

LEVELED READING ASSESSMENTS: WHAT THEY ARE, WHAT THEY ARE NOT, AND HOW WE SHOULD AND SHOULDN'T USE THEM

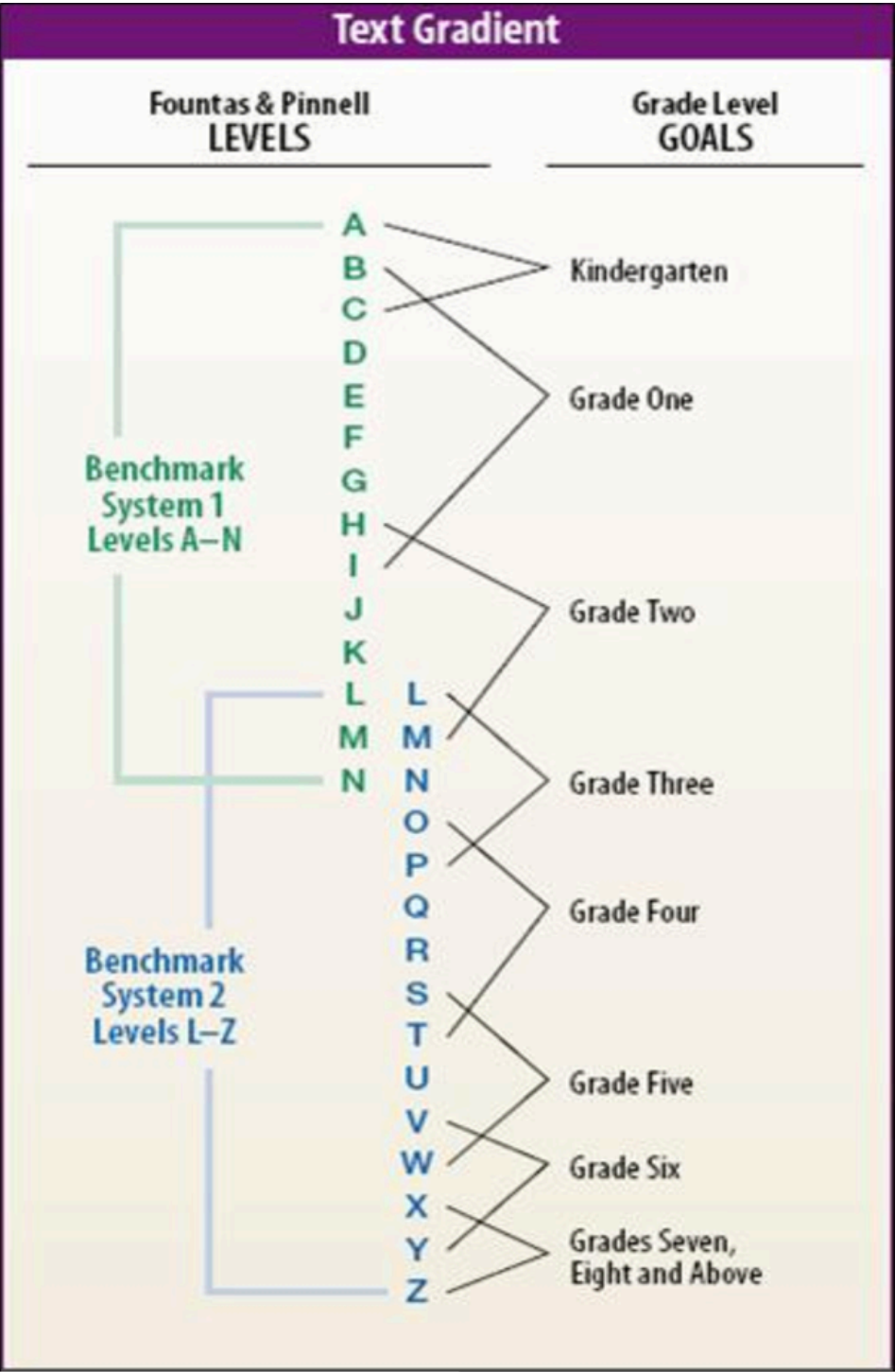
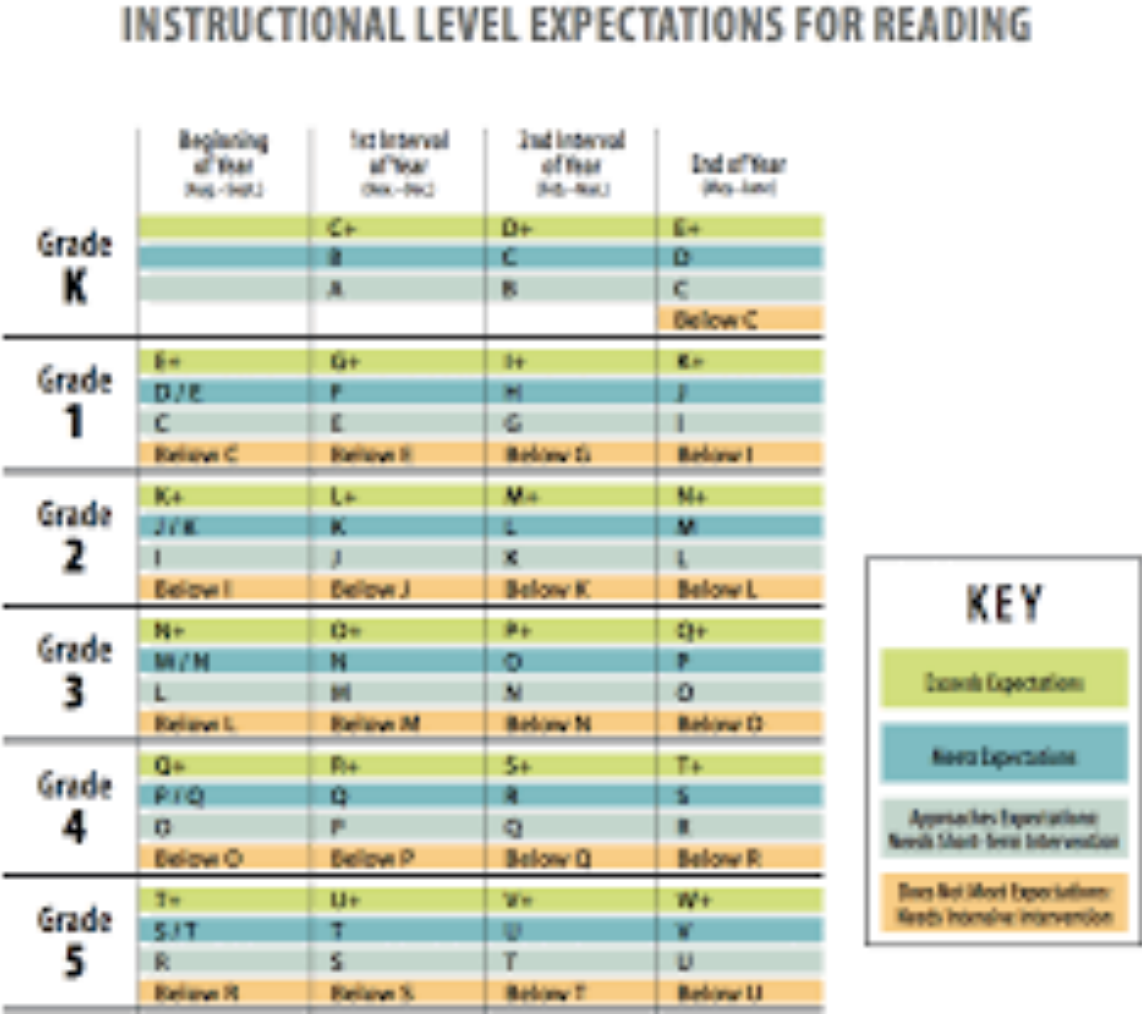
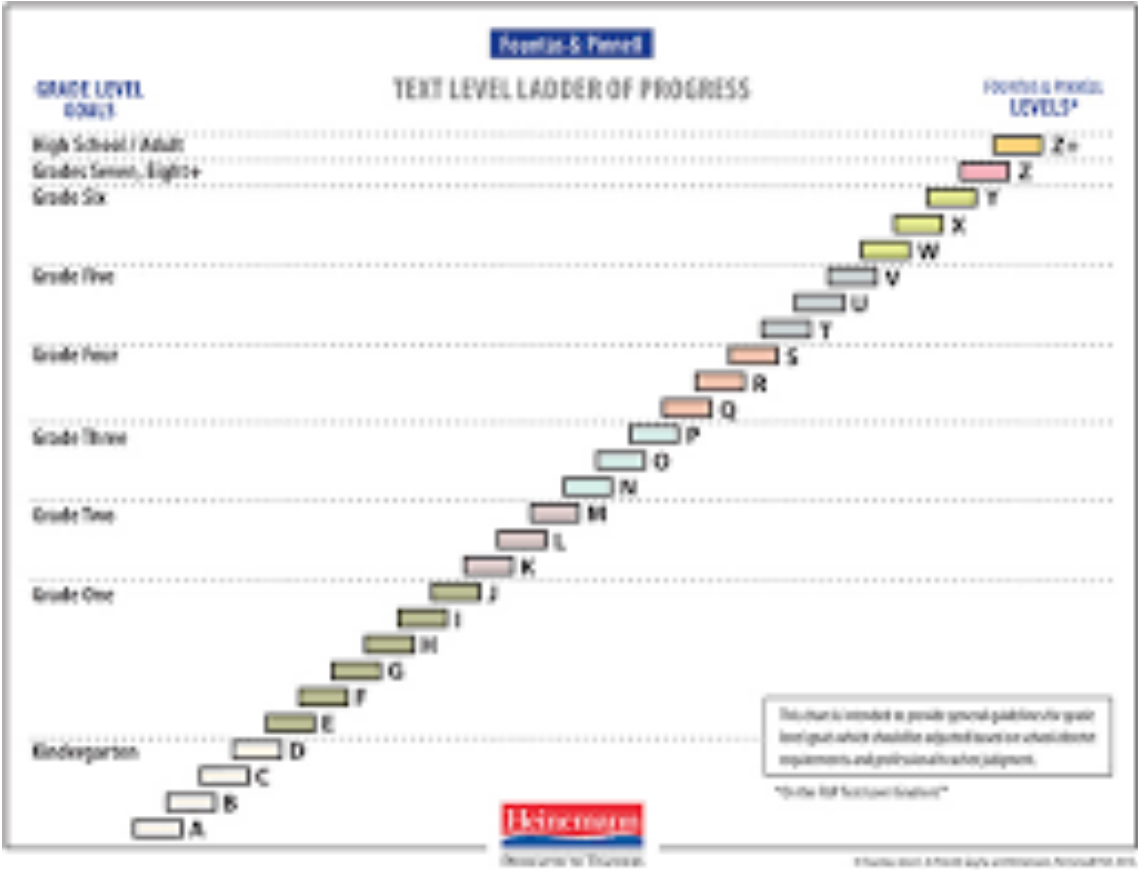


Figure 2. Text Gradient

Reading Level Correlation Chart						
Grade Level	Reading Recovery	Fountas-Pinnell Guided Reading	DRA	Basal Equivilant	Lexile Levels	
Kindergarten	A, B	A	A	Readiness		
	1		1			
	2	B	2	PrePrimer 1		
	3	C	3			
4			4	PrePrimer 2		
Grade 1	5	D	6			
	6					
	7	E	8	PrePrimer 3		
	8					
	9	F	10	Primer		
	10					
	11	G	12			
	12					
	13	H	14	Grade 1	200-299	
	14					
	15	I	16			
	16					
	Grade 2	18	J, K	20	Grade 2	300-399
		20	L, M	28		400-499
	Grade 3	22	N	30	Grade 3	500-599
				34		
24		O, P	38	600-699		
Grade 4	26	Q, R, S	40	Grade 4	700-799	
Grade 5	28	T, U, V	44	Grade 5	800-899	
Grade 6	30	W, X, Y		Grade 6	900-999	
Grade 7	32	Z		Grade 7	1000-1100	
Grade 8	34	Z		Grade 8		

Jesse Steif, Ed.S., NCSP
President, The Reading League Florida
Board Member, International Dyslexia Association- Florida Branch

TODAY'S TOPICS

1. *What is a running record?*

Different definitions: Listening to a student read aloud from a book vs. Levels A-Z

2. *Metrics that go into a level*

3. *Theory behind error analysis*

4. *Reliability*

5. *Validity*

6. *Empirically supported uses for running record levels*

DISTINCTION BETWEEN LEVELED TEXT AND USING LEVELS AS READING ASSESSMENTS

- This presentation **is not** about the uses of leveled text for **instruction**
- This presentation **is** about the use of levels as **assessments** of student skill

USING LEVELED TEXT FOR INSTRUCTION VS. ASSESSMENT

- Text levels as general markers of text difficulty vs. levels as indications of student reading skill.

AS EDUCATORS, WE WANT TO BE IN THE BUSINESS OF....

Valuing ongoing professional development.

Evolving our practices in light of new, empirically-sound, convincing bodies of evidence.

Engaging with information that challenges our own preconceptions and ways of teaching and assessment.

Resisting the urge to dismiss information instead of critically grappling

Getting comfortable in places of discomfort

Publicly modeling shifts in practice based on new information

WHAT IS A RUNNING RECORD?

- An assessment that consists of a student reading aloud from a particular leveled book or passage.
- Teachers observe for reading accuracy and conduct an error analysis
- Teachers also ask for a retell/summary
- A particular book can come out to be independent (95-100% accuracy), instructional (90-94%), or frustration level (below 90%) for the student.
- There are various leveled-text gradients (F&P, Reading Recovery, BAS, DRA).

THUMBS UP, THUMBS DOWN

- Text levels are *equal interval*, meaning there are equal jumps in difficulty from level to level.
- There is no *intra-level variation* in book difficulty. Books within a level are roughly equal in difficulty for all kids.
- If a book rated at a level L (or level 28) is instructional for a student, then a level M will probably be their frustration level.
- We shouldn't listen to children read and analyze the errors they make to guide our instruction.

WHAT METRICS GO INTO A TEXT'S LEVEL?

- According to most leveling structures, 10 characteristics contribute to the level of a book:
 - genre
 - text structure
 - content
 - theme and ideas
 - language and literary features
 - sentence complexity
 - vocabulary
 - word complexity
 - illustrations
 - book and print features
- These are broad and often subjective constructs that lead to variation in difficulty between books within single levels.

WHAT A RUNNING RECORD LEVEL IS AND IS NOT- FROM FOUNTAS AND PINNELL

- A running record refers to the narrow difficulty of a book, NOT the “reading level” of the student.
 - “The truth is that children can read books on a wide variety of levels, and in fact, they experience many different levels of books across the day.” (Fountas and Pinnell, 2017)
 - "In our view, the level of a text has no place on a report card. Although parents do need to know their child’s progress in relation to grade-level expectations, text levels are too narrow to measure.” (Fountas and Pinnell, 2017)
- Fountas and Pinnell have stated that they never intended for children to be limited in their independent reading to a specified level. We should not be limiting students to books at their independent level!
- Because there is often very little difference (or total overlap) in a student’s ability to read books in consecutive levels, it is not an empirically-supported practice to report that the student “reads at a level _____”

INCONSISTENCIES BETWEEN BOOKS AT THE SAME LEVEL, SIMILARITIES BETWEEN BOOKS AT DIFFERENT LEVELS

- When we take a look at another measure of book difficulty like a lexile, we see that books within the same level can often be starkly different.
 - Selecting the book *The Zoo* may lead us to believe that a student “reads at Level E”, but a closer look reveals that the same child might do just as well reading the Level H book, *Trucks*.
- **Key Takeaway:** These designations (level E or Level H) can mean the difference between a student being “at grade level” or “below grade level” and can impact the perception of a student’s ability, both for the student and for the teacher.

Books More Similar Across Than Within Levels

Level	Fiction Book	Lexile	Nonfiction Book	Lexile
A	Best Friends	BR 100	At The Park	BR 80
B	My Little Dog	20L	Playing	20L
C	Socks	190L	Shopping	120L
D	A Nice Little House	100 L	Our Teacher, Mr. Brown	170 L
E	The Loose Tooth	250L	At The Zoo	330L
F	Anna's New Glasses	300L	From Nest to Bird	210L
G	Bedtime for Nick	280 L	Bubbles	330 L
H	The Sleepover Party	480 L	Trucks	350 L
I	The Best Cat	470L	All About Koalas	440L
J	Our New Neighbors	500 L	More Than a Pet	480 L
K	Edwin's Haircut	480 L	Animal Senses	450 L

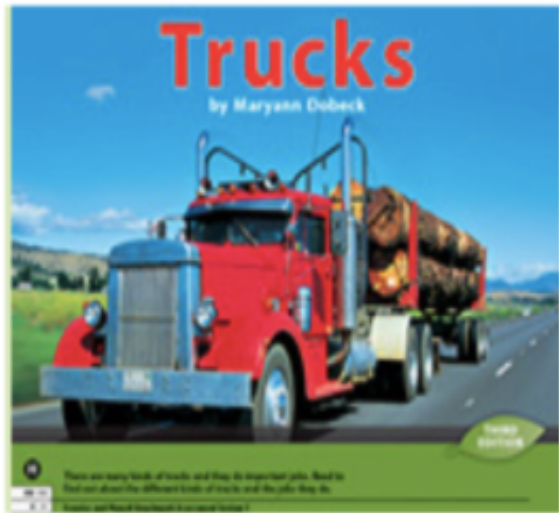


E
Nonfiction
330L

Decoding Indicator: **Low**
Syntactic Indicator: **Medium**
Semantic Indicator: Very Low
Structure Indicator: **Medium**

H
Nonfiction
350L

Decoding Indicator: **Low**
Syntactic Indicator: **Medium**
Semantic Indicator: Low
Structure Indicator: **Medium**



(*Righttoreadproject.com, 2019*)

RUNNING RECORD LEVEL RELIABILITY

- Key Question: Do students read with similar levels of accuracy and comprehension when assessed multiple times on the same level?
- We know that books within the same level can vary significantly in their difficulty. They are not equalized like aimsweb, DIBELS, Star CBM, EZ CBM, etc.
- “The finding that students' scores may vary considerably depending on which passage they read does not bode well for the use of A-Z (and other) leveling structures because text levels may be confounded by failure to consider topical differences that contribute to text difficulty.”
 - Students’ background knowledge and vocabulary on the subject have a significantly larger effect on their accuracy and comprehension than the level of the book.
- “Those techniques **do not appear to produce reliable text levels**, which makes it difficult for one to predict student reading performance.”

Fawson, P., Ludlow, R., & Reutzel, R. (2006). Examining the Reliability of Running Records: Attaining Generalizable Results. *Journal of Education Research*, 67(12), 113-126.

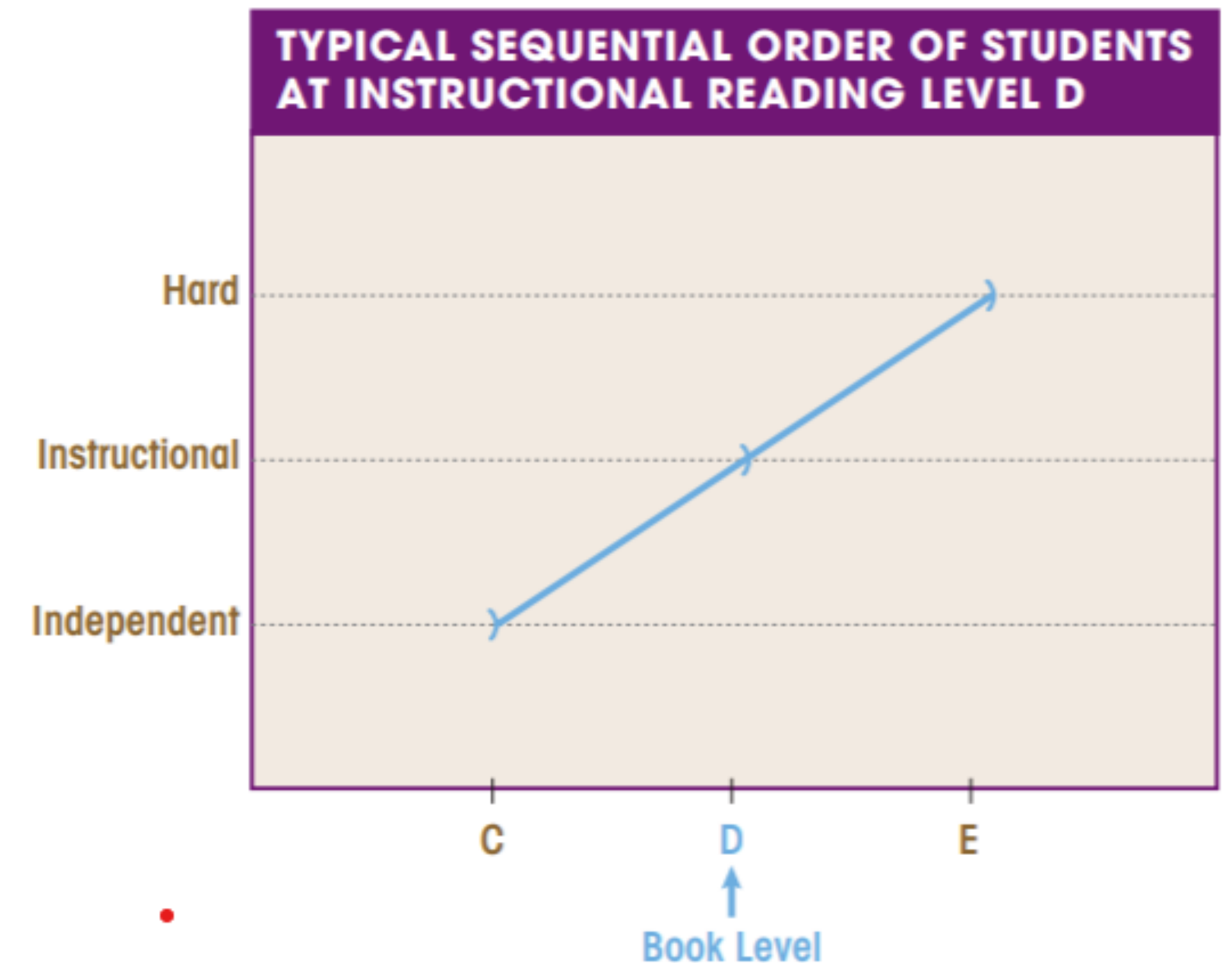
RUNNING RECORD LEVEL RELIABILITY

- Fawson et. al (2006) found that “each student assessed with running records should read a minimum of 3 passages to produce a reliable score”
 - “Inconsistencies between passages of the same level of difficulty” and “inconsistencies between raters” led to unacceptably low levels of reliability when 1 or 2 running record passages were administered.
- This study advises teachers to administer 2 more running records to a student once their instructional level has been found. If a subsequent running record is found to be too easy or difficult, the process must start over until 3 passages all fall at the instructional level.
 - Has anyone observed this practice?

Fawson, P., Ludlow, R., & Reutzel, R. (2006). Examining the Reliability of Running Records: Attaining Generalizable Results. *Journal of Education Research*, 67(12), 113-126.

RUNNING RECORD LEVEL RELIABILITY

- **Key Research Question:** Do students consistently read levels in a sequential and hierarchical manner like the figure to the right?
 - i.e. Is a book at the next level more difficult than one preceding it? Are books in a lower level easier than in the level above it?
- Do students read different books within a level with the same level of accuracy and comprehension?
 - What factors contribute to intra-level variability?



RUNNING RECORD LEVEL RELIABILITY

- A field study (n=497) of a leveled text-based assessment found that 40%-50% of all K-2 students reading at levels A-N did not read books in a sequential and hierarchical order.
 - For some, the level immediately preceding their instructional level was not easier
 - For some, the level immediately succeeding their instructional level was not more difficult
- For older students reading books at levels L-Z, the figure was 20-25% of the study sample.

RUNNING RECORD LEVEL RELIABILITY

VERTICAL TEXT GRADIENT SEQUENTIAL AND HIERARCHICAL PROGRESSION FROM LOWER TO HIGHER LEVELS OF DIFFICULTY		
	Benchmark System 1 (Levels A–N)	Benchmark System 2 (Levels L–Z)
Fiction	60.4%	80.3%
Nonfiction	53.8%	75.4%

Figure 7

What factors contribute to intra-level variability?

- Does the genre of the leveled book contribute to intra-level variability?
 - In other words, does it matter if a student is assessed with a fiction or nonfiction book when given a running record?

IS THERE A DIFFERENCE BETWEEN FICTION AND NONFICTION BOOKS AT THE SAME LEVEL?

- Yes, a very large difference!
- Only 43% of K-2 students whose reading levels fell between A-N had a similar instructional level when assessed with fiction vs nonfiction text.
- Even fewer 3-5 students in levels L-Z consistently had a similar instructional level when assessed with fiction vs nonfiction text (26.1%).
 - Fiction is consistently easier than nonfiction

HORIZONTAL TEXT GRADIENT: STUDENTS READING AT THE SAME LEVEL OF TEXT DIFFICULTY ON FICTION AND NONFICTION TEXTS		
	System 1 (Levels A–N)	System 2 (Levels L–Z)
Fiction–Nonfiction	43.4%	26.1%

Figure 17

THE IMPORTANCE OF PREDICTIVE VALIDITY

Predictive validity refers to the degree to which scores on an assessment are related to performance on a criterion or gold standard assessment that is administered at some point in the future.

An assessment with good predictive validity will tell schools whether or not students **at risk** for failure of a particular future assessment.

- Star Reading had excellent predictive validity with FSA.
- We can tell who is and is not on track to pass FSA and can make instructional changes to try to influence students' trajectories.

THE IMPORTANCE OF PREDICTIVE VALIDITY

Not all assessments need good predictive validity. Many assessments were not created to predict future performance.

- Diagnostic assessments (i.e. phonics surveys, letter-sound inventories) are not meant to tell us who is and isn't on track to pass FSA for example.

We need our progress monitoring assessments to have **very strong** predictive validity

- We need progress monitoring tools to accurately tell us whether or not our instruction is pushing students towards higher likelihood of meeting end of year grade level standards.

PREDICTIVE VALIDITY OF READING LEVELS

- Burns et. al. (2015) compared the diagnostic accuracy of an oral reading fluency task and a leveled reading assessment (BAS) for identifying 2nd and 3rd grade students considered at risk for failing a district-wide end of year criterion-based assessment (MAP)
- Results showed ORF resulted in 86% correct classification of at risk students compared to **31% correct classification base on student BAS level.**

Burns, M., Parker, D., Zaslofsky, A., & Klingbeil, D. (2015). A Brief Report of the Diagnostic Accuracy of Oral Reading Fluency and Reading Inventory Levels for Reading Failure Risk Among Second- and Third-Grade Students, *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 31(1), 56-67

PREDICTIVE VALIDITY OF READING LEVELS– FROM THE DISCUSSION

- **“In a hypothetical school with 100 students needing intervention, 86 of the students who actually need an intervention based on MAP performance would be correctly identified using ORF criteria. Only 31 of those students would be accurately identified using the IRI screening data.”**

Burns, M., Parker, D., Zaslofsky, A., & Klingbeil, D. (2015). A Brief Report of the Diagnostic Accuracy of Oral Reading Fluency and Reading Inventory Levels for Reading Failure Risk Among Second- and Third-Grade Students, *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 31(1), 56-67

RUNNING RECORD PREDICTIVE VALIDITY

- This low classification rate shouldn't be surprising given the low reliability and the more important fact that most assessment creators do not purport them to be pure assessments of readers' skill.
- **Key Takeaway: Just because a student is “on grade level” for their reading level does not mean that they’re not experiencing reading difficulties and are not at risk.**
- **When making MTSS decisions, teachers should weight other measures such as Star or any other norm-referenced measures much more heavily.**

Burns, M., Parker, D., Zaslofsky, A., & Klingbeil, D. (2015). A Brief Report of the Diagnostic Accuracy of Oral Reading Fluency and Reading Inventory Levels for Reading Failure Risk Among Second- and Third-Grade Students, *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 31(1), 56-67

WHAT RUNNING RECORD LEVELS DON'T TELL US

- Anything about the magnitude of the student's reading difficulty. There is significant overlap between sequential text levels and these are very far from norm referenced assessments.
 - Students with different needs/skills can fall in the same level.
Students with similar needs/skills can fall in different levels.
- Levels *alone* tell us very little about what instruction a student needs.

WHAT WE SHOULD CONTINUE DOING

- Conducting flexible, qualitative, observations of student oral reading skill that are meant to inform instruction.
- Listening to a student read aloud from a book remains one of the best assessment tools for teachers. Teacher knowledge matters, levels matter far less.
- Providing students with a variety of text types, depending on what we'd like them to learn to do.

ERROR ANALYSIS ALTERNATIVES TO M-S-V

Error Type	Example	Implication
Grapheme-Phoneme correspondence (GPC) error- Assigning an incorrect sound to a grapheme	Student reads “dad” as “bad”	Review and practice with target GPC
Position pattern error- Leaving off a sound or making an error with a particular part of the word	Student blending “/s/ /l/ /a/ /p/” then saying “lap”	Possible phonological memory issue. Use successive blending to reduce working memory load.
Vowel Sound error- Assigning long vowel where a short vowel or vice versa, difficulty with schwa.	Student reads “wasps” as “/w/ /ahh/ /s/ /p/ /s/” and doesn’t self correct.	Needs explicit instruction in set for variability .
Morphological error- Errors with prefixes and/ or suffixes	Student reads “fries” as “fry”	Explicit instruction in morphology, including prefixes and suffixes.

TAKEAWAYS

- A “level” refers to a book, never the skill of a student. It is one point on a text gradient that can often **vary significantly within and among levels**.
- Just because a student accurately reads one book on a level doesn’t necessarily mean they’ll read another accurately.
 - Similarly, just because they frustrated on one book doesn’t actually mean that’s a frustrational level for the student. They could go on to read another book within that same level or a higher level perfectly well.
- Just because student is “on level” doesn’t mean they’re not exhibiting real problems, **especially in primary grades**. Reading levels simply do not have the same usefulness as a norm referenced assessment of reading skill.
- We shouldn’t cite a reading level as a reason not to move a student up in tiered intervention, not to move forward with an ESE evaluation, or as evidence or lack of evidence of a disability.

TAKEAWAYS

- Just because student is “below level” doesn’t mean they’re incapable of being taught **grade level material**.

TAKEAWAYS

- Given that books vary significantly among and within levels, and that large numbers of students do not read books in sequential and hierarchical orders, the utility of a running record level as a formal assessment of student skill should be called into serious question.
- Books within levels and among successive levels vary too significantly to make the assignment of a single level to a student instructionally meaningful.
 - To be clear, there is certainly a difference in difficulty between, for example, a level C and a level G.

TAKEAWAYS

- Statement we **should NOT** be making
 - “Student reads at a level E, which corresponds to a beginning first grade level. Student is reading at a beginning first grade level.”

TAKEAWAYS

- Of course, books remain excellent **instructional tools** for teachers.
 - Students need a variety of text types.
- Levels should be considered one of many flexible tools to help teachers guide students to books for instructional purposes, **not assessment purposes**.
- Listening to a student read a book and noting the student errors continues to be a very important tool.

TAKEAWAYS

- Given the poor reliability of running record levels, they should not be used as formal assessment or progress monitoring measures within MTSS.
 - We would not be able to determine whether a student's response is due to instructional factors or fluctuation in the measurement instrument.

USING LEVELED TEXT IN INSTRUCTION

- Leveled text can and should be used for instruction, but so should other types of text (i.e. decodable, trade books)
 - One exception is the patterned, predictable early leveled texts (levels A-D)
 - These should not be used for instruction after Pre-K. Starting in K, other text types can be used to teach conventions of print without the pitfalls of predictable, patterned text.
- The text's level is simply a rough guide.
- The students' need guides the book choice. The book level does not guide the instruction
- Explicit word instruction is based on student need, not the graphemes or phonics patterns that incidentally appear in the text.

QUESTIONS- JESSE@FL.THEREADINGLEAGUE.ORG



ADDITIONAL RESOURCES:

- Supporting All Learners with complex Text: <https://achievethecore.org/peersandpedagogy/supporting-all-learners-with-complex-texts/>
- Reading Instruction with the End in Mind: Rethinking “Reading Levels”: <https://achievethecore.org/page/3190/reading-instruction-with-the-end-in-mind-rethinking-reading-levels-2018-december-webinar>
- Fountas and Pinnell Say Librarians Should Guide Readers by Interest, Not Level: <https://www.slj.com/story/fountas-pinnell-say-librarians-guide-readers-interest-not-level>
- Differentiated Reading Instruction: Small Group Alternative Lesson Structures for All Students- Just Read, Florida and FCRR: <https://files.eric.ed.gov/fulltext/ED498777.pdf>