

Understanding Psycho-Educational Evaluations

Tennessee Administrative Law Judge Training

Sara Rich, PhD, NCSP

DECEMBER 1, 2023

1


OBJECTIVES

- Learn a framework for reviewing assessment reports.
- Develop an understanding of important psychometric properties when reviewing assessment data.
- Answer common questions from hearing officers

2

Purpose of Evaluation


Identification, Needs, & Programming



3


Purpose of Evaluation According to IDEA

- To determine if the child is a “**child with a disability**,” as defined by IDEA
- To gather information that will help determine the **child’s educational needs**
- To guide decision making about appropriate **educational programming** for the child




1. 34 CFR § 300.8

4

Medical Mental Health vs Educational “Diagnosis”	DSM (Medical/ Mental Health)	IDEA (Educational)
	<ul style="list-style-type: none"> • Diagnosis. • Meets DSM criteria for disorder • Licensed medical professional: MD, DO, PhD (psychologist) • IEP ≠ Medical Diagnosis  	<ul style="list-style-type: none"> • Categorization for services • Meets criteria for disability (medical or educational) AND has an educational need • IEP Team with a qualified examiner (i.e., school psychologist) • Medical diagnosis ≠ IEP

5




Outside Diagnosis

- DSM-V-TR Codes versus ICD-10 Codes
- Psychiatrist versus Pediatrician versus Psychologist
- (Neuro)Psychological Testing versus Psycho-educational Testing

6

Assessment Procedures

Review data, collect new data, synthesize data,
and report data



7

General Assessment Procedures

-  Review Existing Data
-  Collect New Data
-  Synthesize Data
-  Report Data




8

Full Evaluation Components

"A State educational agency, other State agency, or local educational agency shall conduct a **full and individual initial evaluation** in accordance with this paragraph and subsection (b), before the initial provision of special education and related services to a child with a disability under this subchapter..."

- In conducting the evaluation, the local educational agency shall—
 - **Use a variety of assessment tools** and strategies to gather relevant functional, developmental, and academic information, including information provided by the parent, that may assist in determining—
 - **Not use any single measure or assessment** as the sole criterion for determining whether a child is a child with a disability or determining an appropriate educational program for the child; and
 - **Use technically sound instruments** that may assess the relative contribution of cognitive and behavioral factors, in addition to physical or developmental factors.

34 CFR § 300.304(b).



9


Review, Interview, Observe, Test (RIOT)

- **Record Review:** Review of educational records, background information, discipline records medical information
- **Interview:** Interviews with relevant individuals, including parents/guardians, teachers, and students
- **Observations:** Many different types of observations. Should produce data on frequency, intensity, and context of the problem. Also includes observations to determine validity of assessment results
- **Testing**
 - Rating forms: Measures perception of experiences, not direct behaviors.
 - Standardized assessments: Either direct measurement of skills or measurement of theoretically relevant skills.

10

Tennessee Documentation of Comprehensive Assessments

- The TN Department of Education has provided clear guidelines and checklist for what must be considered in regards to a comprehensive evaluation for each special education category.
- More information can found at the [TN Special Education Evaluation & Eligibility](#) website.




11

Example Assessment

Jacob

- 3rd grade student
- **Reason for referral:** during his speech language evaluation, the examiner noted characteristics of autism, including poor eye contact, repetitive motor movements, and echolalia
- **Evaluation questions:**
 - What are Jacob's currently cognitive processing strengths and weaknesses?
 - What are Jacob's academic skills relative to grade-level language arts and math?
 - How do Jacob's social skills, behaviors, and classroom adaptive skills impact his academic achievement?
 - Do Jacob's current Special Education services (speech-language services) meet his educational needs?




12

Example Assessment-2


Jacob

- **Assessment Procedures:**
 - Review cumulative school records and grades
 - Observe Jacob in math class
 - Conduct Differential Ability Scales (DAS-2)- (IQ Test)
 - Parents and Teachers complete Gilliam Autism Rating Scale (GARS)
- **Results**
 - Grades are average, no deficits in academic performance noted
 - Jacob was on-task during observations, no disruptive behavior observed
 - Jacobs's overall cognitive ability was low-average: Nonverbal IQ average (90), Verbal Ability below average (80)
 - Parents and teachers rated him in the clinically elevated range on the (GARS)




13

Was this evaluation comprehensive?




14



Checklist for Review of Existing Data

- Was information included on current presenting concerns?
- Did it include information about hearing and vision screening, and other relevant health data?
- Was the primary home language documented?
- Was home and school background information included that may affect performance?
- Was there a conclusion regarding referral concern, areas for evaluation, and need for further assessment?



15

Use Technically Sound Instruments

Reliability, Validity, and Usefulness

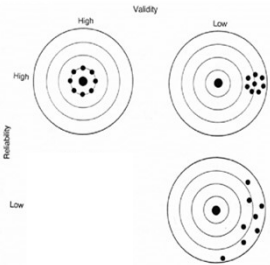



16

Reliability and Validity

Reliability
Consistently and accurately measures something.

Validity
Measures what it is supposed to measure (intent of the measurement).



17

Reliability and Validity

Temperature readings within 10-minute period: 98.3, 96.0, and 100.5


Temperature readings within 10-minute period: 98.4, 98.2, and 98.3

Temperature readings within 10-minute period: 98.4, 98.2, and 98.3, but person is hot to the touch, sweating, and chills

18

How do you know if a test is reliable, valid, and used for the intended purpose?



19

Finding out information on tests


Finding reliability and validity information for a test

1. Technical manual from the test publishers.
2. Mental measurements yearbook (<https://buros.org/mental-measurements-yearbook>).
3. Peer-reviewed research articles.
4. Ask a psychologist.

What do the numbers mean?

Reliability and validity coefficient (correlations) range from 0 to 1. The closer the r-value is to 1 the stronger the reliability or validity.


In general, coefficient numbers above .80 are considered good, but some tests require closer to .90



20

Common Standardized Assessments



<p>Academic/Achievement</p> <ul style="list-style-type: none"> ● KTEA-3, 2014 ● WIAT-4, 2020 ● WJ-IV Ach, 2014 ● WRAT-5, 2017 <p>Communication</p> <ul style="list-style-type: none"> ● CASL-2, 2017 ● CELF-4, 2004 ● EVT-3, 2018 ● GFTA-3, 2015 ● PLS-5, 2010 ● PPVT-5, 2018 	<p>Full Cognitive</p> <ul style="list-style-type: none"> ● K-ABC-II (NU), 2018 ● WISC-V, 2015 ● WPPSH-IV, 2012 ● WJ-IV-Cog, 2014 ● *Stanford-Binet-5, 2003 <p>Nonverbal Cognitive</p> <ul style="list-style-type: none"> ● CTONI-2, 2009 ● Leiter-3, 2013 ● TONI-4, 2010 ● UNIT-2, 2018 	<p>Developmental</p> <ul style="list-style-type: none"> ● Bayley-4, 2019 ● BBCS-4:R, 2022 ● DAYC-2, 2012 ● DIAL-4, 2011 ● WJ-ECAD, 2014 <p>Autism</p> <ul style="list-style-type: none"> ● ADI-R, 2003 ● ADOS-2, 2012 ● CARS-2, 2010
---	--	--



21

Common Standardized Rating Forms

<p>Adaptive Behavior</p> <ul style="list-style-type: none"> ● ABAS-3, 2015 ● Vineland-3, 2016 <p>Autism</p> <ul style="list-style-type: none"> ● ASRS, 2009 ● GARS-3, 2013 ● SRS-2, 2012 	<p>Behavior</p> <ul style="list-style-type: none"> ● BASC-3, 2015 ● BRIEF-2, 2015 ● Conners-4, 2023 ● CBRS, 2008 ● Conners-EC, 2009 <p>Personality</p> <ul style="list-style-type: none"> ● MMPI-A-RF, 2016 ● M-PACI, 2005 	<p>Emotional</p> <ul style="list-style-type: none"> ● BYI-2, 2005 ● CDI-2, 2010 ● MASC-2, 2012 ● RCMAS-2, 2008 <p>Developmental</p> <ul style="list-style-type: none"> ● DP-4, 2020 ● DBC-2, 2018
---	---	---

22

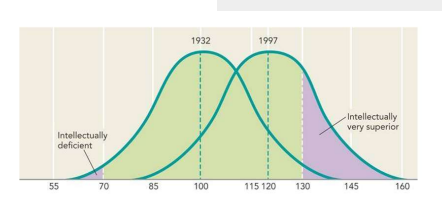

Flynn Effect

Theory that the IQ of the general population increases over time




23

The Flynn Effect


24

Special Factors Consideration


Part of the selection of test is not just if the test has been established as reliable and valid in the development, but if it is appropriate for the individual being evaluating in terms of:

- Language
- Socio-economic factors
- School factors related to access to instruction
- Individual factors that affect ability to perform on tests, such as physical limitation.


Tennessee has a [Checklist for Assessment Instrument Selection](#) to assist evaluators in determining if any of these factors may affect the tests reliability and validity



25



Checklist for Collecting New Data



- Are all administered assessments documented?
- Were up-to-date, commonly used, assessments administered?
- If an uncommon assessment was administered, is there evidence of reliability and validity of this assessment?
- Was consideration given to cultural and linguistic factors when selecting the tests?
- Was more than one method of assessment used within an appropriate time-frame?
- Were the chosen assessments related to the referral concern and useful in determining educational need?

26



Synthesize Data


Interpret Data Based on Norms, Criterion, and Comparison to Previous Performance



27


Interpretation of Standardized Test Scores

Criterion-Referenced and Norm-Referenced



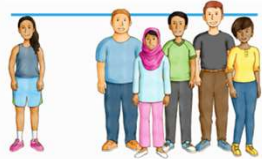
28

Criterion-Referenced



Criterion-referenced tests compare a student's knowledge and skills against a predetermined standard, cut-score, or criterion

Norm-Referenced




Norm-referenced tests compare a student's performance against the performance of their peers

Images from Renaissance: <https://www.renaissance.com/2018/07/11/blog-criterion-referenced-tests-norm-referenced-tests/>


29

Norm-Referenced Score Example 1

On a norm-referenced test, an individual student's percentile rank is calculating according to the performance of their peers.

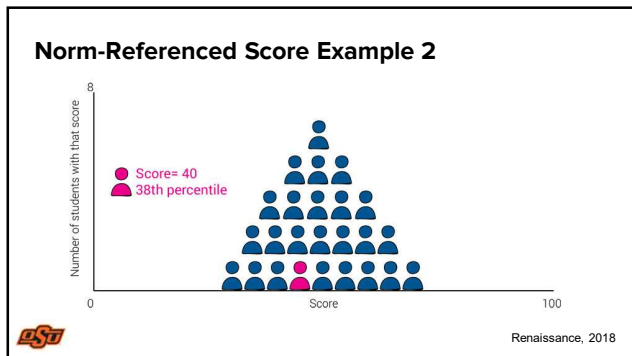


● Score= 40
● 97th percentile

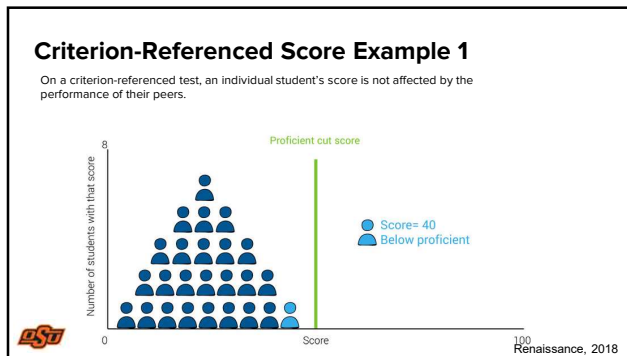


Renaissance, 2018

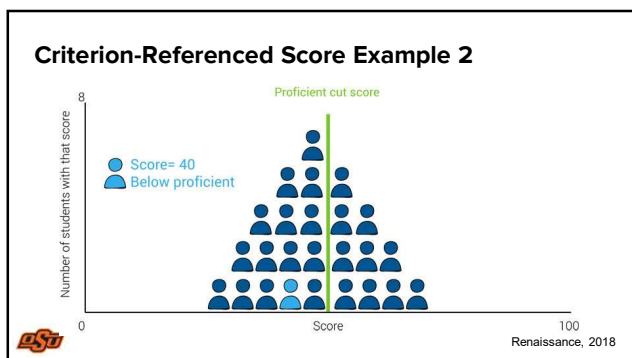
30



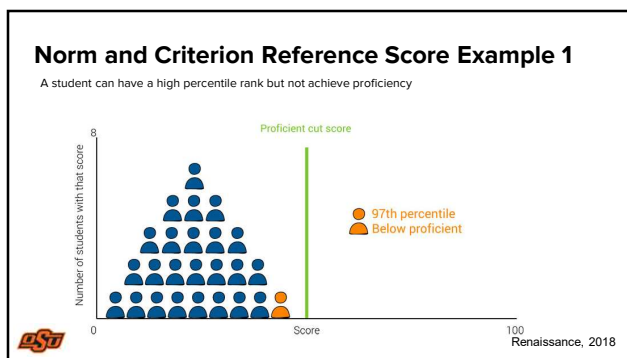
31



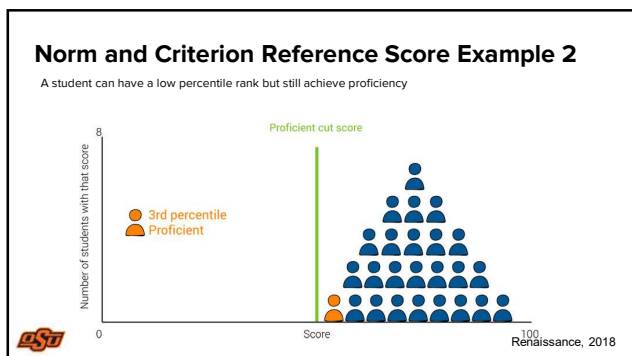
32



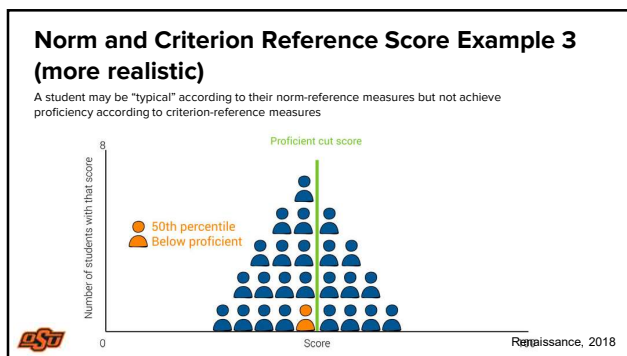
33



34




35



36

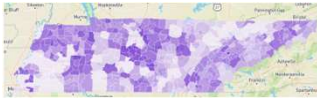
National Norms & Criteria



National **norms** are based on performance on a group that **SHOULD** be similar to the schools' students.

National **criteria** are set based on research and theoretical frameworks

Local Norms & Criteria



Local **norms** are developed using the schools' own population of students.

Local **criteria** are set based on previous performance from the schools' own students.

37

Types of Scores

Score	Definition	Interpretation
Raw Score	Actual number of points or tasks	Not useful for interpretation
Standardized Score	Distance of raw score above or below the mean relative to norm group	"Standard Score of 100 indicates average level of performance"
Percentile Ranks	Describe rank relative to others in the norm group	"Percentile rank of 50, indicates they scored at or above 50% of others in the same age group."
Grade/Age Equivalents	Indicate the same raw score (not skill) as an average student at that same age/grade	"Scored the same number of problems correct as the average 3rd grade student."


38

Standardized Scores

Score	Interpretation	Typical Use
Standard Score (SS)	Mean of 100, SD of 15 Average range: 85-115	Broad summary scores of IQ, achievement, and communication tests
Scale Scores (S)	Mean of 10, SD of 3 Average range: 7 - 13	Individual subtest scores
T-Scores Ranks	Mean of 50, SD of 10 Average range: 40 - 60	On Rating forms of behavior be careful about interpretation because "high scores" may indicate more problems
Stanines	Mean of 5, SD of 2 Average range: 3-7	Individual subtest scores, more rarely used

39

Should grade-equivalents and/or age equivalents be used for decision making?



40

No

There are inherent psychometric problems associated with "equivalent scores" in terms of their reliability and validity


Equivalent scores are not a ratio or interval scale. Therefore, they cannot be added, subtracted, or averaged.

Reynolds, C.R. (1981).

41

Grade Equivalents


- Indicate the same raw score (not skill) as an average student at that same age/grade.
- For example, a GE of 3.9 indicates that the student's raw score for the test was the same as the average raw score for the group of students in the ninth month of third grade on whom the test was normed.
- At best, a grade (or age) equivalent score can tell if that student's raw score is similar to other raw scores in that grade (or age) group. **It does not give you an indicator of actual level of skills.**



42

From KTEA-3 Manual

"Grade Equivalents do not indicate the curriculum level at which a student is working. Grade equivalents are not related to what students at a particular grade ought to be capable of performing or what they are being taught. Similarly, age equivalents do not indicate what students at a particular age ought to be capable of performing. Rather, grade and age equivalents are based entirely on the median performance of students at each grade or age, respectively, on the KTEA-3 subtests."

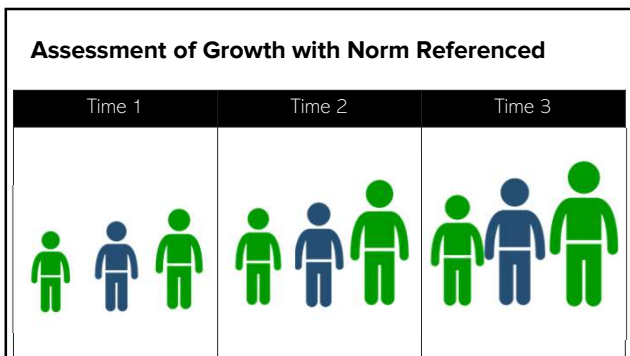


43

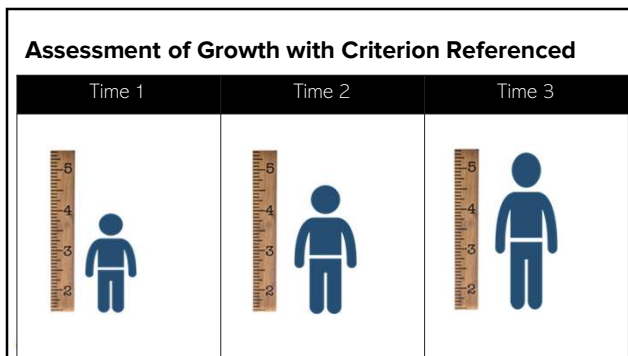
What type of data is best for evaluating progress over time?



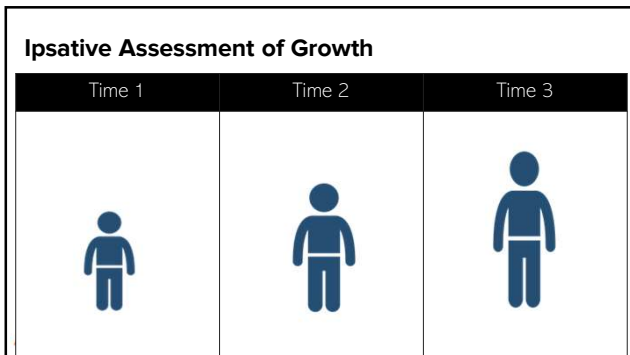
44



45




46



47


Using Data within IDEA Evaluations and IEP Development	Disability Identification	<ul style="list-style-type: none"> • Almost always uses norm-referenced data. • May include criterion reference data. • May include intervention progress monitoring data.
	Needs Identification	<ul style="list-style-type: none"> • Criterion-based and absolute levels of performance • Identify specific skills for intervention/
	Monitoring Progress	<ul style="list-style-type: none"> • Measure a specific skill development over time. • Criterion-based and ipsative comparison.

48



Checklist for Data Synthesis

- Were data reported relative to norms and/or criteria with interpretation guidance?
- Were the norms and/or criteria used appropriate for the student's demographics?
- Were there any noted modifications for standardization that affect the interpretation?
- Does the data converge across multiple methods/sources?



49



Report Data


Components of a Comprehensive Psychoeducational Evaluation Report



50

Sharing Evaluation Reports


- Upon completion an evaluation, a copy of the report should be shared with the parents to ensure meaningful participation (OSEP, 2019).
- Report should include comprehensive summary of the assessment procedures, results, and conclusions.



51

Responsible Reporting of Data

- All qualified professionals who administered assessments should contribute to the narrative in the report.
- All assessments administered should be reported, unless it was deemed invalid. If invalid, then detailed description of why should be reported.
- Report should document any modifications made to the standardized directions
- Part of the narrative should also include how the student's performance in the evaluation relates to the disability category criteria **AND** how this information can be used to develop appropriate services.



52

Essential Components of a Report

Items to Include	Definition/Consideration
Assessment Description	A brief description of the assessments conducted.
Standard Scores	Transformed raw scores with predetermined means and standard deviations. Examples include: Standard Scores, scaled scores, T-Scores, z-scores. It may also be appropriate to include a confidence interval to assist with educational decision-making, as well as an explanation of what a confidence interval means.
Percentile Ranks	Derived score that determines position relative to standardization sample.
Interpretations	Explain what score descriptions mean.
Composite Scores	These scores are more reliable than individual subtest scores, and only reporting these would be a minimum requirement. Best practices would support reporting all subtest scores.
Modifications to Standardization	Include description of any modifications made to a standardized assessment, as well as if a student's behavior and/or attitude during the session negatively impacted the results.

53


Potentially Useful Components of a Report

Items to Include	Definition/Consideration
Raw Scores	In some cases, it would be appropriate to report raw scores; however, if the assessment is standardized, the standard scores should be reported and not the raw scores.
Age and Grade Equivalents	Determined by the average score obtained on a test by members of the same age or grade groups. These scores are psychometrically impure and should never be reported alone . Caution should be used when reporting and interpreting these scores.
Other Score Types from a Specific Assessment	There are some assessments that have score types that do not fit the above classifications. The qualified examiners should use their judgment and training to determine what information should be reported.


54

Components of a full (comprehensive) evaluation report

- Statement of primary purpose of the assessment
- Relevant background
- Procedures/assessments used for the evaluation
- Statement noting limitations of assessment
- Summary of results of all procedures relative to referral question and meaning of the results
- Summary of answers to referral questions
- Recommendations relevant to the referral questions




55



Questions to Ask about Evaluation Reports


- Was data collected to rule-in and rule-out the disability?
- Does the data converge across multiple methods/sources?
- Were relevant skills measured?
- Were the standardized data interpreted within the context of the individual student's environment, background, school history, and observations?
- Were all components of a comprehensive evaluation present?



56

Recommended Resources

- **Books:**
 - [Ethics and Law for School Psychologists, 8th ed.](#) Susan Jacob, Dawn Decker, Elizabeth Timmerman Lugg, and Elena Lilles Diamond. Wiley Publishing
 - [Assessment of disorders in childhood and adolescence.](#) Younstrom, Prinstein, Mash, and Barkley. Guilford Publications
- **Guidance Documents:**
 - [Principles for SLD Eligibility: Practice & Policy Considerations for States and School Districts](#) Procedures/assessments used for the evaluation.



57

References-1


1. 34 CFR § 300.8
2. American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). <https://doi.org/10.1176/appi.books.9780890425787>
3. 34 CFR § 300.304.
4. Principles for SLD Eligibility: Practice & Policy Considerations for States and School Districts: A Comprehensive Evaluation for Special Education for a Child Suspected to Have a Specific Learning Disability.
5. <https://www.nasponline.org/Documents/Resources/20and%20Publications/Resources/Students/PT%20A%20Comprehensive%20Evaluation%20for%20Special%20Education%20for%20a%20Child%20Suspected%20to%20Have%20a%20Specific%20Learning%20Disability.D3.pdf>
6. Sassu, K. A., Gelbar, N. W., Bray, M. A., Kehle, T. J., & Patwa, S. (2015, September). Exploring Neuropsychology: Seeking Evidence of Added Worth to School Psychology Practice. In *School Psychology Forum* (Vol. 9, No. 3), p. 273-34 CFR § 300.304(b).
7. Youngstrom, E. A., Prinstein, M. J., Mash, E. J., & Barkley, R. A. (Eds.) (2020). *Assessment of disorders in childhood and adolescence.* Guilford Publications.



58

References-2

10. American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017). <https://www.apa.org/ethics/code/> (Standard 9.02)
11. National Association of School Psychologists. (2020). Principles for Professional Ethics. Bethesda, MD: Author. Retrieved from <http://www.nasponline.org>. (standards 11.3.3 and 11.5.1)
12. Mental measurements yearbook. (n.d.). EBSCO Publishing.
13. Reynolds, C.R. (1981). The fallacy of "two years below grade level for age" as a diagnostic for reading disorders. *Journal of School Psychology, 19* (4), 350-358.



59

Other Questions?

Sara Rich, PhD, HSP, NCSP
Clinical Assistant Professor, OSU-CHS
OSU BRIDGE Center Co-Director
sara.e.house@okstate.edu

Thank you!

60