Dual Identification of English Learners:
Use of a PSW model for determining SLD with ELs.

COSA 2016 State English Learners Alliance Conference
March 9, 2016
Eugene, OR

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

IDEA Regulations 34 CFR 300.307(a)
Procedures for Identification of Specific Learning Disability

A State must adopt, consistent with 34 CFR 300.309, criteria for determining whether a child has a specific learning disability as defined in 34 CFR 300.8(c)(10). In addition, the criteria adopted by the State:

- Must not require the use of a severe discrepancy between intellectual ability and achievement for determining whether a child has a specific learning disability, as defined in 34 CFR 300.8(c)(10);
- Must permit the use of a process based on the child's response to scientific, research-based intervention; and
- *May permit the use of other alternative research-based procedures for determining whether a child has a specific learning disability, as defined in 34 CFR 300.8(c)(10).*

This includes use of approaches based on a Pattern of Strengths and Weaknesses (PSW)

IDEA Regulations 34 CFR 300.307(a)
Procedures for Identification of Specific Learning Disability

The group described in 34 CFR 300.306 may determine that a child has a specific learning disability, as defined in 34 CFR 300.8(c)(10), if…

- The child does not make sufficient progress to meet age or State-approved grade-level standards in one or more of the areas identified in 34 CFR 300.309(a)(1) when using a process based on the child's response to scientific, research-based intervention; or the child exhibits a pattern of strengths and weaknesses in performance, achievement, or both, relative to age, State-approved grade-level standards, or intellectual development, that is determined by the group to be relevant to the identification of a specific learning disability, using appropriate assessments, consistent with 34 CFR 300.304 and 300.305; and the group determines that its findings under 34 CFR 300.309(a)(1) and (2) are not primarily the result of:
  - A visual, hearing, or motor disability;
  - Mental retardation;
  - Emotional disturbance;
  - Cultural factors;
  - Environmental or economic disadvantage; or
  - Limited English proficiency.

Recognizes that lack of English proficiency or cultural difference cannot be the basis of a disability and cannot be the primary reason for observed academic problems.

Source: IDEA Statute and Regulations. Last retrieved on Feb. 5, 2016 from http://www.wrightslaw.com/idea/law/section1414.pdf

20 U.S.C. 1414 Evaluations, Eligibility Determinations, Individualized Education Programs, and Educational Placements

(b) EVALUATION PROCEDURES –

(3) ADDITIONAL REQUIREMENTS- Each local educational agency shall ensure that—

(A) ~~tests~~ assessments and other evaluation materials used to assess a child under this section—

(i) are selected and administered so as not to be discriminatory on a racial or cultural basis;

(ii) are provided and administered in the ~~child's native language or other mode of communication~~ *language and form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally, unless it is not feasible to so provide or administer;*

(iii) are used for purposes for which the assessments or measures are valid and reliable;

Recognizes that validity is not automatically assured via native language testing.

Source: IDEA Statute and Regulations. Last retrieved on Feb. 5, 2016 from http://www.wrightslaw.com/idea/law/section1414.pdf

General Nondiscriminatory Assessment Processes and Procedures

I. Assess for the purpose of intervention

II. Assess initially with authentic and alternative procedures

III. Assess and evaluate the learning ecology

IV. Assess and evaluate language proficiency

V. Assess and evaluate opportunity for learning

VI. Assess and evaluate relevant cultural and linguistic factors

VII. Evaluate, revise, and re-test hypotheses

VIII. Determine the need for and language(s) of formal assessment

IX. Reduce potential bias in traditional assessment practices

X. Support conclusions via data convergence and multiple indicators

Addresses concerns regarding fairness and equity in the assessment process

Addresses possible bias in use of test scores

Pre-referral procedures (I. - VIII.)
Post-referral procedures (IX. - X.)

Summary of Research on the Test Performance of English Language Learners

Research conducted over the past 100 years on ELLs who are non-disabled, of average ability, possess moderate to high proficiency in English, and tested in English, has resulted in two robust and ubiquitous findings:

1. *Native English speakers perform better than English learners at the broad ability level (e.g., FSIQ) on standardized, norm-referenced tests of intelligence and general cognitive ability.*

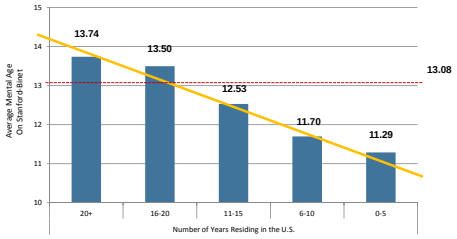
2. *English learners tend to perform significantly better on nonverbal type tests than they do on verbal tests (e.g., PIQ vs. VIQ).*

So what explains these findings? Early explanations relied on genetic differences attributed to race even when data strongly indicated that the test performance of ELLs was moderated by the degree to which a given test relied on or required age- or grade-expected development in English and the acquisition of incidental acculturative knowledge.

Research Foundations for ELL Evaluation

Principle 1: ELLs and non-ELL's perform differently at the broad ability level

Mean Mental Age (MA) from Binet Scales in a non-native English speaking sample from Yerkes' (1921) data as analyzed by C.C. Brigham (1923)

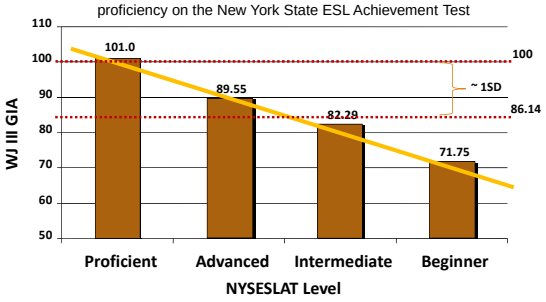


Average score for native English speakers on Beta = 101.6 (Very Superior; Grade A)
Average score for non-native English speakers on Beta = 77.8 (Average; Grade C)

Research Foundations for ELL Evaluation

Principle 1: ELLs and non-ELL's perform differently at the broad ability level

Mean WJ III GIA across the four levels of language proficiency on the New York State ESL Achievement Test

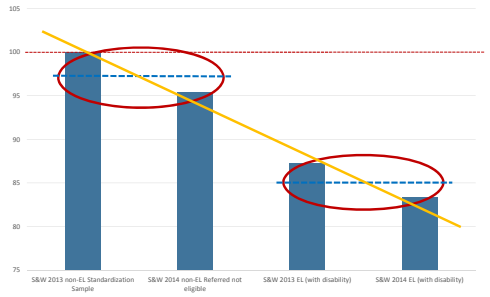


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013).

Research Foundations for ELL Evaluation

Principle 1: ELLs and non-ELL's perform differently at the broad ability level

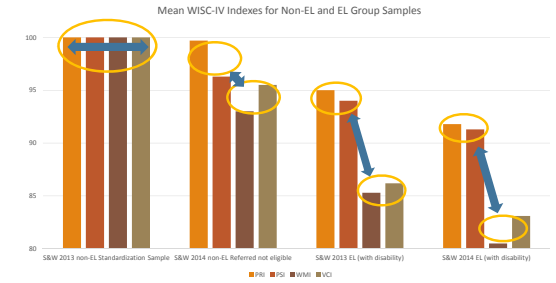
Mean WISC-IV FSIQ for Non-EL and EL Group Samples



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 ELL Evaluation

Principle 2: ELLs perform better on nonverbal tests than verbal tests



Soures: 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. and Styck, K. M. & Watkins, M. W. (2014). Discriminant Validity of the WISC-IV Culture-Language Interpretive Matrix. Contemporary School Psychology , 18, 168-188.

Research Foundations for ELL Evaluation

Historical and contemporary research has tended to ignore the fact that ELLs do not perform at the same level on ALL nonverbal tests any more than they perform at the same level on ALL verbal tests.

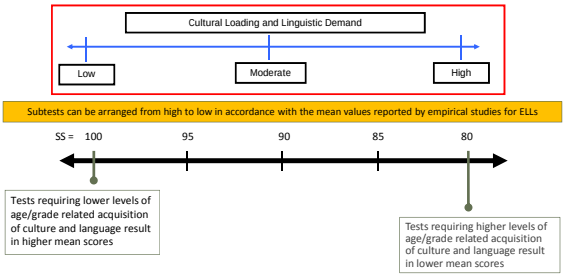
Instead, it appears that test performance of ELLs is not a dichotomy but rather a continuum formed by a linear, not dichotomous, attenuation of performance.

This means, a third principle is evident in the body of research on ELLs but has not been well understood or utilized in understanding test performance:

3. Test performance of ELLs is moderated by the degree to which a given test relies on or requires age- or grade-expected English language development and the acquisition of incidental acculturative knowledge.

Research Foundations for ELL Evaluation

ELL test performance is a linear, continuous pattern, not a dichotomy.



Research Foundations for ELL Evaluation

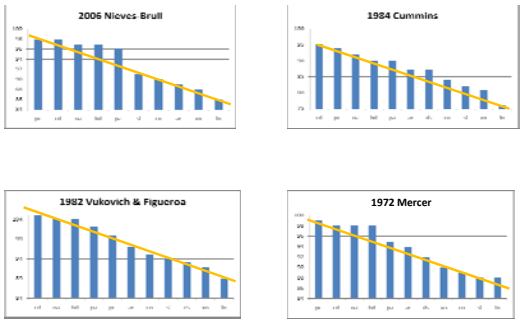
Principle 3: ELL performance is moderated by linguistic/acculturative variables

	Hispanic Group (Mercer) (1972)	Hispanic Group (Vukovich & Figueroa) (1982)	ESL Group (Cummins) (1982)	Bilingual Group (Nieves-Brull) (2006)
Subtest Name	**Mean SS**	**Mean SS**	**Mean SS**	**Mean SS**
Information	7.5	7.8	5.1	7.2
Vocabulary	8.0	8.3	6.1	7.5
Similarities	7.6	8.8	6.4	8.2
Comprehension	7.8	9.0	6.7	8.0
Digit Span	8.3	8.5	7.3	*
Arithmetic	8.7	9.4	7.4	7.8
Picture Arrangement	9.0	10.3	8.0	9.2
Block Design	9.5	10.8	8.0	9.4
Object Assembly	9.6	10.7	8.4	9.3
Picture Completion	9.7	9.9	8.7	9.5
Coding	9.6	10.9	8.9	9.6

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

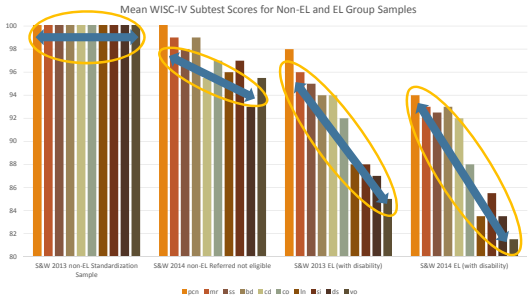
Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables



Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables

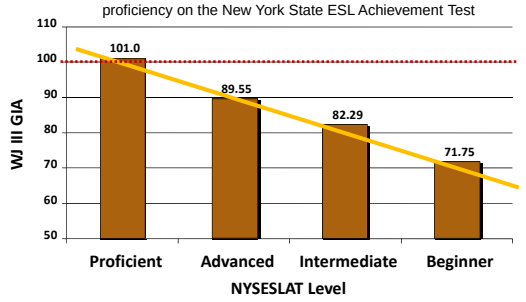


Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382.

Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables

Mean WJ III GIA across the four levels of language proficiency on the New York State ESL Achievement Test



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

Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables

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

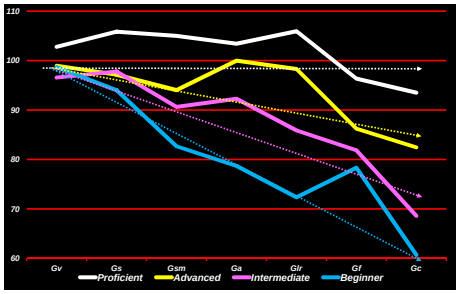
	Variance explained		
Individual test	7-10	11-14	15-18
Verbal Comprehension	.79[a]	.86[a]	.81[a]
General Information	.71[a]	.85[a]	.86[a]
Concept Formation	.67[a]	.71[a]	.67[a]
Visual–Auditory Learning	.40[b]	.37[b]	.41[b]
Delayed Recall Visual–Auditory Learning	.39[b]	.32[b]	.37[b]
Analysis Synthesis	.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]
Numbers Reversed	.17[b]	.26[b]	.30[b]
Pair Cancelation	.17[b]	.11[b]	.11[b]
Rapid Picture Naming	.16[b]	.07[b]	.16[b]
Incomplete Words	.13[b]	.31[b]	.23[b]
Visual Matching	.13[b]	.15[b]	.16[b]
Decision Speed	.12[b]	.15[b]	.19[b]
Auditory Attention	.10[b]	.20[b]	.15[b]
Spatial Relations	.08[b]	.16[b]	.16[b]
Planning	.07[b]	.12[b]	.11[b]
Picture Recall	.02[b]	.06[b]	.10[b]

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

Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables

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

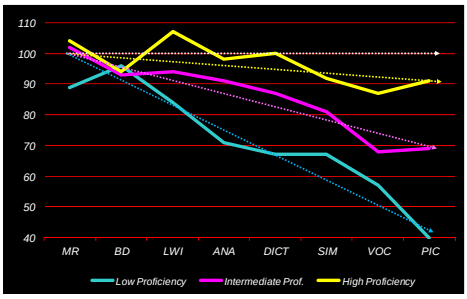


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013). English Language Proficiency and Test Performance: Evaluation of bilinguals with the Woodcock-Johnson III Tests of Cognitive Ability. Psychology in the Schools, Vol 50(8), pp. 781-797.

Research Foundations for ELL Evaluation

Principle 3: ELL performance is moderated by linguistic/acculturative variables

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



Source: Dynda, A.M., Flanagan, D.P., Chaplin, W., & Pope, A. (2008), unpublished data..

Foundational Research Principles of the Culture-Language Interpretive Matrix

Principle 1: EL and non-EL's perform differently at the broad ability level on tests of cognitive ability.

Principle 2: ELs perform better on nonverbal tests than they do on verbal tests.

Principle 3: EL performance on both verbal and nonverbal tests is moderated by linguistic and acculturative variables.

Because the basic research principles underlying the C-LIM are well supported, their operationalization within the C-LIM provides a substantive evidentiary base for evaluating the test performance of English language learners.

- This does not mean, however, that it cannot be improved. Productive research on EL test performance can assist in making any necessary "adjustments" to the order of the means as arranged in the C-LIM.
- Likewise, as new tests come out, new research is needed to determine the relative level of EL performance as compared to other tests with established values of expected average performance.
- Ultimately, only research that focuses on stratifying samples by relevant variables such as language proficiency, length and type of English and native language instruction, and developmental issues related to age and grade of first exposure to English, will serve useful in furthering knowledge in this area and assist in establishing appropriate expectations of test performance for specific populations of ELs.

Main Threats to Test Score Validity for ELLs

NO BIAS

- ***Test items***
  (content, novelty)
- ***Test structure***
  (sequence, order, difficulty)
- ***Test reliability***
  (measurement error/accuracy)
- ***Factor structure***
  (theoretical structure, relationship of variables to each other)
- ***Predictive Validity***
  (correlation with academic success or achievement)

BIAS

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

When a test measures an unintended variable…

- ***Incorrect Interpretation***
  (undermines accuracy of evaluative judgments and meaning assigned to scores)

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

Main Threats to Test Score Validity for ELLs

Acculturative Knowledge Acquisition – Not Race or Ethnicity

"When a child's general background experiences differ from those of the children on whom a test was standardized, then the use of the norms of that test as an index for evaluating that child's current performance or for predicting future performances may be inappropriate."

Salvia & Ysseldyke, 1991

Developmental Language Proficiency – Not Language Dominance

"Most studies compare the performance of students from different ethnic groups…rather than ELL and non-ELL children within those ethnic groups….A major difficulty with all of these studies is that the category Hispanic includes students from diverse cultural backgrounds with markedly different English-language skills….This reinforces the need to separate the influences of ethnicity and ELL status on observed score differences."

Lohman, Korb & Lakin, 2008

Processes and Procedures for Addressing Test Score Validity

IX. REDUCE BIAS IN TRADITIONAL TESTING PRACTICES

Exactly how is evidence-based, nondiscriminatory assessment conducted and to what extent is there any research to support the use of any of these methods in being capable of establishing sufficient validity of the obtained results?

- **Modified Methods of Evaluation**
  - *Modified and altered assessment*
- **Nonverbal Methods of Evaluation**
  - *Language reduced assessment*
- **Dominant Language Evaluation: L1**
  - *Native language assessment*
- **Dominant Language Evaluation: L2**
  - *English language assessment*

Comparison of Methods for Addressing Main Threats to Validity

Evaluation Method	Norm sample representative of bilingual development	Measures full range of ability constructs	Does not require bilingual evaluator	Adheres to the test's standardized protocol	Substantial research base on bilingual performance
Modified or Altered Assessment	✗	✓	✓	✗	✗
Reduced-language Assessment	✗	✗	✓	✓	✗
Dominant Language Assessment in L1: native	✗	✓	✗	✓	✗
Dominant Language Assessment in L2: English	✗	✓	✓	✓	✓

Addressing issues of fairness with respect to norm sample representation is an issue of validity and dependent on a sufficient research base.

Evaluating and Defending Construct ELL Test Score Validity

Whatever method or approach may be employed in evaluation of ELL's, the fundamental obstacle to nondiscriminatory interpretation rests on the degree to which the examiner is able to defend claims of test score construct validity. This is captured by and commonly referred to as a question of:

"DIFFERENCE vs. DISORDER?"

Simply absolving oneself from responsibility of doing so 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 interpretation.

At present, the only manner in which test score validity can be evaluated or established is via use of the existing research on the test performance of ELLs as reflected in the degree of "difference" the student displays relative to the norm samples of the tests being used, particularly for tests in English. This is the sole purpose of the C-LIM.

Practical Considerations for Addressing Validity in Evaluation Procedures for SLD with ELLs

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*

Practical Considerations for Addressing Validity in Evaluation Procedures for SLD with ELLs

Given the preceding considerations, the most practical and defensible general approach in evaluating ELLs would be:

- *Test in English first and if all test scores indicate strengths (average or higher) a disability is not likely and thus no further testing is necessary*
- *If some scores from testing in English indicate weaknesses, re-test those areas in the native language to cross-validate as areas of true weakness*

This approach provides the most efficient process and best use of available resources for evaluation since it permits ANY evaluator to begin and sometimes complete the testing without being bilingual or requiring assistance.

In addition, this approach is IDEA compliant and consistent with the specification that assessments "be provided and administered in the language and form most likely to yield accurate information" because it relies on an established body of research to guide examination of test score validity and ensures that that the results upon which decisions are based are in fact accurate.

A Recommended Best Practice Approach for Using Tests with ELLs

Step 1. Assessment of Bilinguals – validate all areas of performance (exclusion of cultural/linguistic factors)

- *Select or create an appropriate battery that is comprehensive and responds to the needs of the referral concerns, irrespective of language differences*
- *Administer all tests in standardized manner first in English only with no modifications*
- *Score tests and plot them for analysis via the C-LIM*
- *If analysis indicates expected range and pattern of decline, scores are invalid due to cultural and linguistic factors that cannot be excluded as primary reason for poor academic performance*
- *If analysis does not indicate expected range or pattern of decline, apply XBA (or other) interpretive methods to determine specific areas of weakness and difficulty and continue to Step 2*

Step 2. Bilingual Assessment – validate suspected areas of weakness (cross-language confirmation of deficit areas)

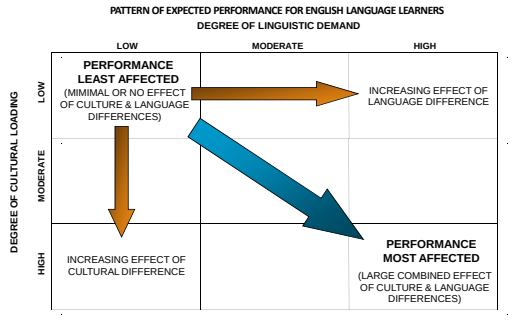
- *Review results and identify areas of suspected weakness or difficulty:*
  a. *For Gc only, evaluate weakness according to high/high cell in C-LIM or in context of other data and information*
  b. *For all other abilities, evaluate weakness using standard classifications (e.g., SS < 90)*
- *Except for Gc, re-test all other areas of suspected weakness using native language tests*
- *For Gc only:*
  a. *If the high/high cell in C-LIM is within/above expected range, consider Gc a strength and assume it is at least average, thus re-testing is not necessary*
  b. *If the high/high cell in C-LIM is below expected range, re-testing of Gc in the native language is recommended*
- *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*
- *Except for Gc, if a score obtained in the native language validates/confirms a weakness score obtained in English (both SS < 90), use/interpret the score obtained in English as a weakness*
- *If a score obtained in the native language invalidates/disconfirms a weakness score obtained in English (native SS ≥ 90), consider it as a strength and assume that it is at least in the average range*
- *Scores for Gc obtained in the native language and in English can only be interpreted relative to developmental and educational experiences of the examinee in each language and only as compared to others with similar developmental experiences*

The Culture-Language Interpretive Matrix (C-LIM) Addressing test score validity for ELLs

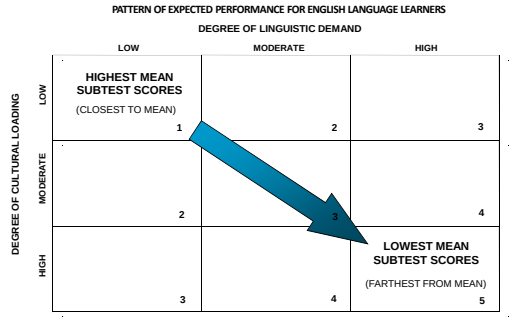
Translation of Research into Practice

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 **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 **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.

Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and C-LIM



Application of Research as Foundations for the Cultural and Linguistic Classification of Tests and C-LIM



The Culture-Language Interpretive Matrix (C-LIM)

Important Considerations for Use and Practice

The C-LIM is not a test, scale, measure, or mechanism for making diagnoses. It is a visual representation of current and previous research on the test performance of English learners arranged by mean values to permit examination of the combined influence of acculturative knowledge acquisition and limited English proficiency and its impact on test score validity.

The C-LIM is not a language proficiency measure and will not distinguish native English speakers from English learners with high, native-like English proficiency and is not designed to determine if someone is or is not an English learner. Moreover, the C-LIM is not for use with individuals who are native English speakers.

The C-LIM is not designed or intended for diagnosing any particular disability but rather as a tool to assist clinician's in making decisions regarding whether ability test scores should be viewed as indications of actual disability or rather a reflection of differences in language proficiency and acculturative knowledge acquisition.

The primary purpose of the C-LIM is to assist evaluators in ruling out cultural and linguistic influences as exclusionary factors that may have undermined the validity of test scores, particularly in evaluations of SLD or other cognitive-based disorders. Being able to make this determination is the primary and main hurdle in evaluation of ELLs and the C-LIM's purpose is to provide an evidence-based method that assists clinician's regarding interpretation of test score data in a nondiscriminatory manner.

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.

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.

The Culture-Language Interpretive Matrix (C-LIM)

RANGE OF POSSIBLE OUTCOMES WHEN EVALUATING TEST SCORES WITHIN C-LIM

Condition A: Overall pattern generally appears to decline across all cells and all cell aggregate scores within or above shaded range—test scores likely **invalid**, cultural/linguistic factors are primary influences, but examinee likely has average/higher ability as data do not support deficits, and further evaluation via testing is unnecessary.

Condition B: Overall pattern generally appears to decline across all cells but at least one cell aggregate (or more) is below shaded range—test scores **possibly valid**, cultural/linguistic factors are contributory influences, and further evaluation, including in the native language, is necessary to establish true weaknesses in a given domain.

Condition C: Overall pattern does not appear to decline across all cells and all cell aggregate scores within or above average range—test scores **likely valid**, cultural/linguistic factors are minimal influences, and further evaluation may be unnecessary if no weaknesses exist in any domain.

Condition D: Overall pattern does not appear to decline across all cells and at least one cell aggregate (or more) is below average range—test scores **possibly valid**, cultural/linguistic factors are minimal influences, and further evaluation, including in the native language, is necessary to establish true weaknesses in a given domain.

The Culture-Language Interpretive Matrix (C-LIM)

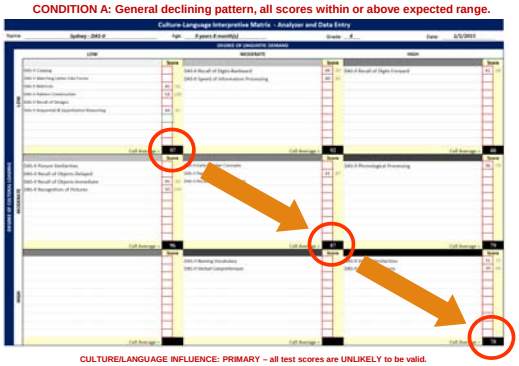
RANGE OF POSSIBLE OUTCOMES WHEN EVALUATING TEST SCORES WITHIN C-LIM

	A general, overall pattern of decline exists?	All scores within or above the expected range?	All scores within or above the average range?	Degree of influence of cultural and linguistic factors	Likelihood that test scores are valid indicators of ability?
Condition A	Yes	Yes	No	Primary	Unlikely
Condition B	Yes	No	No	Contributory	Possibly*
Condition C	No	Yes	Yes	Minimal	Likely
Condition D	No	No	No	Minimal	Possibly*

**Determination regarding the validity of test scores that are below the expected and average ranges requires additional data and information, particularly results from native language evaluation, qualitative evaluation and analysis, and data from a strong pre-referral process (e.g., progress monitoring data).*

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

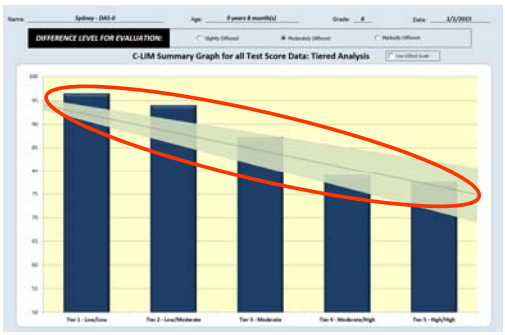
CONDITION A: General declining pattern, all scores within or above expected range.



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

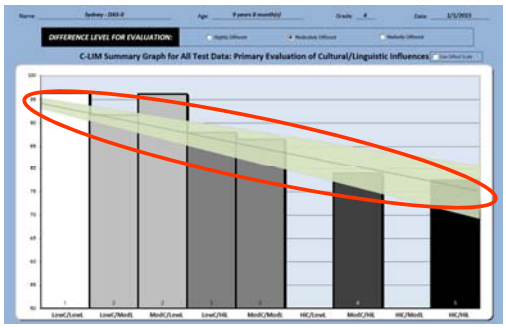
CONDITION A: General declining pattern, all scores within or above expected range.



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

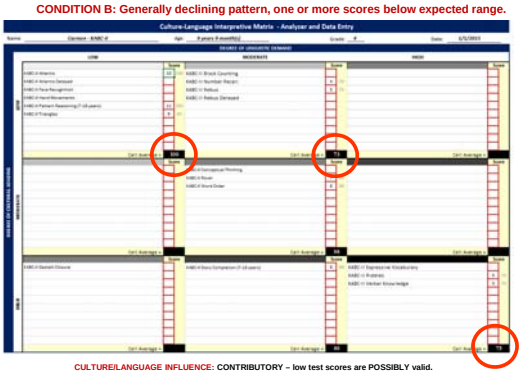
CONDITION A: General declining pattern, all scores within or above expected range.



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

CONDITION B: Generally declining pattern, one or more scores below expected range.



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

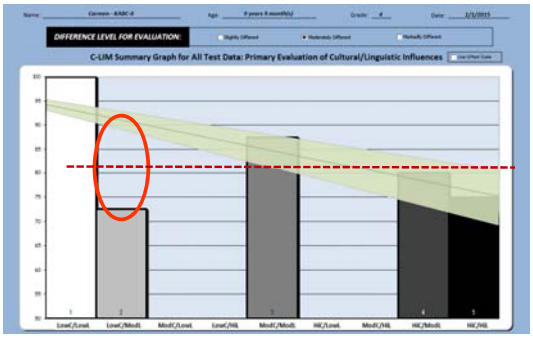
Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

CONDITION B: Generally declining pattern, one or more scores below expected range.

CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

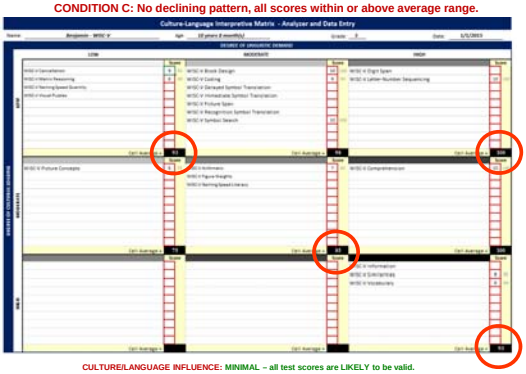
CONDITION B: Generally declining pattern, one or more scores below expected range.



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

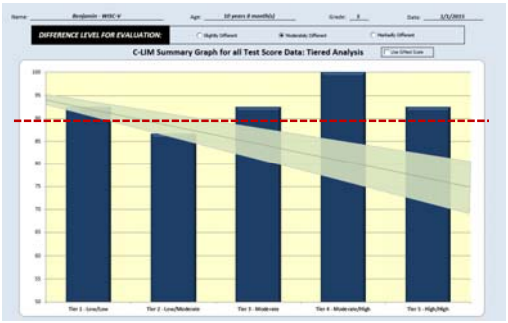
CONDITION C: No declining pattern, all scores within or above average range.



CULTURE/LANGUAGE INFLUENCE: MINIMAL – all test scores are LIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

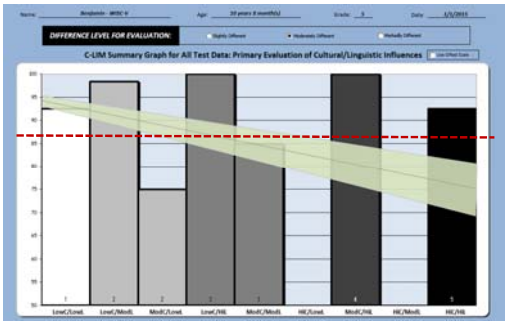
CONDITION C: No declining pattern, all scores within or above average range.



CULTURE/LANGUAGE INFLUENCE: MINIMAL – all test scores are LIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

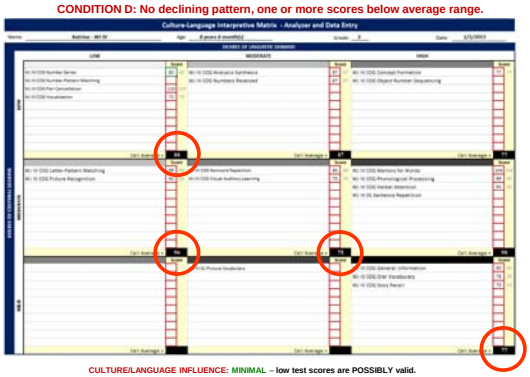
CONDITION C: No declining pattern, all scores within or above average range.



CULTURE/LANGUAGE INFLUENCE: MINIMAL – all test scores are LIKELY to be valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

CONDITION D: No declining pattern, one or more scores below average range.



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

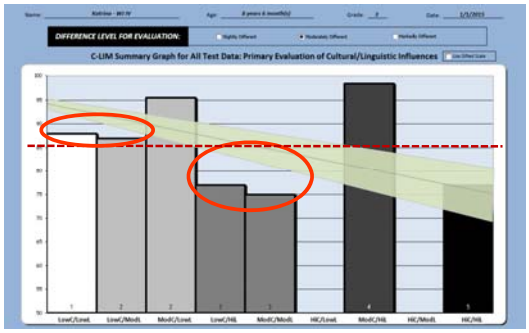
Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

CONDITION D: No declining pattern, one or more scores below average range.

CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Guidelines for evaluating test scores.

CONDITION D: No declining pattern, one or more scores below average range.



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY 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.

Evaluation of the 2013 Styck and Watkins* Study on Use of WISC-IV and C-LIM with English Language Learners

The main finding in the study is stated as follows:

> *"The valid C-LIM profile (i.e., cell means did not decline) emerged in the mean WISC-IV normative sample and the ELL sample." (p. 374). (emphasis added)*

It is clear that the normative sample "did not decline" as their mean on every subtest was invariant,10.3 (SS=102). However, for the ELL sample, the highest mean was on Picture Concepts (SS=98) and lowest was on Vocabulary (SS=85). With minor variation, examination of the data in the following table strongly suggests a clear decline in the ELL sample's means.

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

A Critical Review of Research on the C-LIM: Styck & Watkins

Decline or No Decline? Comparison of Means for WISC-IV Subtests

WISC-IV Subtest	Norm Sample Mean[a]	ELL Mean 2013	Difference[b]	ELL Mean 2014	Difference[b]
Picture Concepts	102	98	4	94	8
Matrix Reasoning	102	96	6	93	9
Symbol Search	102	95	7	93	9
Block Design	102	94	8	93	9
Coding	102	94	8	92	10
Comprehension	102	92	10	88	14
Letter-Number Sequencing	102	88	14	84	18
Similarities	102	88	14	86	16
Digit Span	102	87	15	84	14
Vocabulary	102	85	17	82	20

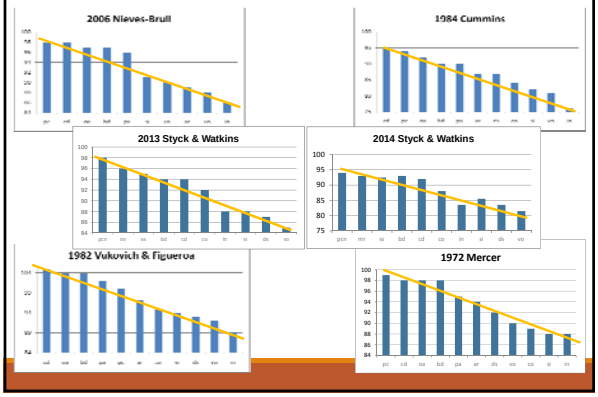
[a] *Means were reported in the study as Scaled Scores (e.g., 10.3). They have been converted here to Deviation IQ metric for the sake of simplicity.*

[b] *The difference between all 15 norm sample and ELL subtest and composite means were found to be statistically significant at the p<.001 level.*

Source: Styck, K. M. & Watkins, M. W. (2013). Diagnostic Utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition Among Referred Students. School Psychology Review, 42(4), 367-382. and Styck, K. M. & Watkins, M. W. (2014). Discriminant Validity of the WISC-IV Culture-Language Interpretive Matrix. Contemporary School Psychology , 18, 168-188.

Declining ELL Test Performance on the WISC

Comparison of 2013 Styck & Watkins data and other WISC studies with ELLs



Evaluation of the Styck and Watkins* Study on Use of WISC-IV and C-LIM with English Language Learners

Main conclusion in the study is stated as follows:

"Thus, neither sample of children exhibited the invalid C-LIM profile when group mean scores were considered" (p. 374) (emphasis added).

The "invalid C-LIM profile" would be indicated by a systematic decline in mean scores in the matrix meaning that the test results were influenced primarily by the presence of cultural and linguistic variables.

The C-LIM is intended to compare individual performance against the group, not evaluate group scores, especially from a population where 97% have identified disabilities. Nevertheless, with a sufficiently large sample such differences in performance are likely to become more and more randomly distributed. Moreover, the C-LIM is certainly subject to modification on the basis of additional quality research.

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

Evaluation of the 2013 Styck and Watkins* Study on Use of WISC-IV and C-LIM with English Language Learners

But the study wasn't conducted with non-disabled ELLs:

"roughly 97% of (n = 83) of participants were identified as meeting criteria for an educational disability (86% as SLD)" (p. 371).

As noted previously, this suggests that individual C-LIM profiles should display ***valid*** results, not invalid, since valid results are needed to support the district's identification of a disability.

When individual C-LIM's for the ELL group were examined, they found that nearly 89.5% of the ELLs did in fact display valid results indicating that any low scores could well reflect a disability and indicating a very high degree of consistency with the clinical decisions made by the district's eligibility team.

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

Evaluation of the 2013 Styck and Watkins* Study on Use of WISC-IV and C-LIM with English Language Learners

		Different (ELL Group)	Standard (Norm Group)
WISC-IV C-LIM Analysis	Invalid Scores	6/3 (7.0%)/(3.5%)	100 (4.9%)
	Valid Scores	77 (89.5%)	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 correct so that the C-LIM was consistent with and supported the placement decision of the child by the district in 93% of the cases.

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

English learners are not a monolithic group with all of the same educational, cultural, and linguistic experiences. Consideration must always be given to these factors and the role they may be playing in setting the context for appropriate expectations of performance.

1. ***Evaluate consistency of score patterns across all graphs***
2. ***Evaluate variability in scores with the same classifications in the matrix***
3. ***Evaluate developmental factors related to education and experience***
4. ***Evaluate developmental factors related to linguistic/acculturative experiences***

Failure to properly account for these issues may result in inequitable expectations of performance and discriminatory conclusions regarding an examinee's true or actual abilities or lack thereof.

Culture-Language Interpretive Matrix (C-LIM): Case Study - Elizabeth

Woodcock-Johnson IV: Tests of Cognitive Ability (English Administration)

	SS	PR		SS	PR		SS	PR
Oral Vocabulary	69	2	General Information	79	8	Number Series	96	39
Concept Formation	87	19	Verbal Attention	84	14	Numbers Reversed	92	30
Letter-Pattern Matching	98	45	Pair Cancellation	94	34	Phonological Processing	81	10
Nonword Repetition	91	27	Story Recall	83	13	Visual-Auditory Learning	89	23
Visualization	102	55	Picture Recognition	91	27			

Wechsler Intelligence Scale for Children - V (English Administration)

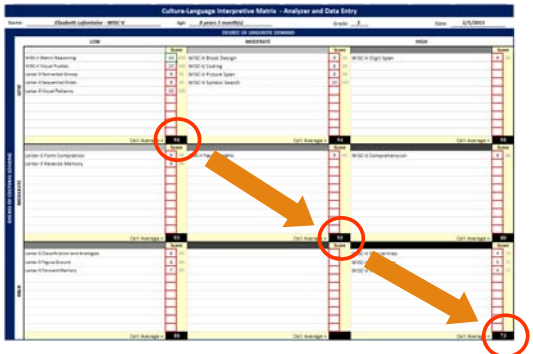
	Scaled Score	PR	Standard Score		Scaled Score	PR	Standard Score
Information	5	9	80	Block Design	9	38	95
Similarities	4	2	70	Matrix Reasoning	10	50	100
Vocabulary	5	2	70	Symbol Search	10	50	100
Comprehension	6	16	85	Coding	8	25	90
Digit Span	9	38	95	Visual Puzzles	10	50	100

Leiter-3 Nonverbal Intelligence Test (Nonverbal Administration)

	Scaled Score	PR	Standard Score		Scaled Score	PR	Standard Score
Nonverbal Stroop	9	38	95	Sequential Order	8	25	90
Visual Patterns	9	38	95	Form Completion	8	25	90
Reverse Memory	10	50	100	Classification & Analg.	8	25	90
Figure Ground	8	25	90	Forward Memory	7	16	85

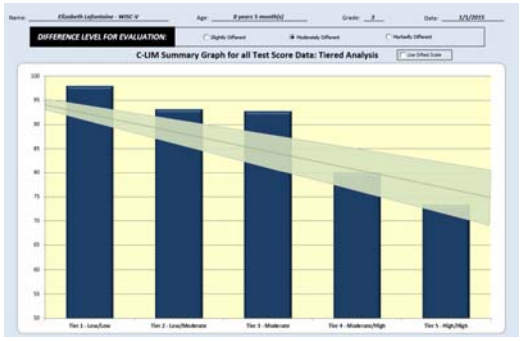
Culture-Language Interpretive Matrix: Case Study

WISC-V & UNIT DATA FOR ELIZABETH (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

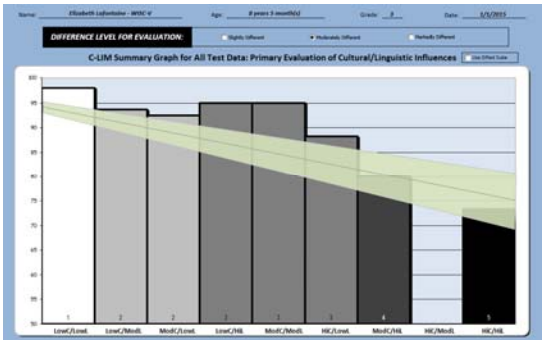
WISC-V & UNIT DATA FOR ELIZABETH (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Case Study

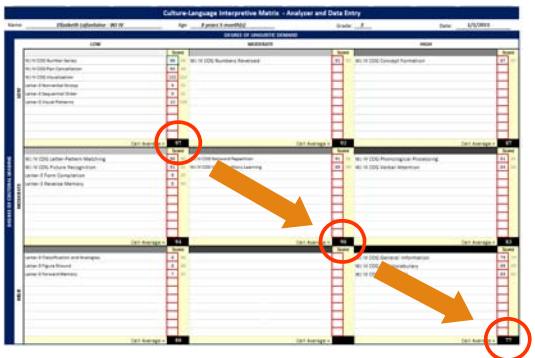
WISC-V & UNIT DATA FOR ELIZABETH (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

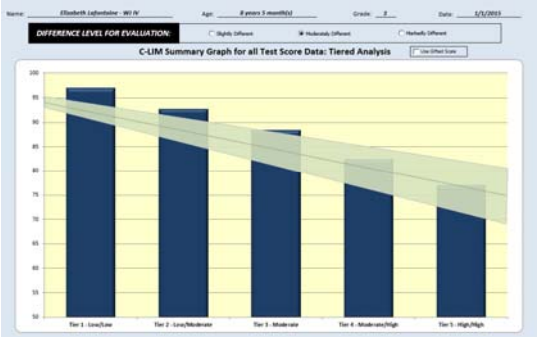
Culture-Language Interpretive Matrix: Case Study

WJ IV COG & LEITER-3 DATA FOR ELIZABETH (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

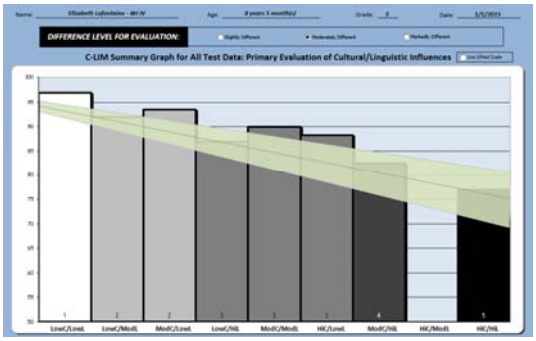
WJ IV & LEITER-3 DATA FOR ELIZABETH (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Case Study

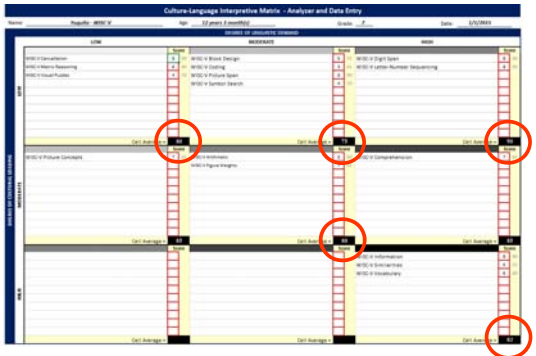
WJ IV & LEITER-3 DATA FOR ELIZABETH (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

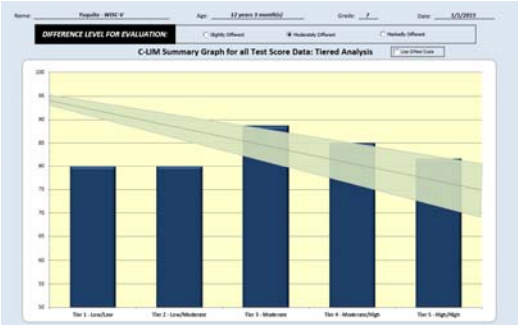
Culture-Language Interpretive Matrix: Case Study

WISC-V ONLY DATA FOR YUQUITA (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

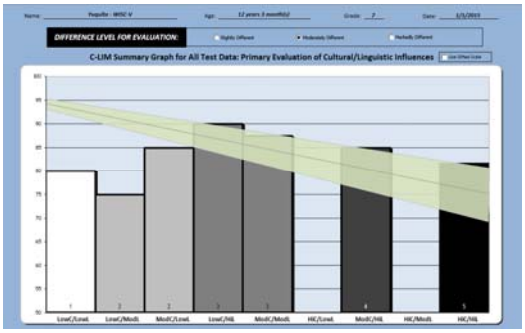
WISC-V ONLY DATA FOR YUQUITA (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Case Study

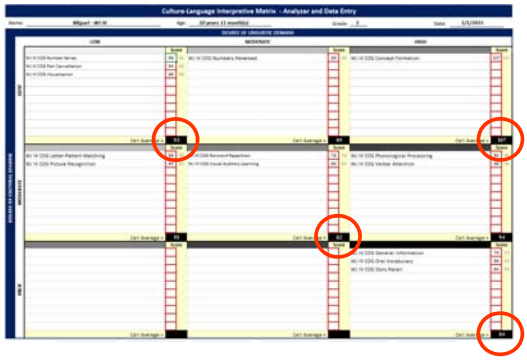
WISC-V ONLY DATA FOR YUQUITA (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

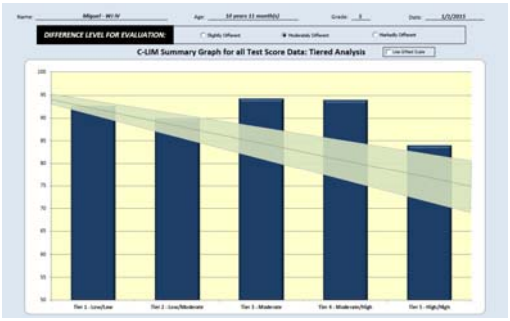
Culture-Language Interpretive Matrix: Case Study

WJ IV COG DATA FOR MIGUEL (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

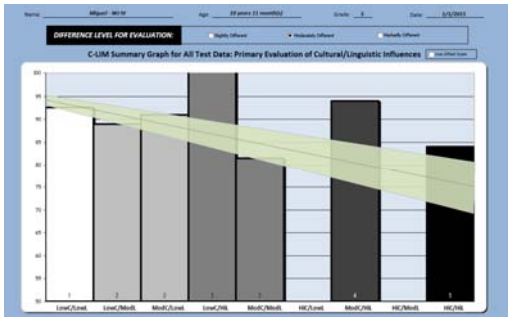
WJ IV COG DATA FOR MIGUEL (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – all test scores are LIKELY to be valid.

Culture-Language Interpretive Matrix: Case Study

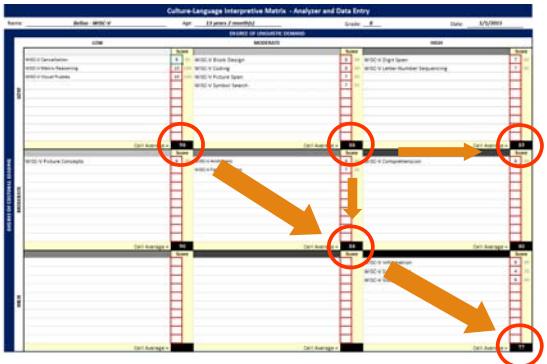
WJ IV COG DATA FOR MIGUEL (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – all test scores are LIKELY to be valid.

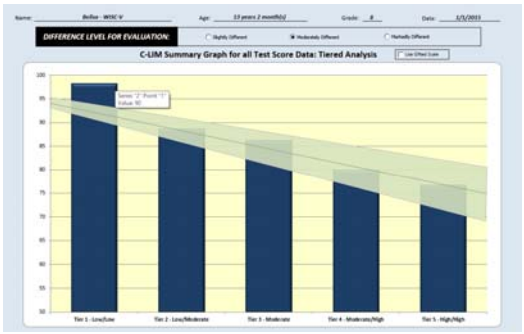
Culture-Language Interpretive Matrix: Case Study

WISC-V DATA FOR BELISA (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

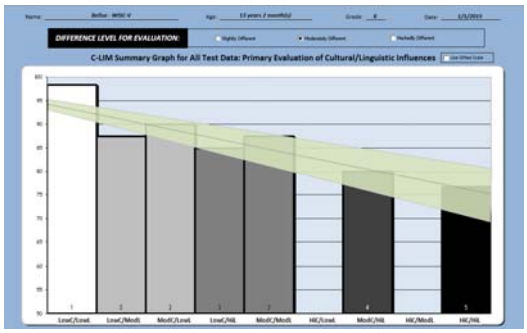
WISC-V DATA FOR BELISA (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

Culture-Language Interpretive Matrix: Case Study

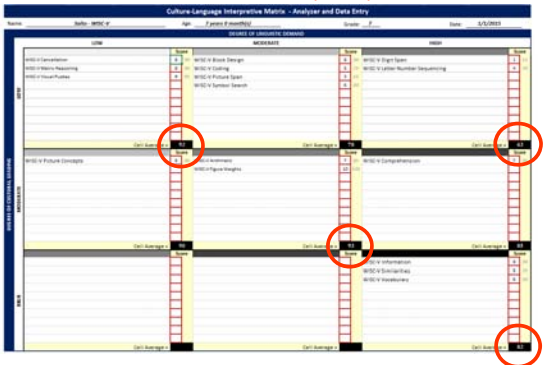
WISC-V DATA FOR BELISA (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: PRIMARY – all test scores are UNLIKELY to be valid.

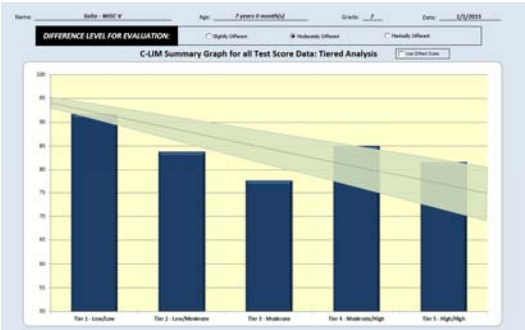
Culture-Language Interpretive Matrix: Case Study

WISC-V DATA FOR SAITO (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

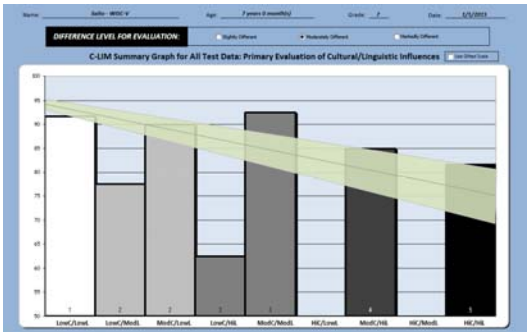
WISC-V DATA FOR SAITO (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Case Study

WISC-V DATA FOR SAITO (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

Nondiscriminatory Interpretation of Test Scores

The Gc caveat for English Language Learners

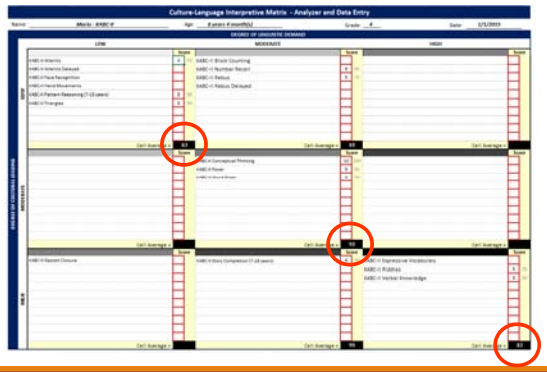
Because Gc is, by definition, comprised of cultural knowledge and language development, the influence of cultural and linguistic differences cannot be separated from tests which are designed to measure culture and language. Thus, *Gc scores for ELLs, even when determined to be valid, remain at risk for inequitable interpretation and evaluation*.

Much like academic tests of manifest skills, Gc scores do reflect the examinee's current level of English language proficiency and acculturative knowledge. However, they do so as compared to native English speakers, not to other ELLs. This is discriminatory and *comparison of Gc performance using a test's actual norms remains unfair when assigning meaning to the value. It is necessary instead to ensure that both the magnitude and the interpretive "meaning" assigned to the obtained value is done in the least biased manner possible to maintain equity*.

For example, a Gc composite score of 76 would be viewed as "deficient" relative to the normative sample where the mean is equal to 100. However, for ELLs, interpretation of a Gc score of 76 should rightly be deemed as being indicative of "average" performance because it falls within the expected range on the C-LIM because it is instead being compared to other ELLs, not native English speakers. Interpreting Gc scores in this manner will 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.

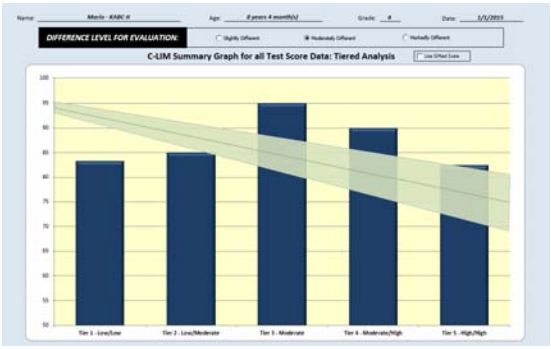
Culture-Language Interpretive Matrix: Case Study

KABC-II DATA FOR MARIO (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

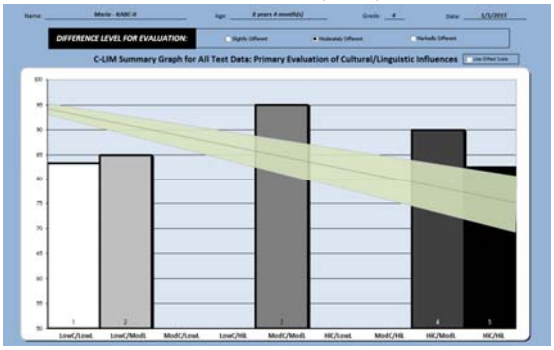
KABC-II DATA FOR MARIO (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Case Study

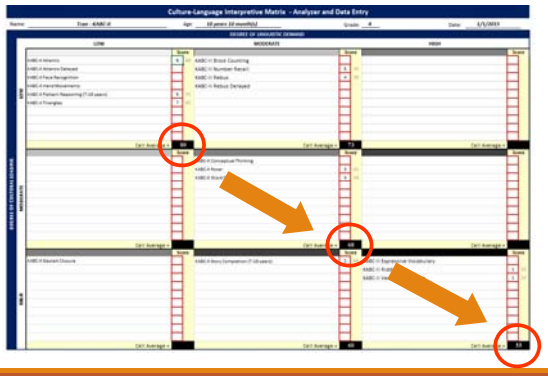
KABC-II DATA FOR MARIO (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: MINIMAL – low test scores are POSSIBLY valid.

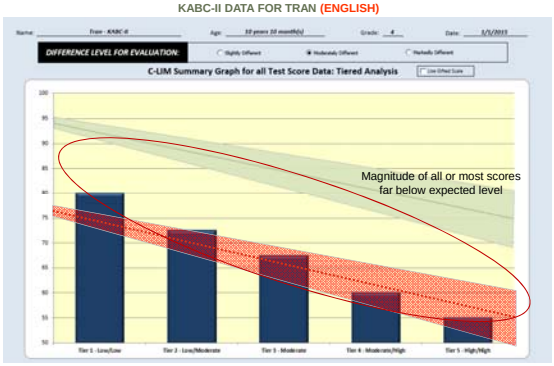
Culture-Language Interpretive Matrix: Case Study

KABC-II DATA FOR TRAN (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

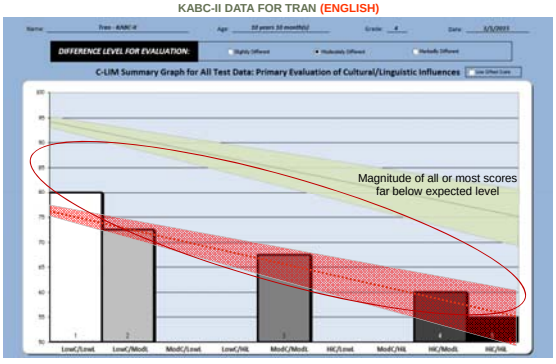
KABC-II DATA FOR TRAN (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Case Study

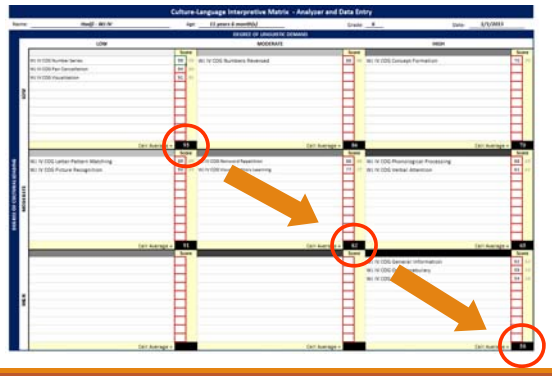
KABC-II DATA FOR TRAN (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

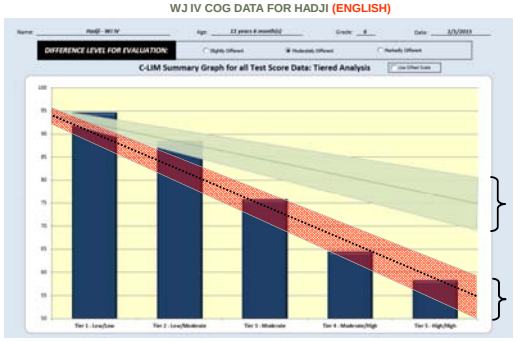
Culture-Language Interpretive Matrix: Case Study

WJ IV COG DATA FOR HADJI (ENGLISH)



Culture-Language Interpretive Matrix: Case Study

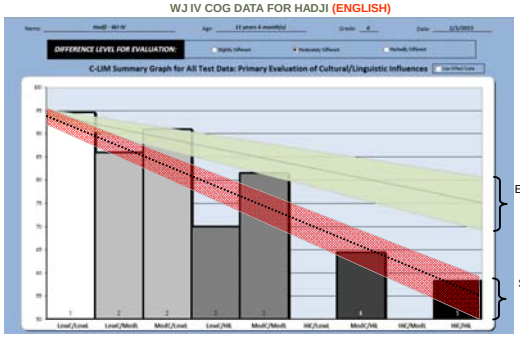
WJ IV COG DATA FOR HADJI (ENGLISH)



CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

Culture-Language Interpretive Matrix: Case Study

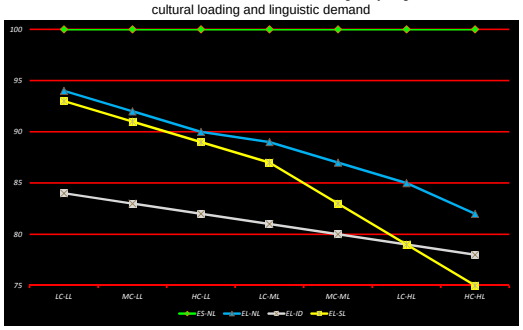
WJ IV COG DATA FOR HADJI (ENGLISH)



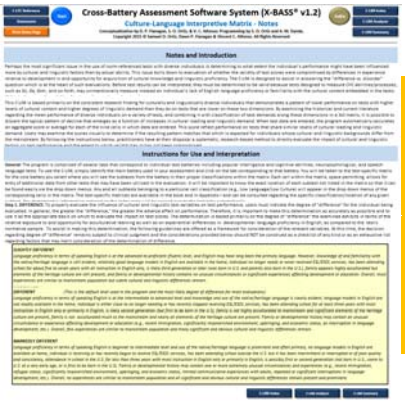
CULTURE/LANGUAGE INFLUENCE: CONTRIBUTORY – low test scores are POSSIBLY valid.

Patterns of Performance Among Monolingual and Bilingual Groups with Learning Disability and Speech Impairment

Mean cell scores on WPPSI-III subtests arranged by degree of cultural loading and linguistic demand



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



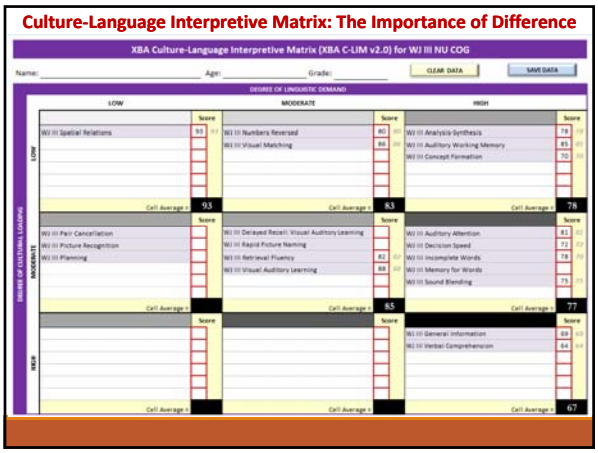
Specific guidelines for determining degree of difference are included on the C-LIM Notes tab and are highlighted in yellow. The guidelines are not meant to be exhaustive or prescriptive but the determination is extremely critical and should be very well considered.

Culture-Language Interpretive Matrix: The Importance of Difference

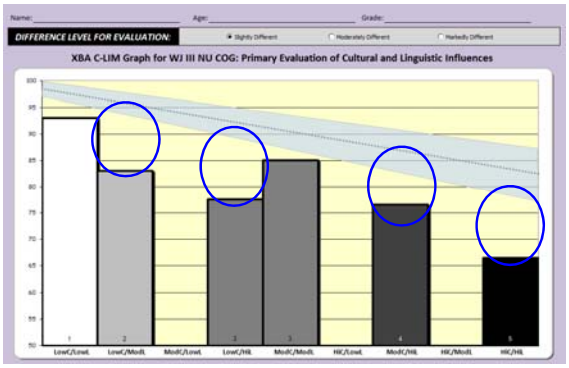
WJ III ONLY DATA

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

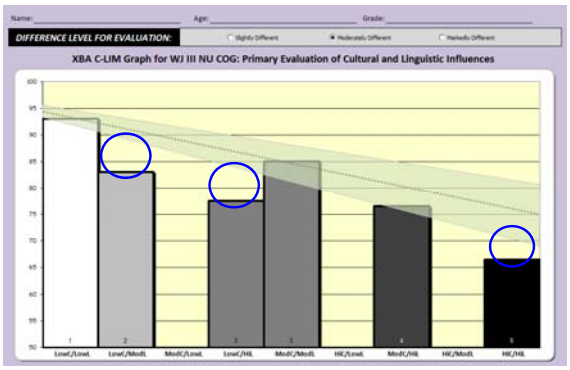
Culture-Language Interpretive Matrix: The Importance of Difference



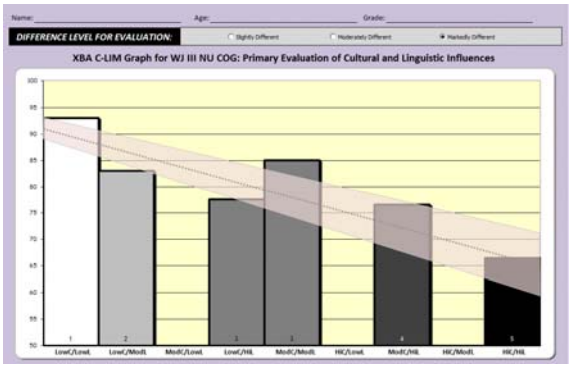
Culture-Language Interpretive Matrix: The Importance of Difference



Culture-Language Interpretive Matrix: The Importance of Difference



Culture-Language Interpretive Matrix: The Importance of Difference



The Culture-Language Interpretive Matrix (C-LIM)

Summary of Important Considerations for Use and Practice

The C-LIM is not a test, scale, measure, or mechanism for making diagnoses. It is a visual representation of current and previous research on the test performance of English learners arranged by mean values to permit examination of the combined influence of acculturative knowledge acquisition and limited English proficiency and its impact on test score validity.

The C-LIM is not a language proficiency measure and will not distinguish native English speakers from English learners with high, native-like English proficiency and is not designed to determine if someone is or is not an English learner. Moreover, the C-LIM is not for use with individuals who are native English speakers.

The C-LIM is not designed or intended for diagnosing any particular disability but rather as a tool to assist clinician's in making decisions regarding whether ability test scores should be viewed as indications of actual disability or a mere reflection of differences in language proficiency and acculturative knowledge acquisition.

The C-LIM's primary purpose is to assist evaluators in ruling out cultural and linguistic influences as exclusionary factors that may have undermined the validity of test scores. Being able to make this determination is the primary and main hurdle in evaluation and the C-LIM can thus guide clinician's in their interpretation of test score data in a nondiscriminatory manner.

Using the XBA Software in SLD Identification: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*	*Fluid Reasoning Index*	*88*	*Visual-Spatial Index*	*95*
Similarities	5	Matrix Reasoning	8	Block Design	9
Vocabulary	6	Figure Weights	8	Visual Puzzles	9
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*
Phonological Processing	99	Story Recall	79
Nonword Repetition	84	Visual-Auditory Learning	75

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

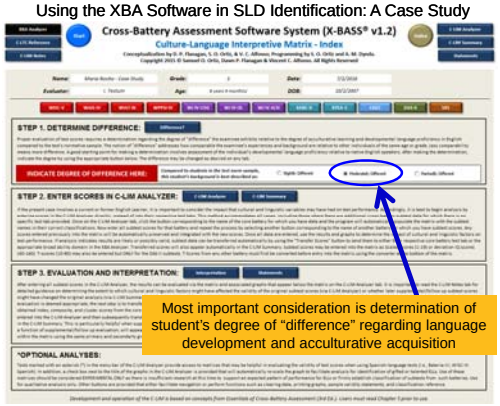
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

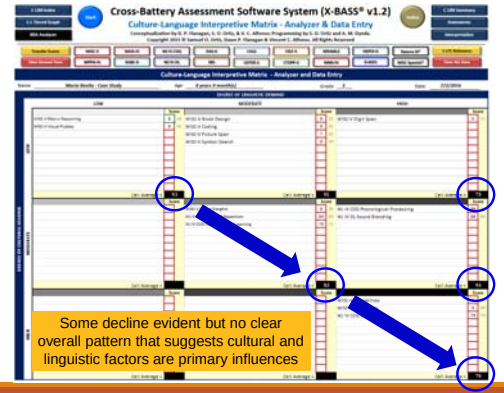
Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study



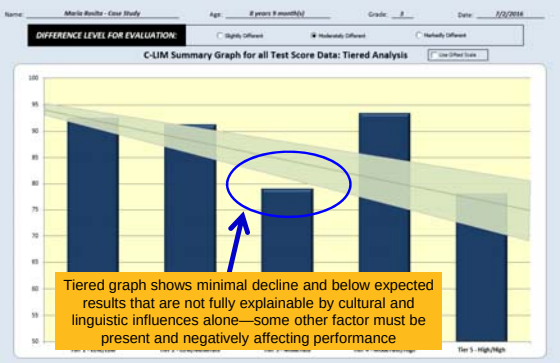
Most important consideration is determination of student's degree of "difference" regarding language development and acculturative acquisition

Using the XBA Software in SLD Identification: A Case Study

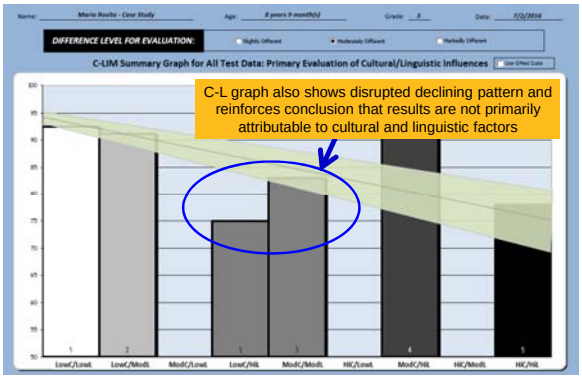


Some decline evident but no clear overall pattern that suggests cultural and linguistic factors are primary influences

Using the XBA Software in SLD Identification: A Case Study

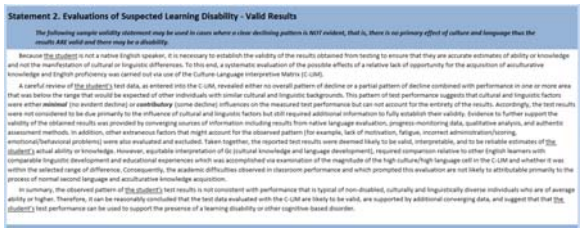


Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study

Sample Validity Statement for ELL Evaluations



The statement above is for use in evaluations of suspected SLD and where it was determined that the obtained test results were NOT likely due primarily to cultural and linguistic factors, either because the influence was only minimal or contributory at best. Note that additional testing and data gathering are necessary to fully support test score validity and that simply excluding culture and language and the primary cause of low scores is not a sufficient basis by itself on which to determine a disability. The wording provided here is intended to create defensible language that explains the process by which these exclusionary factors were evaluated regarding their impact on testing that was conducted.

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

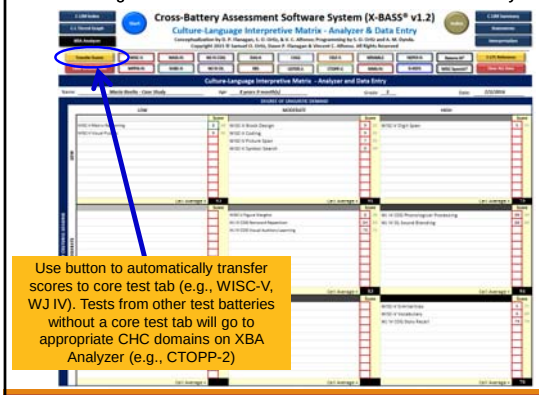
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

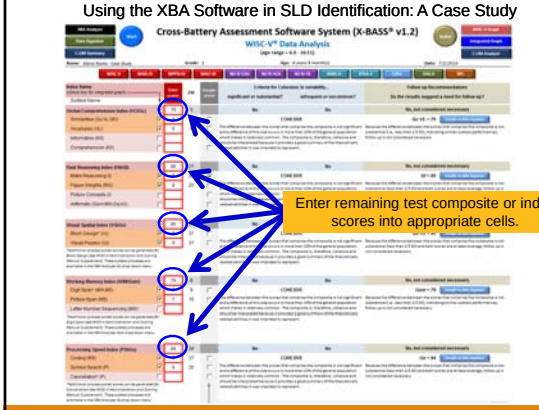
Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study



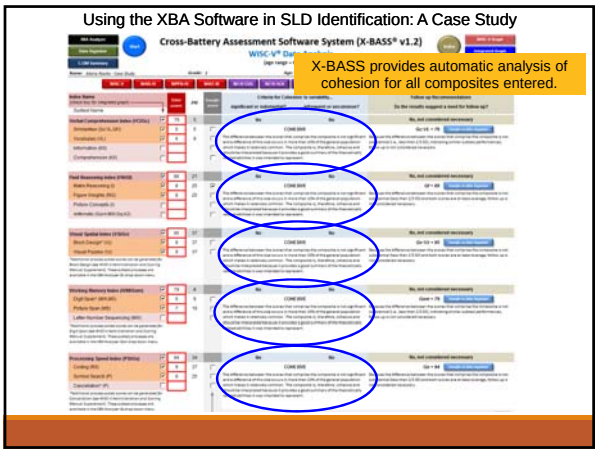
Use button to automatically transfer scores to core test tab (e.g., WISC-V, WJ IV). Tests from other test batteries without a core test tab will go to appropriate CHC domains on XBA Analyzer (e.g., CTOPP-2)

Using the XBA Software in SLD Identification: A Case Study

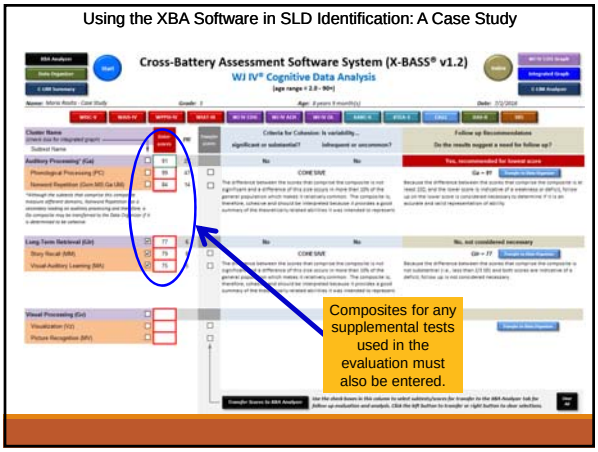


Enter remaining test composite or index scores into appropriate cells.

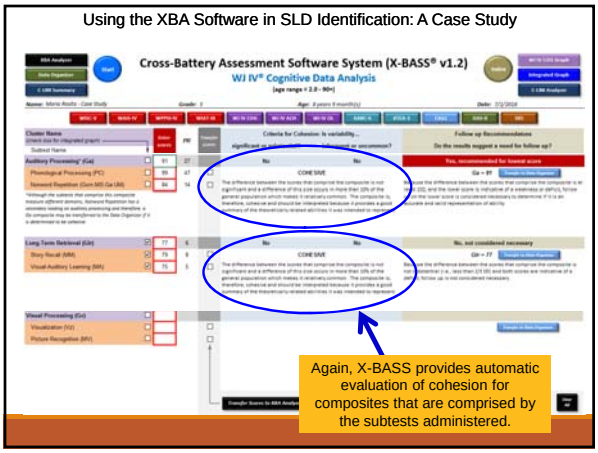
Using the XBA Software in SLD Identification: A Case Study



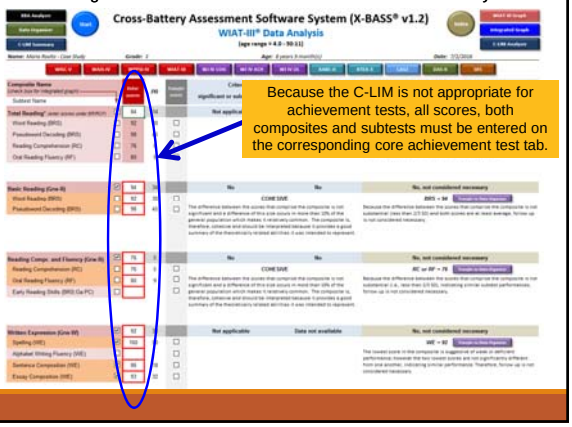
Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

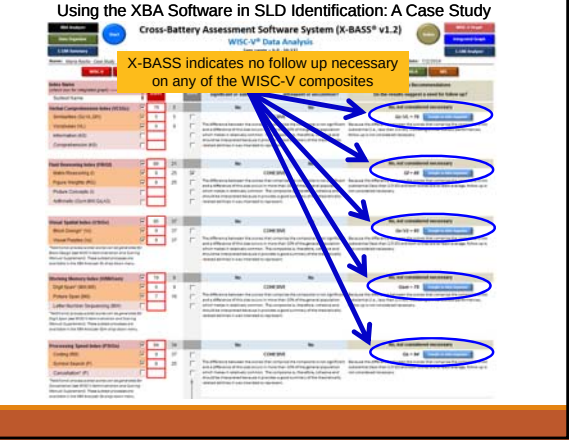
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

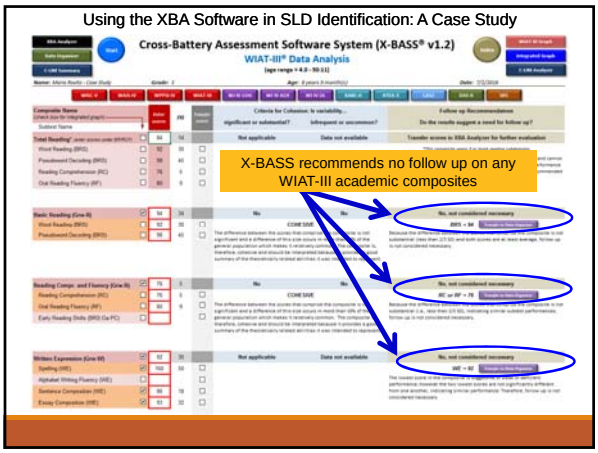
Step 9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

Step 10: Utilize the appropriate validity statement for the evaluation

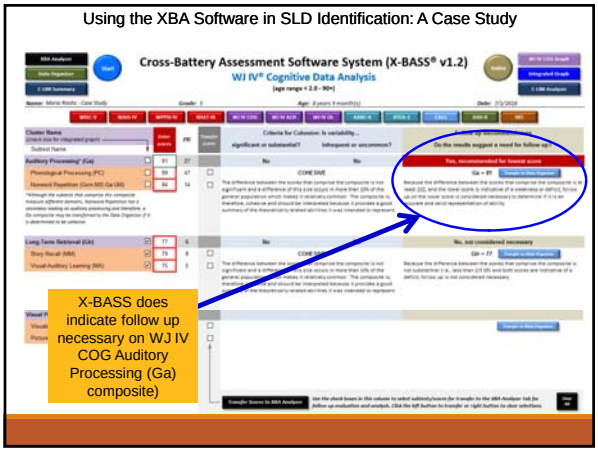
Using the XBA Software in SLD Identification: A Case Study



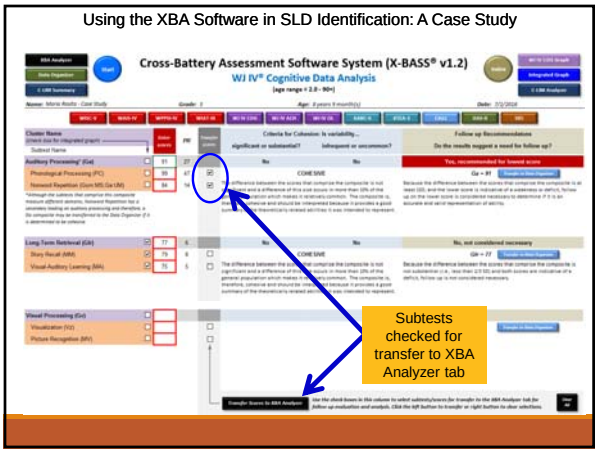
Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study



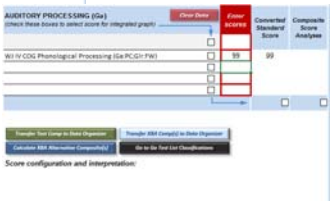
Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: 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 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.



Using the XBA Software in SLD Identification: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*	*Fluid Reasoning Index*	*88*	*Visual-Spatial Index*	*95*
Similarities	5	Matrix Reasoning	8	Block Design	9
Vocabulary	6	Figure Weights	8	Visual Puzzles	9

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*	*Follow Up Testing*	
Phonological Processing	99	Story Recall	79	WJ IV OL Sound Blending	88
Nonword Repetition	84	Visual-Auditory Learning	75		

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

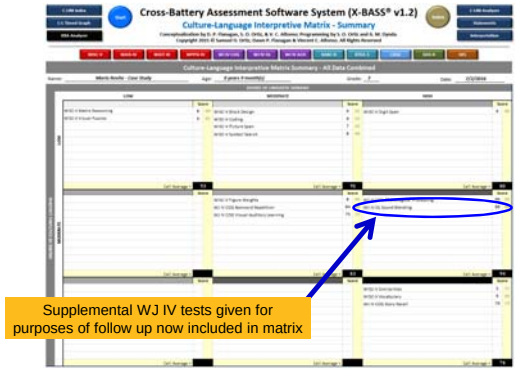
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

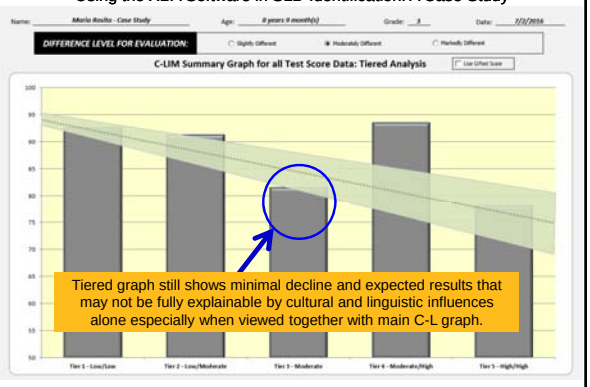
Step 9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

Step 10: Utilize the appropriate validity statement for the evaluation

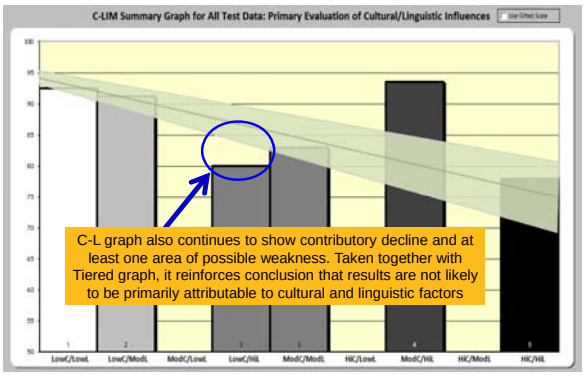
Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

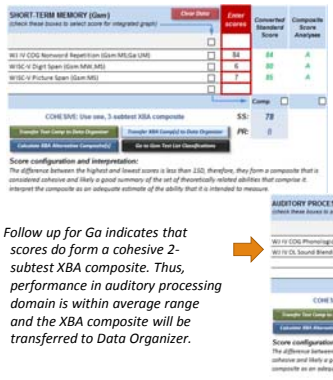
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

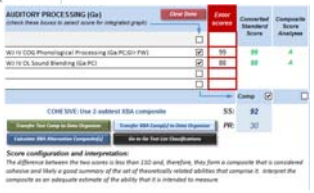
Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study



Combining WISC-V subtests from WMI creates a cohesive 3-subtest XBA composite. 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. Thus, performance in auditory processing domain is within average range and the XBA composite will be transferred to Data Organizer.



Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

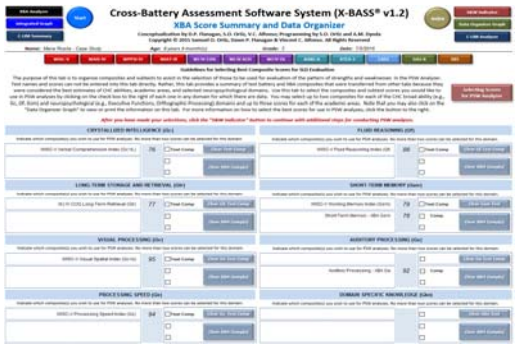
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

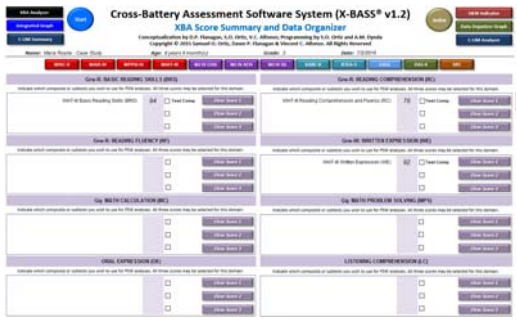
Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study



Data Organizer provides a summary of test and XBA composites for cognitive tests including both test-based composites and any derived XBA composites.

Using the XBA Software in SLD Identification: A Case Study



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

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

Step 10: Utilize the appropriate validity statement for the evaluation

Nondiscriminatory Interpretation of Test Scores

The Gc caveat for English Language Learners

Because Gc is, by definition, comprised of cultural knowledge and language development, the influence of cultural and linguistic differences cannot be separated from tests which are designed to measure culture and language. Thus, *Gc scores for ELLs, even when determined to be valid, remain at risk for inequitable interpretation and evaluation*.

Much like academic tests of manifest skills, Gc scores do reflect the examinee's current level of English language proficiency and acculturative knowledge. However, they do so as compared to native English speakers, not to other ELLs. This is discriminatory and *comparison of Gc performance using a test's actual norms remains unfair when assigning meaning to the value. It is necessary instead to ensure that both the magnitude and the interpretive "meaning" assigned to the obtained value is done in the least biased manner possible to maintain equity*.

For example, a Gc composite score of 76 would be viewed as "deficient" relative to the normative sample where the mean is equal to 100. However, for ELLs, interpretation of a Gc score of 76 should rightly be deemed as being indicative of "average" performance because it falls within the expected range on the C-LIM because it is instead being compared to other ELLs, not native English speakers. Interpreting Gc scores in this manner will 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.

Nondiscriminatory Interpretation of Test Scores: A Case Study

Strengths and Weaknesses: When to re-test Gc

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.

Recommended Guidelines for Using PSW-A with ELLs

Data Entry Guidelines for Using PSW-A with English Learners

Because Maria is an English Learner, it is also necessary to re-administer tests that were possible weaknesses when tested in English. In this case, the following results were obtained:

	English	*Spanish*	*PSW-A Entry*
- Gc (VCI)	**76**	-	
- Gf (FRI)	89	→	**89***
- Glr	**77**	-	
- Gsm (XBA)	**78**	-	
- Gv (VSI)	98	→	**98***
- Ga	92	→	**92***
- Gs (PSI)	94	→	**94***

The obtained Gc score appears to be indicative of a weakness. However, because Maria is an ELL, it is necessary to evaluate Gc in terms of "strength" or "weakness" using an appropriate and nondiscriminatory standard. This can be accomplished by looking at the magnitude of the high/high cell in the C-LIM.

**Note: These scores, obtained from testing in English, were deemed to be in the average range (including the SS=89 for Gf) and therefore did not require further evaluation in the native language. They may be used as obtained for the purposes of PSW analysis.*

Using the XBA Software in SLD Identification: A Case Study

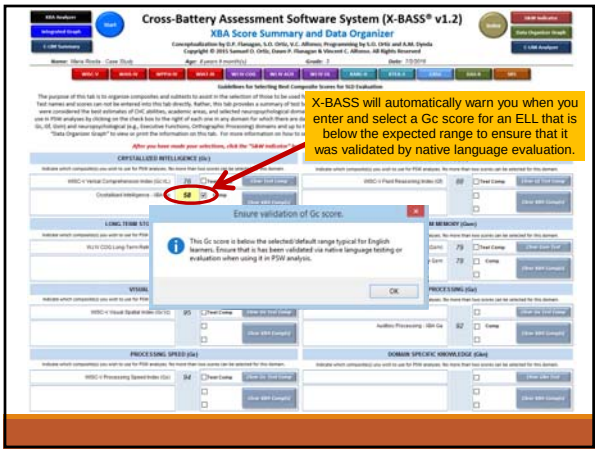
Recommended Guidelines for Using PSW-A with ELLs

Data Entry Guidelines for Using PSW-A with English Learners

Because culture and language cannot be separated from the measurement of culture and language, it is necessary to ensure that Gc for ELLs is interpreted in comparison to other ELLs with similar backgrounds rather than native English speakers. The shaded range of the C-LIM for Tier 5 provides this comparison.

	English	*Spanish*	*PSW-A Entry*
- Gc (VCI)	76	-	76
- Gf (FRI)	89	-	89
- Glr	77	-	
- Gsm (XBA)	78	-	
- Gv (VSI)	98	-	98
- Ga	92	-	92
- Gs (PSI)	94	-	94

Since the aggregate score in the C-LIM for Tier 5 (i.e., the High/High cell where Gc is classified) was within the expected range corresponding to the selected degree of difference deemed most appropriate, it should be considered a strength and indicated as such for the purposes of PSW analysis.



Nondiscriminatory Interpretation of Test Scores: A Case Study

Strengths and Weaknesses: When to re-test 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*

Recommended Guidelines for Using PSW-A with ELLs

Data Entry Guidelines for Using PSW-A with English Learners

Because Maria is an English Learner, it is also necessary to re-administer tests in the native language that were identified as possible areas of weakness when tested in English. In this case, the following domains, Glr and Gsm, should be re-tested:

	English	*Spanish*	*PSW-A Entry*
- Gc (VCI)	**76**	-	**76**
- Gf (FRI)	89	-	**89**
- Glr	**77**		
- Gsm (XBA)	**78**		
- Gv (VSI)	98	-	**98**
- Ga	92	-	**92**
- Gs (PSI)	94	-	**94**

In these cases, the original English scores are deemed to be below the average range and are possible areas of deficit. They must be re-tested in the native language to provide additional support and evidence that they are true deficits and not just spuriously low scores.

Using the XBA Software in SLD Identification: A Case Study

WISC-V/WJ IV/WIAT-III XBA DATA FOR Maria

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-V

Verbal Comprehension Index	*76*	*Fluid Reasoning Index*	*88*	*Visual-Spatial Index*	*95*
Similarities	5	Matrix Reasoning	8	Block Design	9
Vocabulary	6	Figure Weights	8	Visual Puzzles	9
Working Memory Index	*79*	*Processing Speed Index*	*94*		
Digit Span	5	Coding	9		
Picture Span	7	Symbol Search			
WISC IV Spanish WMI	*72*				
Digit Span	5				
Letter-Number Sequencing	4				

WECHSLER INDIVIDUAL ACHIEVEMENT TEST-III

Basic Reading	*94*	*Reading Comprehension*		
Word Reading	92	Reading Comprehension		
Pseudoword Decoding	98	Oral Reading Fluency		
		Essay Composition	93	

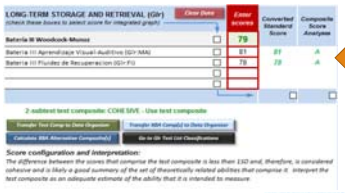
WOODCOCK JOHNSON-IV TESTS OF COGNITIVE ABILITY

Auditory Processing	*91*	*LT Storage/Retrieval*	*77*	*Follow Up Testing*	
Phonological Processing	99	Story Recall	79	WJ IV OL Sound Blending	88
Nonword Repetition	84	Visual-Auditory Learning	75		
		Bateria III LT Retrieval	*79*		
		Visual-Auditory Learning	81		
		Retrieval Fluency	78		

Gsm and Glr needed to be re-tested in the native language to confirm them as weaknesses. The same or similar tests can be used and scores may be generated but the purpose is to observe performance in the domain that validates difficulties even with full comprehension.

Results of native language testing for Gsm (above) and Glr (to the right).

Using the XBA Software in SLD Identification: 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 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 resulted in a similar score and also indicated a possible deficit. Thus, the original score is validated but additional conversing evidence is necessary (e.g., work samples, observations, progress monitoring info, etc.).

Nondiscriminatory Interpretation of Test Scores: A Case Study

Strengths and Weaknesses: Which score to use for PSW analysis

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 English language testing 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*

Recommended Guidelines for Using PSW-A with ELLs

POSSIBLE OUTCOMES WHEN TEST SCORES ARE RE-EVALUATED IN THE NATIVE LANGAUGE

	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 selecting the indicated score
			Original Score (in English)	Follow Up Score (in native lang)	
For ALL domains	SS ≥ 90	n/a	✓		Scores in or above the average range are unlikely to occur by chance and very likely to be valid thus re-evaluation in the native language is unnecessary
For ALL domains	SS < 90	SS ≥ 90		✓	Because a deficit cannot exist in one language only, the original score is invalidated and should be replaced by the follow up average score
For ALL domains (except Gc)	SS < 90	SS < 90	✓		Low scores in both languages suggests true deficits but factors related to lack of native language instruction must also be considered
For Gc domain only	SS < 90 but within the expected range in the C-LIM	n/a	✓		If Gc is within the C-LIM expected range, it should be considered average and native language testing may not be necessary unless there is reason to believe it may be informative

Recommended Guidelines for Using PSW-A with ELLs

Data Entry Guidelines for Using PSW-A with English Learners

To provide cross-linguistic validation and support (or to possibly refute their validity) the domains in which Maria appeared to have possible deficits were re-evaluated using the native language. Re-testing of Maria's abilities in Spanish in the areas of Glr and Gsm produced the following results:

	English	*Spanish*	*PSW-A Entry*
- Gc (VCI)	76	-	76
- Gf (FRI)	89	-	89
- Glr	77	79	77*
- Gsm (XBA)	78	72	78*
- Gv (VSI)	98	-	98
- Ga	92	-	92
- Gs (PSI)	94	-	94

Because the native language scores in these domains also appear to be in the deficit range, they provide confirmation of the likely validity of the original scores obtained in English. Therefore, the original scores remain valid and most appropriate for the purposes of subsequent PSW analysis.

Note: Although the native language scores are slightly higher in one case and slightly lower in the other, both are still indicative of weaknesses and serve to bolster the validity of the obtained scores in the same domains when tested in English. Thus, the original scores from testing in English are even more supported and thus remain the most defensible scores for use in the PSW-A. If, however, any of the native language scores had been found to be average or higher (SS>90), they would then have refuted the original scores obtained in English by because they represent valid indications of at least average ability and would then be the most appropriate score for use in PSW analysis.

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

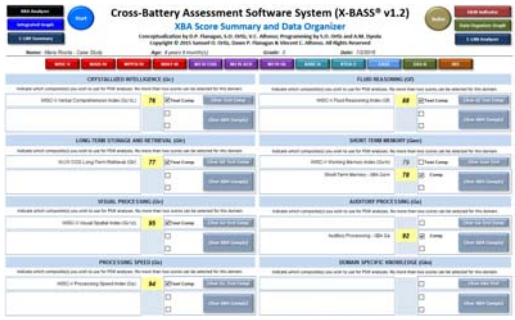
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

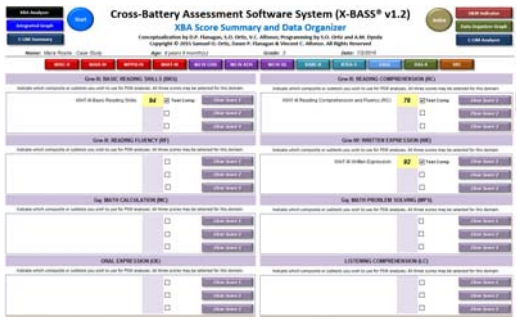
Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study



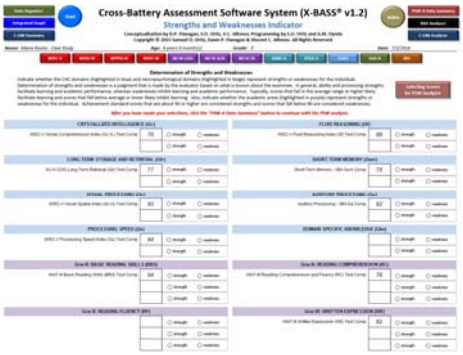
Data Organizer permits selection of specific cognitive composites for use in PSW analysis. Selected scores appear in yellow but a maximum of 2 cognitive scores can be selected (e.g., in cases where there may be both a strength and a weakness or two weaknesses, etc.)

Using the XBA Software in SLD Identification: A Case Study



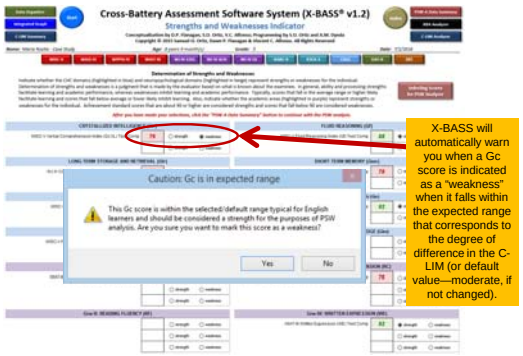
Data Organizer permits selection of specific academic composites or subtests for use in PSW analysis. Selected scores appear in yellow and a maximum of 3 academic scores can be selected including any combination of test composites, XBA composites, or subtest scores.

Using the XBA Software in SLD Identification: A Case Study



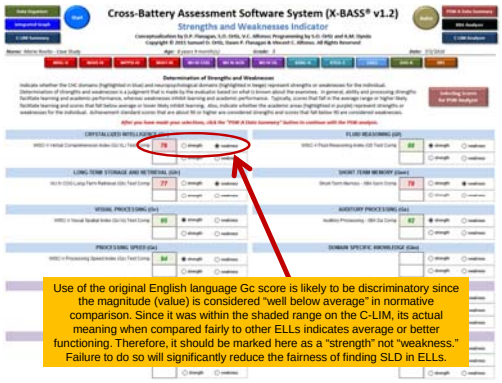
Composites (and any academic subtests) selected on the Data Organizer appear on the Strength and Weaknesses Indicator where they may be designated as "S" or "W" for PSW analysis.

Using the XBA Software in SLD Identification: A Case Study



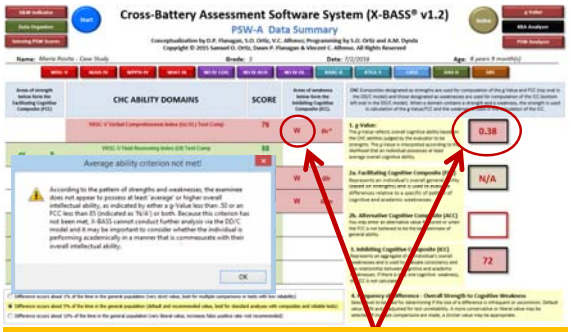
Scores designated as "S" appear in green, those designated as "W" appear in red. When Gc is selected as an area of cognitive weakness, an important cautionary message will appear indicating that Gc should not be used as the sole or only area of cognitive weakness.

Using the XBA Software in SLD Identification: A Case Study



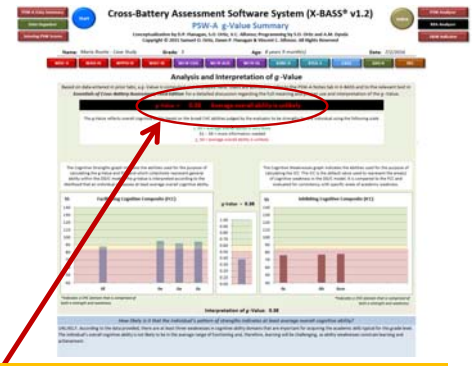
Use of the original English language Gc score is likely to be discriminatory since the magnitude (value) is considered "well below average" in normative comparison. Since it was within the shaded range on the C-LIM, its actual meaning when compared fairly to other ELLs indicates average or better functioning. 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.

Using the XBA Software in SLD Identification: A Case Study



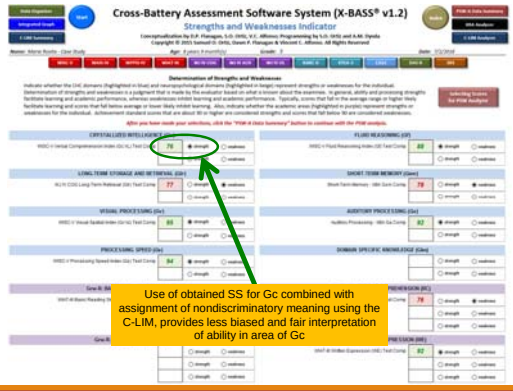
Because Gc is the most important ability related to academic success and accounts for the majority of variance in overall general ability, failure to properly evaluate it against other ELLs with comparable backgrounds may result in highly attenuated g-Values that suggest low ability and mask possible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

Using the XBA Software in SLD Identification: A Case Study



Resulting g-Value suggests that Maria does not have sufficient overall general ability to meet the definition of SLD which requires at least average level of intelligence and halts further analysis of SLD and is discriminatory.

Using the XBA Software in SLD Identification: A Case Study



Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

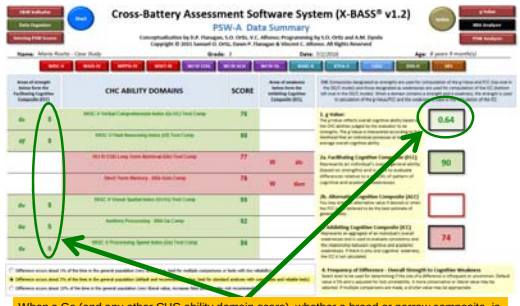
Step 7: Re-evaluate deficits using native language and compare to original scores

Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

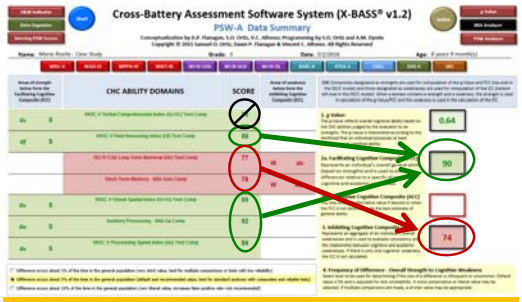
Step 9: Evaluate scores and results from PSW-A Data Summary and PSW Analyzer

Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: A Case Study

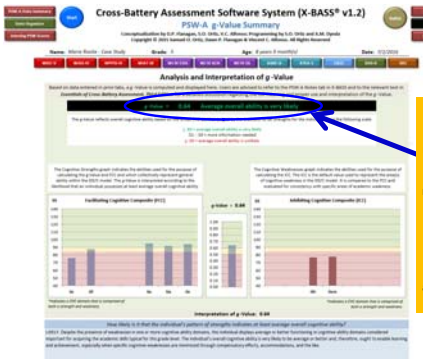


Using the XBA Software in SLD Identification: A Case Study



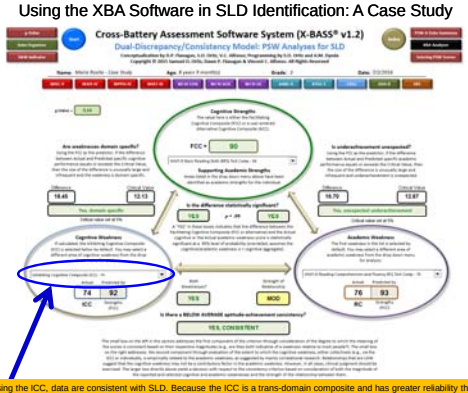
When a Gc (and any other CHC ability domain score), whether a broad or narrow composite, is marked as a "strength," (typically SS ≥ 90), X-BASS will always include its value in calculation of the FCC. Likewise, any scores marked as "weakness" are always factored into calculation of the ICC. However, to prevent discriminatory attenuation in the case of ELLs, the Gc score alone is not included in calculations for the FCC when Gc has been designated as a "strength" but is a SS < 90.

Using the XBA Software in SLD Identification: A Case Study



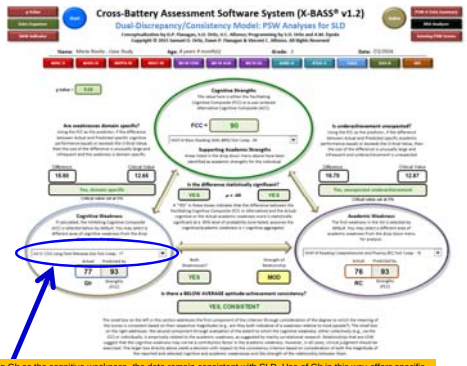
Unlike when Gc was indicated as a "weakness," the g-Value now correctly reflects a true and equitable estimate of Maria's overall cognitive ability and does not unfairly represent her as lacking general intelligence. The g-Value is not affected by the magnitude of the standard score since it is based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

Using the XBA Software in SLD Identification: A Case Study



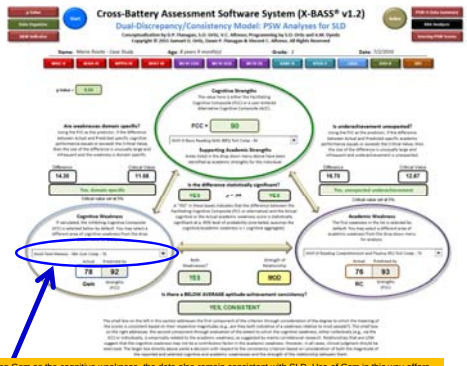
Using the ICC, data are consistent with SLD. Because the ICC is a trans-domain composite and has greater reliability than a domain specific composite and is more likely to reveal a significant difference. The ICC, however, does not provide specific information regarding the nature of the cognitive deficit 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.

Using the XBA Software in SLD Identification: A Case Study



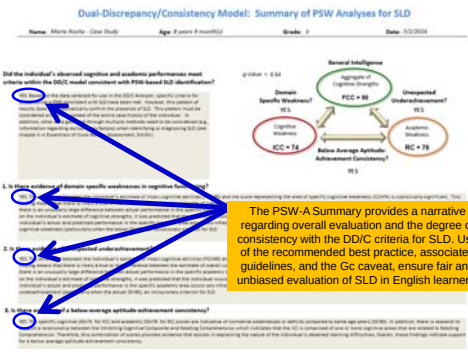
Using Glr as the cognitive weakness, the data remain consistent with SLD. Use of Glr in this way offers specific information regarding the nature of the cognitive deficit and can provide valuable information regarding learning needs and problems as well as suggest appropriate avenues for improving intervention and instruction.

Using the XBA Software in SLD Identification: A Case Study



Using Gsm as the cognitive weakness, the data also remain consistent with SLD. Use of Gsm in this way offers specific information regarding the nature of the cognitive deficit and can provide valuable information regarding learning needs and problems as well as suggest appropriate avenues for improving intervention and instruction.

Using the XBA Software in SLD Identification: A Case Study



The PSW-A Summary provides a narrative regarding overall evaluation and the degree of consistency with the DD/C criteria for SLD. Use of the recommended best practice, associated guidelines, and the Gc caveat, ensure fair and unbiased evaluation of SLD in English learners.

Using the XBA Software in SLD Identification: A Case Study

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

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

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

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

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

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

Step 7: Re-evaluate deficits using native language and compare to original scores

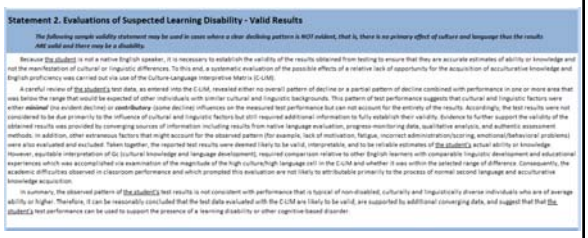
Step 8: Select and designate appropriate scores for PSW Analysis as strength or weakness

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

Step 10: Utilize the appropriate validity statement for the evaluation

Using the XBA Software in SLD Identification: 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 that 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).

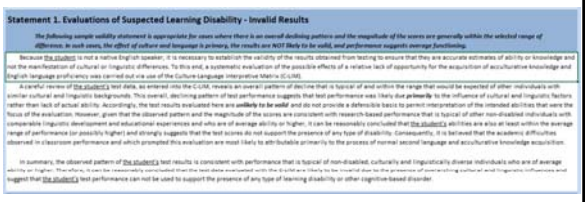
A careful review of the student's test data, as entered into the C-LIM, revealed either no overall pattern of decline or a partial pattern of decline combined with performance in one or more area that was below the range that would be expected of other individuals with similar cultural and linguistic 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 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 now considered to be fully supported and is appropriate for this case where the evaluation focused on suspected SLD and where it was determined that the obtained test results were NOT due primarily to cultural and linguistic factors, albeit they remained contributory. Additional native language testing was conducted in this case to further support test score validity and to systematically exclude culture and language and the primary cause of low scores and the observed academic difficulties. These statements 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.

Using the XBA Software in SLD Identification: A Case Study

Sample Validity Statement for ELL Evaluations



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.

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 language 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, reveals an overall pattern of decline that is typical of and within the range that would be expected of other individuals with similar cultural and linguistic backgrounds. This overall, declining pattern of test performance suggests that test performance was likely due **primarily** to the influence of cultural and linguistic factors rather than lack of actual ability. Accordingly, the test results evaluated here are **unlikely to be valid** and do not provide a defensible basis to permit interpretation of the intended abilities that were the focus of the evaluation. However, given that the observed pattern and the magnitude of the scores are consistent with research-based performance that is typical of other non-disabled individuals with comparable linguistic development and educational experiences and who are of average ability or higher, it can be reasonably concluded that the student's abilities are also at least within the average range of performance (or possibly higher) and strongly suggests that the test scores do not support the presence of any type of disability. Consequently, it is believed that the academic difficulties observed in classroom performance and which prompted this evaluation are most 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 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 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 or other cognitive-based disorder.

The C-LIM "Statements" tab provides four sample validity statement narratives, including two for evaluation of SLD (one where results are likely invalid, and one where results are possibly valid), and two others that are relevant to the evaluation of intellectual disability and speech-language impairment. Note that these statements pertain only to the use of the C-LIM and the manner in which the obtained scores were deemed to be valid or invalid relative to cultural and linguistic factors.

The statement above is for use in evaluations of suspected SLD and where it has been initially determined that the obtained test results were likely due primarily to cultural and linguistic factors. The wording provided is intended to create defensible language that explains the process by which these exclusionary factors were evaluated regarding their impact on testing that was conducted.

The Culture-Language Interpretive Matrix (C-LIM)

Summary of Important Facts for Use and Practice

The C-LIM is not a test, scale, measure, or mechanism for making diagnoses. It is a visual representation of current and previous research on the test performance of English learners arranged by mean values to permit examination of the combined influence of acculturative knowledge acquisition and limited English proficiency and its impact on test score validity.

The C-LIM is not a language proficiency measure and will not distinguish native English speakers from English learners with high, native-like English proficiency and is not designed to determine if someone is or is not an English learner. Moreover, the C-LIM is not for use with individuals who are native English speakers.

The C-LIM is not designed or intended for diagnosing any particular disability but rather as a tool to assist clinician's in making decisions regarding whether ability test scores should be viewed as indications of actual disability or a mere reflection of differences in language proficiency and acculturative knowledge acquisition.

The primary purpose of the C-LIM is to assist evaluators in ruling out cultural and linguistic influences as exclusionary factors that may have undermined the validity of test scores. Being able to make this determination is the primary and main hurdle in evaluation and the C-LIM can thus guide clinician's in their interpretation of test score data in a nondiscriminatory manner.

The Culture-Language Test Classifications and Interpretive Matrix: Caveats and Conclusions

Used in conjunction with other information relevant to appropriate bilingual, cross-cultural, nondiscriminatory assessment including...

- level of acculturation
- language proficiency
- socio-economic status
- academic history
- familial history
- developmental data
- work samples
- curriculum based data
- intervention results, etc.

...the C-LTC and C-LIM can be of practical value in helping establish credible and defensible validity for test data, thereby decreasing the potential for biased and discriminatory interpretation. Taken together with other assessment data, the C-LTC and C-LIM assist practitioners in answering the most basic question in ELL assessment:

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

Assessment of English Language Learners - Resources

BOOKS:

Rhodes, R., Ochoa, S. H. & Ortiz, S. O. (2005). Comprehensive Assessment of Culturally and Linguistically Diverse Students: A practical approach. New York: Guilford.

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

Flanagan, D.P. & Ortiz, S.O. (2012). Essentials of Specific Learning Disability Identification. New York: Wiley & Sons, Inc.

Ortiz, S. O., Flanagan, D. P. & Alfonso, V. C. (2015). Cross-Battery Assessment Software System (X-BASS v1.0). New York: Wiley & Sons, Inc.

ONLINE:

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

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

Cross-Battery XBA Assessment