

Equity in Assessment: Considerations for EL Students – Considering ALL of the Data

Tammy L. Stephens, PH.D.

1

Presenter

Tammy L. Stephens, Ph.D.

- » Senior Manager, Clinical Product Support & Training
- » Senior Clinical Assessment Consultant
- » University Professor
- » Educational Diagnostician
- » Special Education Teacher
- » Author of C-SEP
- » Creator of Beyond The Score Webinar Series

2

Agenda

- » Understanding Disproportionality
- » Contributors to Disproportionality
- » Data Referral Concerns – Underlying Issues
- » Importance of Collecting and Using Multiple Sources of Data in Decision-making
- » Importance of Considering and Ruling Out Exclusionary Factors
- » Importance of Investigating Language for All Students
- » Recommendations

3

4

Title VI of the Civil Rights Act (1964), 42 U.S.C. § 2000d et seq. and 34 C.F.R. pt. 100 provide that **no program or activity receiving Federal financial assistance from the Department of Education may discriminate on the basis of race or national origin.**

4

5

Defining Disproportionality
What is it?


5

6

So, What is Disproportionality?

The National Education Association (NEA) and the National Association of School Psychologists (NASP) define disproportionality as the **overrepresentation or underrepresentation of groups of people in special education services or gifted-talented programs by comparison to their representation in the total school population** (Peterson, 2019; Sullivan & Osher, 2019).

6



7

Contemplate Some Statistics

- » Nearly half of all students enrolled in special education are **students of color** (Fish, 2019b).
- » 30% of children with SLD also experience emotional and behavioral problems (Cristofani et al., 2023);
- » Students with emotional and behavioral disorders perform below grade level standards in literacy and math (Kern et al., 2019); half fail to meet expectations on standardized testing (Kern et al., 2019);
- » 54% of students with a disability drop out of high school (Carney, 2021);
- » Early diagnosis of SLD improves outcomes (Cristofani et al., 2023).

7




8

Some Statistics of Disproportionality

- » **Native Americans** are **four times** more often referred for developmental delays by comparison to other groups (Peterson, 2019).
- » **African Americans** are **twice** as likely to meet the requirements of special education services related to **emotional disturbance (ED)** and **intellectual disability (ID)** by comparison to other groups (Grindal et al., 2019; Peterson, 2019)

8



10

Earliest Documentation of Overrepresentation

- » In 1968, Lloyd Dunn was among the first to realize that there was a **disproportionate number of minority students** in special education classrooms.
- » According to Dunn, about **60-80%** of special education students with mental retardation came from "**low status background**" which mainly meant they were Latino, African American or Native American

10

12

A History Lesson

Let's examine some reasons for why disproportionality may exist.

12

15

Engagement in Poor Assessment Practices- Resulting in Overrepresentation

- » Overreliance on IQ scores
- » Overreliance on Standard Scores
- » Overreliance on cut-scores
- » Limited focus on triangulating all the data
- » Failure to consider and rule out exclusionary factors

15

16

Overrepresentation

Questions the efficacy of our professional practices, our methods of testing, and our take on honoring diversity.

16

17

Poor Assessment Practices

Traditional assessment practices have been identified as one of the major culprits to overidentification and underidentification.

17

18

A History of Overrepresentation

- » Around 1976, the OCR (Office of Civil Rights) in Ohio started collecting data on the educational system. Their results further cemented the idea of overrepresentation.
- » During the period of 1976-1977, **black students** were placed in **EMR** (educable mentally retarded) classes **3.4 times** as much as white students.
- » Data from 1978-1979 indicated the ratio increased even further, to **3.5**.
- » They also found that a lot of **limited English proficiency** students were placed in special education programs **without proper assessment** simply because they weren't good at English.

Source: [Overrepresentation of minority students in special education – Vernex Cognition \(vernex-cognition.com\)](http://www.vernex-cognition.com)

18

19

A History of Overrepresentation

- » The OCR dug even further, specifically into 148 school districts, over the period of **1975-1979**
- » Some students were being assigned **EMR classes** without examination for visual/ auditory problems; some based on **outdated IQ scores**; and some were assigned these classes despite having **IQ scores that surpass the EMR range**
- » At the time, being a minority student was enough to earn placement in special education classes.

Source: [Overrepresentation of minority students in special education – Vernex Cognition \(vernex-cognition.com\)](http://www.vernex-cognition.com)

19

22

The Role of Implicit Bias & Stereotypes

- » *"the disability-cultural diversity analog"*
- » We couple things together in our mind like bed and sleep, food and drink
- » People have subconsciously learned to link cultural diversity to disability.
- » It's a stereotype that has become so embedded in our thoughts that even some members of minority groups share this preconception.
- » **Different ethnicity, race, gender, language or social class shouldn't equal disability.**

Source: [Overrepresentation of minority students in special education – Verney Cognition \(verneycognition.com\)](#)

22

23

Historical Causes of Overrepresentation

- » The first aspect is **litigation**. Before the passage of public law 94-142 in 1975, litigation was mainly concerned with protecting minority students against **unfair placement based on inadequate assessment methods**, such as the cases of Diana (1970) and Guadalupe (1972).
- » After 1975, litigation was more concerned with **defining mild retardation and the fairness of intelligence tests**. A lot of money, time and effort were spent to reevaluate children with mild mental retardation to make sure they weren't unfairly placed. All these litigation cases made people question a lot of things like what exactly was the definition of mental retardation? What's "intelligence"? How can the diagnosis of mental retardation be dependent on something like IQ scores if the IQ cutoff point is constantly changing from time to time?
- » Most of these tests **weren't actually measuring intelligence** but were in fact, measuring how much of the dominant culture the student had accumulated. That was definitely unfair for students from ethnic backgrounds.

Source: [Overrepresentation of minority students in special education – Verney Cognition \(verneycognition.com\)](#)

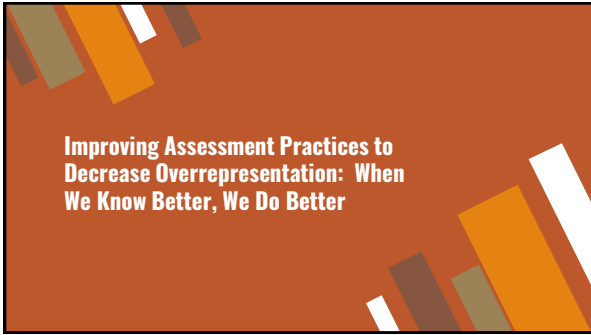
23

24

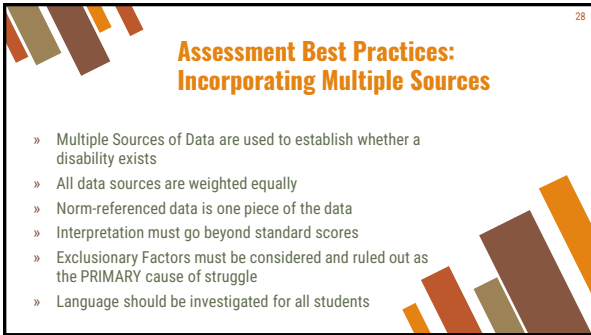
Other Factors that Contribute

- » Our education system is influenced by our social, cultural, and historic experiences and may need to be revised and updated (Cavendish et al., 2018; Sullivan & Osher, 2019; Tefera & Fischman, 2020).
- » **Poor data collection during the pre-referral and referral process resulting in inappropriate referrals to special education.**
- » **Systemic inequalities** exist in some areas for certain races, classes, genders, etc. (Biddanda et al., 2018; Fish, 2019b; Grindal et al., 2019).
- » Some school districts or staff may be influenced by implicit or explicit bias (Carney, 2021; Grant, 2020).

24



27



28




29

33

Data-Based Referral Decisions

- » Referral decisions must be made based on all the data collected prior to and part of the referral process
- » When important data is lacking, we do not obtain a complete picture of what the student can and cannot do
- » This leads to inappropriate referrals and decisions

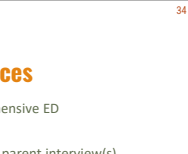


33

34

Poor Data Collection Practices

- » Five data sources are recommended for a comprehensive ED evaluation:
 - Classroom observations, teacher interview(s), parent interview(s), student interview, and normative data from rating scales completed by at least two different informants.
- » One study, however, shows that only 28% of school psychologists consistently include all 5 sources and nearly 30% include only four of the five sources (Allen & Hanchon, 2013).
- » Sadly, 5% do not consistently include any of the critical data sources listed; and 13% only consistently include one of the five.



34

35

Gleaning Insight from Such Research

- » Some in our profession are failing to collect enough relevant data to make legally defensible decisions.
 - Poor referral process
- » Collectively, we are inconsistent in our interpretation and recommendations.
- » These weaknesses are concerning in a profession that relies heavily on data collection and interpretation.
- » These MAY also attribute slightly to disproportionality

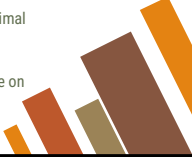


35

36

Poor Data Collection = Poor Decisions

- » Leads to students being referred who shouldn't be referred
- » Students who should be referred are not – resulting in ongoing academic struggles and sometime manifestations of behavioral issues
- » Leads to inaccurate and poor decisions based on minimal data
- » Once referral is made, lack of data that should be integrated into the assessment results on overreliance on NRT scores



36

Legal Mandates – Assessment

- » *May not use any single measure or assessment as the sole criterion* for determining whether a child is a child with a disability and for determining an appropriate educational program for the child.
- » Use a *variety of assessment tools and strategies* to gather relevant functional, developmental, and academic information about the child, including information provided by the parent, that may assist in determining whether a child has a disability; and use it for individualized educational planning.


Norm-Referenced Standardized Tests are One Source of Data

37

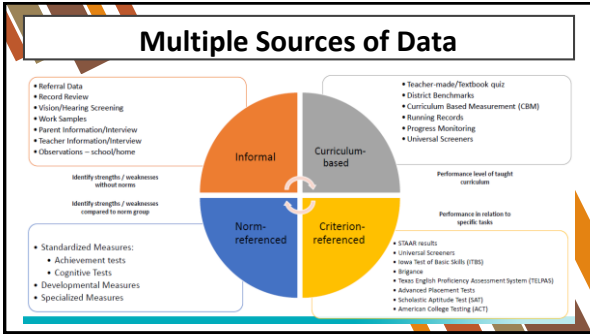
38

Rethinking the Use of Norm-Referenced Standardized Tests

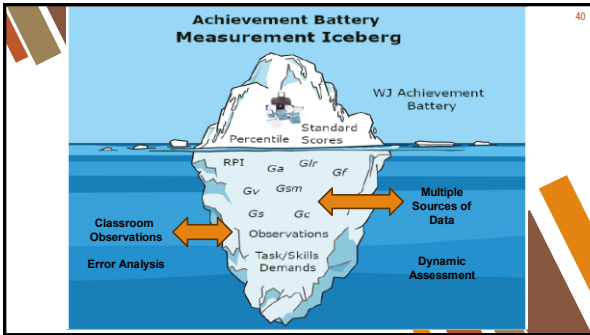
One piece of the data about the student



38



39



40

Considering Student's Performance Beyond a Standard Score

- » Standard score are only one data point obtained from an NRT
- » Standard scores are **ordinal measures** and only indicate an **individual's place in line** (normal curve)
- » Standard scores **DO NOT** indicate skill proficiency or functioning
- » One-time **snapshot** of student's performance
- » Must consider student's behaviors/strategies when completing tasks

41

41

Historical Uses of Norm-Referenced Standardized Tests

A focus on interpreting student's performance through standard scores

42

Norm-Referenced Tests Provide Us With an Abundance of Information about Each Student – Don't Leave Data Behind

43



LEVELS OF INFORMATION PROVIDED BY THE WJ IV AND/OR OTHER NORM-REFERENCED TESTS

Level 1	Level 2	Level 3	Level 4
Qualitative, Informal, error analysis Useful for instructional planning • Test Session Observations Checklist • Useful for behavioral observations	Level of Development Age Equivalent • Level of Instruction • Grade Equivalent	Level of Proficiency Relative Proficiency Index, CALP • Easy to Difficult Range • Developmental/Instructional Zone	Relative Standing in Group Standard Scores • Rank Order • Percentile Ranks • Significantly high or low scoring • Discrepancy PR, SD

44

44

Level 1: Qualitative Data
Important when interpreting student's performance on tasks.

45

Level	Type of Information	Basis	Information and Scores	Uses
1	Qualitative (Criterion Referenced)—Describes content or supports a clinical hypothesis	Observations during testing and analysis of responses	<ul style="list-style-type: none"> Description of the individual's behavior during testing Patterns of errors and correct responses within specific tasks Strategies (correct or erroneous) used to perform specific tasks 	<ul style="list-style-type: none"> Consideration of the possible effect of the individual's behavior on the obtained test scores Prediction of the individual's behavior and reactions in instructional situations Analysis of an individual's strengths, misunderstandings, and limitations regarding specific academic skills, procedures, knowledge, and cognitive abilities

Level 1: Qualitative Data

46

Take a lot of notes on the test record during the testing session/sessions

47



LEVEL 1: Qualitative Data - Example

Observations made during the testing session specific to behaviors

- » Tapping pencil
- » Excessive movement/fidgety behaviors
- » Out of seat
- » Frequently asks for items to be repeated
- » Frequently requires redirection

48



LEVEL 1: Qualitative Data - Example

Observations made during the testing session specific to task performance

WJ IV Word Attack

- » Child initially sounds out each letter in the word or chunks the parts of the word several times, but then after a bit of wait time, the child pronounces the word correctly.

Math Applied Problems

- » Child constantly asks that items be repeated.
- » Child works problems in his head or verbally talks through the problem
- » Counts on fingers
- » Grips pencil awkwardly when writing

49

Level 2: Level of Development

53

LEVEL	TYPE OF MEASUREMENT	SCORE	CONSTRUCTION AND USE	NOTE
2	Level of Development (Norm Referenced)—Indicates an individual's level of development, such as age or grade equivalents	Sum of item scores Age or grade level in the norming sample at which the median score is the same as the individual's score	<ul style="list-style-type: none"> Raw score Test or cluster <i>W</i> score Age equivalent (AE) Grade equivalent (GE) 	<ul style="list-style-type: none"> Reporting an individual's general level of development in a skill, ability, or area of knowledge compared with others of the same age or in the same grade in the norming sample Monitoring an individual's progress within a specific skill or ability Basis for describing the implications of developmental strengths and weaknesses Basis for initial recommendations regarding instructional level and materials

Level 2: Level of Development

54



The Most Misunderstood Scores of All...

Age/Grade Equivalents

55

Age/Grade Equivalent – Interpretation	
»	GE reflects the examinee's performance in terms of the grade level in the norming sample at which the average score is the same as the examinee's score <ul style="list-style-type: none"> ◊ If average raw score for students in grade 2 (the 6th month) is 14, then any examinee who scored 14 would receive 2.6 as a grade equivalent score
»	GE 2.6 ≠ Your student is reading at the mid-second grade level.
»	GE 2.6 = On reading tasks, your student is performing the same as the average student in the norm sample who is in the 2 nd grade, 6 th month

56

Remove from your vocabulary (and reports):

According to the WJ IV Achievement, *Student* is performing at the ____ grade level.

57

Level 3: Proficiency (Criterion-Referenced)

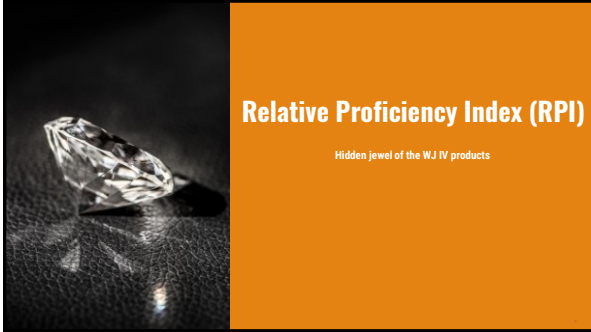
Proficiency and Functioning

58

Level 3: Proficiency (Criterion-Referenced)

3	Proficiency (Criterion-Referenced)—indicates the quality of performance on criterion tasks of a given difficulty level	Distance of an individual's score on the W scale from an age or grade reference point	<ul style="list-style-type: none"> Quality of performance on assessed skills and abilities compared to that of age or grade peers in the norming sample Test or cluster <i>W</i> difference (WDIF) Relative proficiency index (RPI) Cognitive-academic language proficiency (CALP) level Institutional or developmental zone 	<ul style="list-style-type: none"> Degree of proficiency on tasks mastered by average age or grade peers Developmental level at which the individual will perceive typical tasks to be easy, mildly challenging, or very difficult Placement decisions based on a criterion of significantly strong or weak proficiency Prediction of performance with similar task
---	--	---	---	---


59



Relative Proficiency Index (RPI)

Hidden jewel of the WJ IV products

60



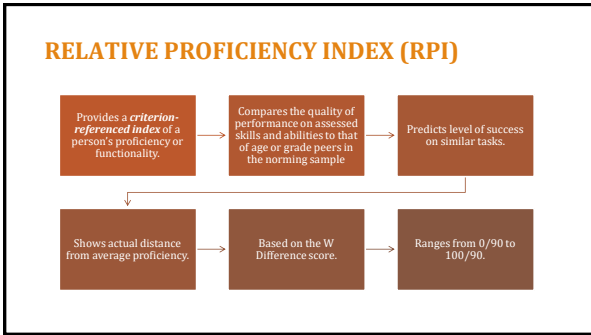
THE MANY USES OF THE RPI

The RPI has several important uses based on the information it provides. These uses include:


- Informing users about how difficult an examinee will find age/grade appropriate tasks
- Describing the quality of an examinee's performance on tasks
- Offering criterion-referenced information
- Helping to monitor progress
- Indicating where on the range of development or instruction the examinee falls ("Developmental Zone")

61

61



62



RELATIVE PROFICIENCY INDEX (RPI)

RPI scores are represented as fractions (e.g., 75/90)

- The **numerator** represents the examinee's predicted proficiency if given similar tasks
- The **denominator** is fixed at 90, indicating the proficiency on average same-age or same-grade peers
- For example, if an examinee obtains an RPI of 75/90 on Test 8: Oral Reading, it indicates that the examinee was 75% successful on an oral reading task that average people at the examinee's same age or grade reference group would perform with 90% success.

63

RELATIVE PROFICIENCY INDEX (RPI)

Reflects the individual's proficiency on tasks that the average age or grade mate would have 90% proficiency.

↓

Examples:

When average grade mates would have 90% success in spelling, Sandy is predicted to have only 4% success (RPI = 4/90). Her proficiency on spelling tests would be very limited.

Bennett's RPI of 98/90 on the Math Problem Solving cluster indicates his performance would be very advanced compared to his grade peers.

64

W Difference Values	Reported RPI	Proficiency	Implications
+31 and above	100/90	Very Advanced	Extremely Easy
+14 to +30	98/90 to 100/90	Advanced	Very Easy
+7 to +13	95/90 to 98/90	Average to Advanced	Easy
-6 to +6	82/90 to 95/90	Average	Manageable
-13 to -7	67/90 to 82/90	Limited to Average	Difficult
-30 to -14	24/90 to 67/90	Limited	Very Difficult
-50 to -31	3/90 to 24/90	Very Limited	Extremely Difficult
-51 & below	0/90 to 3/90	Extremely Limited	Nearly Impossible

INTERPRETATION OF RPI SCORES

65

RPI and Instructional Zone

The instructional zone is a special application of the RPI score.

It is based on a range along a developmental scale that indicates and examinee's present level of functioning.

It ranges from easy (the Independent Instructional level to difficult (the Frustration Instructional level)

RPI	Instructional Level
96/90 to 100/90	Independent
76/90 to 95/90	Instructional
75/90 & below	Frustration

66

RPI and Instructional Zone

- An examinee with an RPI of 80/90 is expected to be at the instructional level and should find similar tasks developmentally appropriate.
- An examinee with an RPI of 60/90 is expected to demonstrate frustration on similar tasks when compared to average same-age or same-grade peers. Similar tasks are expected to be developmentally challenging.

RPI	Instructional Level
96/90 to 100/90	Independent
76/90 to 95/90	Instructional
75/90 & below	Frustration

67

Level 4: Relative Standing in a Group

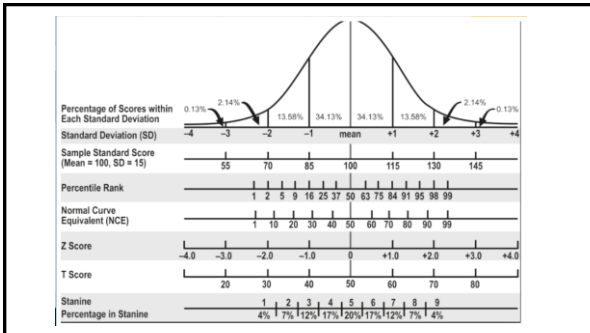
The Position or "Place in Line" of the student's performance in relation to the normative sample (placement on the normal curve)

68

4	<p>Relative Standing in a Group (Norm Referenced)—Provides a basis for making peer comparisons (percentile ranks or standard scores)</p>	<p>Relative position (A transformation of a difference score, such as dividing it by the standard deviation of the reference group)</p>	<ul style="list-style-type: none"> Rank order Standard score (SS) (including T score, z score, NCE, discrepancy SD/DF) Percentile rank (PR) (including discrepancy PR) 	<ul style="list-style-type: none"> Statement of the relative (ordinal) position of an individual's score, based on the standard deviation (SD), within the range of scores obtained by age or grade peers in the norming sample Placement decisions based on a criterion of significantly high or low standing in a group
---	---	--	---	---

Level 4: Relative Standing in a Group

69



70


Standard Scores are Not Equivalent to Functioning

The fundamental **misunderstanding** and **common interpretive error** with **standard scores** being equivalent to **functioning** or **performance**. This leads to faulty generalizations.

For example, a standard score of 90 on a memory test could be **Misinterpreted** to mean the student has "average" functioning in memory when in fact a more accurate description of this score is that it **represents an individual's relative position** or "place" in line as it is ordinal data (Jaffe, 2009; Adeyemi, 2010).

71

Looking beyond the standard score and considering the student's performance through other lenses can provide a richer understanding of the learner



Integration of other data sources is also mandatory. Test scores should never be interpreted in isolation

72

72

PROFICIENCY VS. POSITION

- At times, the proficiency information provides insights into performance that are not revealed by standard scores or percentile ranks.
- If only the standard score is considered, we may miss important information about the student's functioning.

READING FLUENCY	473	8-3		SS	92 (86-97)	92	29
Oral Reading	484	8-2			94 (89-98)	94	33
Sentence Reading Fluency	461	8-3			92 (85-99)	92	29

73

PROFICIENCY VS. POSITION

- At times, the proficiency information provides insights into performance that are not revealed by standard scores or percentile ranks.
- If only the standard score is considered, we may miss important information about the student's functioning.

READING FLUENCY	473	8-3	R/FI	63/90	Limited	SS	92 (86-97)	92	29
Oral Reading	484	8-2		78/90	Limited to Average		94 (89-98)	94	33
Sentence Reading Fluency	461	8-3		45/90	Limited		92 (85-99)	92	29

74

Language Type	BICS (Social Language)	CALP (Academic Language)
Description	<ul style="list-style-type: none"> • Everyday use • Social interaction • Less specialized • Less cognitive demanding • Used in a social setting • Requires an understanding of cultural and social norms, including nonverbal cues 	<ul style="list-style-type: none"> • Academic • Used in the classroom for reading and writing tasks • More cognitive demanding due to complex vocabulary and grammar structures • Requires specialized knowledge
Acquisition	<ul style="list-style-type: none"> • Can develop in 3-5 years 	<ul style="list-style-type: none"> • Proficiency can develop in minimum of 4-7 years
Example	<ul style="list-style-type: none"> • Engage in an informal, face-to-face conversation • Writing a social media post • Texting • Reading a menu 	<ul style="list-style-type: none"> • Defining a scientific term • Explaining how to solve a math problem • Comparing and contrasting art • Summarizing a research paper

79

CALP Sci

Table A-2
 2021 Update for Safety
 CALP Level

Reading Domain	Writing Domain	Open-Ended Domain
Reading	Writing	Open-Ended
Non-Reading Skills	Basic Writing Skills	Academic Applications
Reading Comprehension	Writing Comprehension	Academic Knowledge
Reading Comprehension Content		Basic Knowledge

Table A-3
 CALP Levels and
 Corresponding Specifications

CALP Level	WPI Difference	WPI	Subtest(s)	
6	+12 and above	9000	Literacy very	
5	Advanced	8000 to 9000	Very easy	
4.5 (4.5)	Fluent to Advanced	+7 to +12	8000 to 9000	Easy
4	Fluent	+5 to +7	6000 to 8000	Strategic
3.5 (3.5)	Linked to Fluent	+3 to +5	4000 to 6000	Effort
3	Linked	+1 to +3	2000 to 4000	Very effort
2	Very Limited	+0 to +1	000 to 2000	Literacy effort
1	Extremely Limited	-1 and below	000 to 000	Heavily supportive

80

An Important Component of a Comprehensive Assessment & Assurance of Adequate Identification Requires Investigating Exclusionary Factors

81

88

Definition of SLD

Specific Learning Disability:

Means a **DISORDER** in one or more of the basic psychological processes involved in **understanding** or in using **LANGUAGE**, spoken or written, that may manifest itself in the imperfect ability to **listen, think, speak**, read, write, spell, or to do mathematical calculations... 34 CFR,300.8 (c) (10)

88

Importance of Language Development

89

90

Language and Reading

Theories of Reading

Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990)

Word Recognition × Language Comprehension

↓

Reading Comprehension

90

91

Language and Reading

Theories of Reading
Rope Model of Reading (Hollis S. Scarborough, 2001)

FIGURE 1. THE MANY STRANDS THAT ARE WOVEN INTO SKILLED READING*

Language comprehension

Word recognition

91

92

The Language Literacy Network

The many language components that unify into skilled reading and writing (Wolowicz, 2021)

READING

WRITING

The speech-to-print advantage

92

93

Language & Early Math Learning

- » Math learning starts with concepts of quantity, size and comparisons, and the words that represent numbers. According to Mazzocco and Thomas (2005), math learning starts with counting physical objects with parents or caretakers, understanding concepts of greater than and less than, full and empty by playing with food or toys, and general ideas of mass with bigger or smaller. It's important to note that none of these math skills require numerals, instead, they require language. The learning of early math skills is based on **creating a connection between language and physical objects**. Examples include:
 - » A child being asked if they would like more snacks.
 - » A parent counting a young child's toes.
 - » A child presented with a big toy car and a small toy car and asked, "Which toy car is the biggest?"

93

94

Language & Math Word Problems

» Connections between language and symbolic representations of numerals and math operations is key in understanding a word problem. Examples of ways language deficits may impact student's performance on word problems include:

- Lack of understanding what operation or operations (e.g., addition, subtraction, multiplication, or division) are represented by the language of the problem, when it is not explicitly stated.
- Lacking language skills to identify past, present, future, possession, direction, pronouns, and verbs used within the context of the word problem.
- **Example:** *Jack had two apples, he ate one, he plans to buy another tomorrow morning. How many apples will Jack have tomorrow?*
- Inability to link characters in word problem to pronoun usage (he, she, they).
- The concept of "less than" is presented in many ways (e.g., smaller than, fewer than, lower than) to indicate one quantity is less than another.

94

Impact of Low SES on Student's Language

95

Classroom Language Demands Observation

Classroom Language Demands Observation

Instruction: Record words used in instruction (by teacher, peer, or video) in the appropriate column during the observation. Ask target student what each word means. Record % for each column.

Common Words (e.g., today, clock, time)	Academic Words (e.g., summarize, transfer, variable, independent)	Content Specific Words (e.g., nucleus, hypotenuse, isosceles)
% Correct	% Correct	% Correct

107

108

Recommendations

Manage disproportionality through these best practices as extracted from the literature.

108

109

When Conducting Evaluations

- (1) Be aware of bias;
- (2) Recognize that statistics are not neutral;
- (3) Appreciate that categories are neither "natural" nor given;
- (4) Provide clear voice and insight to findings (since data cannot "speak for itself"), and;
- (5) Pursue equity in your work (see also Biddanda et al., 2018; Blanchard et al., 2021).

Gillborn, Warmington, & Demack (2018)

109

120

Make Multiple-Sources-of-Data (MSD) Decisions

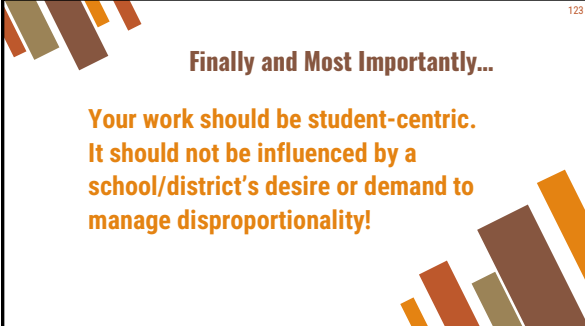
- » Always make decisions based on an aggregation of the data (Biddanda et al., 2018).
- » Use MSD to properly identify eligibility, areas of concern, and recommend services and/or accommodations (Blanchard et al., 2021; Kern et al., 2019).
 - Check for bias (yours and others) when reviewing the data and making decisions (as a collective) (Blanchard et al., 2021).
 - Use the data collected to identify the characteristics or deficits and then design accommodations that suit those (Kern et al., 2019).

120

123

Finally and Most Importantly...

**Your work should be student-centric.
It should not be influenced by a
school/district's desire or demand to
manage disproportionality!**



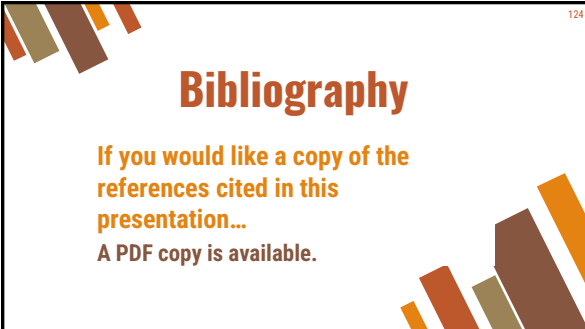
123

124

Bibliography

**If you would like a copy of the
references cited in this
presentation...**

A PDF copy is available.



124



Tammy.Stephens@RiversideInsights.com

125
