Examining reading assessments: Increasing the knowledge, power and prerogative of the classroom teacher.

Peter Dewitz, PhD

Professional Decision Making

	Group Size	Time
Physicians	One person at a time	Often delayed / sometimes Instantaneous
Psychologists	One person or small group	Instantaneous
Classroom teacher	Large group, small group, individual	Instantaneous / sometimes delayed

Shulman, 1987

We believe that data driven instruction is:

- Precise; takes the guess work out of understanding students.
- Identifies students with reading problems and their needs.
- Efficient; directs instruction appropriately
- Raises achievement and test scores.

We believe that data driven instruction:

- Is synonymous with interim assessments: Benchmark tests and progress monitoring
- Administrator, place great faith in pencil and paper or computer driven tests.
- **Yet**, We have narrowed what we consider data often ignoring the flow of information that comes from students' daily language and work.

What data are important?

- We privilege data that:
  - Come from a publishing company or are part of a technological system.
  - Is produced by a computer analysis of the results.
  - Is integrated into online instructional program.
- We undervalue data that comes from classroom discussions, reading conferences, and students written work.

Is data drive instruction effective?

- When interim or benchmark tests are the sole source of data (Shepard, 2010)
  - Curriculum is narrowed – less is taught.
  - Curriculum is fractured into small bits.
  - We teach the test.
  - Instruction becomes more teacher-centered, not student centered.

Is data driven instruction effective?

- If data driven instruction raises test scores, but not general reading ability, is it effective?
- If data driven instruction causes schools and teachers to focus on the wrong components of reading, is it effective?
- If data driven instruction consumes excessive amounts of instructional time is it effective?

My Agenda for Today

- Explore the currently popular assessment systems:
  - Benchmark Test,
  - CBM – Progress monitoring
- Delve into the basics of assessment − a topic we have neglected
- Consider how we can use students' work and conversations as a basis for assessment.

A Primer on Assessment

- An assessment is a sample of students' knowledge, skills or attitudes. From this sample we make inferences about achievement and instruction.
- The validity of our inferences depends on several factors:
  - The reliability of the test or our judgment of the student's behavior
  - The validity of the test; does it measure what it purports to measure?
  - Our interpretations of the scores or student responses.

Reliability

- Reliability is a measure of the consistency or dependability of a test.
- All test scores contain error; scores vary from one administration to another.
- Reliability is influenced by:
  - Length of a test
  - Quality of the passages and test items
  - Number of students tested

Reliability

- All measurements have error. The score reported is really just a number in a range of possible scores.
- Reliability of a test, internal consistency, is expressed by a coefficient 0 - .99.
- The higher the reliability the less error.
- If you ask a test to do more, the reliability must be higher.

Percentage of students whose pass/fail status is indeterminate (Koretz, 2008)

	Reliability			
Passing Score	.60	.70	.80	.90
90	13%	11%	9%	6%
70	27%	22%	17%	12%
50	31%	26%	21%	14%
30	27%	22%	18%	13%
10	14%	11%	9%	6%

Reliability of Test Scores

- Fluctuation in test scores is a function of the number of students tested.
  - Grades with fewer students (< 100) will show greater fluctuation in scores than grade levels with more students (+200).
  - Greater fluctuation of scores for small subgroups – ELL, students with disabilities.
  - Fluctuation in grade scores from one year to another is a result of sampling error: the Leo effect.

Reliability of Test Scores

- Because of measurement and sampling error no important educational decisions should be based on one score or one test. (Koretz, 2008)

Validity of the Assessment

- Does it look like a reading test? face validity
- Does it cover the content of reading? content validity
- Do the results predict performance on other assessments? – concurrent validity
- Does the assessment reflect a known construct of reading? Construct validity
- Will the giving of this assessment lead to a positive outcome? Consequential validity

The Validity of Assessments

- The results of a given assessment may support a wide range of inferences − be careful.
- Validity is undermined by tasks only tangentially related to the reading process, e.g. maze test .
- Don’t push a test beyond what it is meant to measure, e.g. using a survey test for diagnostic purposes.
- Scores on a test must be validated against performance in the classroom.

Challenges to Validity

- A fluency assessment depends on passage difficulty, motivation or purpose, and the length.
- A vocabulary assessment depends on the sample of the words and the type questions we ask.
  - Recognition or usage or comparison
- A comprehension assessment depends on the passages, question types, the number of sub-processes assessed.

What is a Benchmark Test?

- Resembles the length, genres, standards of the district with same question types of high stakes assessment.
- Employs multiple-choice and open ended responses.
- Given 3 or 4 times a year.
- Provides summative data but often used in a diagnostic fashion.

How We Interpret Benchmark Test Results

1. We identify students, grades, classrooms and subgroups at risk, on the bubble.
2. We use scores to identify skills or concepts that require additional instruction.
3. We predict future high stakes test performance.

Problems in Benchmark Test Construction

- We assume all tests are created equal.
- Many tests are not created by professional test developers.
- Errors stem from:
  - Passage difficulty
  - Clarity of the questions
  - The similarity of the distracters
  - The students' writing

Why We Can't Use Benchmark Test Diagnostically

- Each skill, strategy, or concept requires multiple items to measure them reliably
- We confuse item difficulty with the skill that is being measured.
  - Is the skill difficult or the wording of the item?
- Vocabulary in the stems and answers compromises students' thinking.

Distinguishing the wording of the item from the skill being assessed

1. What is the main idea of the passage?
2. The article is mostly about? (Main idea, reaching some form of a generalization)
3. The best heading for the third section would be:(Main idea, a generalization made through the application of a label)
4. Which of the following statements should be included in a summary of section 1?

Predicting Future Test Performance

- Given the reliability of the benchmark test our predictions about future performance can be wrong.
- We can’t assume that a low benchmark score predicts subsequent poor performance on a high stakes test.
- Few correlation or anchor studies have been conducted between benchmark tests and the high stakes tests.

Reading Strategies vs. Test taking strategies

- The act of comprehension while taking a test is not the same as comprehension when reading for enjoyment, enlightenment or information. (Rupp, 2006)
- Test taking requires:
  - Search strategies
  - Comprehending the questions
  - Rationalizing and justifying the answers
  - Persistence and stamina

Looking deeper into benchmark results

- Examine the five factors that make up reading comprehension.
  - How did the students perform on the vocabulary items?
  - How did the students perform on the inferential items?
  - How did the students perform on the text structure item?
  - Which genre caused the students the most difficulty?
  - How did motivation effect reading comprehension?

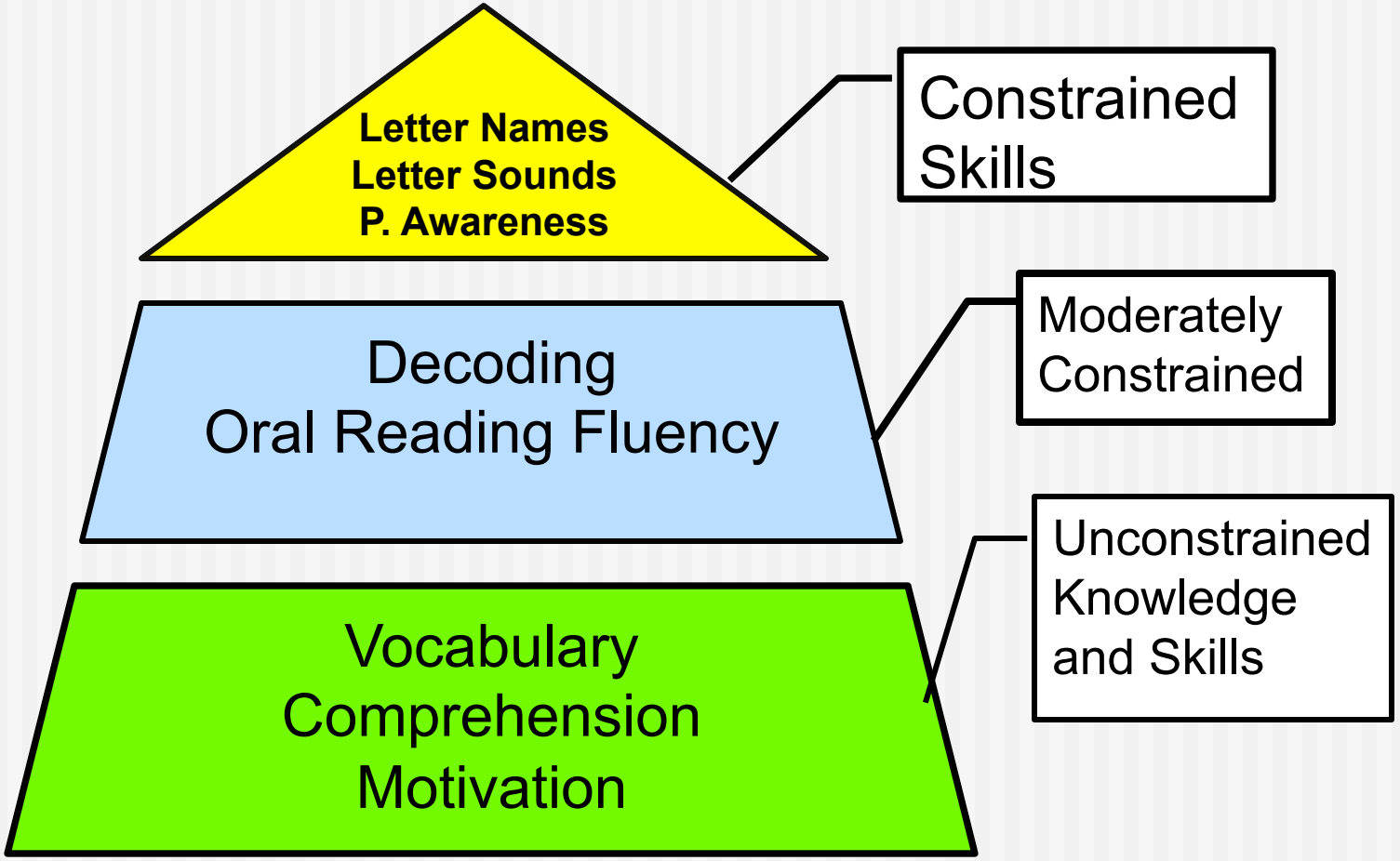
A Broad "Item Analysis"

Passage	Vocabulary	Inferential Comp.	Text Structure	Passage Totals
Information	2/2	1/3	1/2	4/7
Narrative	2/3	2/5	3/4	7/12
Narrative	1/2	2/5	3/3	6/8
Information	1/3	1/4	1/2	3/9
Totals	6/10 60%	6/17 35%	8/11 73%	

Cautions about Interpretations

- Support benchmark conclusions with other data: Student's written work, small group discussions, reading conferences.
- Note, not all standards or strategies can be assessed in one test.
- Don't let the test define your curriculum.

Progress Monitoring – What We Assess



Progress Monitoring

- We tend to measure those components of reading that are easy to measure – constrained skills.
- We find expedient, but sometimes invalid, ways to measure unconstrained skills – comprehension – maze test
- We ignore vital aspects of reading and literacy – vocabulary and motivation

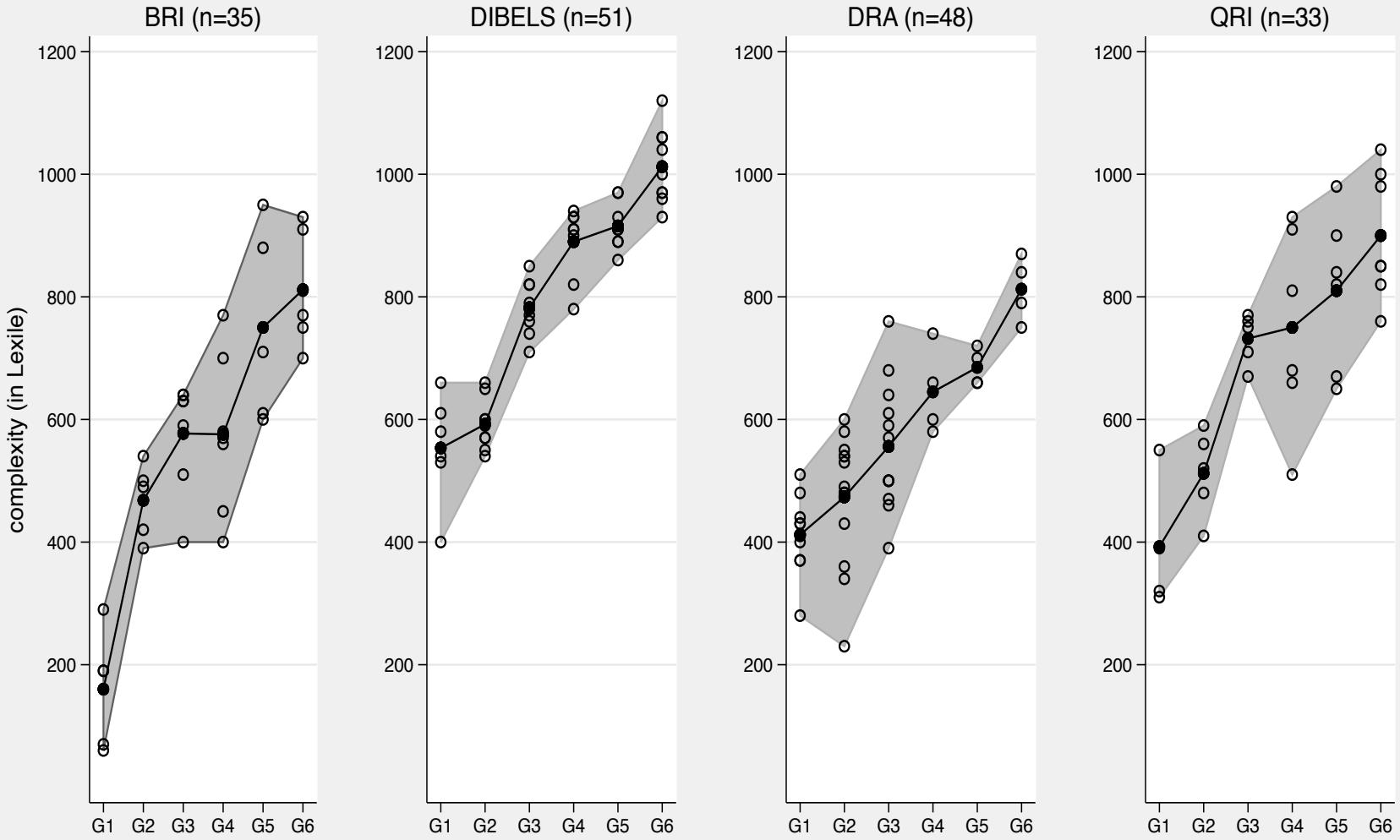
Progress Monitoring

- Progress monitoring is a form of dynamic assessment that attempts to measure a student's rate of progress. (Fuchs & Fuchs, 1999)
- Progress monitoring also measures the student's current level of performance.
- Assessments for placement and progress monitoring are used interchangeably.

Progress Monitoring and Passage Difficulty

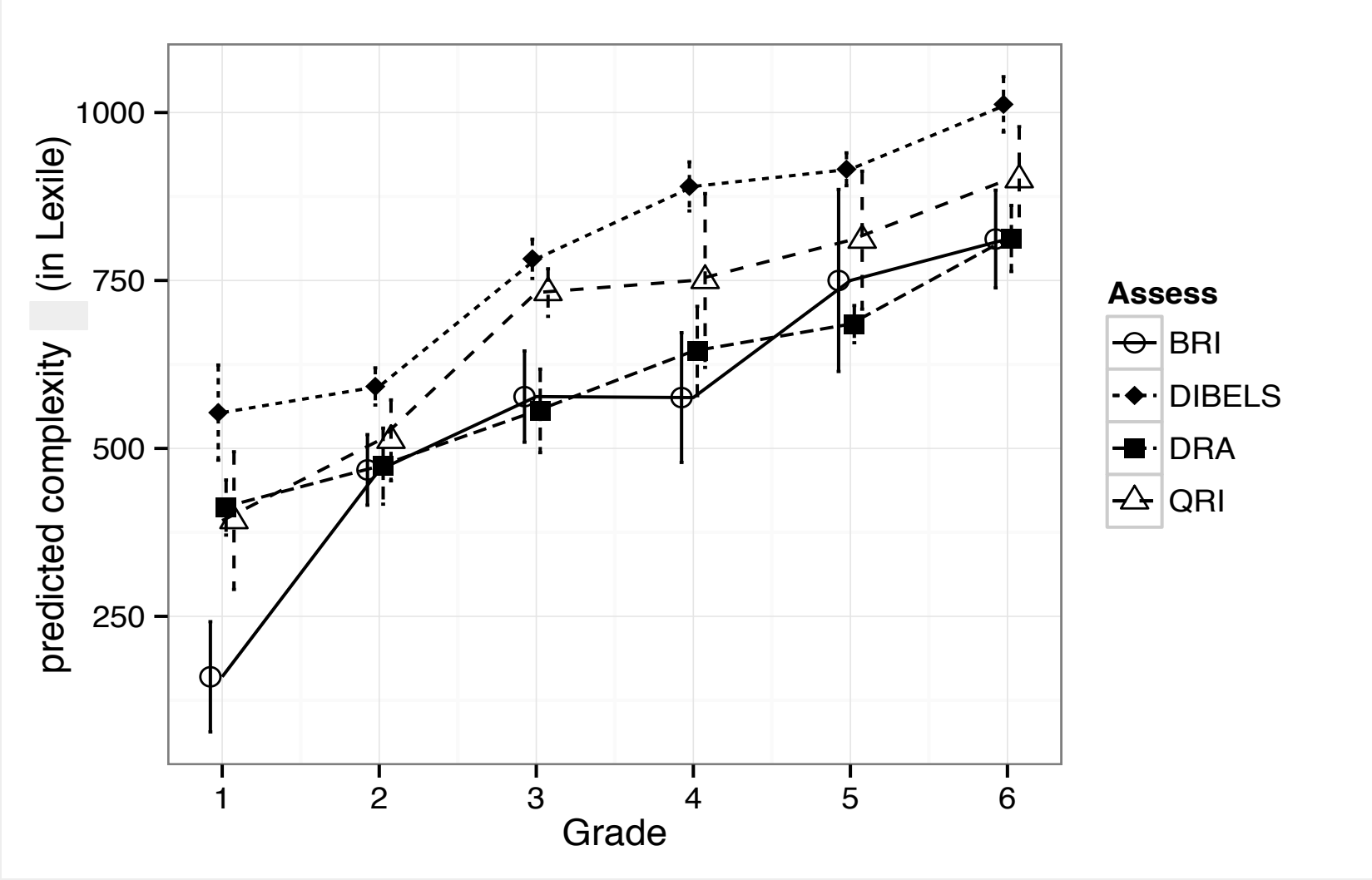
- Readability formulae are used to construct passages for:
  - Oral reading fluency
  - Reading accuracy
  - Reading comprehension
- Not all passages are created equal and this influences our judgments.

Lexile levels of four reading assessments



Toyama, Pearson & Hiebert, 2017

Text complexity of four reading inventories



Toyama, Pearson & Hiebert, 2017

Factors that Influence Fluency

- Prior knowledge
- Cohesiveness or clarity of the writing
- Interest
- Criteria – rate, prosody, comprehension or all three
- Error rates can be larger than a students' growth in fluency
  - DIBELS: Aim = 1-2 WCPM; Error = .69 - .91

Using Progress Monitoring to Predict Performance

	Words Correct Per Minute			
	High Risk <80	Some Risk 80 – 109	Low Risk >110	Totals
Adequate (Meet State Benchmark)	42	188	531	741 (67%)
Inadequate (Didn't Meet Benchmark)	178	130	53	361 (33%)
Number (Percent of Total)	220 (20%)	318 (29%)	564 (51%)	1102

Progress Monitoring

- Narrows the curriculum to what we can easily assess.
- CBM are not the curriculum, but they become the curriculum.
- Over-emphasizes speed versus thought.
- Leads to multiple inferences about the results. What causes poor fluency?

The Components of Formative Assessment (Black & Wiliams, 2009)

	Where the Learner is going	Where the learner is right now	How to get there
Teacher	1 Clarifying learning intensions and criteria for success.	2 Effective discussions and tasks that elicit evidence of students' understanding.	3 Providing feedback that moves learners forward.
Peer	Understanding and sharing these criteria.	4 Activating students as instructional resources for each other.	
Learner		5 Activating students as the owners of their own learning.	

What is Formative Assessment?

(Black & Wiliams, 2009)

- Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers and learners, to make decisions about the next steps in instruction that are likely to be better founded, than the decisions they would have taken in the absence of the evidence.

Formative Assessment

- Information is collected and digested in real time.
- Benchmark tests are too late; we don't catch students in the act of thinking.
- Teachers provide feedback to the students.
- Students provide feedback to each other.

Formative Assessment

- Teachers are the primary managers of assessment within the classroom.(Kennedy & Wilson, 2007)
- Assessments should be seamlessly embedded within instruction and indistinguishable from classroom activities.
- Assessments should be based on a developmental perspective.

Formative Assessment

- The assessments are integrated into the instructional materials and activities.
- Potential for greater validity, relevance and assessment of a wider range of skills, knowledge and attitudes.
- Assessments are organized developmentally trying to capture students' growth as they move through instruction or the release of responsibility model.

A case study: Assessing the Development of Inferential Thinking

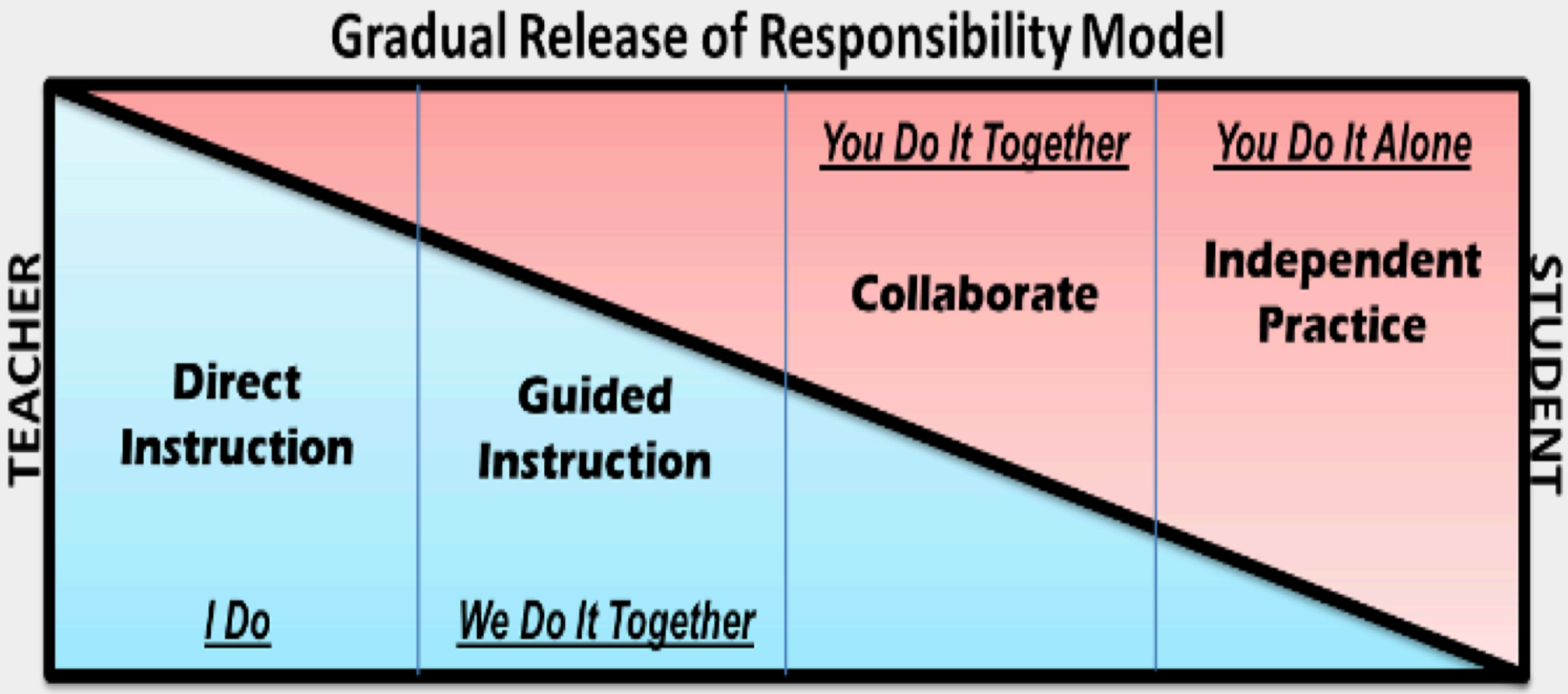
- Inferential thinking is key to reading comprehension.
  - Readers connect ideas within a text – relational inferences.
  - Readers add information that the author implied – character traits, feelings, motives.
  - Readers infer themes and author's purpose.
  - Readers reach beyond the text to form judgments.

Formative Assessment Within the Instructional Cycle

- Direct Explanation – Do the students understand the strategy or concept?
- Guided practice – Can the students apply the strategy or concept with teacher or peer support?
- Independent practice – Can the students employ the strategy or concept on their own?

Assessing the Development of Inferential Comprehension

Direct instruction	**Oral responses from the students**			
	Guided Practice	**Oral responses & limited written responses**		
		Guided & Collaborative Reading	**Oral responses & limited written responses**	
			Independent Reading	**Written responses**
	Character Feelings	**Character Traits**	**Character Motives (C&E)**	**Theme**



Shared Reading
Reading Aloud with Student Interactions

Independent Reading
Book Clubs

Assessing Inferential Thinking during Discussions

- We can note:
  - What kinds of inferences were the students able to generate?
  - Could they do it with or without teacher prompting?
  - Were the inferences on track, within the bounds of the text?
  - Could the students cite evidence to support their inferences?

"Go ahead, Mr. Palmer said to his wife. 'I'll help Abby."
When the text said _____, I thought this was an important detail because _____. This makes me think _____ .

- When the text said that _____, I though this was an important detail because it told me that the family all cared about each other. This makes me think that they, especially Jonathan was considerate.
- When the text said that _____, I thought this was an important detail because the dad helped Abby who could not walk very well. This makes me think that they have to help her.

GROUP SNAPSHOTS:
COMPREHENSION STRATEGIES AND RESPONSES

Strategies	Student Names					
Makes and Justifies Predictions						
Infers character's feelings						
Infers character's traits						
Infers character motives						
Infers theme						
Connects ideas to prior knowledge						
Connects ideas between paragraphs						
Generates questions						
Create a summary						

Use the marking below to characterize the students' use of the strategies.

⃠ The student is unable to use the strategy even when prompted by the teacher.

✓ The student is able to use the strategy when prompted by the teacher.

+ The student initiates the use of the strategy without teacher prompting.

Assessing with Inferential Thinking with Limited Written Tasks

- *What type of inferences can you make?*
  - *Character feelings, traits, motives etc.*
- *What clues did the author leave in the text?*
- *What inferences did you make?*
- *Did your inferences make sense?*

Assessing Inferential Thinking with Limited or Guided Written Tasks

- *The team boarded the school bus and started out for the big game. If they won this game they would be champions! Suddenly, fifteen miles from the site of the game the bus broke down. There they sat, waiting. Nobody seemed to know what to do, and it was getting closer and closer to game time.*

Assessing inferential thinking during guided reading

- What didn't the author tell us?
- What inferences can you make?
- What are the clues in the text?
- What inferences id you make?
- Did your inferences make sense?

- Why the bus broke down.
- How the kids are feeling.
- Big game; miles away.
- The kids are upset and mad.
- Cell phones don't work.

Assessing inferential during independent reading

- At the end of each chapter of Hatchet (Paulsen, 1987) the students write a diary entry, in the main character's voice that includes:
  - What happened to Brian? (Summary)
  - How is Brian feeling? (inferences)
  - What will he do next and why? (Motives and prediction)

Rubric for Scoring Diary Entries

1. The reader makes inferences about the main character's feelings and motives that are consistent with and can be supported from the chapter.
2. The reader makes inferences about the main character that are logical but cannot be fully supported from the chapter.
3. The reader makes inferences about the main character that are not logically derived from the story. The reader relies on prior knowledge and not text information.
4. The reader fails to make any inferences about the main character's motives and feelings.

Comparing Assessments

- Case study of four 5th grade students reading and responding to *Kensuke's Kingdom* (Morpurgo, 1999)
  - What set of data is more valid and reliable?
  - What can a teacher learn from each set of data?
  - What set of data is most useful to the school's administrator?

Making more valid inferences

- Integrate test results to form profiles of students
- Confirm all test results with students' written work and oral performance in the classroom.
- Tests tell you what students can do, observation tell you how they do it.
  - How did you infer the main idea?
  - What is your summary of the passage?
  - How did you conclude that James was demented.?

Developing Profiles of Students

- Profiles present a portrait of a reader that begins a narrative about his knowledge, strategies, skills and motivation.
- Profiles can be developed by integrating data from:
  - Benchmark or interim assessments
  - Standardized tests
  - CBM
  - Written and oral work in the classroom

Profiles of Struggling Readers

(based on Valencia & Buly, 2004)

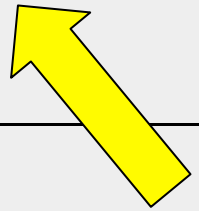
	Word ID	Fluency	Meaning
Automatic Word Callers	+ +	+ +	–
Struggling Word Callers + Vocabulary difficulties	–	+ +	–
Compensatory Readers	–	–	+
Slow Comprehenders	+	–	+ +
Slow Word Callers	+	–	–
Disabled Readers	– –	– –	– –

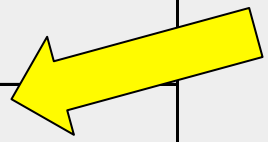
Organizing Your Data

- Divide the chart/graph vertically into two columns to organize your comprehension data.
- Divide the chart/graph horizontally by oral reading fluency scores (WCPM).
  - Hasbrouck & Tindal Fluency Norms
- Organize students by both comprehension results and fluency scores on the chart.

Plotting Comprehension and Fluency Scores

%ile	WCPM	Benchmark > 70 % (Pass)	Benchmark < 70% (Fail)
	110	Michael Landon (95%)	
	100		John Smith (53%)
	90		
75	80		
	70		
	60		
50	50	Marcia Hampton (85%)	
	40		
	30		Mary Jones (57%)





Benchmark and Fluency Results 2nd Marking Period
5th Grade

Oral Reading Fluency

%ile	WCPM	Benchmark > 70 % (Pass)	Benchmark < 70% (Fail)
	180	James (90%) Taylor (97.5%)	
	170	Jenna (95%) Coleby (97.5%) Casey (95%) Alejandra (100%)	Erica (68%)
	160	Ashley (85%)	Cole (61%)
	150		Rachel (57%)
75	140	Ryshae (75%)	Linda (55%)
	130		
	120		
50	110	Ryan (72.5%)	
	100		Brock (47.5%)
	90	Alexus (70%)	Brashawn (60%) Desean (45%)
25	80	Dejon (72.5%) Leilani (90%) Zachary (80%)	Tyrell (60%) Marcus (57.5%) Megan (62.5%)
	70		ShaCarola (67.5%)

Slow Word Callers; disabled readers (lower right quadrant)

- What you know
  - Students below the 25%ile in ORF and weak comprehension.
  - Likely have significant decoding problems.
  - Students between the 25% and 50%ile likely have fluency problems.
- What you don't know
  - What lies beneath – problems with phonemic awareness, orthographic processing, metacognitive processing.
  - Motivation, learned helplessness.

Automatic word callers (upper right quadrant)

- What we know
  - Student read with adequate reading rate and accuracy.
  - Weaker performance on a mix of comprehension questions.
- What we don't know
  - Level of vocabulary and prior knowledge.
  - Strategies that students employ.
  - Metacognitive thinking.

Slow word callers; compensatory readers (lower left quadrant)

- What we know
  - Relatively weak oral reading fluency, but not affecting comprehension
- What we don't know
  - Is comprehension strong beyond the test?
  - Should fluency be a focus of intervention?
  - How much and how engaged are these readers?

Fluent readers with strong comprehension (upper left quadrant)

- What we know
  - Strong comprehenders and strong oral reading fluency.
- What we don't know
  - Is test performance indicative of reading within and outside the classroom?
  - Are the students engaged, thoughtful and avid readers?
  - What guidance and support do they still need?

The Mismatch: Test Results and Learning to Read

- Tests and test analysis leads us to focus on individual skills and strategies.
- Some important factors are not addressed by benchmark or progress monitoring tests.
- Undervalued by most assessments
  - Vocabulary knowledge
  - Prior knowledge
  - Motivation
  - Oral language skills
  - Metacognition strategies
  - Test taking strategies

Conclusions about Data Driven Instruction

- Schools need a well organized assessment system.
- Use a wide variety of assessment tools with multiple points of view:
  - standardized tests, informal assessments, written work, observations – embedded assessments.
- Triangulate information from many sources to support your inferences.
- Involve students in the data collection and interpretation process.

More information on data driven instruction

For further information on a test taking unit that teaches the logical reasoning required on high-stakes tests, or a unit on inferential comprehension.

- Contact:
  - pdewitz@cstone.net,
  - padewitz@marybaldwin.edu
  - www.readingbydesign.com
  - 434-981-1696