

Understanding Leveled, Decodable and Authentic Text: Assessment and Instruction

Paige C. Pullen, Ph.D., Chief Academic Officer, UF Lastinger Center

Jesse Steif, Ed.S., Literacy Implementation Manager, UF Lastinger Center
President, The Reading League Florida

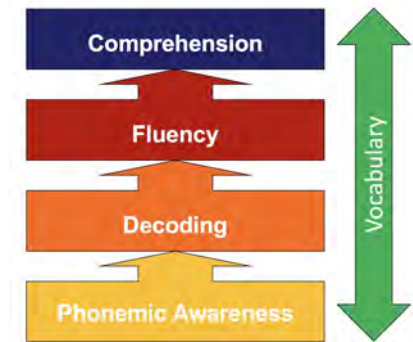
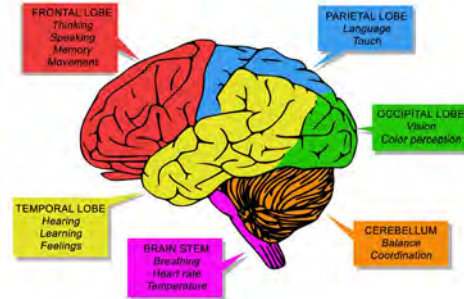
UF | Lastinger Center for Learning
UNIVERSITY of FLORIDA

Agenda

- Who We Are
- Reading Assessments and Leveling Systems
 - Using Texts for Assessment
 - Running Records Versus Reading Observations
 - Metrics Included in Levels
 - Reliability and Validity
 - Empirically Supported Uses for Running Record Levels
- What Research Says about Authentic and Decodable Text
- Selecting the Right Text
- Research to Practice
- Questions and Comments

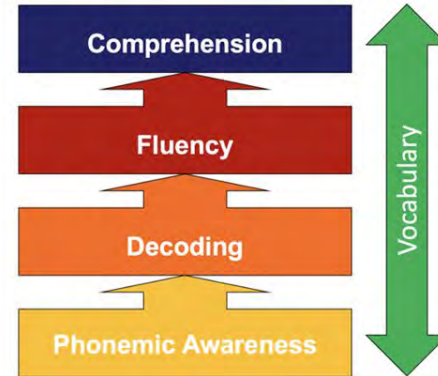
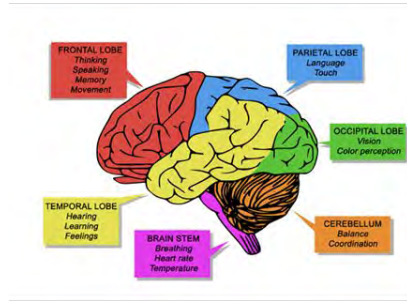
The Foundation of Our Literacy Work

The **Science of Reading** is the foundation of our work. It is a body of knowledge comprising 40+ years of research on reading development, reading difficulties, instruction and content.



What the Science of Reading Is NOT

- It is not about reading wars.
- It is not about teachers being wrong.
- It is not about saying practices are Science of Reading that do not have an evidence base.



As Educators, We Want to Be in the Business of...

- Valuing ongoing professional learning;
- 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; and
- Publicly modeling shifts in practice based on new information.

Distinctions Between Text Usage: Using Levelled Text for Instruction and for Assessment

- Distinction between:
 - ▶ “Leveling students” by assigning them to a level on a text gradient
 - ▶ Using leveled texts for instruction



Using Leveled Text for Assessment

Issues and considerations

What Do We Mean When We Say “Running Record?”

- A running record is an assessment that consists of a student reading aloud from a particular leveled book or passage.
- Teachers observe for reading accuracy and conduct a miscue analysis.
- Teachers also ask for a retell/summary.
- The 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, such as Fountas and Pinnell, Reading Recovery, Benchmark Assessment System (BAS), Developmental Reading Assessment (DRA).

In the chat

True or False?

- Text levels are *equal interval*, meaning there are equal jumps in difficulty from level to level.
- There is *intra-level variation* in book difficulty. Books within a level are not equalized in difficulty for all kids.
- If a book marked at a level L (or level 28) is instructional for a student, then a level M will be their frustration level.
- We should not listen to children read and analyze the errors that they are making.

Metrics that Determine Text Level

Ten characteristics that most commonly 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.

Book Levels as Assessment According to 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” or limiting their choices from the book bin.
- 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.”

Importance of Test Reliability

Test reliability refers to the consistency of scores students would receive on alternate forms of the same test.

A reliable test yields the same results when given multiple times with all other factors held equal. An unreliable test yields different results when given multiple times with all other factors held equal.

- Fictional example: “The PCS Test” is an unreliable assessment. The PCST has 5 alternate forms (PCST-A, PCST-B, PCST-C, etc.) that all have similar content on them.
- As an experiment, a teacher gives the PCST- A through E to a student throughout the day. The results come back and there are 5 very different scores, ranging from “meeting expectations” to “below expectations.”
- Because it is unreliable, the same student can receive different scores based on nothing but the test’s construction.

Inconsistencies and Similarities Between Books

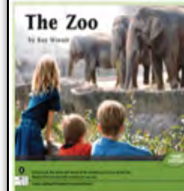
When we take a look at another measure of book difficulty like a lexile, we see that books within the same level can often be starkly different.

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

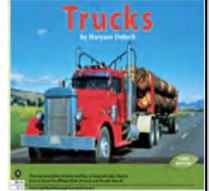


E
Nonfiction
330L

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**



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.

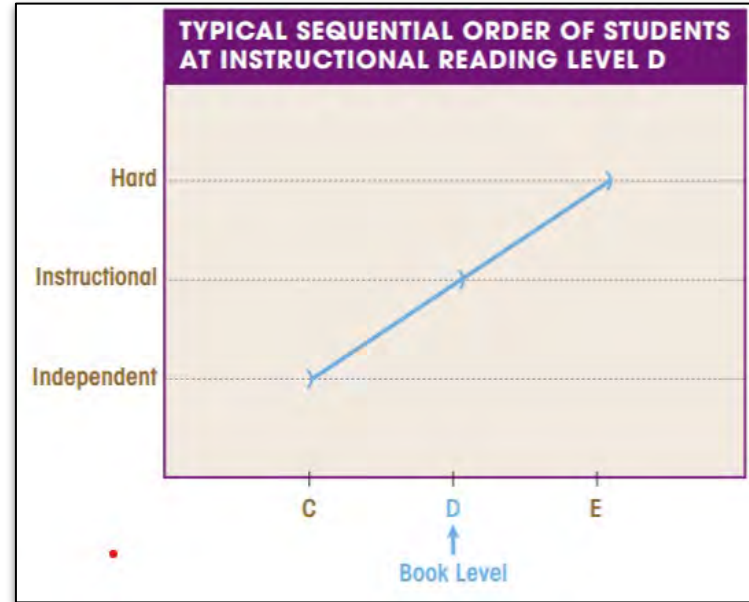
- “The finding that students' scores may vary considerably depending on which passage they read does not bode well for the use of A-Z (and other) leveling structures because text levels may be confounded by failure to consider topical differences that contribute to text difficulty.”
- Students’ background 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 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?

Reliability of Progression Through Levels

Key Research Question:

- Do students consistently read levels in a sequential and hierarchical manner like the figure to the right?
- Is a book at the next level more difficult than the 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.

Factors Contributing to Intra-level Variability

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 versus nonfiction text.
- Even fewer, 3-5 students in levels L-Z consistently had a similar instructional level when assessed with fiction versus nonfiction text (26.1%).
- Fiction is consistently easier than nonfiction.

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

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

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 are at risk for failure of a particular future assessment.

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

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 an outcome measure, for example.

We need our screening and 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 (ORF) 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 (Measures of Academic Performance (MAP)).
- Results showed ORF resulted in 86% correct classification of at risk students compared to **31% correct classification base on student BAS level.**

Predictive Validity of Reading Levels - *from 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)

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.

What Running Records 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.
- An overemphasis on level can cause us to inadvertently hold students back from more complex text.

What We Should Continue Doing

- Conducting flexible, qualitative observations of students' oral reading skills that are meant to inform instruction.
- Analyzing students' reading errors to help 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.

What is MSV Error Analysis?

- M-S-V refers to meaning, structure or visual.
- These are sources of information that students use in order to identify words in a running record passage.
- Error analysis is conducted in which the assessor indicates which source of information the student used when they misread a word.
 - A meaning-based error would be a word that is misread, but the meaning of the sentence is preserved. For example, the student reads “happy” instead of “glad.”
 - A structure-based error would be a word that is misread, but the error conforms to the rules of grammar and syntax of the sentence. For example, the student reads “the dog jumped” as “the dog *jumps*.”
 - A visually-based error would be a word that is misread, but the error was visually similar to the correct word. For example, the student reads “we go to the park” as “we *got* to the *play*.”

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 blends "/s/ /l/ /a/ /p/" then says "lap."	Possible phonological memory issue. Use successive blending to reduce working memory load.
Vowel Sound Error- Assigning a long vowel where a short vowel is or vice versa, difficulty with schwa	Student reads "wasps" as "/w/ /ahh/ /s/ /p/ /s/" and doesn't self correct.	Needs explicit instruction in <i>set for variability</i> .
Morphological Error- Errors with prefixes and/or suffixes	Student reads "fries" as "fry."	Provide explicit instruction in morphology, including prefixes and suffixes.

Analyzing Student Errors

Error Type	Example	Instructional Implication
Grapheme-Phoneme Correspondence (GPC) Error	Student reads “plain” as “plan.”	Review and practice with the target GPC—in this case, focus specifically on the ‘ai’ spelling of long a.
Position Pattern Error	Student reads ‘slap’ as ‘lap’ leaving off the initial sound.	Possible phonological memory issue, teach successive blending.
Vowel Sound Error	Student mispronounces irregular words without self-correcting (e.g., reads ‘was’ as /wăs/).	Needs explicit instruction in lexical flexibility to improve <i>set for variability</i> .

Analyzing Behaviors That Are Errors

Errors	Error Type	Notes
Wipes for whips	Vowel substitution	Hesitated but did not have a strategy to self-correct.
Make for makes	Morphological error	Dropped inflectional ending.
Möther for mother	Vowel sound error in irregular word	Focus on set for variability; lexical flexibility.
Plant for planted	Morphological error	Dropped inflectional ending.

Analyzing Behaviors That Are Not Errors

Non-error Observations	Analysis	Reinforcement
Corrected said from /s-ă-ĭ-d/	Demonstrated lexical flexibility (indicator of set for variability).	Yes, said looks like /s-ă-ĭ-d/, but that isn't a word. You know that it's "said."
Reread phrase	After decoding a word, the student reread the phrase; reread was prosodic.	Reinforced rereading strategy.
Shed	Read shed only after looking at the picture.	Yes, that's shed. Let's say this word sound by sound.
Self-corrected /made/	Blended the sounds and said ă, then self-corrected.	

Takeaways from Assessment

- A “level” refers to a book, never the skill of a student.
- Accuracy on one book doesn’t mean accuracy on another book at the same level.
- Frustration on one book doesn’t mean frustration on another book at the same level.
- A student who is “on level” may exhibit problems, especially in primary grades. Reading levels do not have the same usefulness as a norm-referenced assessment of reading skill.
- Just because student is “below level,” it does not mean they’re incapable of being taught grade-level material.
- Text levels **should not** be used to make high stakes decisions (e.g., movement in instructional tiers or assessment for reading disability) within a Multi-Tiered System of Supports (MTSS).

Takeaways

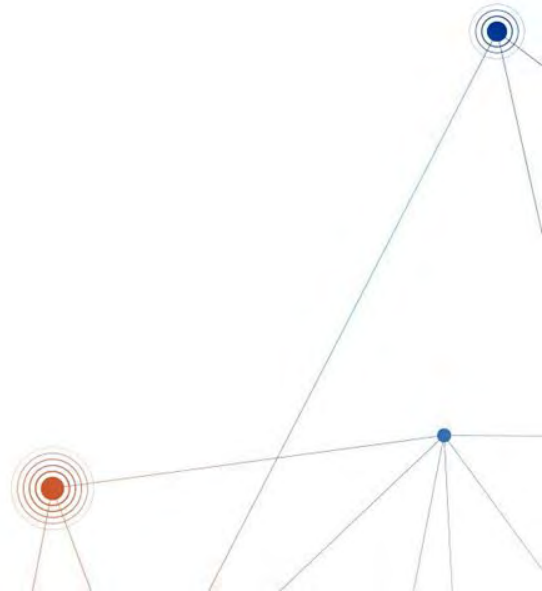
Statement we should NOT be making...

- “Student reads at a level E, which corresponds to a beginning first-grade level. Student is reading at a beginning first-grade level.”

However...

- 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.

Using Different Text Types for Instruction



Using Leveled Texts in Instruction

- Leveled texts can and should be used for instruction; but so should other types of texts.
- The level is simply a rough guide to help determine the difficulty level of the text.
- Texts should be selected based on student needs beyond the level.
- The student need guides the book choice. The book level does not guide the instruction.
- Explicit word instruction is based on student need, not the graphemes that happen to appear in the text.
- In other words, text selection is intentional, purposeful and based on various data points.

Beyond Leveled Text: What Types of Texts Should We Be Using?

Decodable	<ul style="list-style-type: none">• Large portion of phonetically regular words + high frequency irregular words (e.g., the, said)• Less complex sentence structure• May contain vocabulary that fits a targeted phonetic pattern (e.g., jig)• Tends to be more accessible (easier to read)
Leveled	<ul style="list-style-type: none">• Variable proportion of phonetically regular words; this will depend on the leveling system• Texts are often predictable at lower levels-may contain repetitive sentence patterns composed of mostly high frequency words• Accessibility (readability) will depend on the level of text
Authentic	<ul style="list-style-type: none">• Use of phonetically regular words and irregular words is not purposeful, thus there may be a smaller proportion of phonetically regular words• More complex sentence structure• Use of sophisticated vocabulary• Tends to be less accessible (harder to read)

What Is a Decodable Text?

Decodability is the proportion of words that are phonetically regular.

Contrived decodability uses only the letters and sounds that students have learned.

<i>Sentence</i>	<i>% Decodable</i>
<u>Ben</u> <u>can</u> <u>sit</u> <u>on</u> <u>it</u> .	100%
<u>Ben</u> <u>can</u> put <u>it</u> <u>on</u> .	80%
<u>Ben</u> put the <u>cap</u> <u>on</u> .	60%
<u>Ben</u> is the <u>best</u> one.	40%
<u>Ben</u> is the first one.	20%

Research on Decodable Versus Authentic Text

<p>Study 1 Jenkins et. al., 2004</p>	<p>N=79 1st-grade students randomly assigned to two groups</p>	<p>Both groups received core phonics instruction and tutoring with either more or less decodable books.</p>	<p>The more and less decodable text groups did not differ on any posttest (decoding, passage reading, passage comprehension).</p>
<p>Study 2</p>	<p>N=36 young children randomly assigned to low or high phonically decodable texts</p>	<p>Instruction in either high or low decodable text; both groups played games prior to reading to introduce new vocabulary.</p>	<p>Reading comprehension was statistically higher for the low phonically decodable text. Word identification and sound detection approached significance and had large and medium effect sizes respectively.</p>

Research on Decodable Versus Authentic Text

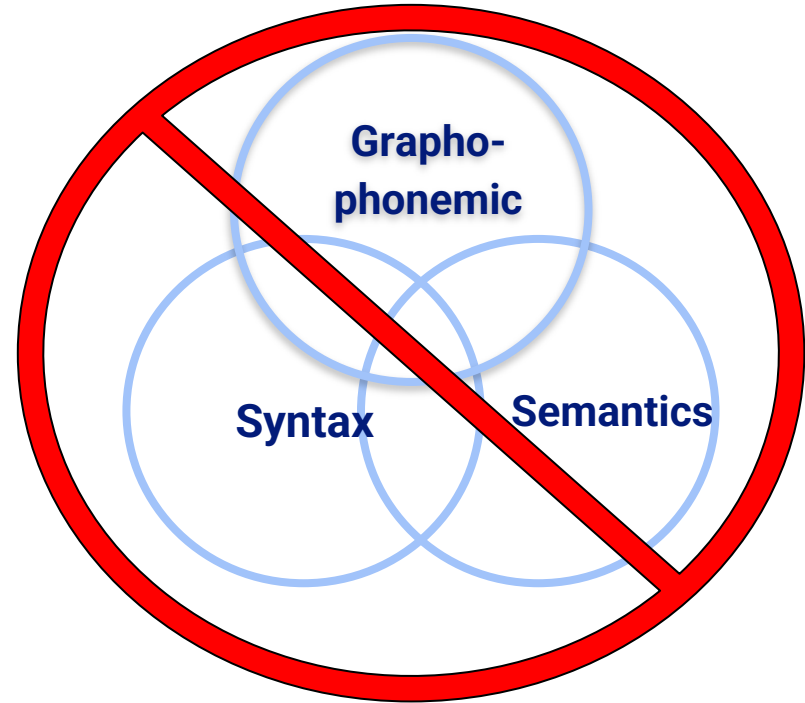
Study 3	Systematic review of studies that examined student performance when reading texts of varying decodability levels. Studies compared the reading performance of students after participation in a treatment that manipulated decodable text as an independent variable.		Collectively the results indicate that decodability is a critical characteristic of early reading text as it increases the likelihood that students will use a decoding strategy and results in immediate benefits, particularly with regard to accuracy.
Study 4	1st-grade students assigned to highly decodable text versus less decodable text.	Highly decodable text with coordinated phonics instruction; all students received the same phonics instruction.	Participants in highly decodable text applied letter/sound knowledge to a greater extent than control participants and had higher decoding accuracy.

So, Are Decodable Texts “Research-Based?”

- While decodable text has not been studied as much as other instructional tools, this text type rests on sound theoretical and pedagogical grounds (Petscher et al., 2021).
- Giving beginning readers the opportunity to read decodable texts provides practice applying the grapheme-phoneme relations they have learned to successfully decode words.
- Lack of research doesn’t mean lack of effectiveness. In the future, researchers will need to “disentangle the active ingredients of effective interventions to specify what to use, when, how often, and for whom” (Petscher et al., 2021).
 - Many evidence-based curricula incorporate decodable text, but research has not isolated decodable text as an “active ingredient” in these effective treatments.

Why Shouldn't We Use Predictable Books For K-2?

- Large differences between early leveled texts (roughly A-D) and later levels.
- Predictable books promote the reading habits of students at the partial alphabetic phase of reading.
- Good readers do not predict upcoming words based on previous semantic and syntactic information.
- Graphophonic information is not the lowest importance. In fact, letters are the starting point from which meaning is made.



Read a Variety of Texts with Corrective Feedback

- How do we select texts?
- How do we apply the research about types of texts?

Decodable



"Frog is here!" clucked Hen. Hen was glad to see Frog. She had a big hug for him.

2



"We will have a snack," Hen said. "We can sit on the grass." Frog saw it was a bit wet.

3

Auth

Using Gravity

It is easier to go downhill than it is to go uphill. Most people go faster when they are going downhill because of gravity, the force that pulls toward the earth's center. When people move downhill, they are moving with gravity, not against it.

Skiers use poles with spikes to push themselves forward.

Skiers try to slide down a hill as quickly as possible. They wax their skis to help them move smoothly over the snow.

A photograph of a skier in blue and green gear sliding down a snowy hill. The skier is wearing a white helmet and is leaning forward. The background is a clear blue sky with some clouds.

Informational

"My goodness," said the mouse. "That was a very short wedding. Did you have fun?"

"Oh, yes!" said the cat. "The bride looked just beautiful!"

"And what is your cousin's name?" asked the mouse.

"Top-Off," replied the cat.

"How interesting," said the mouse. "I've never heard that name before."

After this, the two friends lived together happily for a while. But the cat could not stop thinking about the tub of peanut butter. Again she went to the mouse.

"Dear friend," said the cat, "my second cousin is getting married today. Will you take care of the house while I go to the wedding?"

"Of course," said the mouse.

An illustration of a white cat with a red collar and a white mouse. The cat is standing and looking at the mouse. The mouse is sitting on the floor. There is a small table with a bowl of peanut butter and a jar of peanut butter. The background is a light green wall.

Decodable Texts

- Practice skills that correspond to the phonics instruction you are providing.
- Align texts to the scope and sequence of your core phonics instruction.

Decodable



"Come here, Min!" Rob said.
Min ran and ran!

4



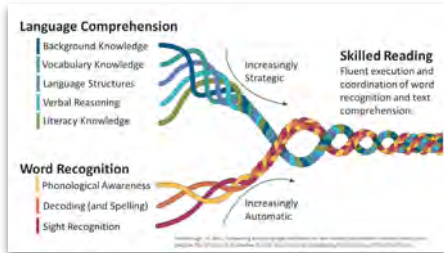
Min got a big bag.
Min ran with it.

5

CVC words
Short a, i,
and o

Authentic and Leveled Texts

- As texts become more challenging, they provide opportunities for teachers to focus on decoding as well as other important skills.
- Look at this text. What skills can you teach?



**Authentic or
Leveled Text**

**Fiction
Vocabulary:
fright**



It was a dark and stormy night.
Horse heard a windy roar.
He woke up with a fright
and kicked his stable door.

2

Child Friendly Explanation

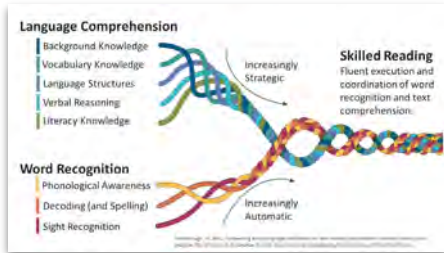
Something that gives you a fright makes you suddenly scared.

What gave the horse a fright?



Authentic and Leveled Texts

- Vocabulary: child friendly explanations
- Fluency: rate, accuracy, prosody
- Decoding: r-controlled vowels; open syllable, stable final syllable; stormy/windy



Authentic or
Leveled Text

Fiction
Vocabulary:
fright



It was a dark and
stormy night.

Horse heard a windy roar.

He woke up with a fright
and kicked his stable door.

2

Child Friendly Explanation

Something that gives you a
fright makes you suddenly
scared.

What gave the horse a fright?



Decodable or Leveled Text: Which is Better For Decoding and Fluency Practice?

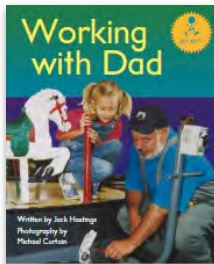
- No research-based answer to this question.
- Choose a leveled text with a high **lesson-to-text-match (LTTM)**.
 - Lesson-to-text match refers to the proportion of words that feature previously taught letter patterns.
 - For example, if a book has a 70% lesson-to-text match, 70% of the words in the book are comprised of phonics patterns that have been previously taught.
 - Research has not identified an optimal percentage of lesson-to-text match for any grade level or student profile.
- Several studies have found that word reading accuracy improves and decoding skill usage increases when students are taught with curricula with a high LTTM (e.g., 68%) compared with curricula in which LTTM is not taken into consideration (Juel & Roper/Schneider, 1985; Mesmer, 2005).

Teaching Decoding Skills Regardless of Text Type



I helped my dad fix the fan
in a house.

“I will hold the ladder for you,”
I said.



- What is the temptation here?
- Don't refer children to visual cues. Teach decoding skills.
- Mark the vowels.
- Find the syllables.
- Teach syllable types.
- **Take away? Purposeful and intentional instruction!**

l a d | d e r

Closed Syllable

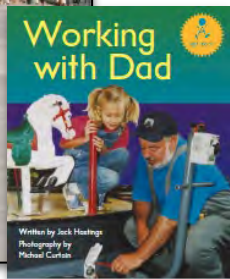
r-controlled
Syllable

I helped my dad
fix a light
at the beach.

"I will hold
the lightbulb
for you," I said.



6



h e l p e d

l i g h t

l i g h t b u l b

Thank You! Contact Us to Learn More



Jesse Steif, EdS
Literacy Implementation Manager
steifj@ufl.edu



Reach Dr. Paige Pullen via Jessica Richards,
UF Lastinger Language and Literacy
Partnerships Manager
jessica.richards@coe.ufl.edu