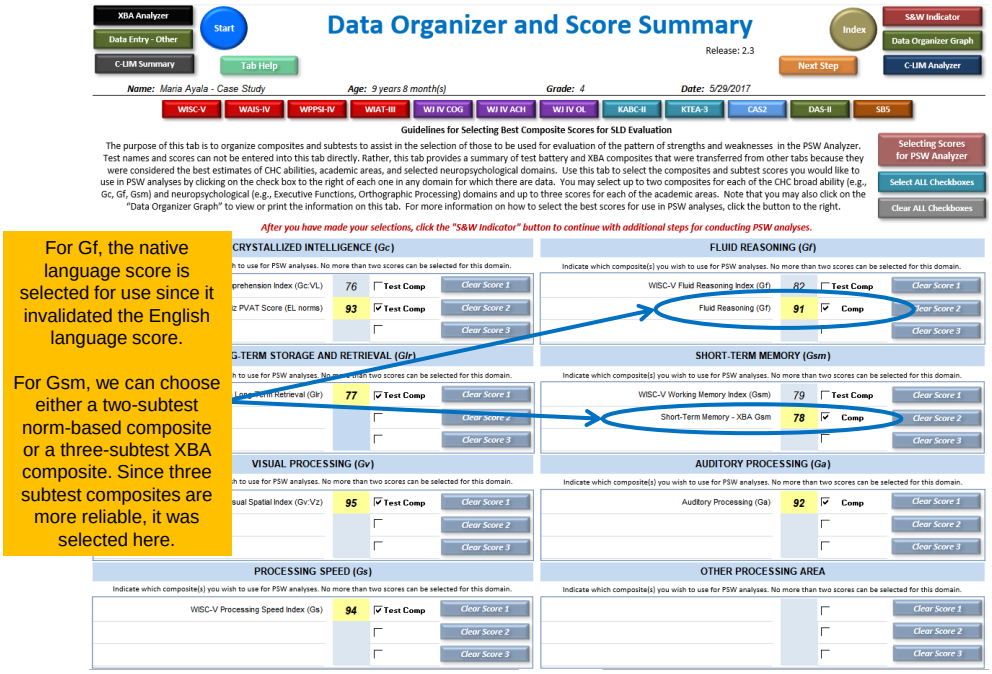
Determining which scores are valid and interpretable

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**	**S**
- Glr	**77**	(79)	**77 - Yes**	**W**
- Gsm	**78**	(72)	**78 - Yes**	**W**
- Gv	**98**	-	**Yes**	**S**
- Ga	**92**	-	**Yes**	**S**
- Gs	**94**	-	**Yes**	**S**
- **Gc** (Ortiz PVAT)	**93**	-	**Yes**	**S**

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.

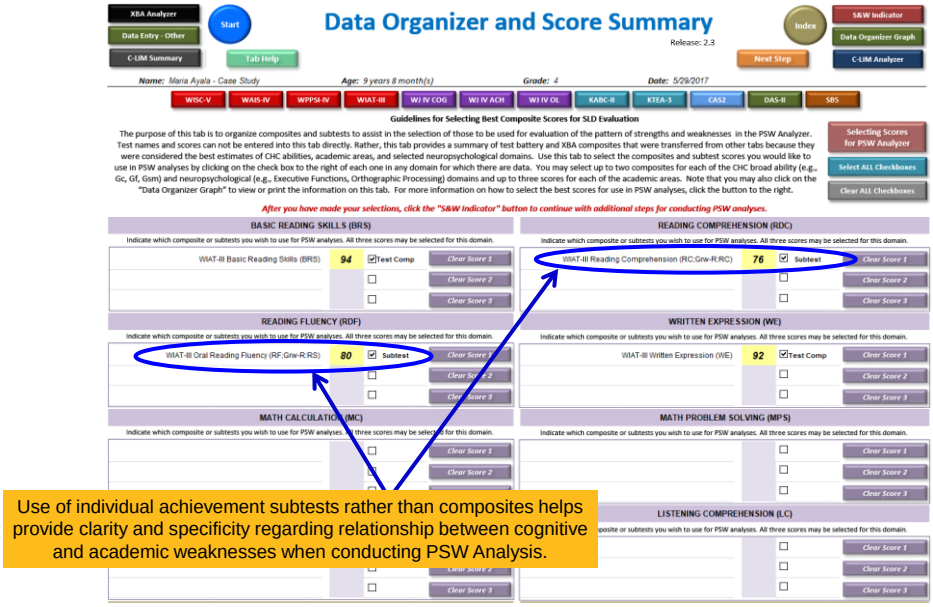
SLD Identification with an English Learner: A Case Study



For Gf, the native language score is selected for use since it invalidated the English language score.

For Gsm, we can choose either a two-subtest norm-based composite or a three-subtest XBA composite. Since three subtest composites are more reliable, it was selected here.

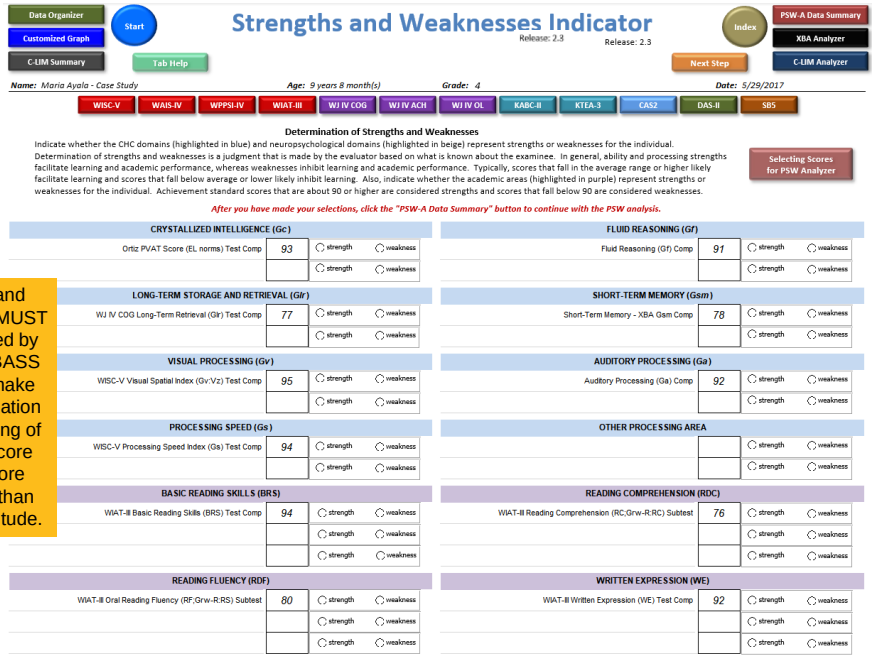
SLD Identification with an English Learner: A Case Study



Use of individual achievement subtests rather than composites helps provide clarity and specificity regarding relationship between cognitive and academic weaknesses when conducting PSW Analysis.

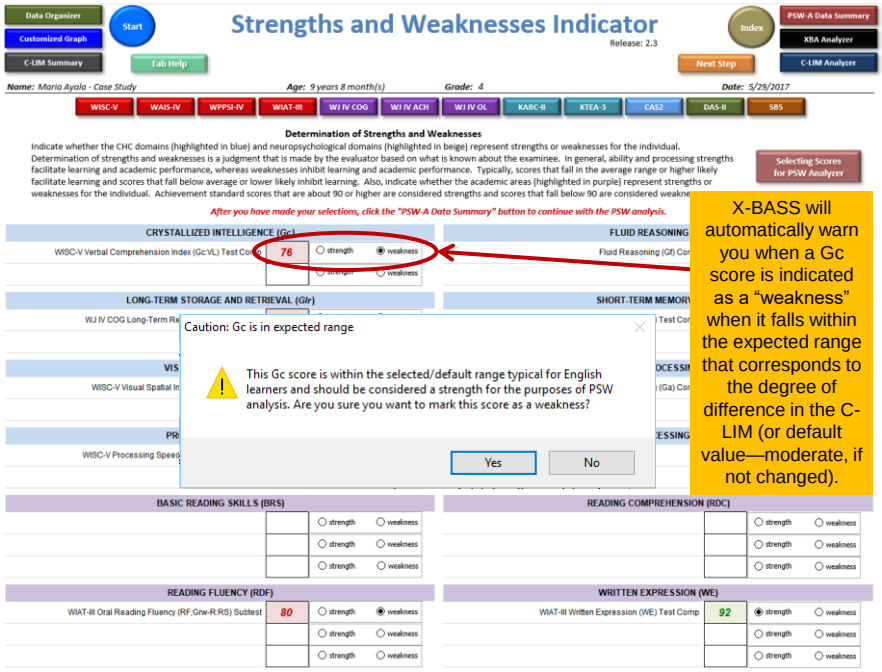
Selected scores appear in yellow and a maximum of 3 academic scores can be selected including any combination of test composites, XBA composites, or subtest scores.

SLD Identification with an English Learner: A Case Study



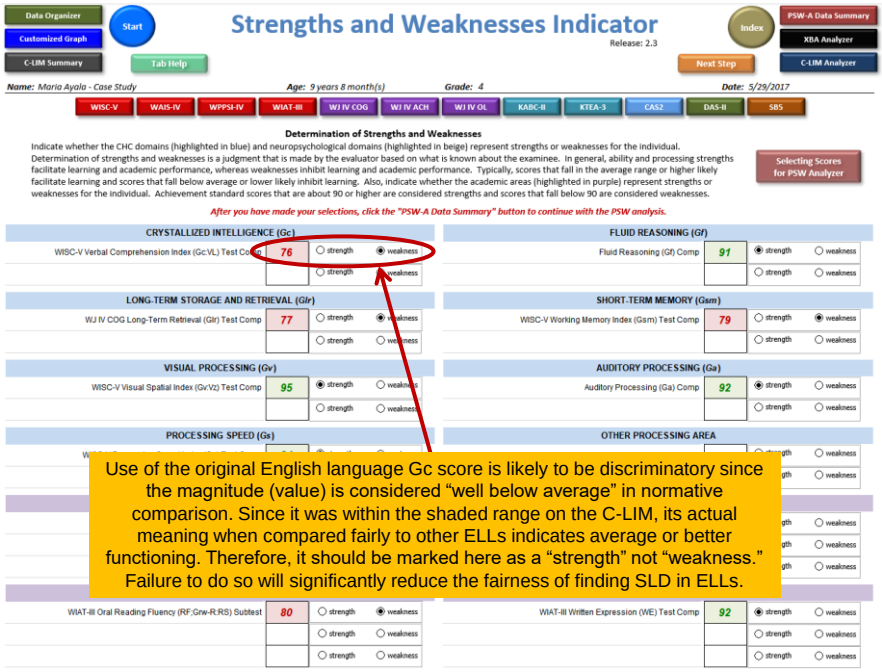
Strengths and weaknesses MUST be designated by the user. X-BASS does NOT make this determination as the meaning of any given score requires more information than just its magnitude.

SLD Identification with an English Learner: A Case Study



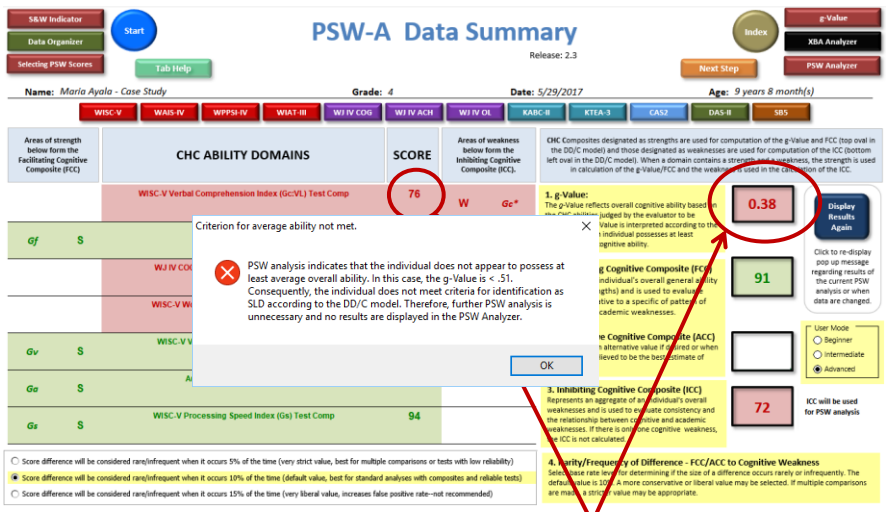
X-BASS will automatically warn you when a Gc score is indicated as a "weakness" when it falls within the expected range that corresponds to the degree of difference in the C-LIM (or default value—moderate, if not changed).

SLD Identification with an English Learner: A Case Study



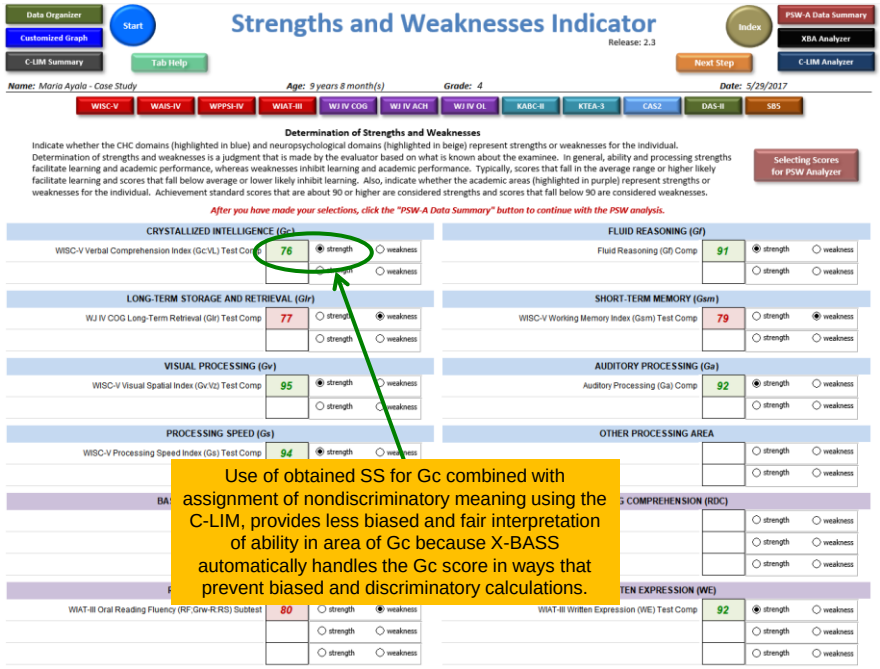
Use of the original English language Gc score is likely to be discriminatory since the magnitude (value) is considered "well below average" in normative comparison. Since it was within the shaded range on the C-LIM, its actual meaning when compared fairly to other ELLs indicates average or better functioning. Therefore, it should be marked here as a "strength" not "weakness." Failure to do so will significantly reduce the fairness of finding SLD in ELLs.

SLD Identification with an English Learner: A Case Study

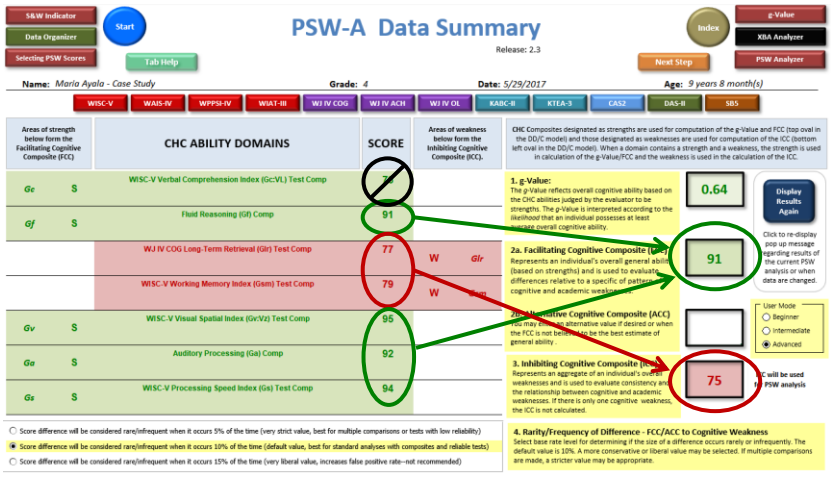


Because Gc is the most important ability related to academic success and accounts for the majority of variance in overall general ability, failure to properly evaluate it against other ELLs with comparable backgrounds may result in highly attenuated g-Values that suggest low ability and mask possible 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: A Case Study

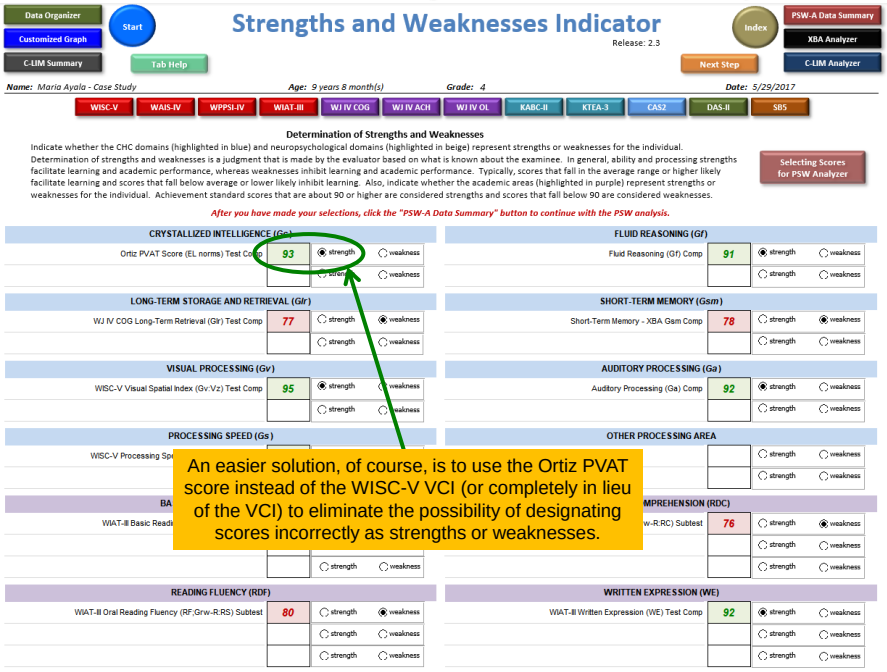


SLD Identification with an English Learner: A Case Study



To prevent discriminatory attenuation 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. Use of the Ortiz PVAT eliminates the need for this corrective action.

SLD Identification with an English Learner: A Case Study



Multilingual Assessment of ELs: Step by Step

Step 1. Test first in English (L2) and evaluate construct validity in all areas in English (exclusion of cultural/linguistic factors)

- *If all scores indicate normative strengths (SS ≈ 90 or higher) when tested in English (L2), scores are valid to the extent that a disability is not likely, thus no further testing is necessary.*
- *If some scores are normative weaknesses (SS < ≈ 90) evaluate test score validity in a research-based manner, e.g., via the C-LIM.*
- *If C-LIM indicates primary influence of language/culture, test scores are likely invalid and indicate average ability in all areas and a disability is not likely, thus no further testing is necessary.*
- *If C-LIM indicates contributory or minimal influence of language/culture, test scores are likely to be valid and the evaluation should continue.*

Step 2. Re-evaluate areas of weakness in native language (L2) to provide additional supporting evidence of validity (cross-linguistic confirmation)

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

Step 3. Further cross-validate L1 and L2 test scores with contextual factors and pre-referral data and academic concerns (ecological validity for disability)

- *Use all other case data and information to serve as the context by which to evaluate the test scores and ensure ecological validity to conclusions*

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

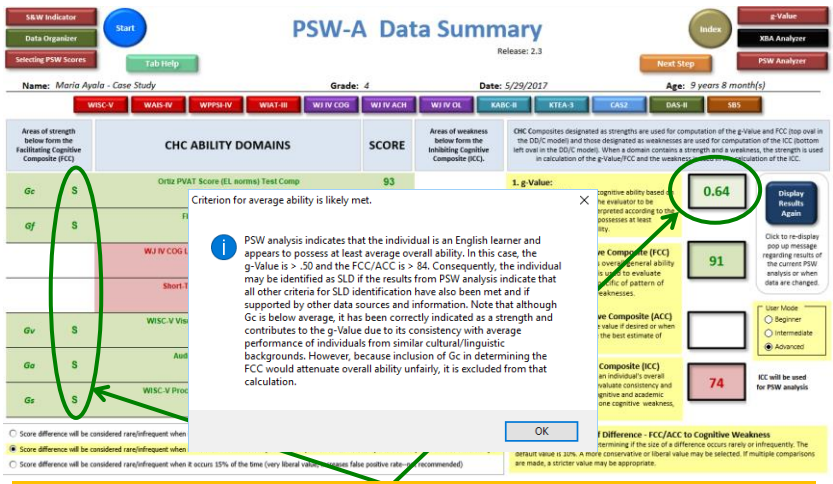
8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

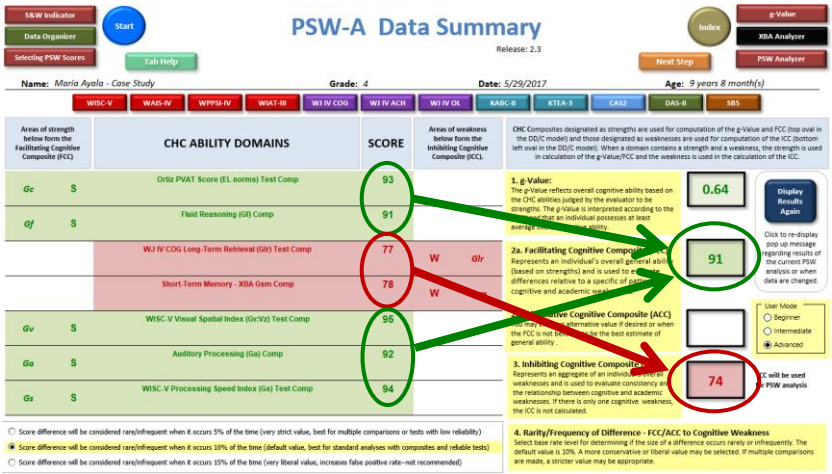
Procedures for Step 3

SLD Identification with an English Learner: A Case Study



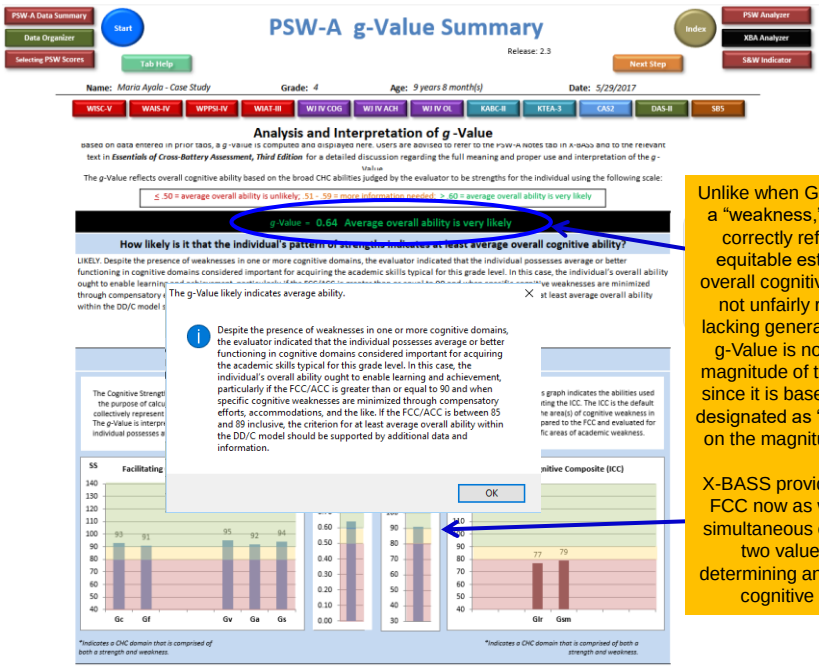
When a Gc (and any other CHC ability domain score), whether a broad or narrow composite, is marked as a "strength," it is included in calculations for determining the g-Value. Likewise, any score marked as a "weakness" are not used in deriving the g-Value. This keeps the g-Value free from the influence 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: A Case Study



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.

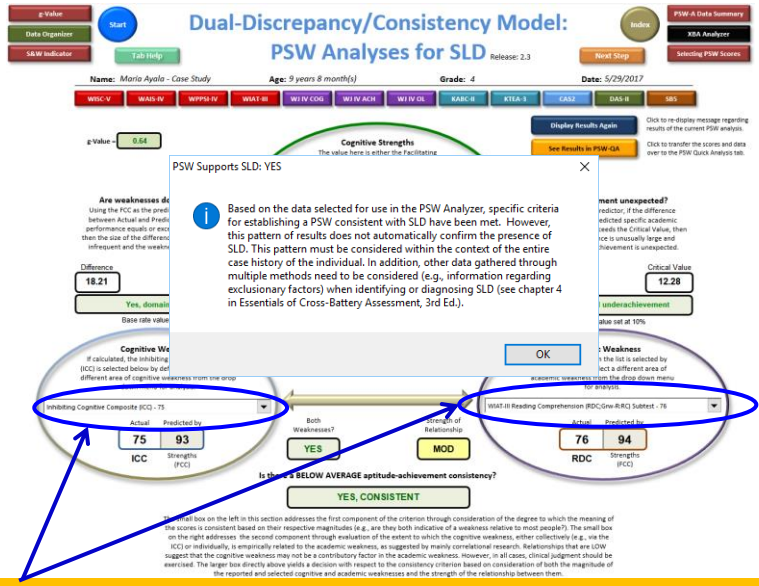
SLD Identification with an English Learner: A Case Study



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.

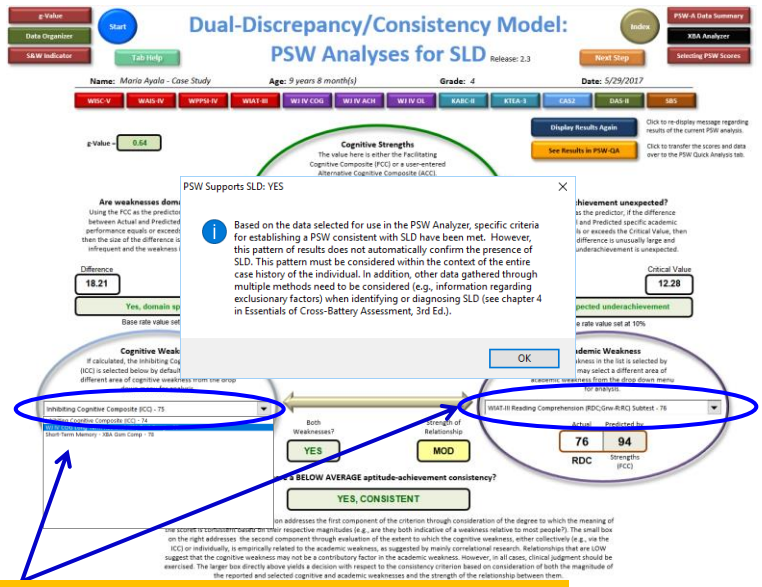
X-BASS provides a graph of the FCC now as well which allows simultaneous comparison of the two values as a part of determining an otherwise normal cognitive ability profile.

SLD Identification with an English Learner: A Case Study



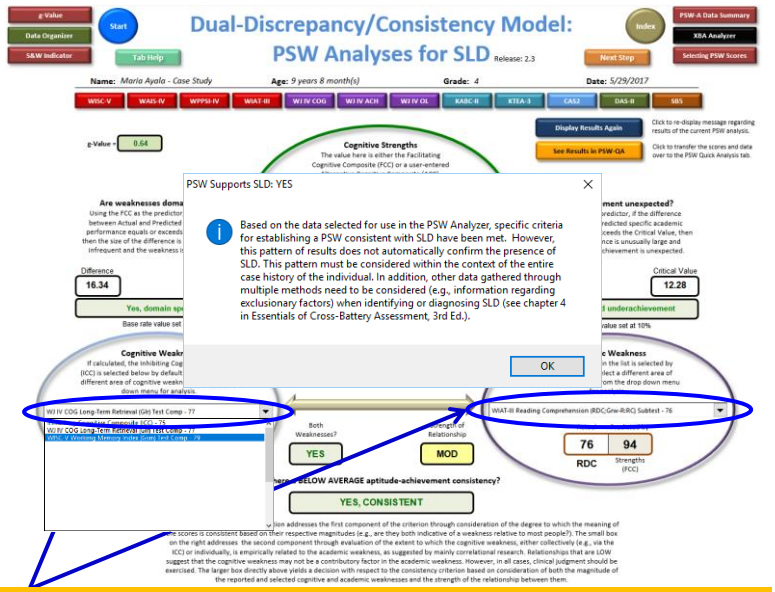
Using the ICC, data are consistent overall with SLD. But because the ICC is a trans-domain composite with greater reliability than a domain specific composite, it is more likely to reveal a significant difference. In addition, the ICC does not provide specific information regarding the nature of the 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: A Case Study



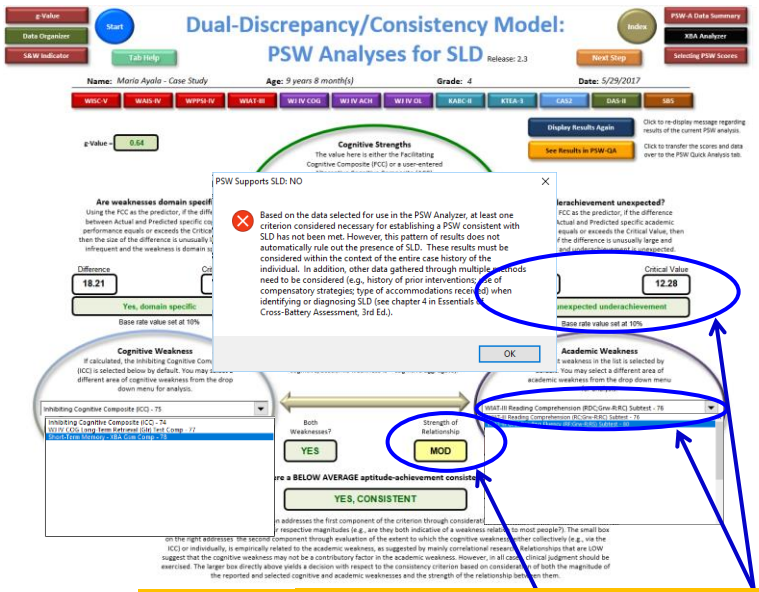
Changing the cognitive weakness to Glr also reveals a PSW consistent with SLD. It also provides a better indication that the academic problems are likely the result of deficits in Glr.

SLD Identification with an English Learner: A Case Study



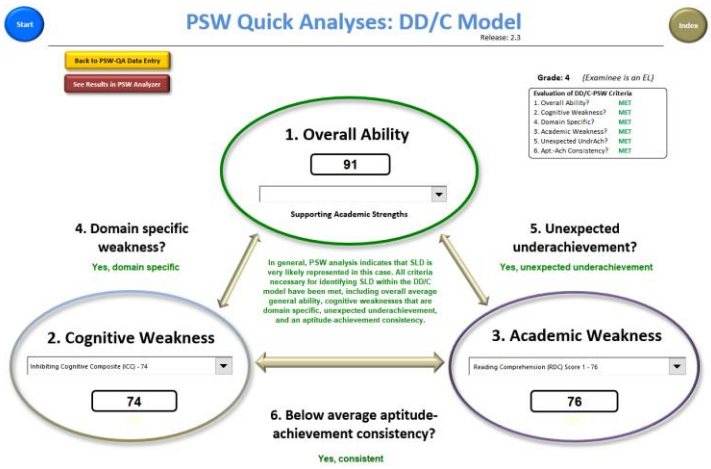
In this case, changing the cognitive weakness to Gsm also results in a PSW consistent with SLD and provides additional information regarding the likely cause of the academic problems as having a basis, at least in part, to deficits in Gsm.

SLD Identification with an English Learner: A Case Study



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: A Case Study



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

SLD Identification with an English Learner: A Case Study

1: Enter all available subtest scores in C-LIM Analyzer to determine validity

2: When likely/possibly valid, transfer data and enter remaining composite scores

3: Use XBA to conduct follow up testing where indicated and as necessary

4: Enter follow up tests and re-evaluate pattern with C-LIM Summary

5: If still likely/possibly valid, evaluate follow up testing results via XBA Analyzer

6: Transfer cohesive composites (and academic subtests) to Data Organizer

7: Identify deficits for native language re-evaluation and compare to original scores

8: Select best scores for PSW Analysis and designate each as strength or weakness

9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

10: Use additional data and information to support interpretations and conclusions

Procedures for Step 3

SLD Identification with an English Learner: A Case Study

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 two forms of evidence, a third should be added to fully support conclusions and interpretation of the obtained 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.

SLD Identification with an English Learner: A Case Study

Sample Validity Statement for ELL 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.

SLD Identification with an English Learner: A Case Study

Sample Validity Statement for EL Evaluations

Simplified Validity Statement for LIKELY SLD 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 suspected SLD; 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. This statement has been placed in the public domain and may be freely copied, modified, and distributed for non-profit purposes without the need to secure permission.

SLD Identification with an English Learner: A Case Study

Sample Validity Statement for EL Evaluations

Simplified Validity Statement for UNLIKELY SLD 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.

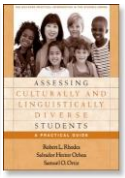
However, because the scores were compared to other individuals from research studies who were of average ability and who had not been identified as having a disability, it suggests that XXXX's performance is also average (possibly higher) and that it is not likely that a learning disability is present 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.

Assessment and Related Resources

TESTS:

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

Ortiz PVAT Free 30-Day Trial and 2 Free Uses
http://info.mhs.com/OrtizPVATfreetrial

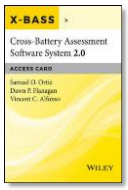


BOOKS:

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, S. O., Flanagan, D. P. & Alfonso, V. C. (Winter 2019--coming soon). Intervention Library: Finding interventions, resources and supports for students with learning difficulties (IL:FIRST v1.0). New York: Wiley & Sons, Inc.

Flanagan, D. P., Ortiz, S.O. & Alfonso, V.C. (2013). Essentials of Cross-Battery Assessment, Third Edition. New York: Wiley & Sons, Inc.



ONLINE:

Competency-based XBA Certification Program
https://www.schoolneuropsych.com/xba/

CHC Cross-Battery Online
http://www.crossbattery.com/

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

Cross-Battery XBA Assessment