Assessing Children’s Reading Comprehension in K-5

Scott Paris
University of Michigan/CIERA
With the collaboration of Rob Carpenter, Ali Paris, Ingham ISD and many colleagues at
With SPECIAL Thanks to

- Bill Loyd and teachers at Paddock School in Milan
- Chris Butler, Judy Kelley, Char Woemer, LAP coaches, and teachers in Monroe
- Diane Schroeder (Willow Run) and teachers at Kettering School
- Bob Galardi, Luther Corbitt, and teachers at Pattengill and Bryant schools
Reading Comprehension

- Needs better definition(s)
- Not a stable or unitary ability
- Interactive and depends on text difficulty, prior knowledge, vocabulary, genre, and interest
- Interactive and depends on decoding skills and strategies for constructing meaning
- Confounded with differences in intelligence, language, and experience
Key Features of Children’s Reading Comprehension

- Builds on prior knowledge & experiences
- Includes narrative & paradigmatic knowing
- May require effort & strategies
- Easier when decoding demands are low
- Better when reading is a tool for learning, rather than a goal in itself
- Second grade slump = word calling
- Fourth grade slump = literal comprehension
Assessment Issues

- High stakes tests define comprehension STILL
- Assessments vary in sensitivity by student’s skill, age, and motivation
- Assessments vary in sensitivity by text difficulty, format, and items
- Comprehension is low priority until grade 3
- Designed for group administration, fast scoring, and minimal teacher interpretation
- Limited test formats
Limited Range of Assessments

- Few assessments for nonreaders, struggling readers, ESL students
- Limited variety of genre
- Disconnected from disciplinary thinking/knowledge
- Few performance measures of verbal discussion, debate, critique, synthesis
- One-time snapshots, usually brief
The Usual Informal Practices

Teachers assess children’s comprehension with:

- Observation
- Anecdotal records
- Answering questions about text
- Retelling
- Writing in response to text
- Discussion groups
Humbug! Say the critics

- Not uniform
- No standards
- Too subjective, not quantitative
- Need hard data
  - To measure growth
  - To show achievement
  - To compare schools
Can assessments be....?

- Authentic
- Linked to instruction
- Diagnostic and summative both
- Used with teachers’ discretion & control
- Sensitive to student differences
- Based on scientific research
- Scaled up
- Acceptable to policymakers
Some Things We Learned About Assessing Comprehension

- Informal Reading Inventories
- Comprehension of Narrative Picture Books
- MLPP & TPRI
Informal Reading Inventories

- Examples: QRI, BRI, DRA

- IRIs can include:
  - Oral reading accuracy measures
  - Fluency ratings
  - Reading rate
  - Retelling
  - Comprehension questions
  - Graded Word lists
Need to Collect Information

- From multiple passages & levels
- From multiple genres
- From multiple word lists
- From silent reading for children reading beyond grade 3 because
  - Comprehension and accuracy unrelated for many children
  - Silent reading allows look backs & strategies
  - Silent reading avoids social anxiety
What Data Should Be Collected for Children in Grades K-3?

- Reading rate
- Oral reading accuracy
- Fluency ratings
- Comprehension questions
- Retelling
IRIs Are Diagnostic When Teachers

- Interpret patterns of oral reading miscues & self-corrections
- Identify difficulties answering specific questions or retelling information
- Use results for conferences with children & parents
- Align reading materials and instruction with children’s needs
IRIs Are Effective When

- Teachers are trained well
- School or district-wide use of IRIs
- Parents and administrators value them
- Data are collected on Palm Pilots, e.g., Milan and Monroe research studies
IRIs Are Difficult to Use for Measuring Growth

Paris (2002) in Reading Teacher describes problems with comparing data across leveled passages, e.g.,
- Basic confounds between passages in difficulty, interest, familiarity, etc.
- Levels are debatable
- Scores may be skewed
How Can We Report Reading Progress & Achievement?

- Pre-Post gain scores on same passages
- Increasing levels of text mastery
- Categories of development (e.g., Frustration, Instructional, Independent)
- IRT growth scores

All can be aggregated by classroom and school and reported as Gain/No Gain or Text Level/Standard met.
IRT Analyses Can Reveal Growth & Provide Summative Data

Two research examples:
- Michigan Summer Reading programs
- Livonia Benchmark tests
Reading Accuracy on the BRI
Comprehension on the BRI
Conclusion

Were there summer school effects?

– Yes!

– …but only statistically significant when we could use IRT to put different passages on a common scale.
Livonia Benchmark Tests

- One narrative and one expository passage at each grade level
- About 10 multiple choice questions per passage so 20 data points
- Given to students in grades 3, 4, 5, 6 in Fall, Winter, Spring
- Problem: How to measure growth?
- Solution: IRT
Livonia Benchmark Tests

2001 and 2002 cohort means
Gender Gap Narrowed - Maybe
Good News About IRIs

- Authentic tasks
- Linked with instruction
- Provide valuable data quickly on several measures
- Data are reliable and valid
- Data can be BOTH diagnostic and summative
BUT - IRIs have problems too

- Too much focus on oral reading accuracy
- Accuracy and rate DO NOT measure comprehension
Caution: Fluency Does Not Mean Good Comprehension

- **Word callers** - High accuracy, low comprehension
- **Gap fillers** - Low accuracy, high comprehension
- More of both kinds of readers after grade 3 so silent reading and comprehension assessments are needed
Figure 3. Posttest Correlations Between Oral Reading Factor and Comprehension Factor
The Fluency-Comprehension Disjunction Hypothesis

Accurate oral reading and comprehension may be positively correlated best ($r = .4-.6$) when:

- Text is easy to decode
- Text is brief and familiar
- Questions are factual or easy inferences
- Assessed among young/beginning readers
The Fluency-Comprehension Disjunction Hypothesis

- Oral reading accuracy may become less related to comprehension among:
  - Older readers in grade 2 and above
  - More skilled and faster decoders

Because automaticity of decoding provides necessary but not sufficient foundation for understanding
Implications

- Measures of oral reading accuracy, like running records and miscue analyses, may be limited beyond grade 2 as assessment tools for word identification.

- Need to compare oral and silent reading comprehension as functions of text difficulty, reading skill, and age.
How Can We Assess Comprehension In Children Who Cannot Decode Text?

The Narrative Comprehension Task
(Paris & Paris, RRQ in press)

- Measures young readers’ narrative comprehension skills
- Does not require decoding skills
- Based on wordless picture books
A Wordless Picture Book

Robot-Bot-Bot, by Fernando Krahn

- Clear story line with main elements of stories, black line drawings, no words

- Adapted version:
  - Spiral bound 18-page book
  - Omits unnecessary pictures
Robot-Bot-Bot

by

Fernando Krahn
NC Task Part 1: Picture Walk

Observational scheme: Five types of behaviors
1. Book handling skills
2. Engagement
3. Picture comments
4. Storytelling comments
5. Comprehension strategies

Scoring
0-1-2 point rubric
0-10 point Picture Walk scale
Part 2: Retelling

Story elements included in retellings:

1. Setting
2. Characters
3. Goal/initiating event
4. Problem/episodes
5. Solution
6. Resolution/ending

Scoring

0-6 point Retelling scale
Part 3: Prompted Comprehension

Explicit Information
- Characters
- Setting
- Initiating event
- Problem
- Outcome resolution

Implicit Information
- Feelings
- Causal inference
- Dialogue
- Prediction
- Theme
NC Task Research Shows

- Developmental improvement with age
- Readers score better than nonreaders
- Easy and reliable to administer
- Predicts reading comprehension scores 1-2 years later
## Predictive Validity

<table>
<thead>
<tr>
<th></th>
<th>ITBS Comp</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC Picture Walk</td>
<td>.30*</td>
</tr>
<tr>
<td>NC Retelling</td>
<td>.46**</td>
</tr>
<tr>
<td>NC Comp Total</td>
<td>.52**</td>
</tr>
</tbody>
</table>
Implications of NC Task

- Is developmentally appropriate and has diagnostic sensitivity for young children & nonreaders
- Yields quantitative data for measuring growth and achievement in comprehension
- Can be used interchangeably for both assessment and instruction & is a bridge to pictures + text
- Can be extended to writing - NP task
- Can be extended to expository genre - EC & EP tasks
- Strategy and genre instruction
Important Parts of a Story

5 Fingers Trick

Who is in the story?

Who is the main character?

What is the problem the story?

What was wrong in the story?

How was the problem solved?

How did the story end?

Where does the story happen?

Where does the story take place?

What does the main character want?

What was the goal of the story?
State Designed Reading Batteries

- Michigan Literacy Progress Profile (MLPP)
  - K-3 Milestone and Enabling tasks

- Texas Primary Reading Inventory (TPRI)
  - K-2 tasks focus on:
    - Book and Print Awareness
    - Phonemic Awareness
    - Oral Reading Accuracy & Fluency
    - Listening & Reading Comprehension
**Test-Retest Reliability of MLPP**

**Strong Correlations**
- Letter Identification
- Letter/Sound Identification
- Hearing and Recording Sounds
- Phonemic Awareness
- Known Words
- Reading Rate
- Sight Words

**Modest Correlations**
- Concepts About Print
- Oral Reading Fluency
- Oral Reading Accuracy
- Comprehension
- Retelling
Validity of MLPP

- MLPP tasks were correlated highly with similar tasks on the TPRI. Phonemic Awareness, Letter Naming, and Letter/Sound correspondence were all very highly correlated. Word Identification, Oral Reading Accuracy, and Comprehension tasks in the MLPP and TPRI were correlated at lower but acceptable levels.

- QRI and DRA comprehension scores correlated between $r = .38-.71$ with Gates McGinitie Reading Comprehension scores.
Consequential Validity

- Rob’s dissertation
MLPP & TPRI Are

- Valid and reliable instruments
- Child-focused & teacher controlled
- Designed for individual, diagnostic, early assessment of multiple reading skills
- Not easily used for summative reporting
Teachers’ Choices of Reading Assessments Depend on:

- Purpose and use of information
- Familiarity with task and training
- Alignment with instruction
- Usefulness for parents and students