Through the Looking Glass:
Evaluation and testing of bilinguals in a monolingual world.

**"'and what is the use of a book,' thought Alice,
'without pictures or conversations?'"**

Keynote Presentation
for
Washington State Psychological Association
October 16, 2020

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

The Bilingual-Bicultural Experience

Through the looking glass of two languages and two cultures

Children raised in two cultures naturally split their time and experience across each of them. Unfortunately, parents are usually able to mediate aspects of only their native culture to their children and cannot do so with the new culture because of their own lack of familiarity with it.

This means that children from bilingual-bicultural backgrounds must often navigate both the new language and culture almost on their own, a term I refer to as “cultural pioneering.”

The process of learning a new language and culture thus becomes very dependent on the experiences a child has, particularly while in school. If something is not explicitly taught in school, the chances that it may be taught and learned informally outside the school decrease. This often results in hit-or-miss learning that although occurs for children in general, becomes a much more frequent occurrence for children from diverse cultural and linguistic backgrounds. Areas that are highly susceptible to this influence include cultural knowledge, especially the subtle, idiosyncratic, and less frequent aspects of it as well as language, particularly correct grammar, pronunciation, usage, and pragmatics including idioms and humor.

Cultural and Linguistic Experiences Mediate Development:
Formal and Informal Learning Experiences

Old Bay and Michigan

Cultural and Linguistic Experiences Mediate Development:
Opportunity for Learning

Assessment of a student's academic skills and abilities must directly examine the student's skills and abilities with respect to the actual materials and content used for instruction. Thus, authentic assessment seeks to uncover whether learning difficulties can be ascribed to experiential differences rather than ability differences. Not only does this ensure greater validity of the assessment, it provides valuable information necessary to develop specific and effective instructional strategies. In general, evidence of lack of opportunity for learning, ineffective prior instruction, and linguistically inappropriate curricula, are all factors that increase the likelihood that no disability exists.

For example –

According to the manual (1993) for the Nelson-Denny Reading Test, the 80 vocabulary words and their definitions were drawn from :

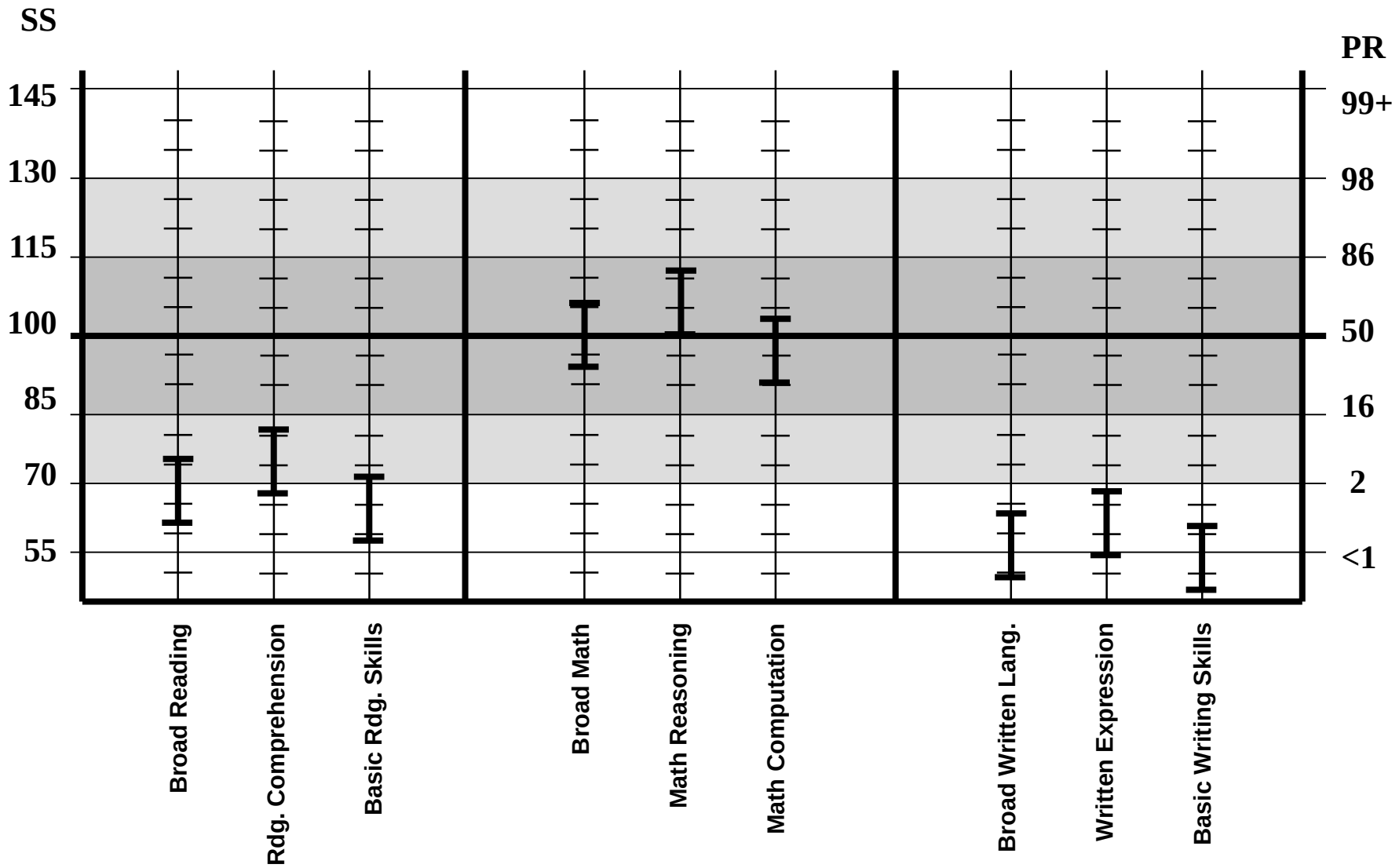
*"current, widely used high school and college texts, including words that **must** be known by students in order to cope successfully with school assignments."* *(emphasis added)*

Cultural and Linguistic Experiences Mediate Development:
Vocabulary Exposure and Development

- It is fruitless to attempt to indoctrinate a superannuated canine with innovative maneuvers.
- Scintillate, scintillate, asteroid minified.
- Members of an avian species of identical plumage congregate.
- Pulchritude possesses solely cutaneous profundity.
- It is fruitless to become lachrymose over precipitately departed lacteal fluid.
- Eschew the implement of correction and vitiate the scion.
- All articles that coruscate with resplendence are not truly auriferous.
- Where there are visible vapors having their prevalence in ignited carbonaceous materials, there is conflagration.
- A plethora of individuals with expertise in culinary techniques vitiate the possible concoction produced by steeping certain comestibles.
- Individuals who make their abodes in vitreous edifices should be advised to refrain from catapulting petrous projectiles.

Homes where English is not the primary or native language results in linguistic experiences that shape the perceptions and views of the speakers particularly in reference to vocabulary but also what might constitute acceptable ways of communicating that can include comfort with basic grammatical errors, use of code-switching, frequent use of slang or colloquial terminology, uncommon or unusual pragmatics, and variances in general language usage.

Cultural and Linguistic Experiences Mediate Development:
Academic Skills Acquisition and the "Bilingual Bermuda Triangle"



Cultural and Linguistic Experiences Mediate Development: Classroom Behavior and Performance

Characteristics and behaviors often associated with various learning problems	Common manifestations of English Language Learners (ELLs) during classroom instruction that may mimic various disorders or cognitive deficits.
Slow to begin tasks	ELLs may have limited comprehension of the classroom language so that they are not always clear on how to properly begin tasks or what must be done in order to start them or complete them correctly.
Slow to finish tasks	ELLs, especially those with very limited English skills, often need to translate material from English into their native language in order to be able to work with it and then must translate it back to English in order to demonstrate it. This process extends the time for completion of time-limited tasks that may be expected in the classroom.
Forgetful	ELLs cannot always fully encode information as efficiently into memory as monolinguals because of their limited comprehension of the language and will often appear to be forgetful when in fact the issue relates more to their lack of proficiency with English.
Inattentive	ELLs may not fully understand what is being said to them in the classroom and consequently they don't know when to pay attention or what exactly they should be paying attention to.
Hyperactive	ELLs may appear to be hyperactive because they are unaware of situation-specific behavioral norms, classroom rules, and other rules of social behavior.
Impulsive	ELLs may lack the ability to fully comprehend instructions so that they display a tendency to act impulsively in their work rather than following classroom instructions systematically.
Distractible	ELLs may not fully comprehend the language being being spoken in the classroom and therefore will move their attention to whatever they can comprehend appearing to be distractible in the process.
Disruptive	ELLs may exhibit disruptive behavior, particularly excessive talking—often with other ELLS, due to a need to try and figure out what is expected of them or to frustration about not knowing what to do or how to do it.
Disorganized	ELLs often display strategies and work habits that appear disorganized because they don't comprehend instructions on how to organize or arrange materials and may never have been taught efficient learning and problem solving strategies.

Cultural and Linguistic Experiences Mediate Development:
Listening Comprehension and Receptive Language

"I pledge a lesson to the frog of the United States
of America, and to the wee puppet for witches hands.
One Asian, under God, in the vestibule,
with little tea and just rice for all."

Source: In the Year of the Boar and Jackie Robinson by Bette Bao Lord, © 1986, Harper Trophy.

Children who are learning a second language hear and interpret sounds in a manner that conforms to words that already exist in their vocabulary. This is a natural part of the first and second language acquisition processes and should not be considered abnormal in any way. It represents the brain's attempt to make sense and meaning of what it perceives by connecting it to what it already knows.

Songs are a good example of this linguistic phenomenon even for native English speakers. Consider these classic misheard lyrics:

"*There's a bathroom on the right*"
"*Excuse me while I kiss this guy*"
"*Doughnuts make my brown eyes blue*"
"*Midnight after you're wasted*"

Cultural and Linguistic Experiences Mediate Development:
Oral and Expressive Language

'Twas the night before Christmas, y por todo la casa,
Not a creature was stirring—Caramba! Que Pasa?
Los niños were tucked away in their camas,
Some in camisas, some in pijamas.
While hanging the medias with mucho cuidado,
In hopes that old Santa would feel obligado.
To bring all children, both buenos y malos,
A nice batch of dulces y otros regalos.

A Visit From St. Nicolas – Anonymous, 1823

Bilinguals/bicultural individuals are perfectly happy with two languages existing side by side. It provides an ability to use code switching and dual-mode communication not available to monolinguals. For bilinguals, it doesn't matter what language is used in conversation because it all makes sense—and mutual comprehension is the goal of all language and communication.

Cultural and Linguistic Experiences Mediate Development:
Reading Comprehension

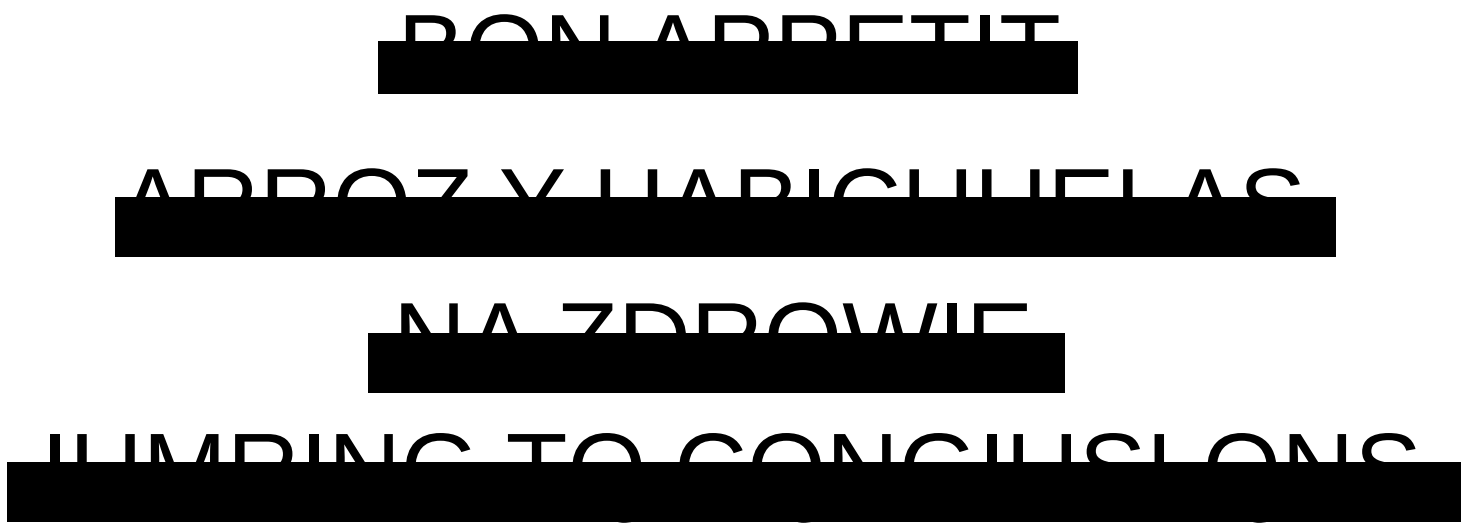
'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe.
All mimsy were the borogroves,
And the mome raths outgrabe.

Jabberwocky by Lewis Carroll

Questions: 1) What things were slithy? 2) What did the toves do in the wabe? 3) How were the borogroves? 4) What kind of raths were there?

Meaning in print is not derived solely from word knowledge. Mature and advanced readers eventually discard “decoding” as the primary means for developing reading abilities in favor of orthographic processing of letters, words, sentences, and grammatical structure. Meaning is often inferred from our cultural knowledge and experience with the language. More experience equals clearer meaning and better comprehension.

Cultural and Linguistic Experiences Mediate Development:
Orthographic Processing



As before, comprehension in print is not derived solely from actual word or letter identification or recognition. English is extremely irregular in morphology and mature and advanced readers eventually discard “decoding” as the primary means for developing reading abilities in favor of orthographic processing of letters, words, sentences where even small surface features are sufficient to derive meaning. Similar to grammatical structure, the ability to understand printed text in the absence of such structure, is accomplished via knowledge of the morphological rules and experience with vocabulary that comes from formal and informal sources. Comparatively speaking, ELLs have less experience and thus less ability to generate meaning automatically, fluently, or transparently.

Cultural and Linguistic Experiences Mediate Development:
Orthographic Processing

Finished files are the result of years of scientific study combined with the the experience of years...

How many times does the letter “f” appear in the sentence above?

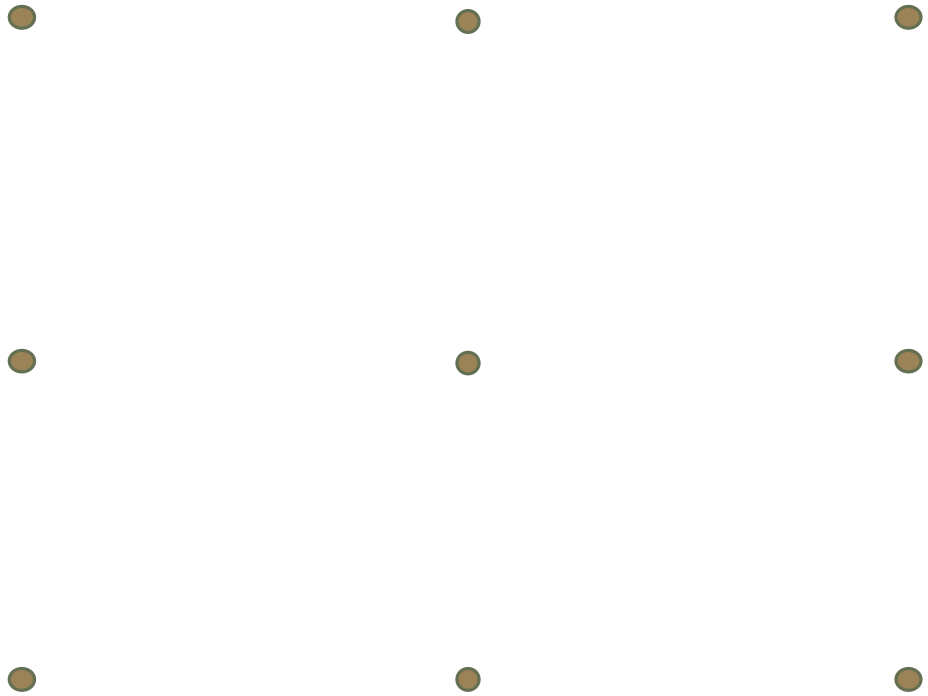
Cultural and Linguistic Experiences Mediate Development:
Verbal and Mathematical Reasoning

What day follows the day before yesterday if two days from now will be Sunday?

Paul makes $25.00 a week less than the sum of what Fred and Carl together make. Carl's weekly income would be triple Steven's if he made $50.00 more a week. Paul makes $285.00 a week and Steven makes $75.00 a week. How much does Fred make?

The ability to engage successfully in verbal reasoning tasks and mathematical word problems presumes the existence of a developmentally proficient level of fluency with the language since it is not the language that is being tested, but the ability to reason. When the native language development is interrupted, bilingual/bicultural individuals may not have the necessary command of the language and the task is confounded by simple comprehension issues and degrades into a test of language, not reasoning.

Cultural and Linguistic Experiences Mediate Development:
Cultural Perspective and Reasoning Ability



Rules: Connect all 9 dots above using only 4 straight lines. You may cross lines, but you cannot lift your pencil.

Cultural and Linguistic Experiences Mediate Development:
General Knowledge and Cultural Artifacts

Cultural and Linguistic Experiences Mediate Learning:
Acculturative Knowledge Acquisition

What I thought	The reality
Tabasco – Mexican hot sauce	Made by McIlhenny Co., USA
Kahlua – Hawaiian liquor	Coffee liqueur made in Mexico
Enfamil – Puerto Rican baby formula	Made by Meade-Johnson, USA
Amoco – Bilingual reference to mucous	Brand of British Petroleum gas
Chiclet – Mexican chewing gum	Made by Cadbury/Adams, USA
Toto – Strange name for a dog	Dorothy's dog's real name

Acculturation to the mainstream plays a significant role in linguistic development and learning in and out of the classroom. The presence and interaction of dual cultural contexts with which to embed certain culturally-specific words or ideas in English may lead to a failure to comprehend or acquire the true meaning of the word or the concept. Idioms are another example of this problem, for example: "I think it's cool the way you don't get on my case about everything."

Linguistic-Cultural Difference vs. Intrinsic Disorder-Disability:
How the bilingual/bicultural experience affects test score validity.

- *Bilinguals are not two monolinguals in one head*
- *Attainment of developmental proficiency in language and acculturation is multifaceted and complex and requires linguistically appropriate education*
- *Both language acquisition and acculturative knowledge acquisition are and must be understood as developmental processes*
- *Use of normative standards developed on monolingual, native-English speakers are not representative of bilinguals and their varying language experiences*
- *Once a bilingual, always a bilingual—individuals do not suddenly cease to be bilingual/bicultural simply because they have become English dominant or English proficient*
- *Bilingual/bicultural experiences differ significantly from monolingual/monocultural ones and have important implications for the development of academic skills and cognitive abilities across the lifespan*
- *Influences on early language development can have profound and lifelong effects that are manifested in nearly all types of test results due to experiences that are different in terms of time and opportunity for learning.*

Understanding First and Second Language Acquisition

Basic Interpersonal Communication Skills (BICS)

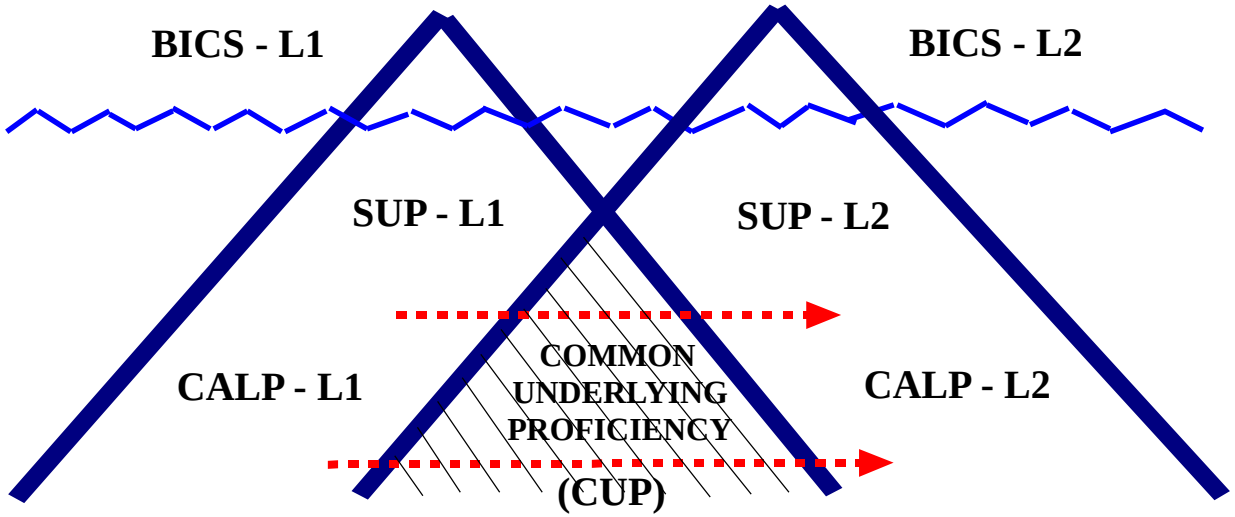
- *ability to communicate basic needs and wants, and ability to carry on basic interpersonal conversations*
- *takes 1 - 3 years to develop and is insufficient to facilitate academic success*

Cognitive Academic Language Proficiency (CALP)

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

Cummins' Developmental Interdependence Hypothesis ("Iceberg Model")

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



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

Understanding First and Second Language Acquisition

Emergence of CALP – Requires 5 to 7 years of formal education.

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

Basic Interpersonal Communicative Skills (BICS)

- We learn more common and frequent words first—particularly those related to general communicative proficiency in as little as 1-3 years.

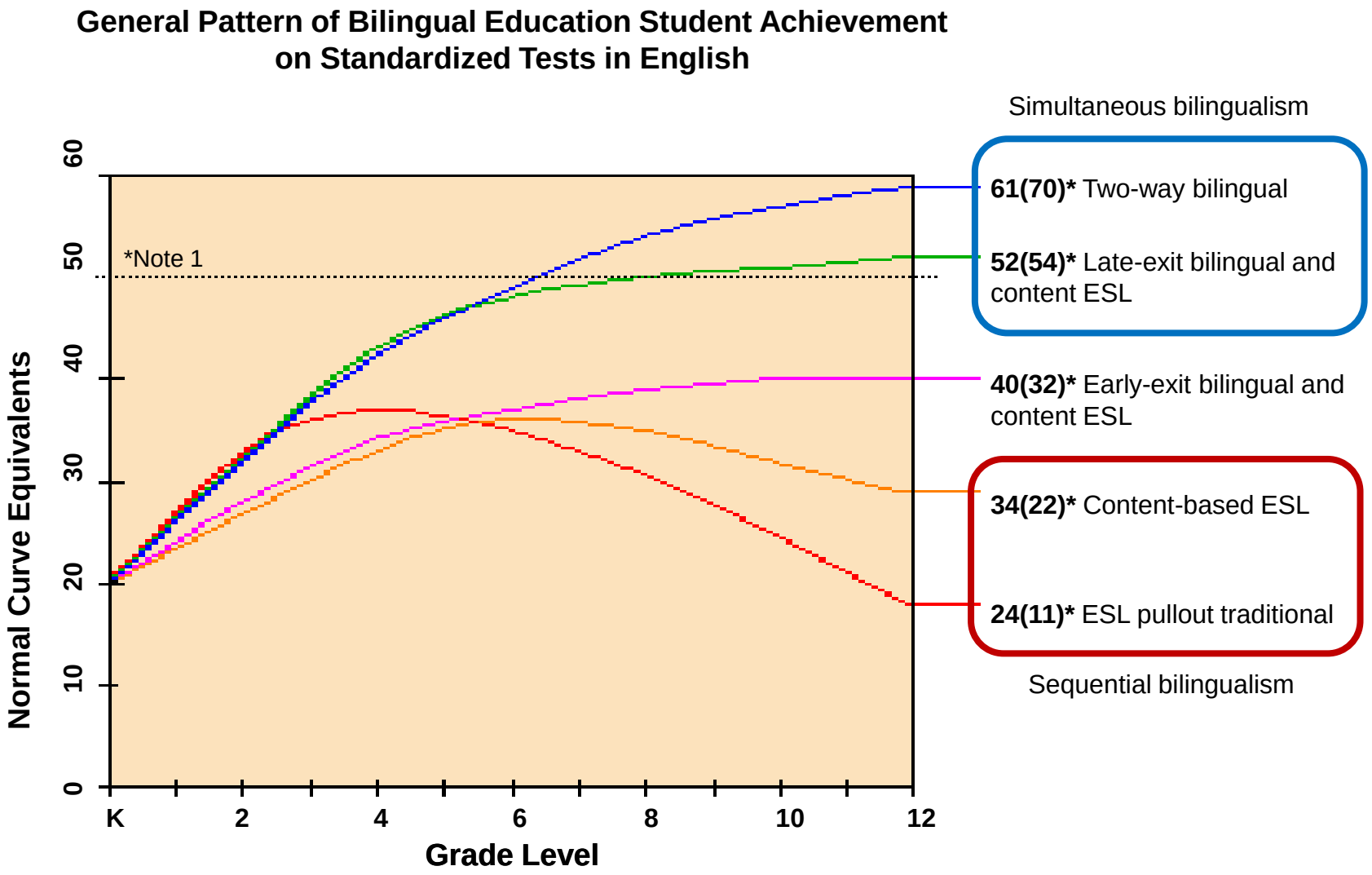
Cognitive Academic Language Proficiency (CALP)

- Education expands proficiency by including formal, academic words (e.g., via reading, advanced language models, extended discourse) and CALP begins to emerge after 5-7 years of schooling.

English learners cannot maintain age/grade language development if given English-only instruction due starting at least 5 years behind monolingual peers unless…

…they have already reached CALP via education at least up to 5th grade in their heritage language.

Achievement Outcomes for ELs: A developmental perspective.



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

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

What is a "Bilingual School Psychologist," and what is a "Bilingual Evaluation?"

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

Flanagan, McGrew & Ortiz, 2000

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

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

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

Development Varies by Experience – Not necessarily by race or ethnicity

"The key consideration in distinguishing between a difference and a disorder is whether the child's performance differs significantly from peers with similar experiences." (p. 105)

- Wolfram, Adger & Christian, 1999

For ELs, Test Score Validity is Not Established by Age Alone

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

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

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

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

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

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

For ELs, the Problem in Testing is Test Score Validity

NO BIAS

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

POTENTIAL BIAS

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

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

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

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

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

Example of Potential Construct Invalidity:

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

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

Example of Potential Interpretive Invalidity:

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

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

Score Validity Requires Construct Validity Not "Caution"

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

"DIFFERENCE vs. DISORDER?"

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

Current and typical approaches for addressing test score validity are superficial, at best, and there is no research that indicates that any of them produce "valid" scores for English learners. Therefore, because the question regarding "difference vs. disorder" centers on the concept of validity, note of the current approaches provide anywhere near a satisfactory or complete solution to the problems inherent in testing ELs.

Current Methods Fail to Establish Test Score Validity

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

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

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

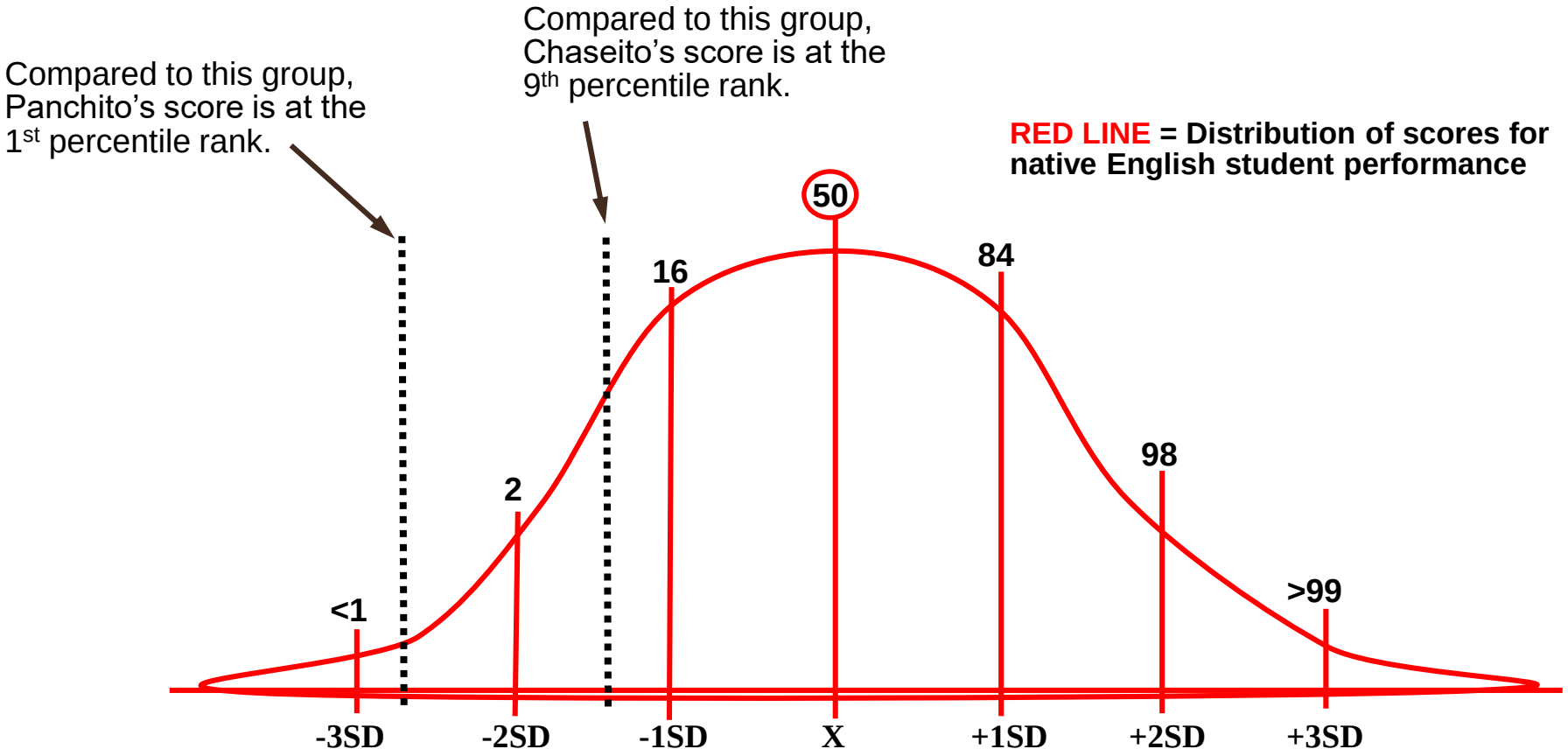
For native English speakers, growth of cognitive abilities and knowledge acquisition are tied closely to age and assumes normal educational experiences. Thus, age-based norms effectively control for variation in development and provide an appropriate basis for comparison. However, this is not true for English learners who may neither live in a "mainstream" culture nor benefit to an equivalent degree from formal education as native English speakers.

Development Varies by Experience With and Exposure to English and Mainstream Culture

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

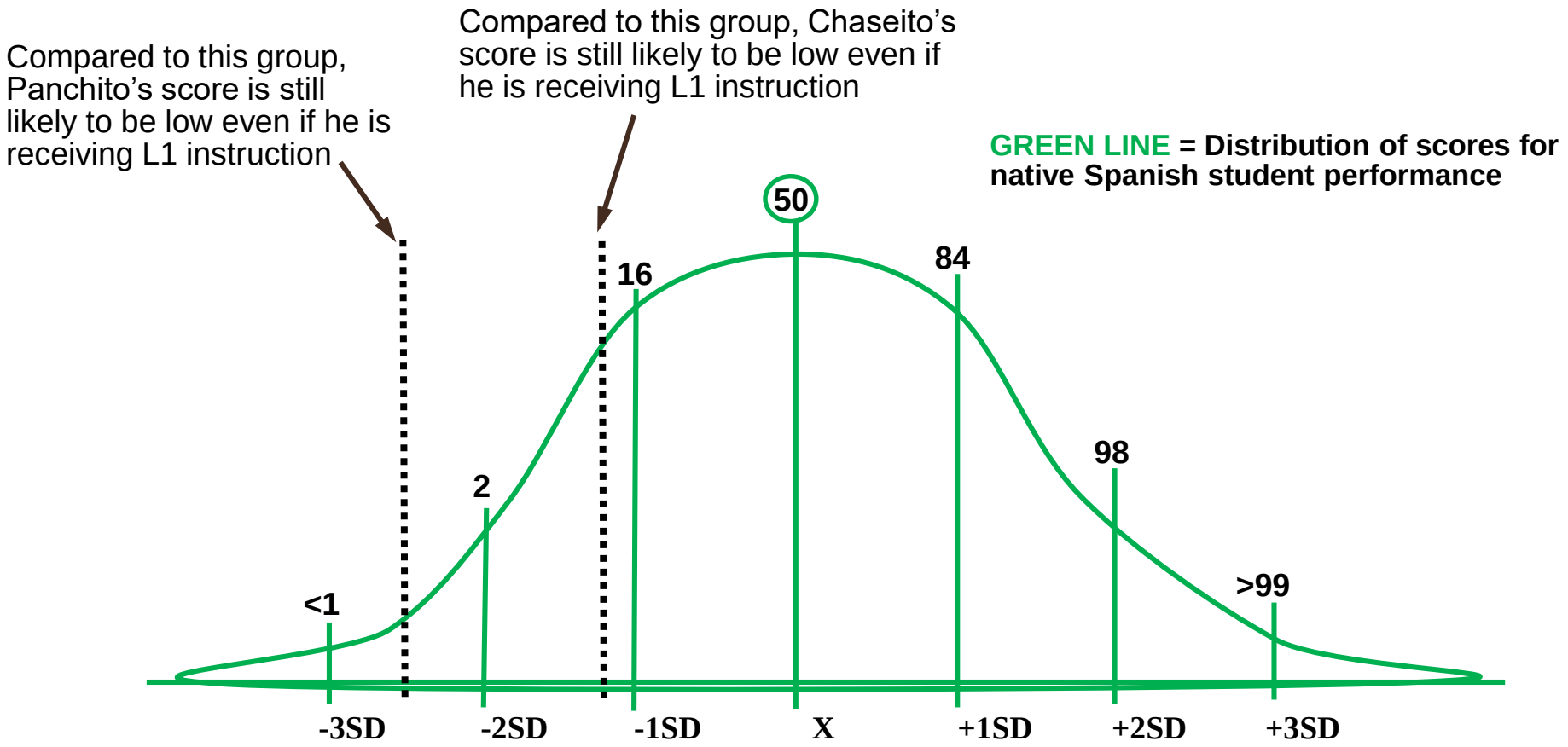
Test score validity must be evaluated or established via use of a "true peer" comparison standard and, with some limited exceptions, the only way to accomplish this task is to utilize research on EL test performance.

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



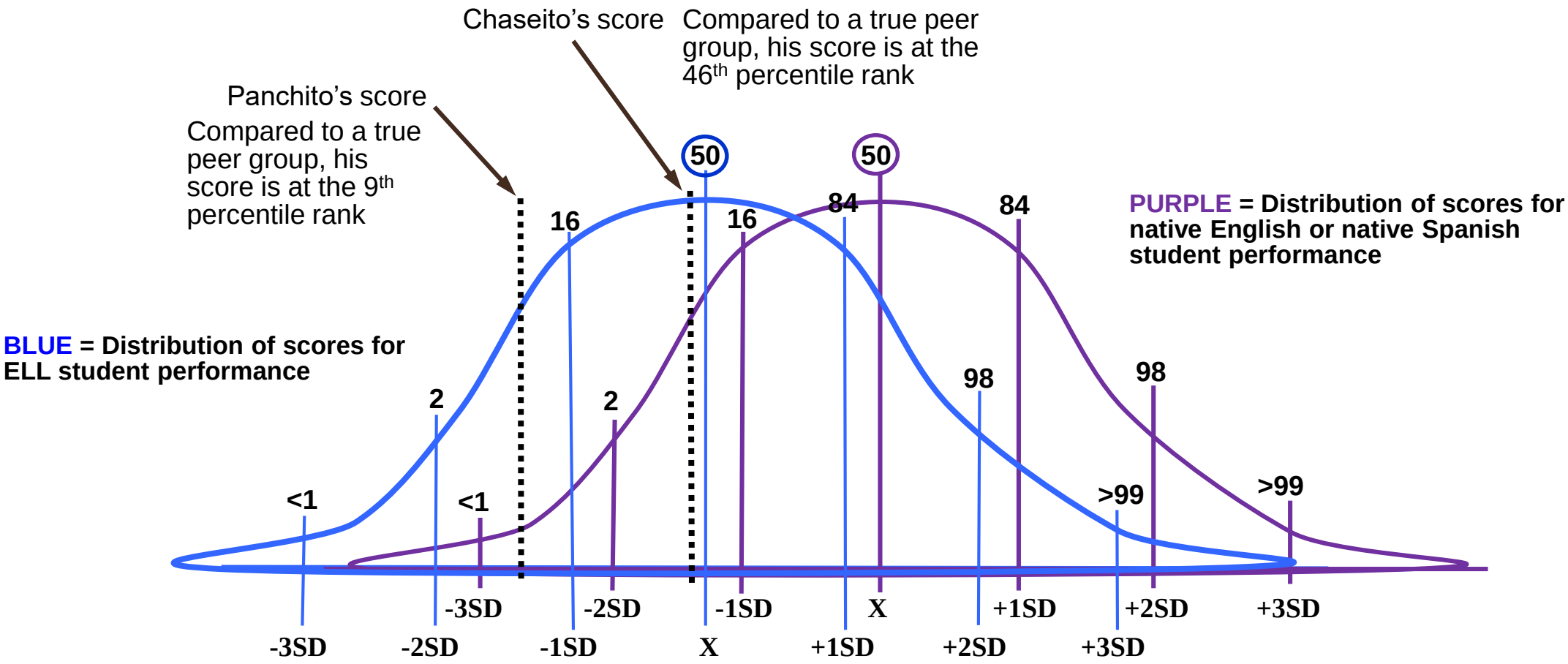
For the purposes of determining whether a disability exists, use of a monolingual English speaking comparison group is discriminatory and makes it appear incorrectly that both students might have some type of disability.

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



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

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



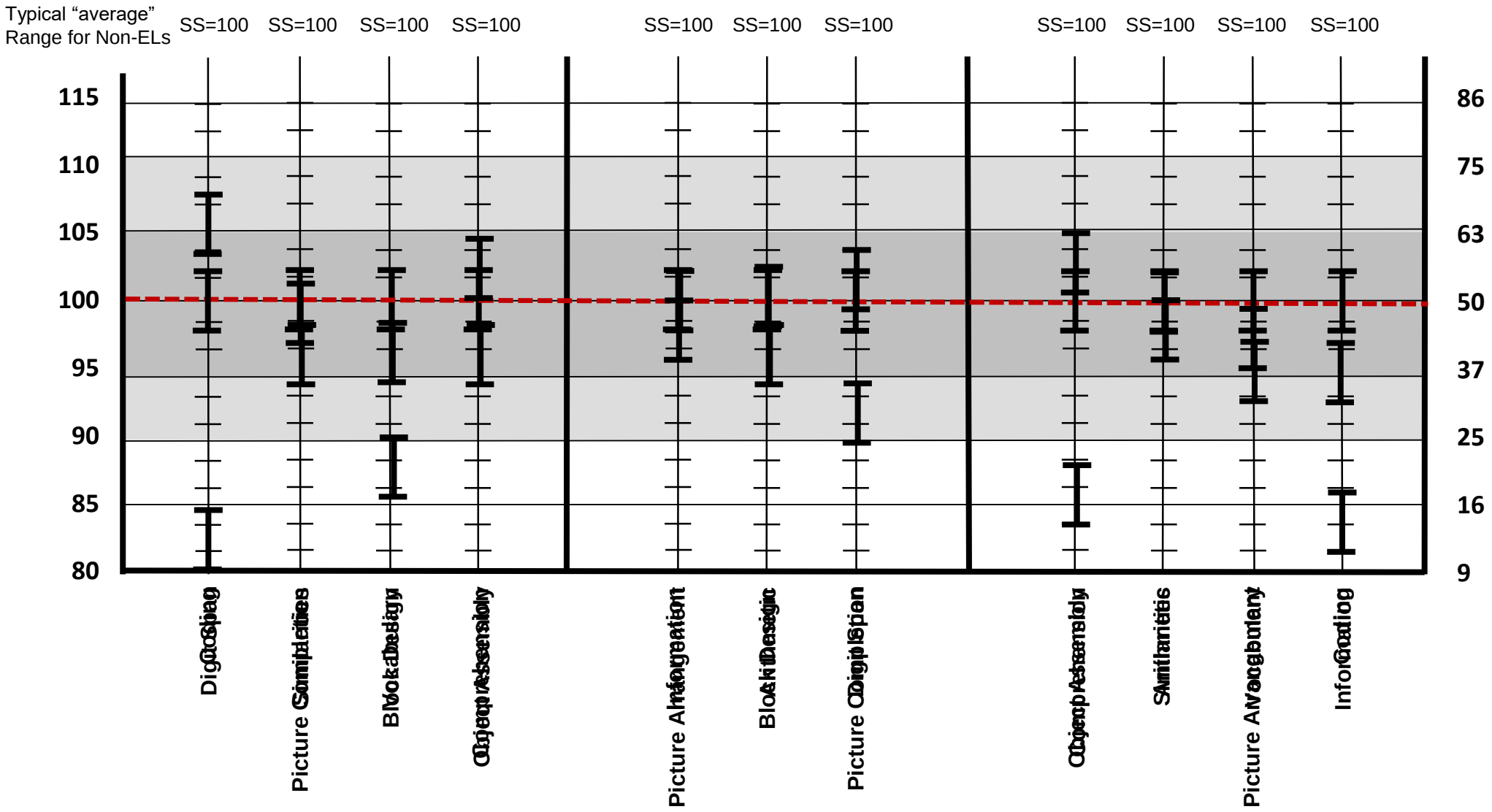
Whether conducted through RTI/MTSS or testing, only use of a "true peer" comparison group provides the basis for making non-discriminatory diagnostic decisions as long as there is control for developmental language differences between English learners and English speakers and among English learners and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

Current Approaches Fail to Establish Test Score Validity

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

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

Fairness in Determining "Average" Performance: ES to ES



Research Foundations for EL Evaluation: EL to ES

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

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

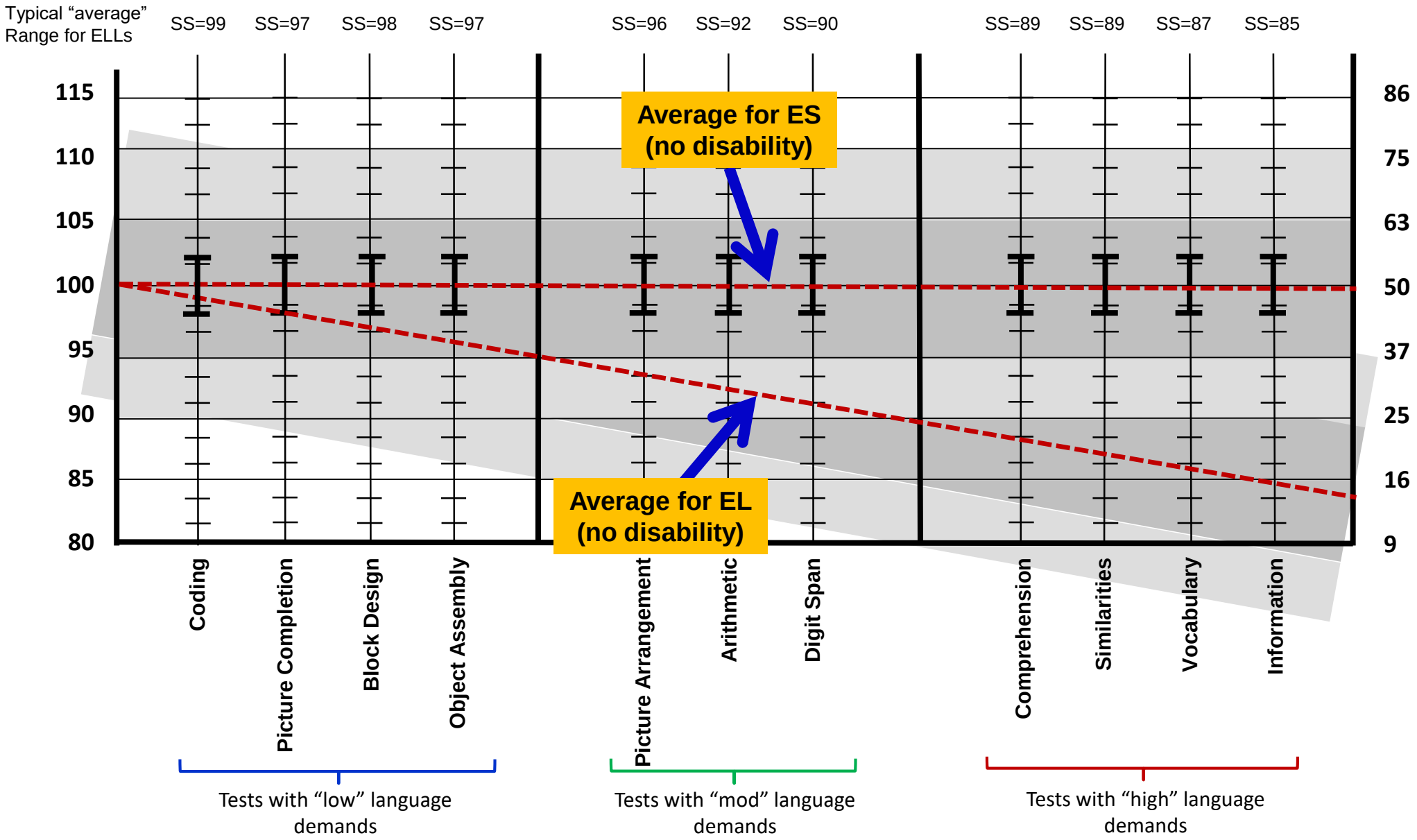
Tests with **"high"** language demands

Tests with **"mod"** language demands

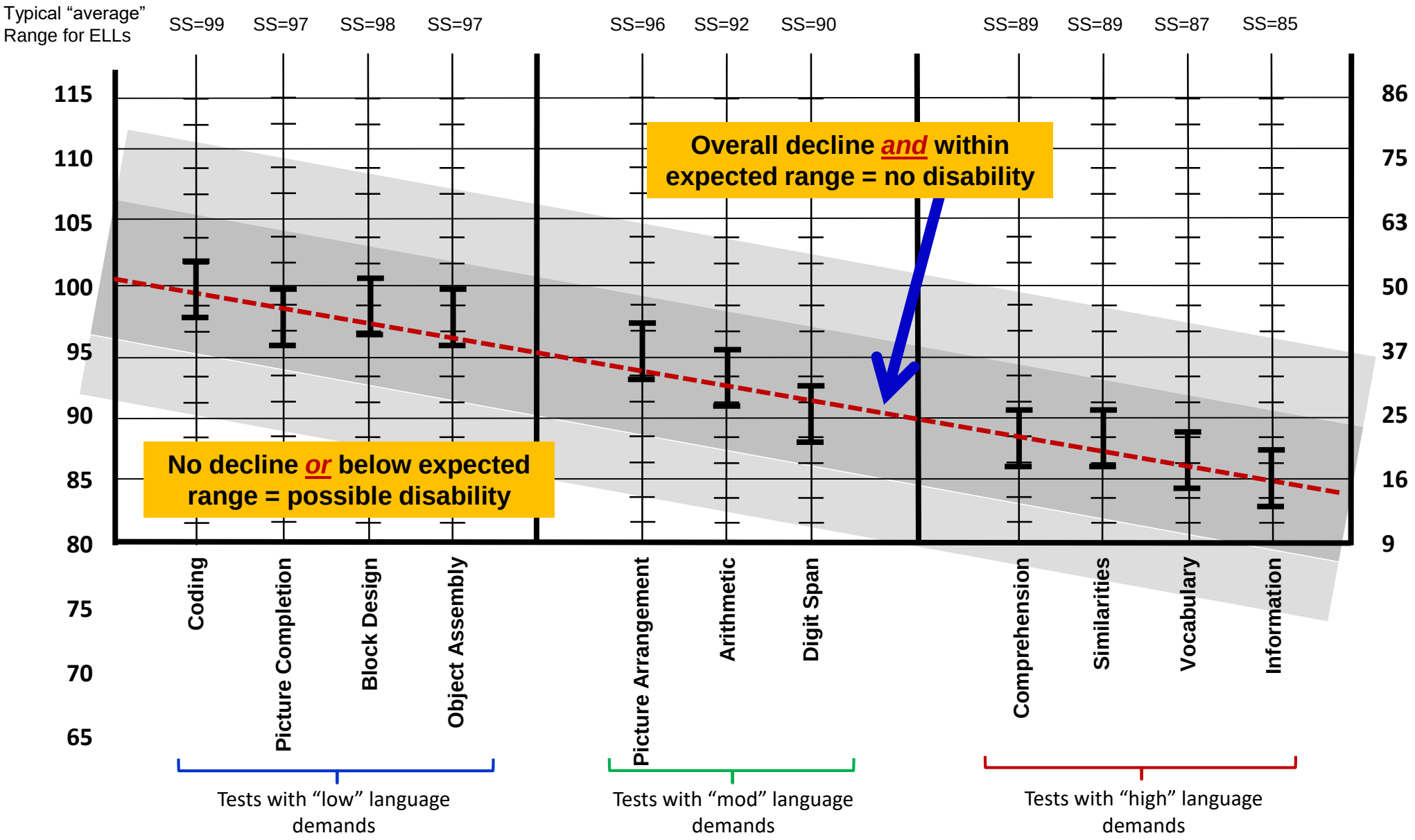
Tests with **"low"** language demands

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

Fairness in Determining "Average" Performance: EL to ES



Fairness in Determining "Average" Performance: EL to ES



Foundations of the Culture-Language Interpretive Matrix

Research Foundations for EL Evaluation: EL to ES

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

Tests with "high" language demands

Tests with "mod" language demands

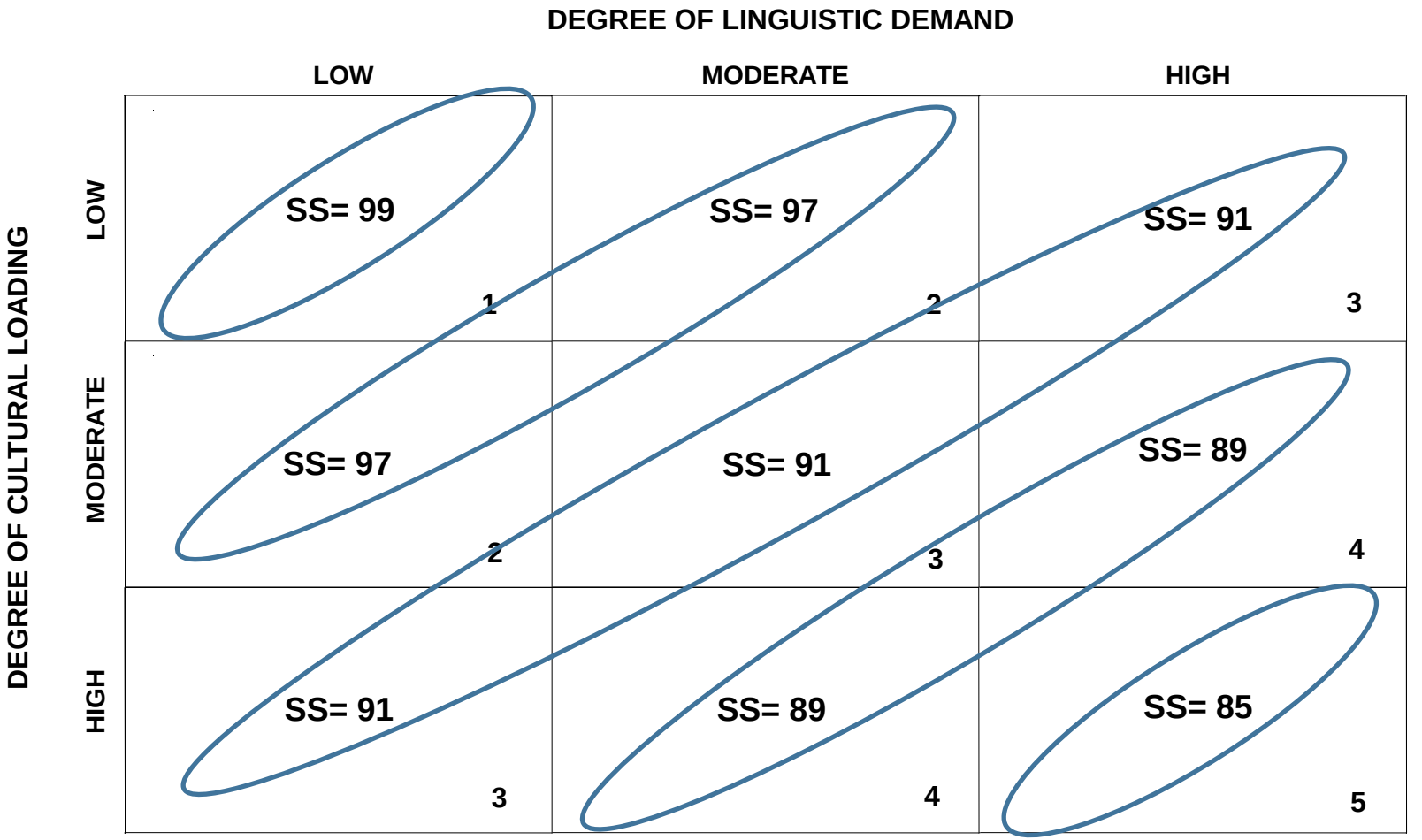
Tests with "low" language demands

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

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

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

SAMPLE OF RESEARCH-BASED MEANS REGARDING EXPECTED PERFORMANCE FOR ENGLISH LEARNERS



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

Fairness in Determining "Average" Performance: EL to ES

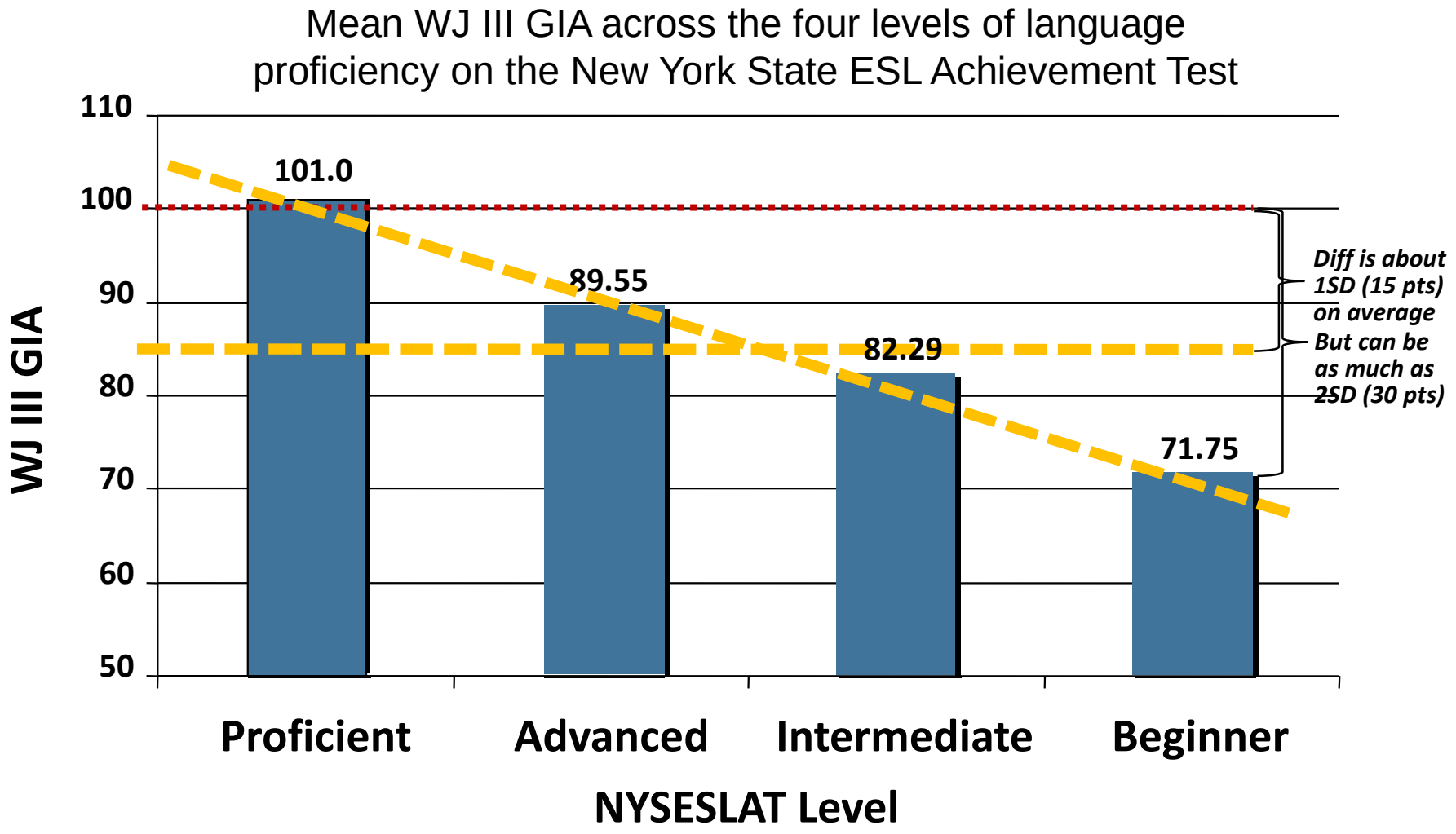
Matrix of WISC subtest means arranged by EL vs. ES test performance

DEGREE OF LINGUISTIC DEMAND

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

Research Foundations for EL Evaluation: EL to EL

General ability level performance as compared to other English learners

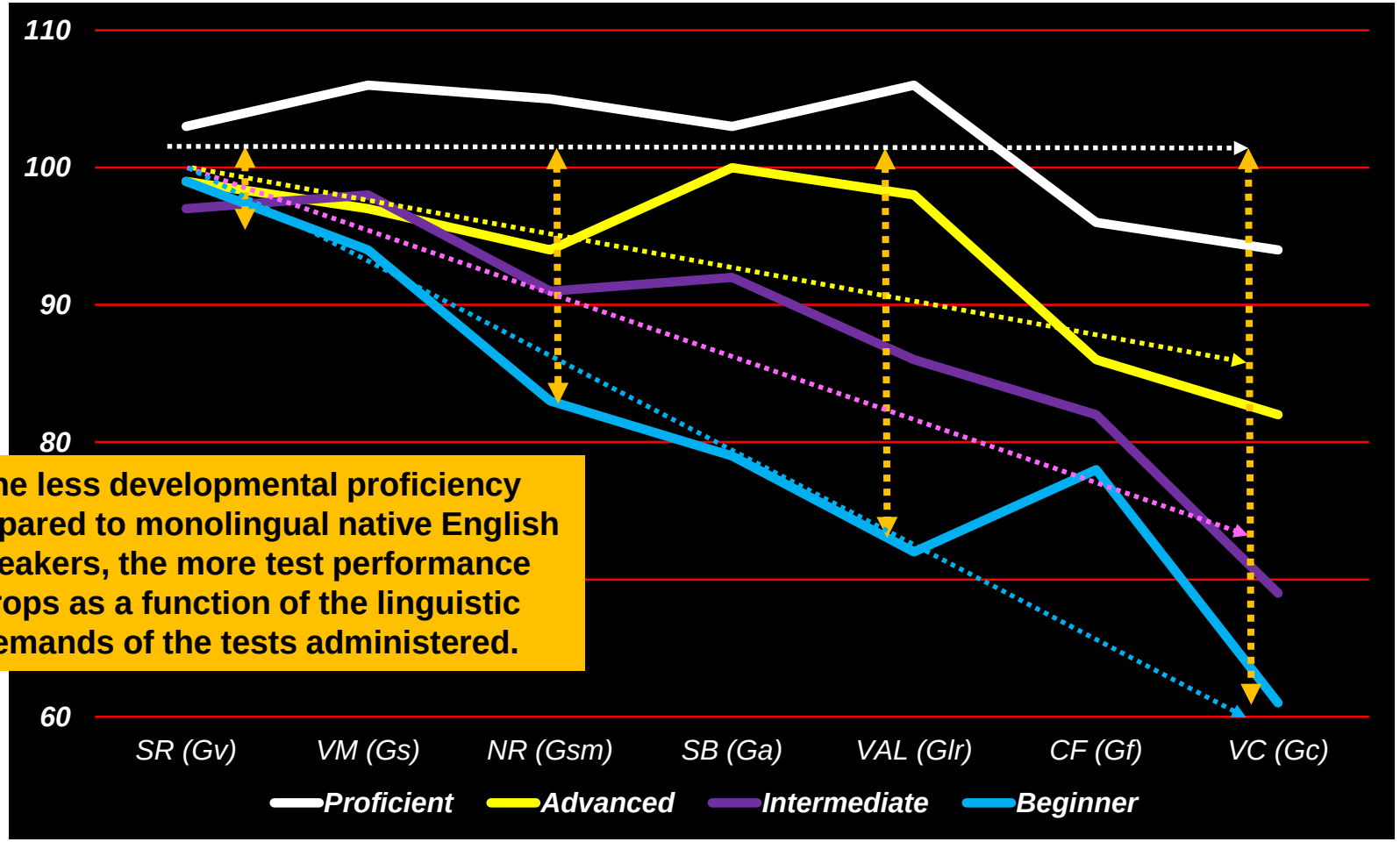


Source: Sotelo-Dynega, M., Ortiz, S.O., Flanagan, D.P., Chaplin, W. (2013).

Research Foundations for EL Evaluation: EL to EL

Subtest level performance as compared to other English Learners

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



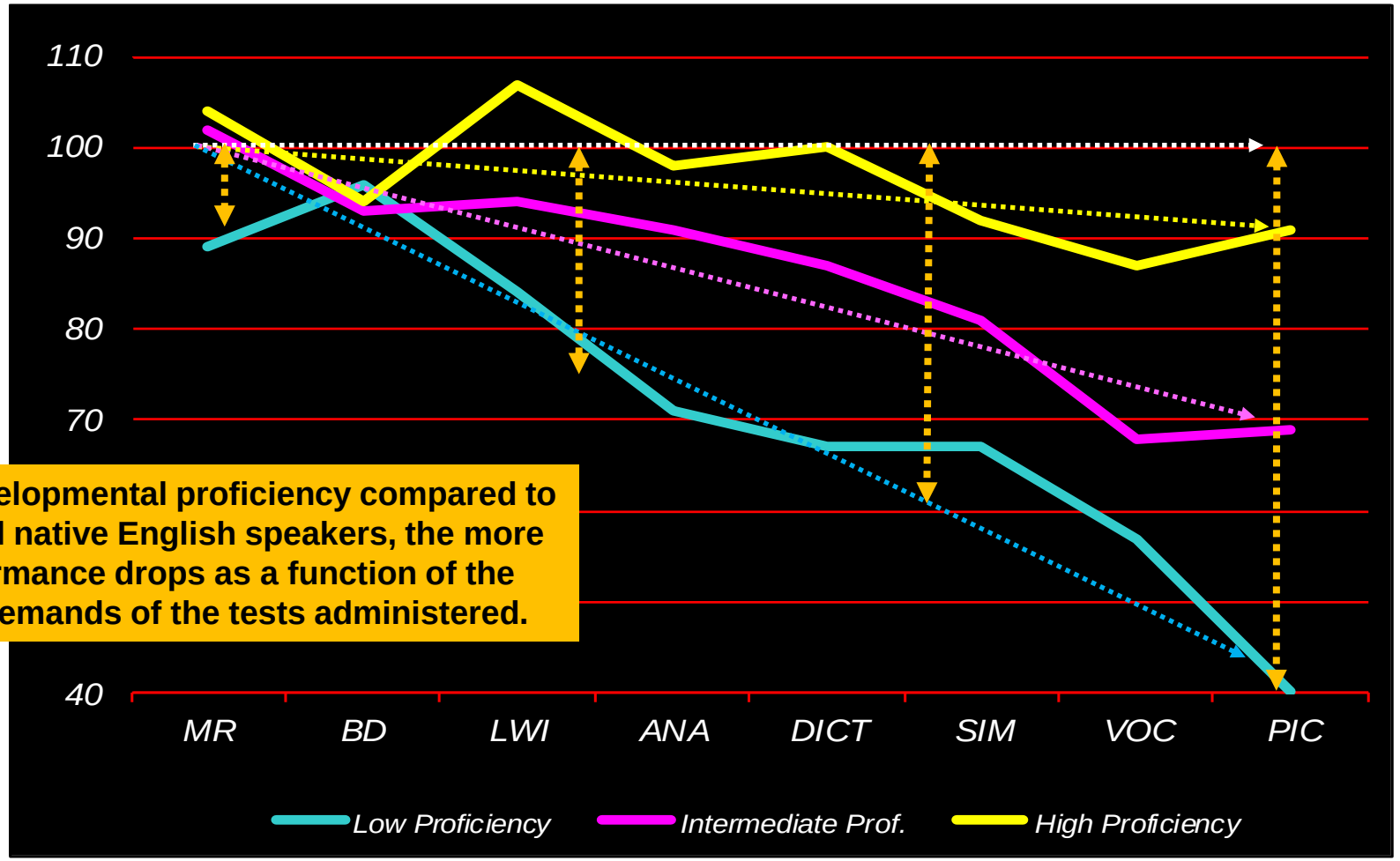
The less developmental proficiency compared to monolingual native English speakers, the more test performance drops as a function of the linguistic demands of the tests administered.

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

Research Foundations for EL Evaluation: EL to EL

Subtest level performance as compared to other English Learners

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



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

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

Summary of Research Foundations for EL Evaluation

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

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

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

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

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

The Culture-Language Interpretive Matrix (C-LIM)

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 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.

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

The Culture-Language Interpretive Matrix (C-LIM)

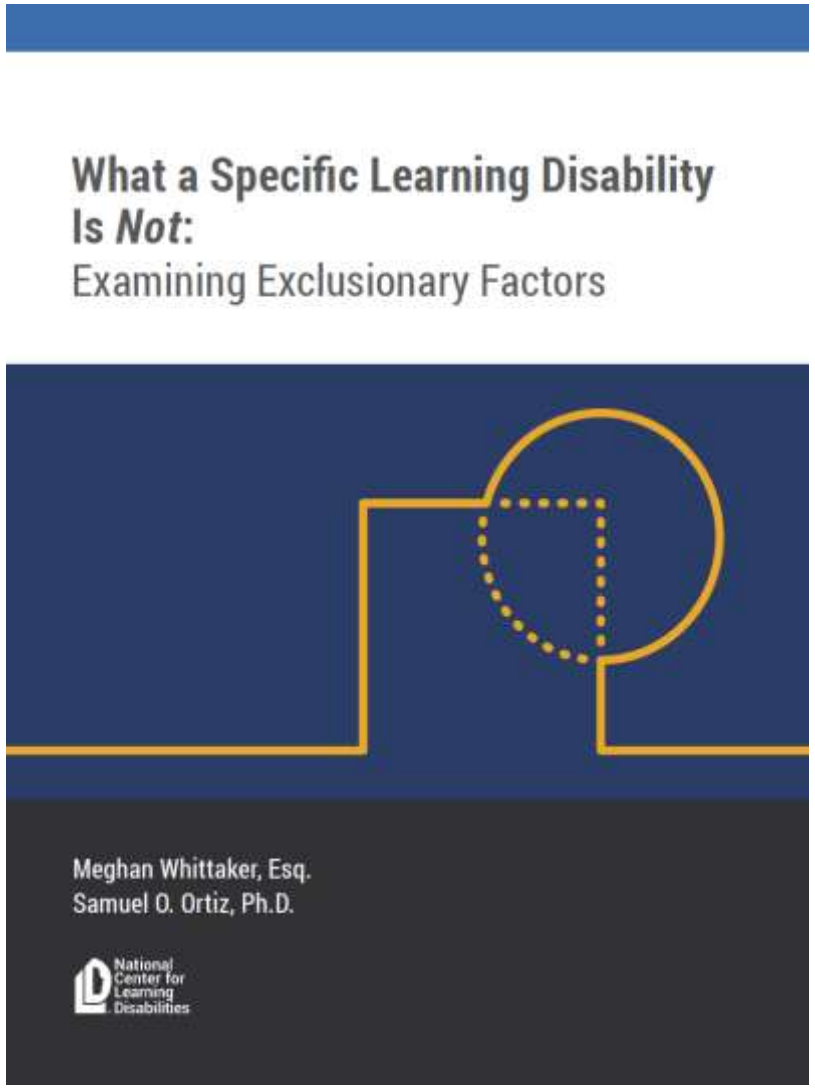
Addressing exclusionary factors via examination of test score validity

Translation of Research into Practice

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

The Culture-Language Interpretive Matrix (C-LIM)

Addressing exclusionary factors via examination of test score validity

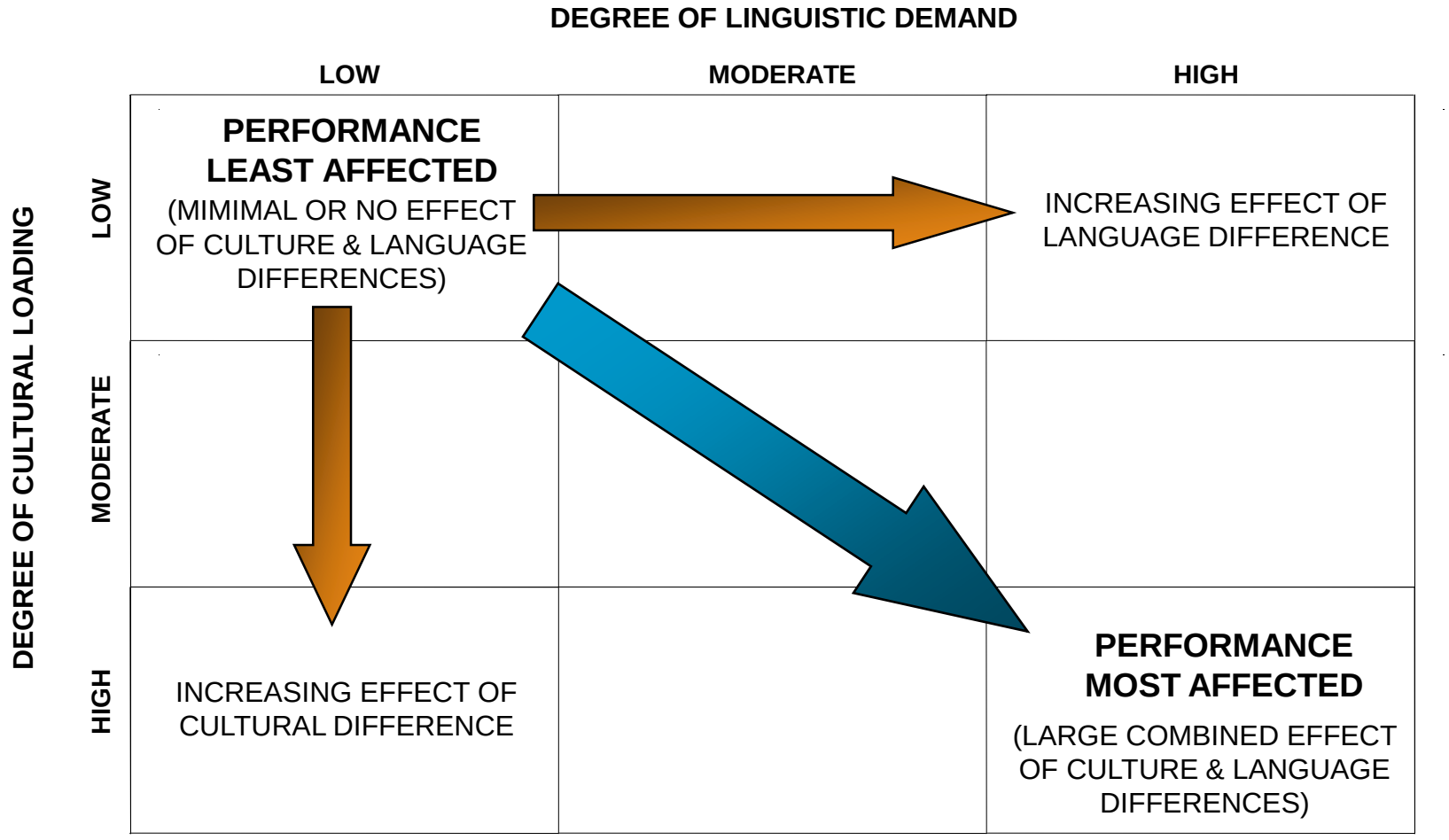


"To properly meet the definition and its exclusionary criteria, LEAs would first have to identify the primary cause(s) of a student's low achievement. For instance, if a child has limited English language proficiency, and it influences behavior and learning, it could appear as though the child has SLD.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

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

Fairness in Determining “Average” Performance: EL to ES

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



Fairness in Determining "Average" Performance: EL to EL

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

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

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

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

Evaluation MUST account for EL vs. ES differences

Evaluation MUST also account for EL vs. EL differences

Fairness in Determining "Average" Performance: EL to EL

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

Degree of Linguistic Demand

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

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

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

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

The Culture-Language Interpretive Matrix – Basic Version 4.0

Clear ALL Data in Matrix

Culture-Language Interpretive Matrix - Basic Version 4.0

Clear Unused Subtests

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

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

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

Print C-LIM Matrix

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

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING	LOW	MODERATE	HIGH
LOW	Score Cell Average =	Score Cell Average =	Score Cell Average =
MODERATE	Score Cell Average =	Score Cell Average =	Score Cell Average =
HIGH	Score Cell Average =	Score Cell Average =	Score Cell Average =

The Culture-Language Interpretive Matrix – Basic Version 4.0

Culture-Language Interpretive Matrix - Basic Version 4.0

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

Print C-LIM Graph

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

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

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

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

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

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

Cross-Battery XBA Assessment

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

The Culture-Language Interpretive Matrix (C-LIM)

GENERAL RULES AND GUIDANCE FOR EVALUATION OF TEST SCORE VALIDITY

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

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

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

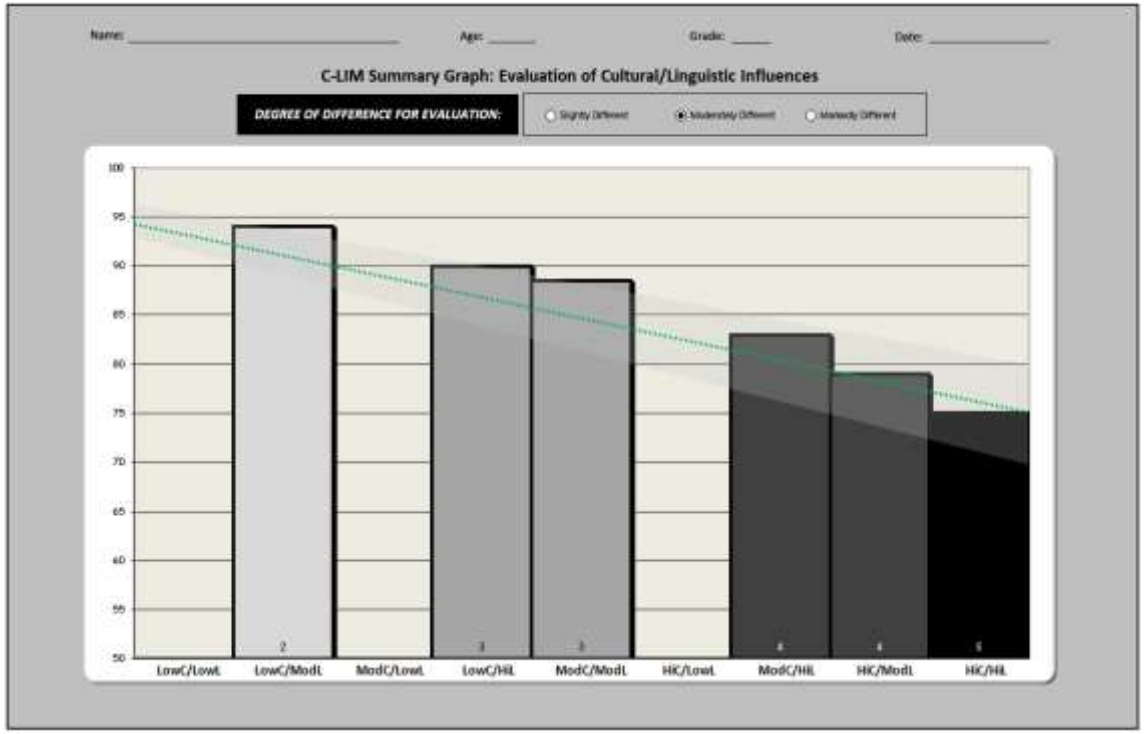
3. No Significant Score Variability: *There is no variability in the scores that form the aggregate in any one cell or any variability between or among cells in the same level where high score performance may be masking the presence of low performance? Variability is defined as one score below average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

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

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

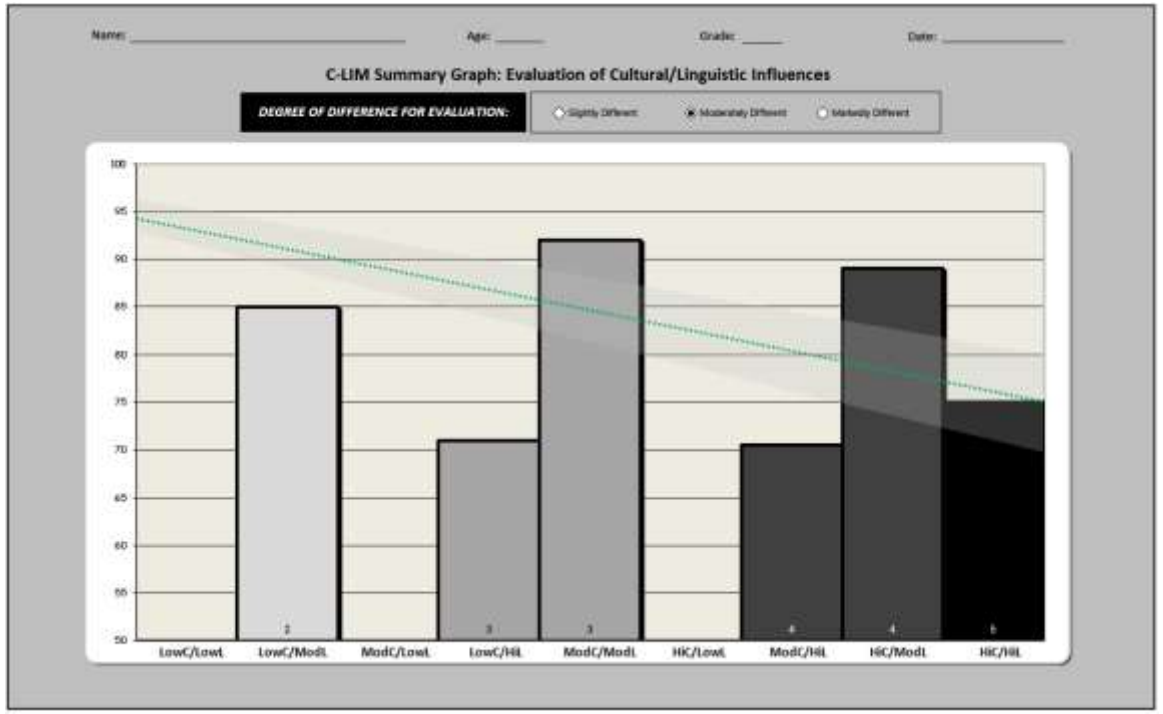
When all three criteria are **MET,** it may be concluded that the test scores are INVALID because they were likely influenced primarily by the presence of cultural/linguistic variables and should not be interpreted. When any single criterion is **NOT MET**, the results can be assumed to be **VALID** and may be interpreted.

Interpreting Test Score Validity with the C-LIM



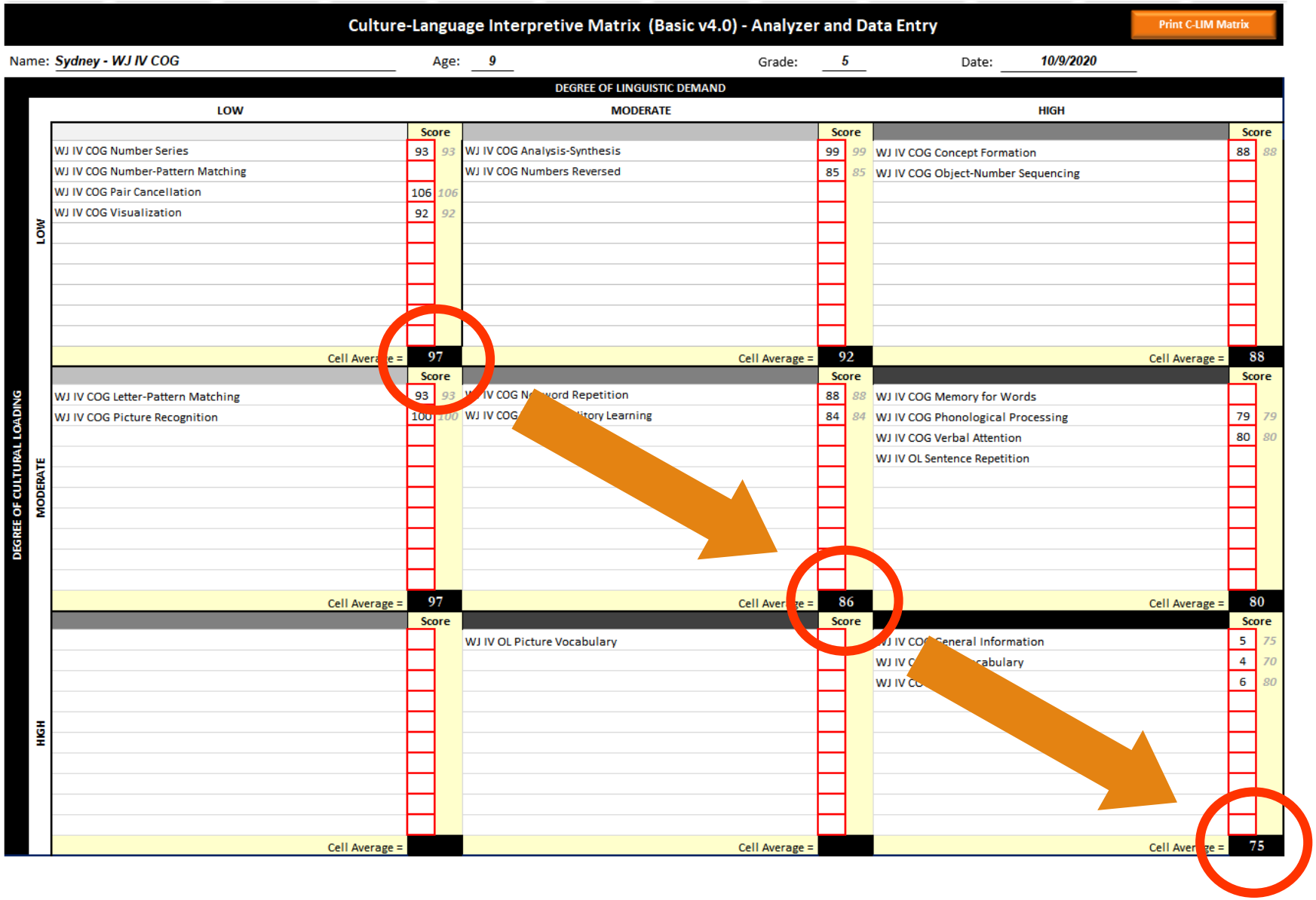
Example of "likely invalid" score pattern—overall general decline **AND** scores within or above expected (average) range **AND** no important variability. Performance is primarily due to linguistic and cultural factors: CANNOT interpret or assign specific meaning to scores.

Example of "likely valid" score pattern—no overall decline **OR** scores below expected (average) range **OR** important variability. Performance is NOT due primarily to linguistic and cultural factors: OK to interpret scores except for Gc (language or knowledge).



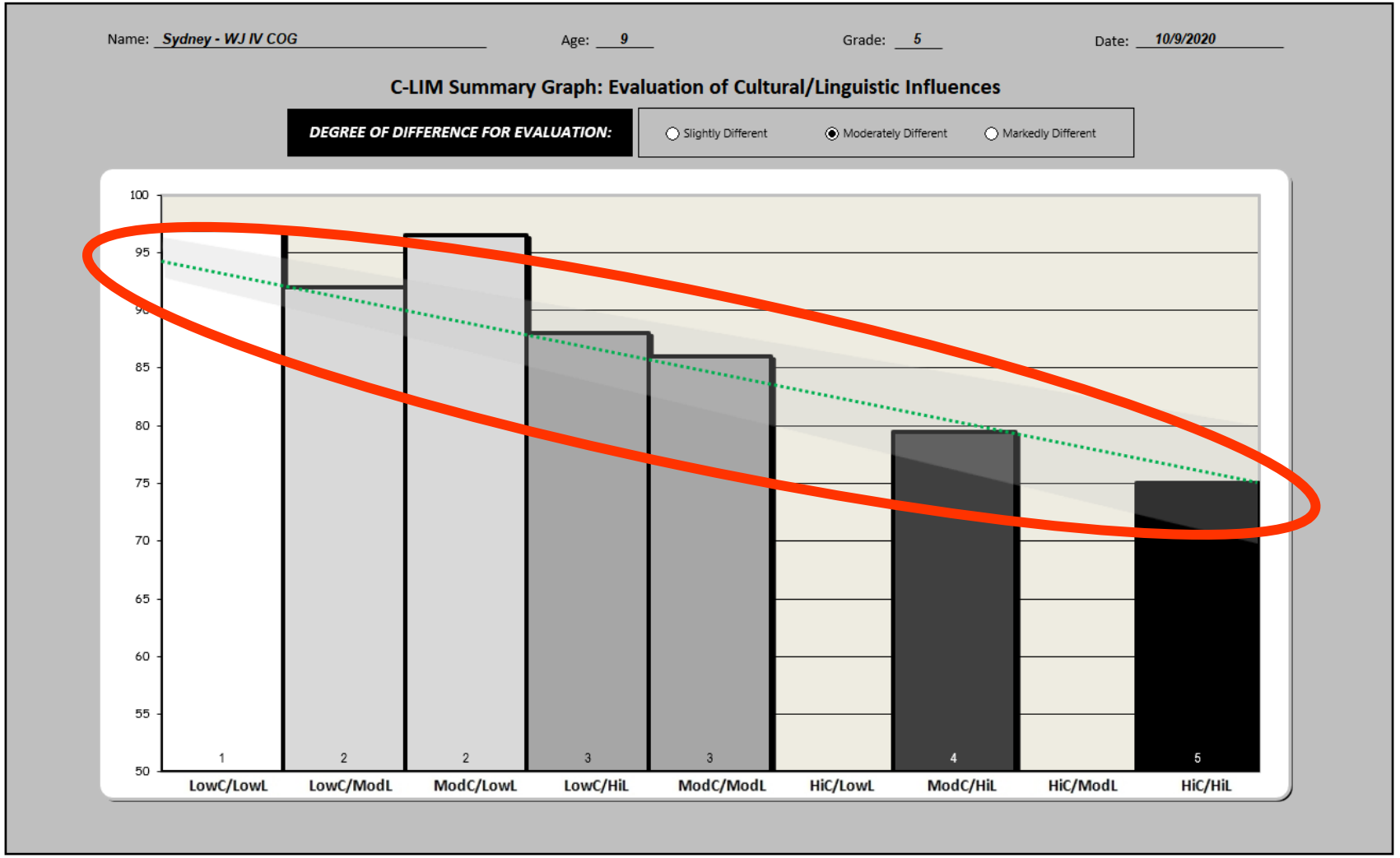
C-LIM Guidelines for Evaluating Test Scores

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



C-LIM Guidelines for Evaluating Test Scores

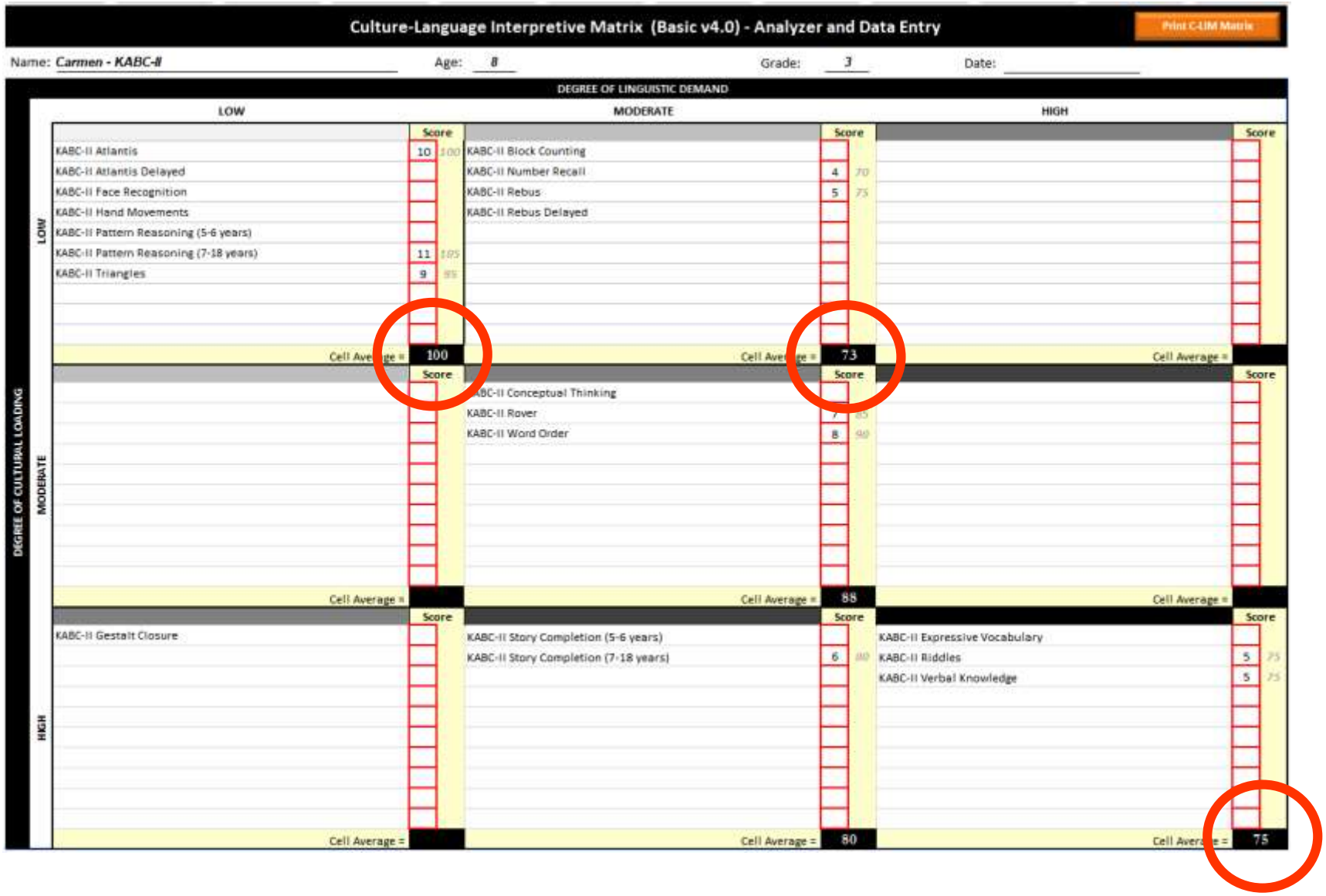
General pattern of decline AND all scores within or above the expected range for ELs.



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

C-LIM Guidelines for Evaluating Test Scores

General pattern of decline <u>OR</u> one or more scores below expected range for ELs.



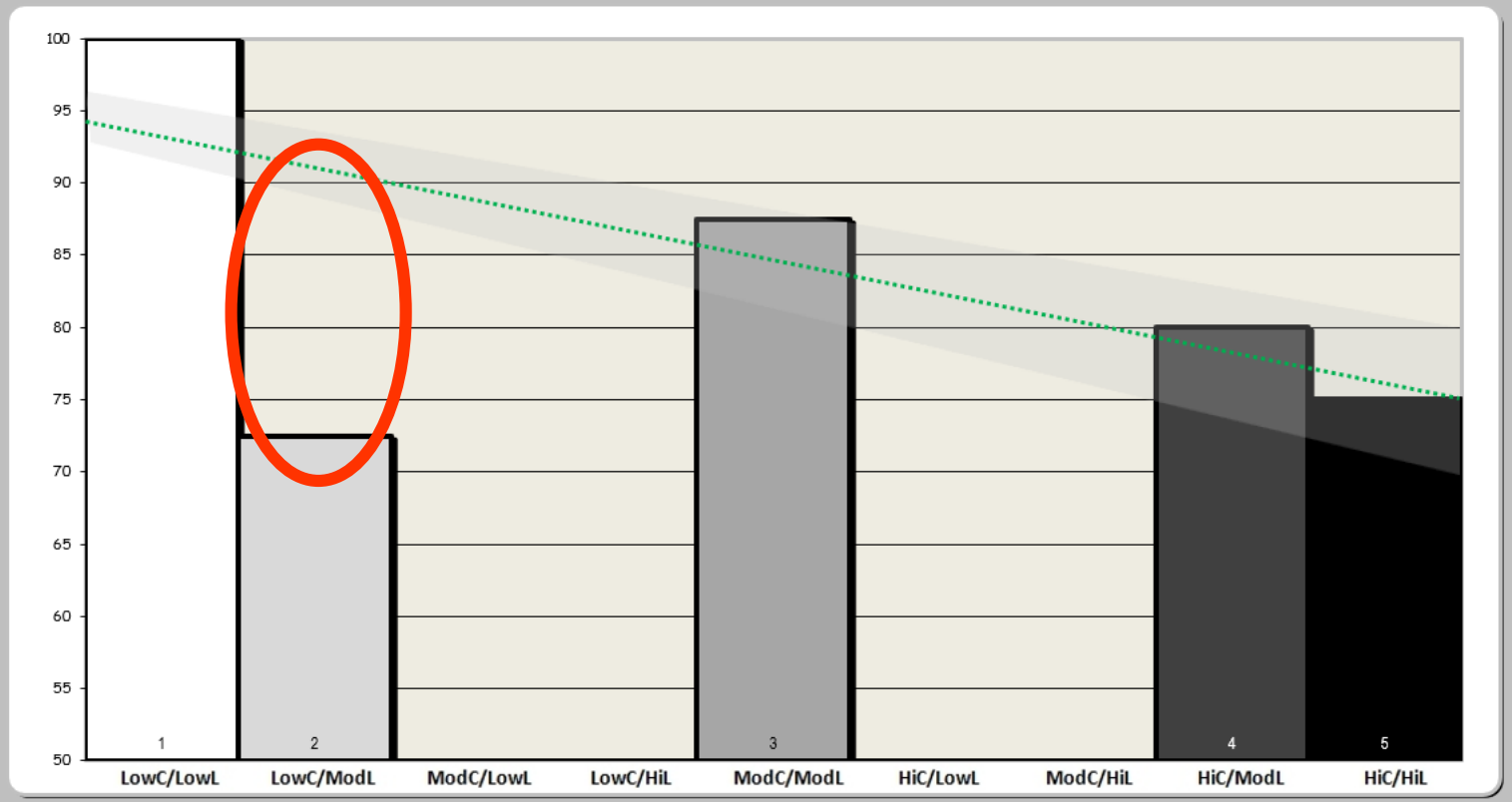
C-LIM Guidelines for Evaluating Test Scores

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

Name: *Carmen - KABC-II* Age: *8* Grade: *3* Date: _____

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

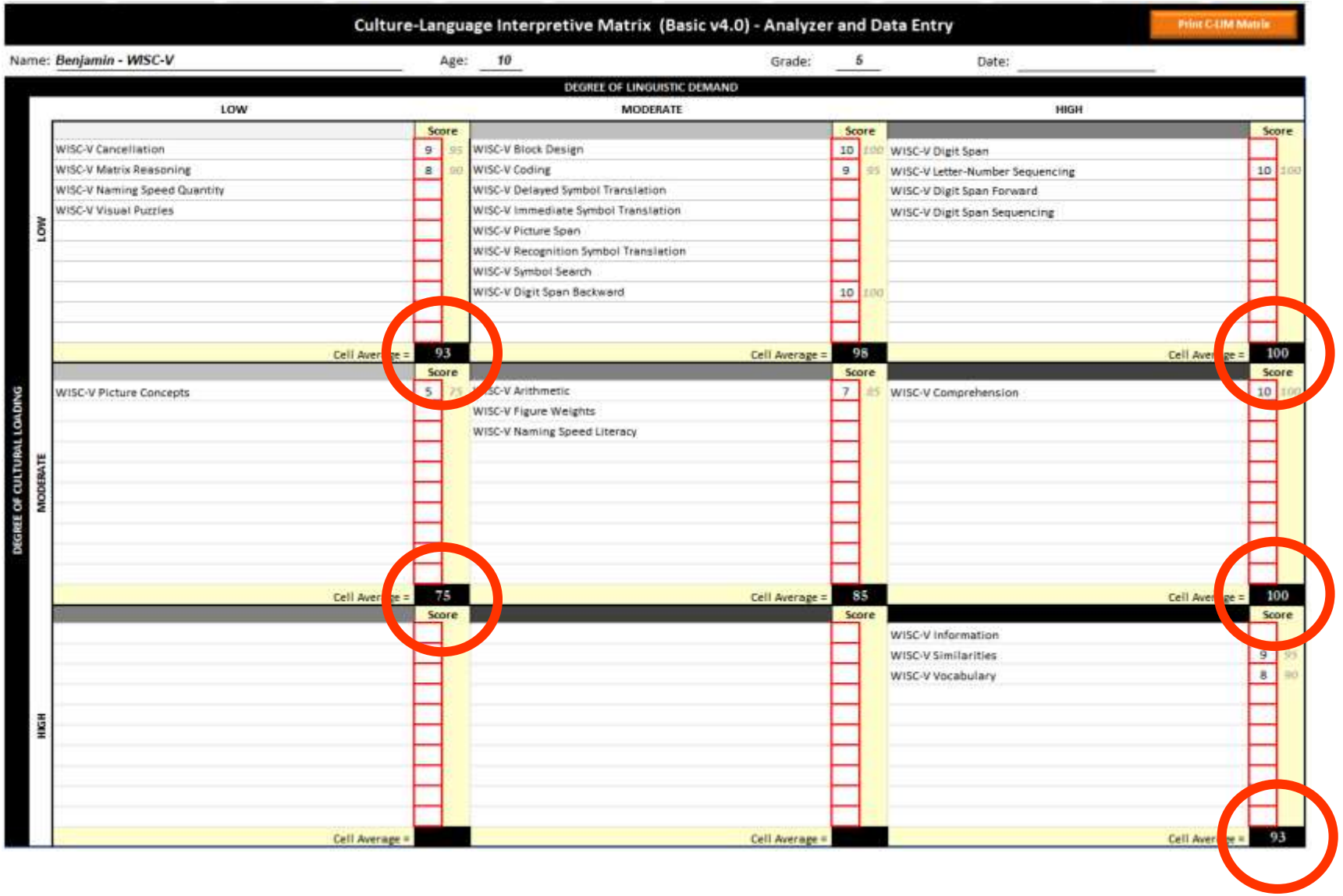
DEGREE OF DIFFERENCE FOR EVALUATION: ○ Slightly Different ◉ Moderately Different ○ Markedly Different



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

C-LIM Guidelines for Evaluating Test Scores

No general pattern of decline.



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

Print C-LIM Matrix

Name: *Benjamin - WISC-V* Age: *10* Grade: *5* Date:

DEGREE OF LINGUISTIC DEMAND

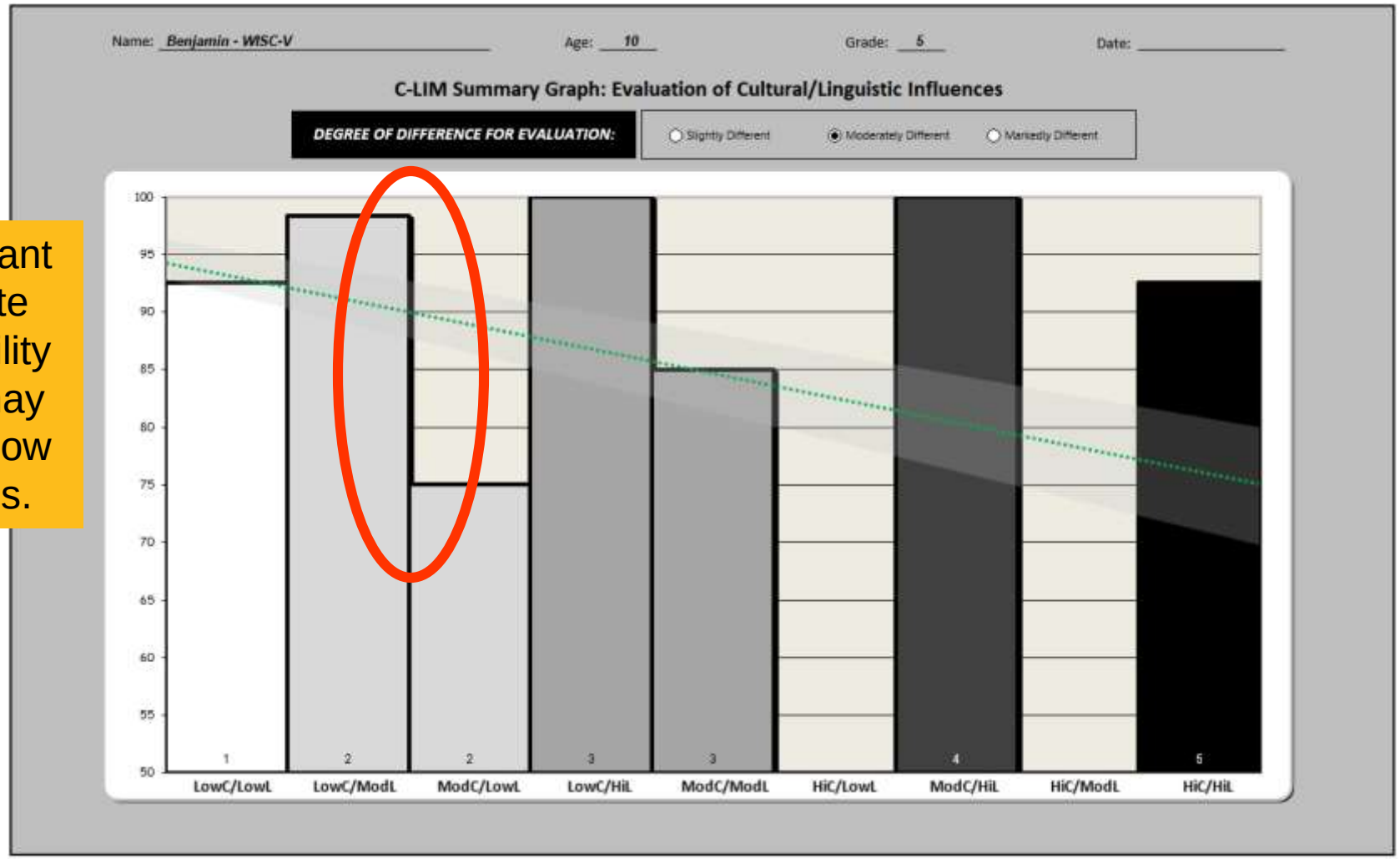
DEGREE OF CULTURAL LOADING

	LOW		MODERATE		HIGH	
		Score		Score		Score
LOW	WISC-V Cancellation	9 (95)	WISC-V Block Design	10 (100)	WISC-V Digit Span	
	WISC-V Matrix Reasoning	8 (90)	WISC-V Coding	9 (95)	WISC-V Letter-Number Sequencing	10 (100)
	WISC-V Naming Speed Quantity		WISC-V Delayed Symbol Translation		WISC-V Digit Span Forward	
	WISC-V Visual Puzzles		WISC-V Immediate Symbol Translation		WISC-V Digit Span Sequencing	
			WISC-V Picture Span			
			WISC-V Recognition Symbol Translation			
			WISC-V Symbol Search			
			WISC-V Digit Span Backward	10 (100)		
	Cell Average =	93	Cell Average =	98	Cell Average =	100
MODERATE		Score		Score		Score
	WISC-V Picture Concepts	5 (75)	WISC-V Arithmetic	7 (85)	WISC-V Comprehension	10 (100)
			WISC-V Figure Weights			
			WISC-V Naming Speed Literacy			
	Cell Average =	75	Cell Average =	85	Cell Average =	100
HIGH		Score		Score		Score
					WISC-V Information	
					WISC-V Similarities	9 (95)
					WISC-V Vocabulary	8 (90)
	Cell Average =		Cell Average =		Cell Average =	93

C-LIM Guidelines for Evaluating Test Scores

No general pattern of decline.

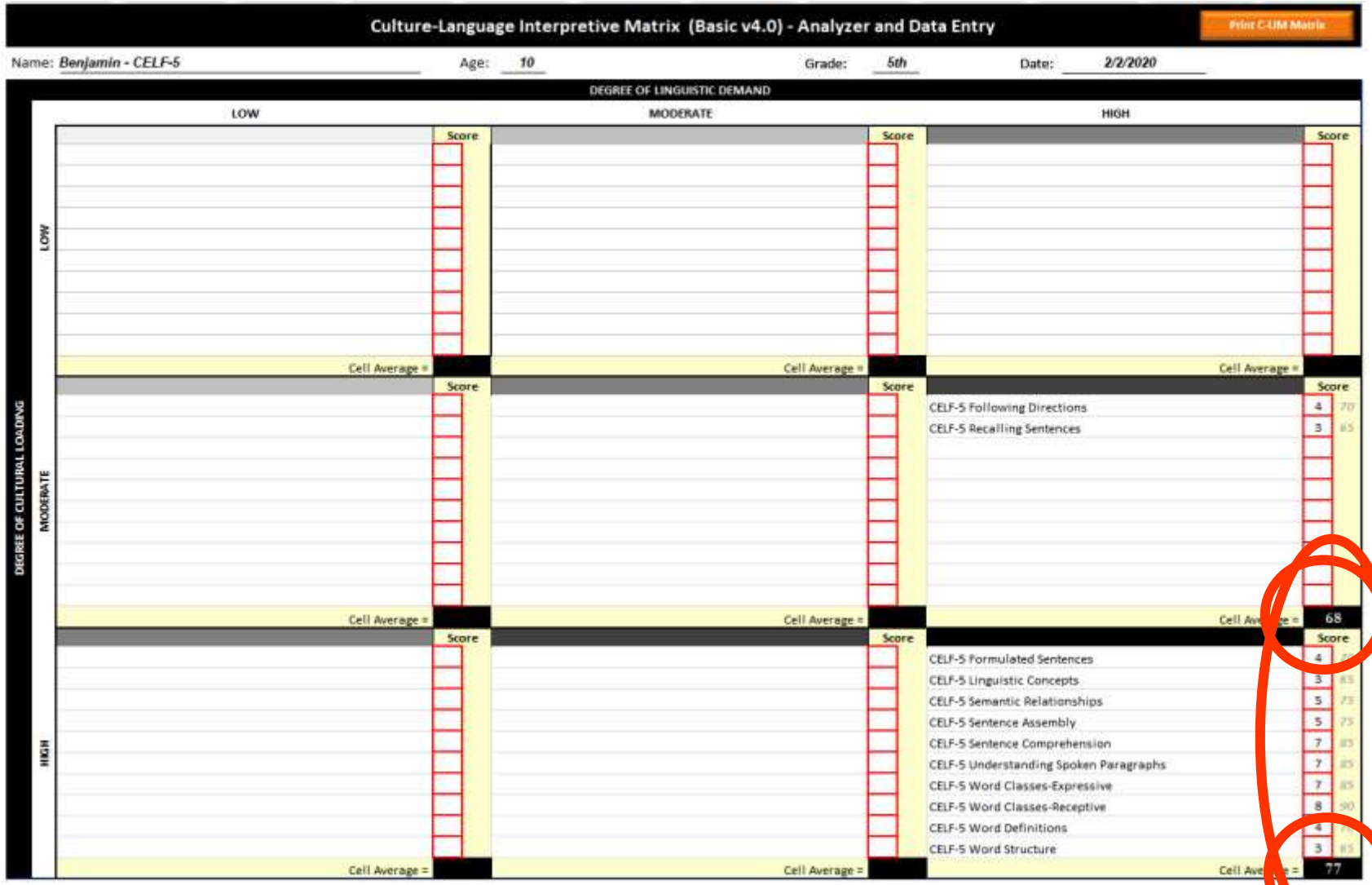
Important to note variability that may mask low scores.



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

C-LIM Guidelines for Evaluating Test Scores

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



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

Print C-LIM Matrix

Name: *Benjamin - CELF-5* Age: *10* Grade: *5th* Date: *2/2/2020*

DEGREE OF LINGUISTIC DEMAND

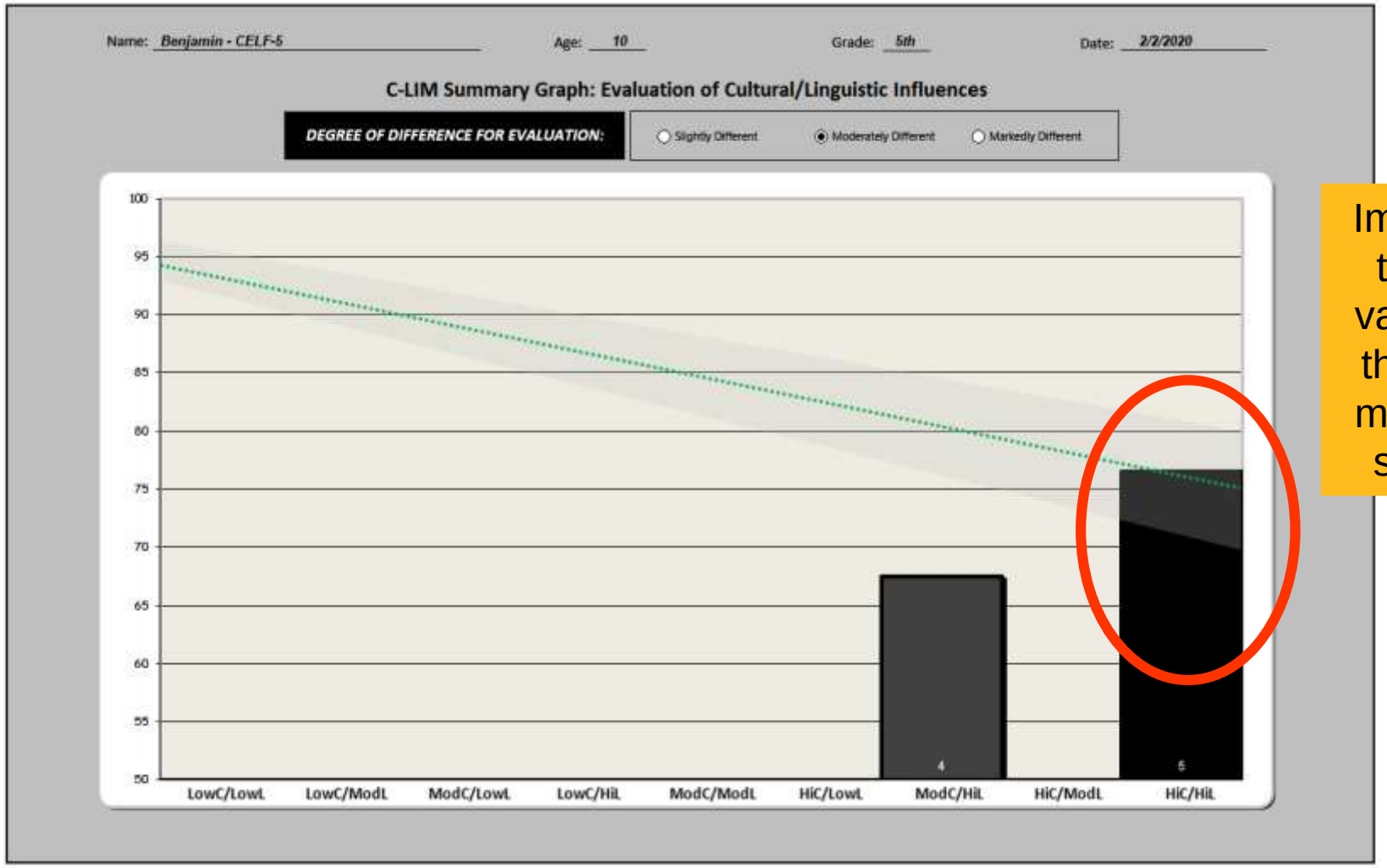
DEGREE OF CULTURAL LOADING

	LOW	MODERATE	HIGH
LOW	Score Cell Average =	Score Cell Average =	Score Cell Average =
MODERATE	Score Cell Average =	Score Cell Average =	Score CELF-5 Following Directions 4 70 CELF-5 Recalling Sentences 3 65 Cell Average = 68
HIGH	Score Cell Average =	Score Cell Average =	Score CELF-5 Formulated Sentences 4 70 CELF-5 Linguistic Concepts 3 65 CELF-5 Semantic Relationships 5 75 CELF-5 Sentence Assembly 5 75 CELF-5 Sentence Comprehension 7 85 CELF-5 Understanding Spoken Paragraphs 7 85 CELF-5 Word Classes-Expressive 7 85 CELF-5 Word Classes-Receptive 8 90 CELF-5 Word Definitions 4 70 CELF-5 Word Structure 3 65 Cell Average = 77

Important to note variability that may mask low scores.

C-LIM Guidelines for Evaluating Test Scores

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

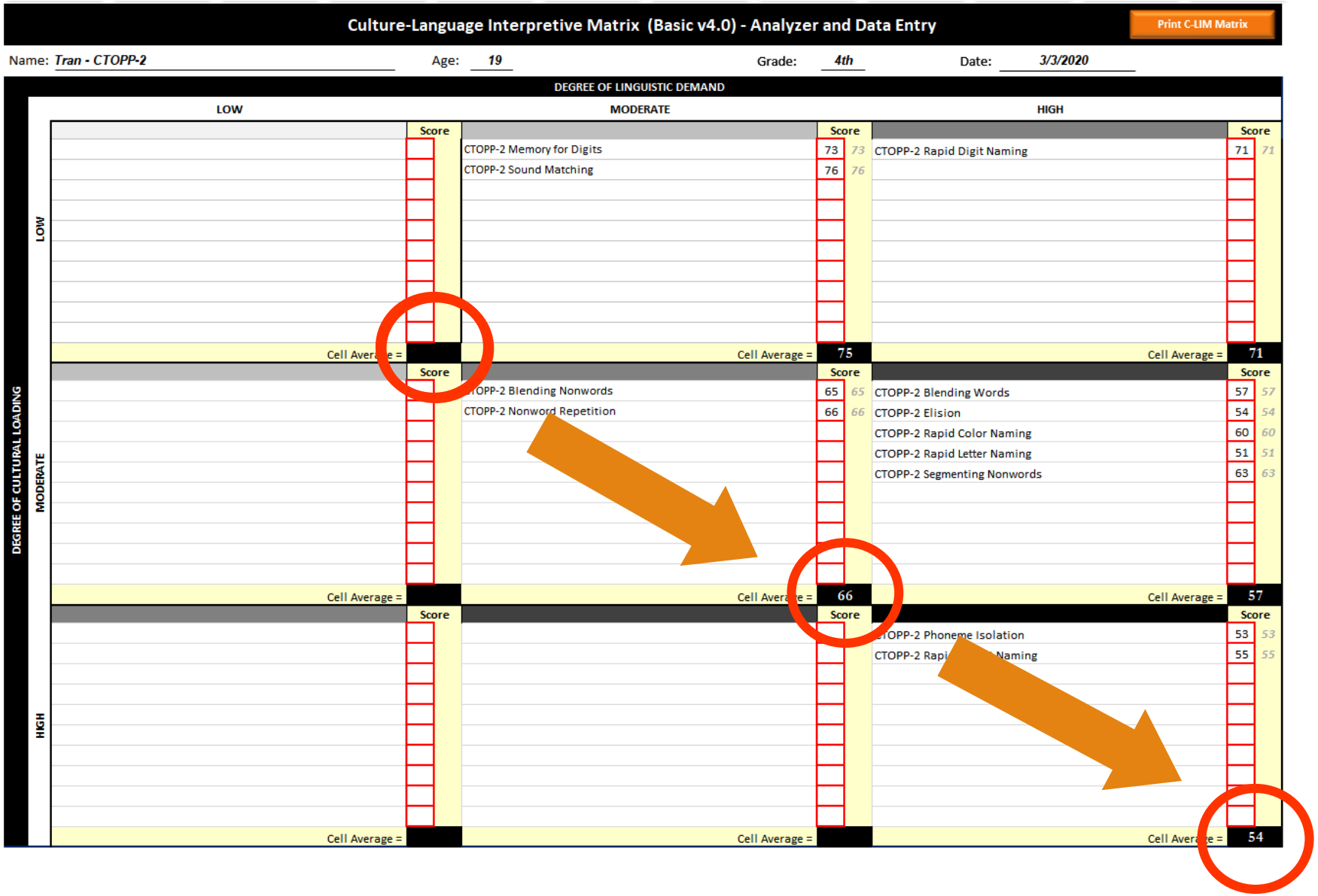


Important to note variability that may mask low scores.

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

C-LIM Additional Interpretive Issues

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



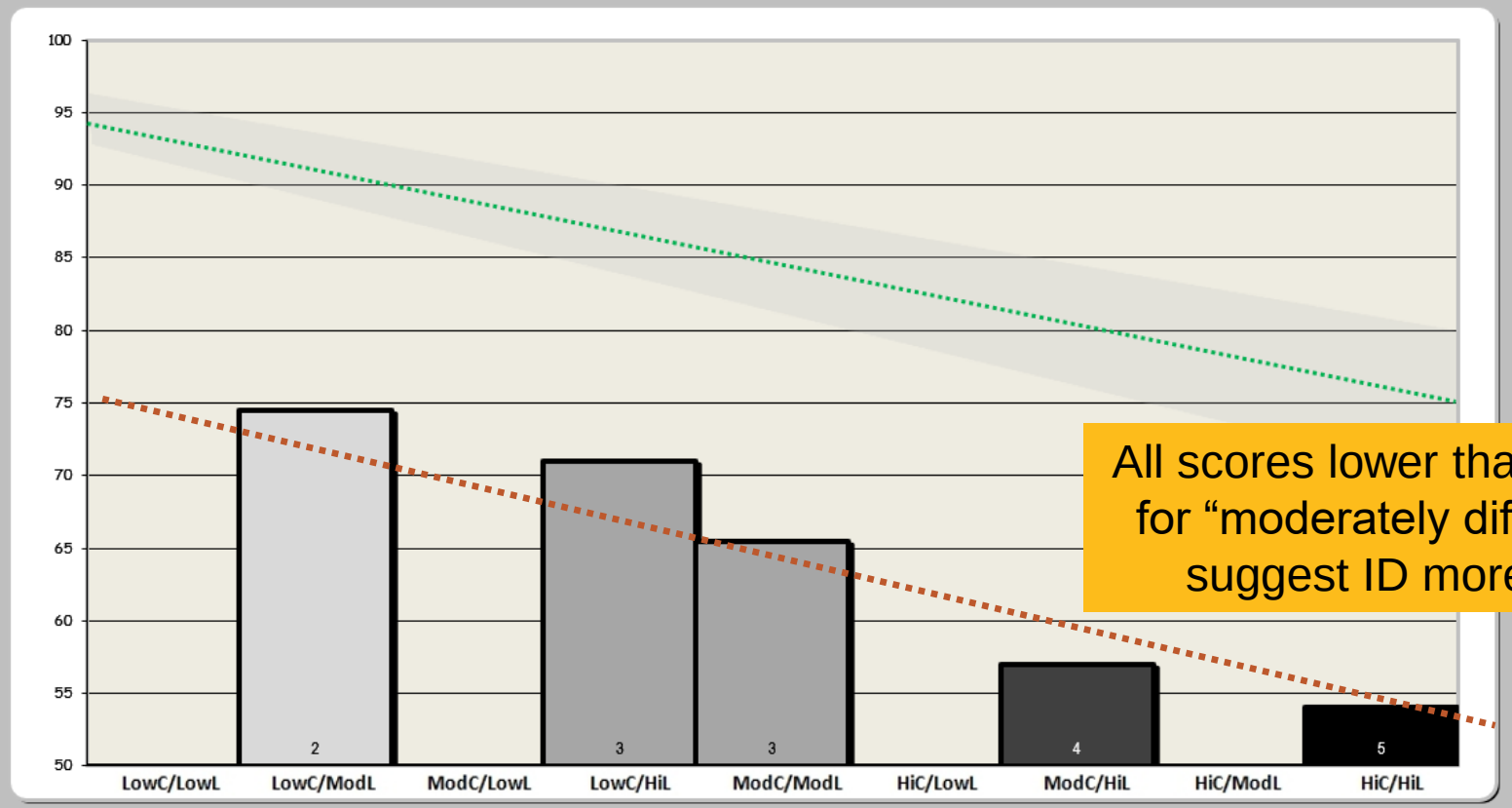
C-LIM Additional Interpretive Issues

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

Name: *Tran - CTOPP-2* Age: *19* Grade: *4th* Date: *3/3/2020*

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

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

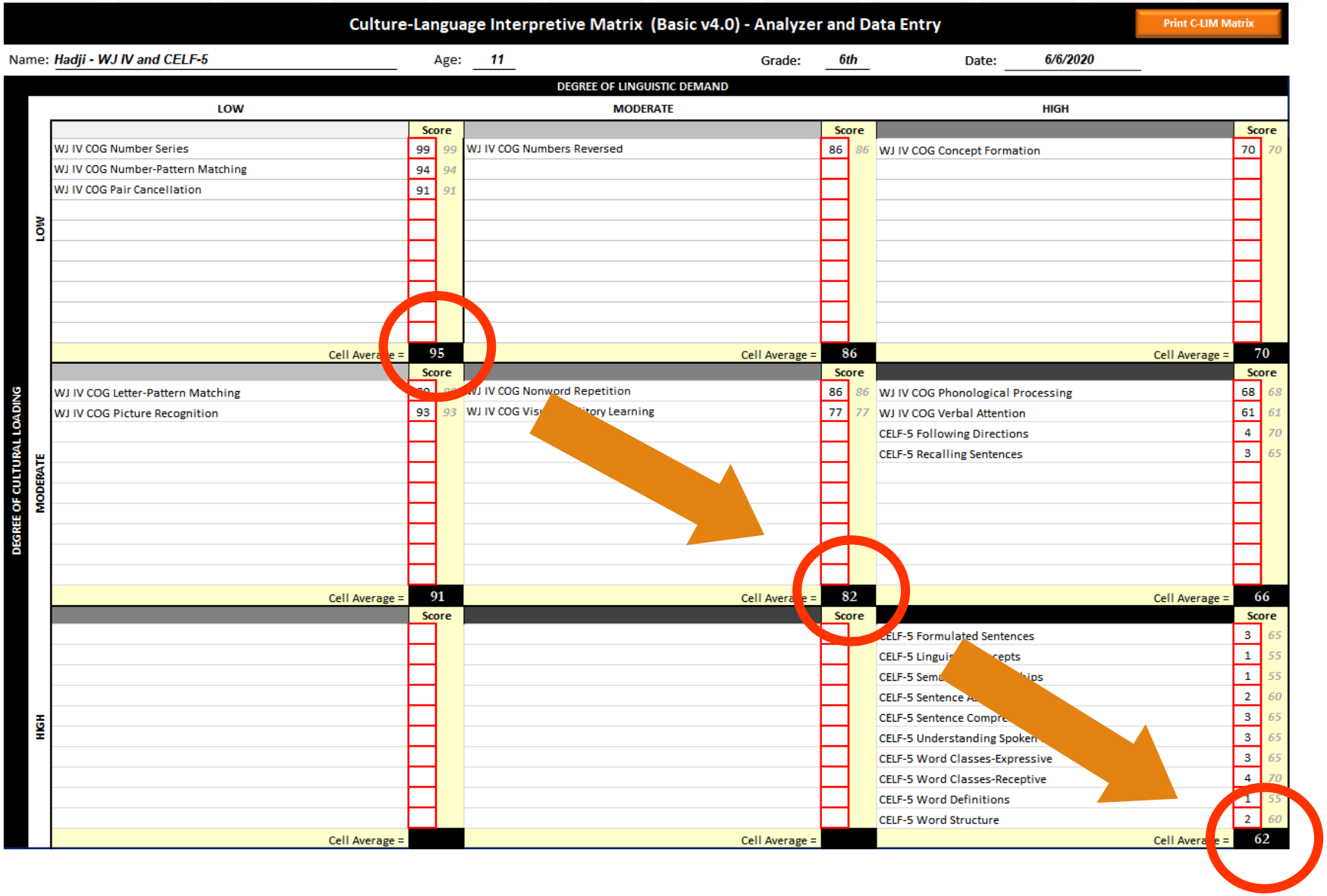


All scores lower than expected range for “moderately different” ELs. May suggest ID more so than SLI.

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

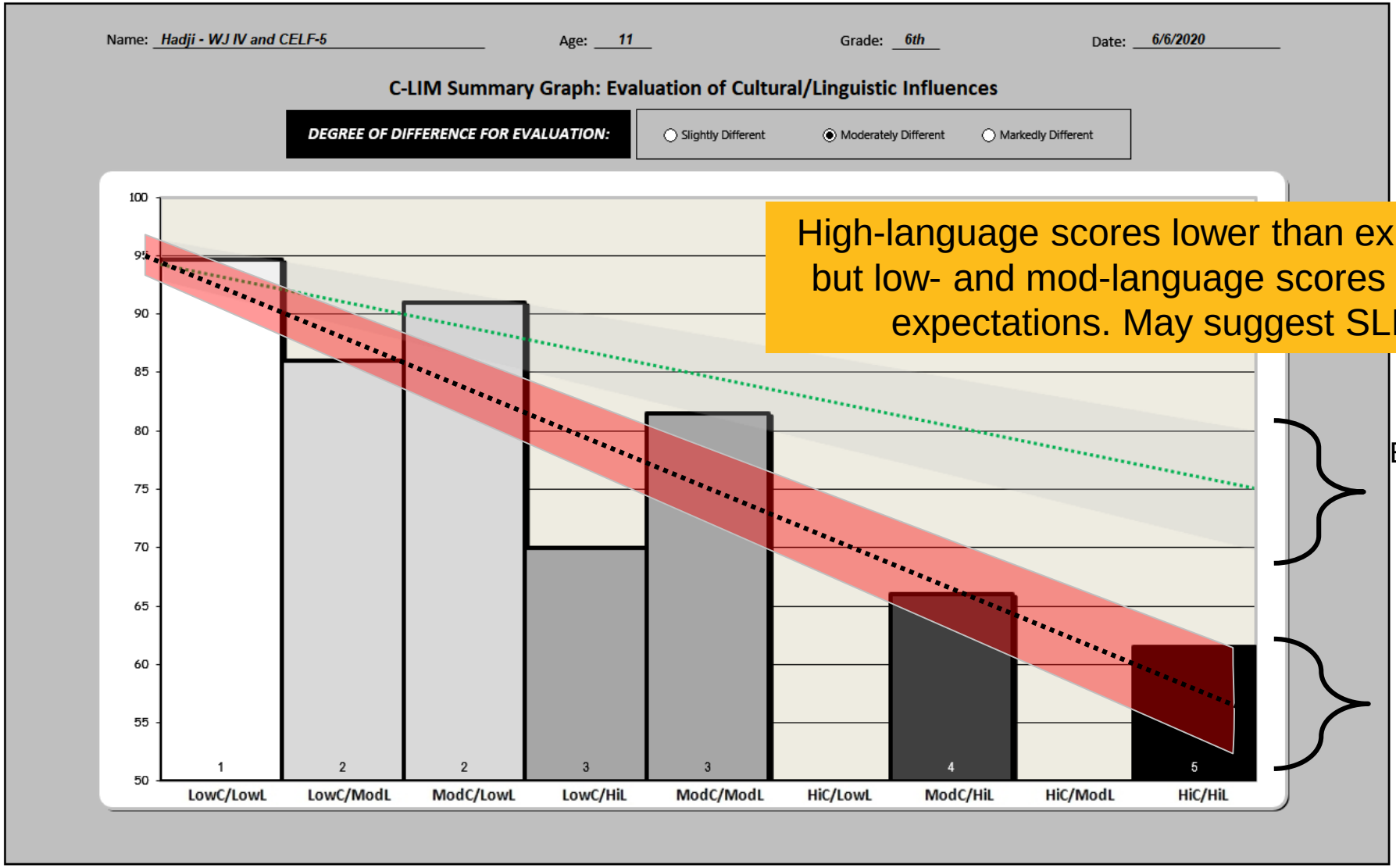
C-LIM Additional Interpretive Issues

General pattern of decline BUT not all scores within expected range



C-LIM Additional Interpretive Issues

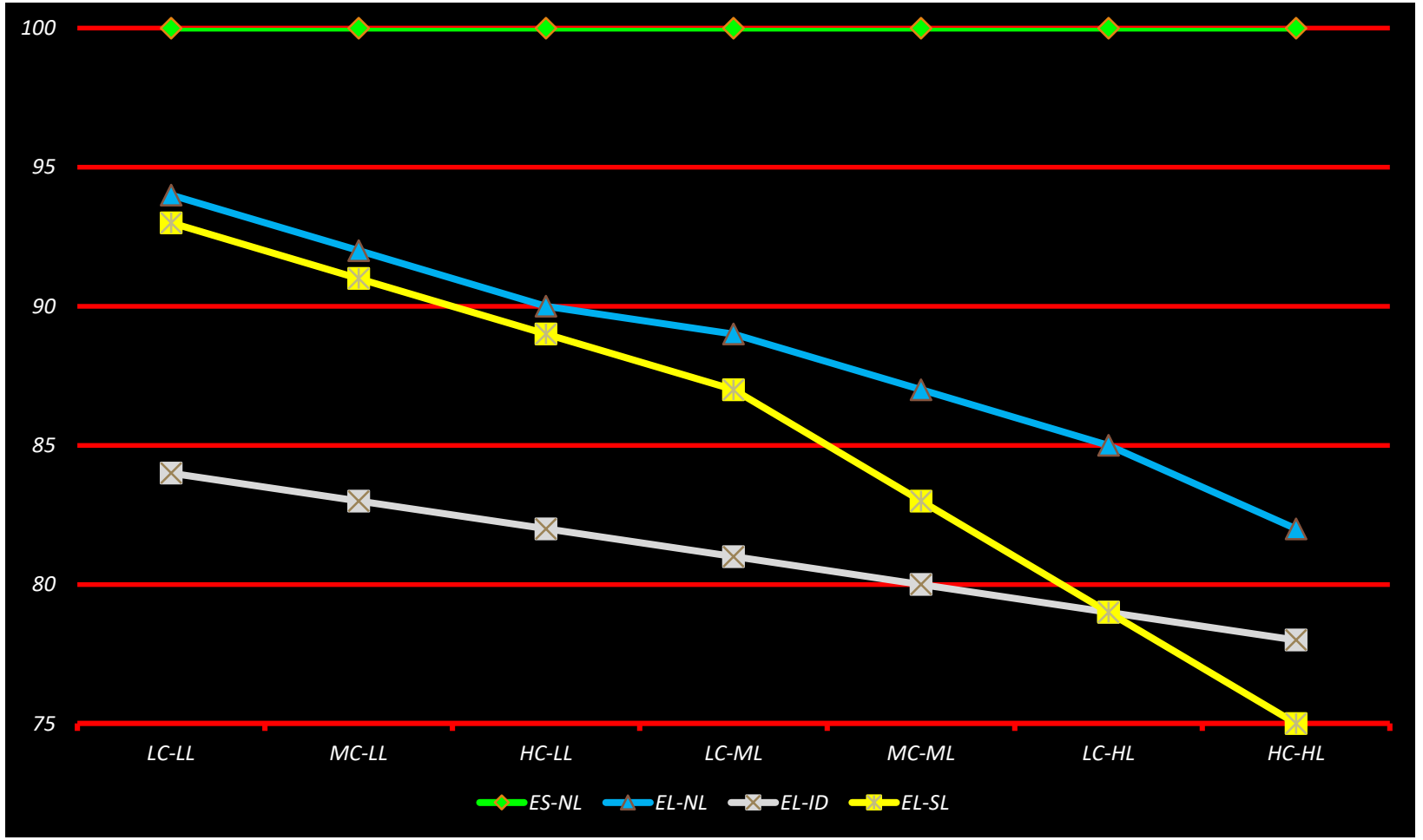
General pattern of decline BUT not all scores within expected range



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

C-LIM Additional Interpretive Issues

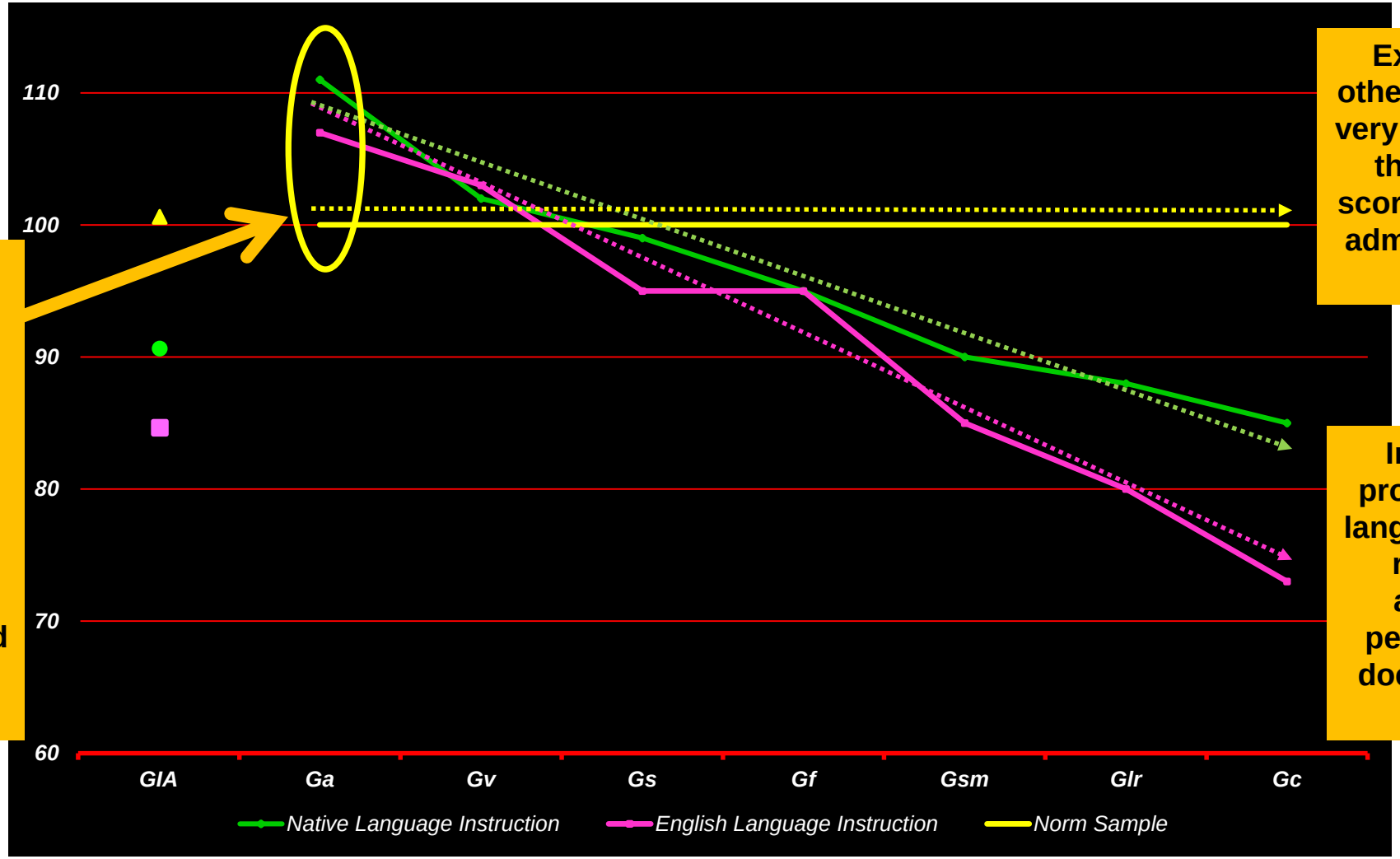
Mean C-LIM cell aggregates for WPPSI-III subtests arranged by degree of cultural loading and linguistic demand for ELs identified with language impairment, learning disability, and intellectual disability.



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

C-LIM Additional Interpretive Issues

Comparison of Bateria III Scores for EL's by Language of Instruction (English vs. Spanish)



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

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

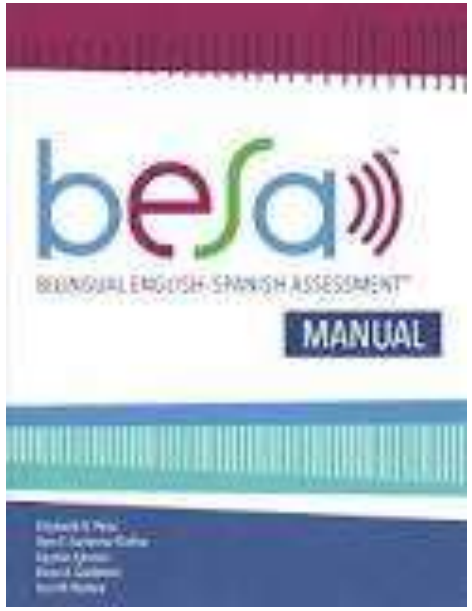
In addition, the provision of native language instruction results in less attenuation of performance than does instruction in English only.

Although limited, research does indicate that a similar pattern of decline also exists for ELs on tests administered in the native language which would also be evident in the C-LIM and is also moderated by language development differences among ELs.

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

The Bilingual English-Spanish Assessment:

Sampling bilinguals—categorical levels (3) of exposure



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

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

Table 5.2 Sample Distribution by Age and Language Exposure

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

The Ortiz Picture Vocabulary Acquisition Test

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

Author: Samuel O. Ortiz

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

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

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

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

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

Translating Research into Practice

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

An accessible, evidence-based approach to evaluation and testing with English learners must consider issues beyond test score validity and include attention to psychometric, practical, legal, and competency issues.

Additional Considerations in Conducting Evaluations and Testing with ELs

1. The usual purpose of testing is to identify deficits in ability (i.e., low scores).
2. Validity is more of a concern for low scores than average/higher scores because:
    - Test performances in the average range are NOT likely a chance finding and strongly suggests average ability (i.e., no deficits in ability)
    - Test performances that are below average MAY be a chance finding because of experiential or developmental differences and thus do not automatically confirm below average ability (i.e., possible deficits in ability)
3. Therefore, testing in one language only (English or native language) means that:
    - It can be determined that a student DOES NOT have a disability (i.e., if all scores are average or higher, they are very likely to be valid)
    - It CANNOT be determined if the student has a disability (i.e., low scores must be validated as true indicators of deficit ability)
4. Testing in both languages (English and native language) is necessary to determine disability.
    - Testing requires confirmation that deficits are not language-specific and exist in both languages (although low performance in both can result from other factors)
5. All low test scores, whether in English or the native language, must be validated.
    - Low scores from testing in English can be validated via research underlying the C-LIM
    - Low scores from testing in the native language cannot be validated with research

A Best Practice Framework for Comprehensive Evaluation of ELs: Multilingual Assessment (L1+L2)

Pre-referral Activities

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

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

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

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

Addresses concerns regarding fairness and equity in the assessment process

Post-referral Testing

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

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

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

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

Addresses possible bias in use of test scores

Decision Making

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

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

Multilingual Testing of ELs: Step by Step

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

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

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

- Conduct native language evaluation to the extent feasible based on availability of native language tests, professional competency, and utilize all options including translators/interpreters, modified/altered testing, nonverbal administration, dynamic assessment, error analysis, process assessment, etc., to generate both quantitative and qualitative evidence and data.
- If data indicate an area is a strength (i.e., average), then original L2 score is invalid, use the L1 score.
- If data indicate an area is still a weakness, then original L2 score is valid, use the L2 score/data.

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

- Use pre-referral data and any other case data and information to serve as the context by which to evaluate the test scores, ensure consistency in the findings, and provide defensible, causal explanations and conclusions supported by ecological validity.

Meeting the Standards for Fairness in Evaluation of ELs

Although there are no professional or legal standards that specify actual procedures for evaluation of English learners, there are consensus recommendations that provide some guidance in being able to document and establish that a given evaluation has been conducted in compliance with standards necessary to demonstrate and establish fairness. The following are standards that may be used to assess the extent of fairness and validity of any evaluation.

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

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

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

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

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

Meeting the Standards for Fairness in Evaluation of ELs

Used in conjunction with other information relevant to appropriate bilingual, cross-cultural, nondiscriminatory assessment including knowledge and information regarding…

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

…the framework presented here (along with the C-LIM and Ortiz PVAT) represents an evidence-based method for evaluating English learners and addressing the issue of test score validity. This process can assist **all practitioners** in decreasing the potential for biased and discriminatory interpretation by creating the ability to answer the most basic question in EL assessment:

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

Assessment and Related Resources

RESOURCES:

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

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

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

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

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

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

Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT) Jodi Kennis – jodi.kennis@mhs.com
https://www.mhs.com/ortizpvat

Additional References and Related Resources

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