



The following presentation by Donald L. Compton, Ph.D. of Vanderbilt University, was given at the Response to Intervention Symposium in Austin, Texas in April 2006. This PowerPoint is provided as a resource material by the Center on Instruction.

The Center on Instruction is operated by RMC Research Corporation under cooperative agreement grant S283B050034 with the U.S. Department of Education, and in partnership with the Florida Center for Reading Research at Florida State University, RG Research Group, the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; and the Vaughn Gross Center for Reading and Language Arts at the University of Texas at Austin.

The views expressed herein do not necessarily represent the policies of the U.S. Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, or enterprise in this product is intended or should be inferred.

**The original author and the Center on Instruction request that no changes be made to the content or appearance of this product.**

2006

# Utilizing CBM to Predict Placement Status

Donald L. Compton, Douglas Fuchs, and Lynn S. Fuchs

Vanderbilt University

# Typical RTI Procedure

---

- All children in a class, school, district are tested once in the fall to identify student at risk for long-term difficulties.
- The response of at-risk students to GE (Tier1) is monitored to determine whose needs are not met and therefore require more intensive tutoring (Tier 2).
- For at-risk students, research-validated Tier 2 tutoring is implemented. Student progress is monitored throughout intervention. Students are re-tested following intervention.
- Those who do not respond to the validated tutoring are identified
  - As LD
  - For multi-disciplinary team evaluation for possible disability certification and special education placement.

# Advantages of RTI Approach

---

- Provides assistance to needy children in timely fashion. It is NOT a wait-to-fail model.
- Helps ensure that the student's poor academic performance is not due to poor instruction.
- Assessment data are collected to inform the teacher and improve instruction. Assessments and interventions are closely linked.

# Within RTI Identification

---

- Tier 2 tutoring is viewed as the “test” to which at-risk students respond to determine disability.
- That response needs to be measured and categorized as “responsive” (not LD) or “unresponsive” (LD) using an appropriate tool for such measurement.

# Implementing RTI

---

Utilizing CBM (Progress Monitoring) to Predict  
Placement Status

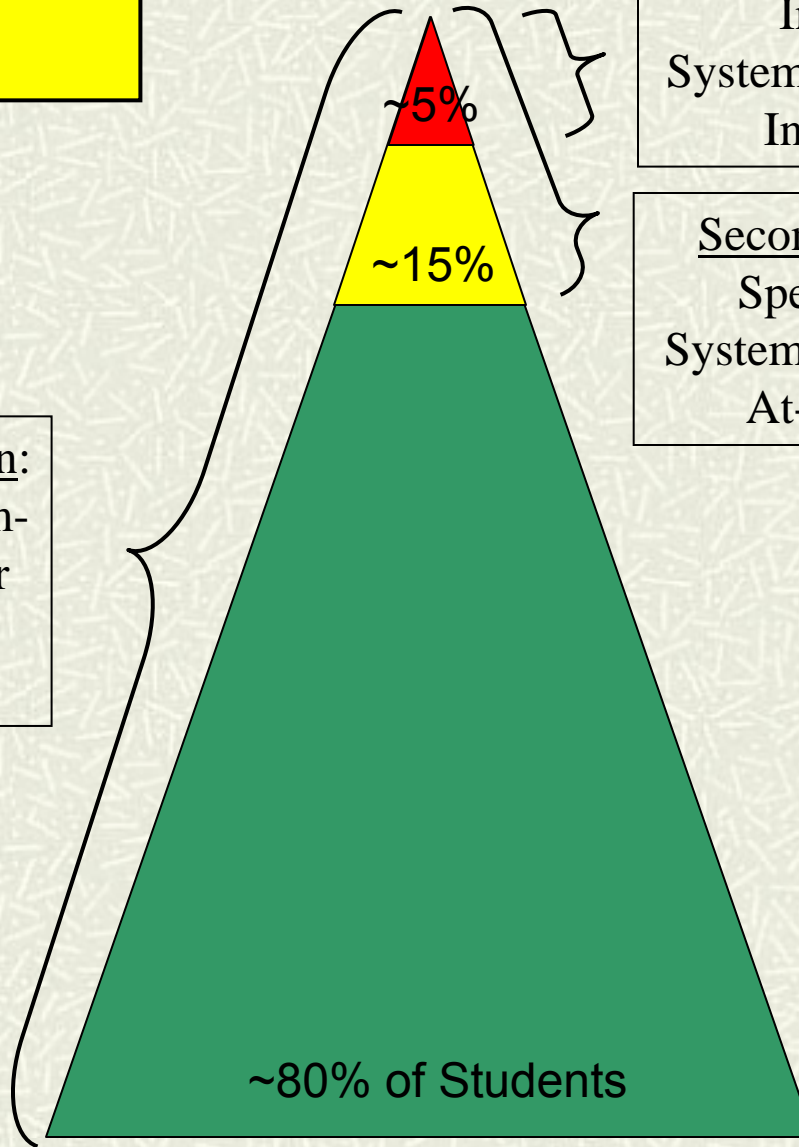
# RTI: Three Tiers

---

- Tier 1
  - General education
    - Research-based program
    - Faithfully implemented
    - Works for vast majority of students
    - Screening for at-risk pupils, with weekly monitoring of at-risk response to general education
- Tier 2
  - Small-group preventative tutoring
  - Weekly monitoring of at-risk response to tier 2 intervention
- Tier 3
  - Special education

**CONTINUUM OF  
SCHOOL-WIDE  
SUPPORT**

Primary Prevention:  
School-/Classroom-  
Wide Systems for  
All Students,  
Staff, & Settings



Tertiary Prevention:  
Specialized  
Individualized  
Systems for Students with  
Intensive Needs

Secondary Prevention:  
Specialized Group  
Systems for Students with  
At-Risk Behavior



# The Core Concepts of RTI Tier 1

---

- Students receive high quality instruction in their general education setting
- General education instruction is research-based
- General education instructors and staff assume an active role in students' assessment in that curriculum
- Universal screening of academics and behavior

# Components of Effective Tier 1 Instruction

(National Reading Panel Report, 2000)

---

- Phonemic Awareness Instruction
- Word Identification and Decoding Instruction
- Fluency Instruction
- Vocabulary Instruction
- Comprehension Instruction

# RTI: Screening in Tier 1

---

- Children are assessed to specify who enters the RTI process.
- RTI success depends on accurate specification of this risk pool.
- Perfect screening would result in 100% accurate identification of “True Positives” (those who will develop RD) who will go into Tier 2 interventions and “True Negatives” (those who will not develop RD) who will be excluded from Tier 2 intervention.

# Two Types of Screening Errors

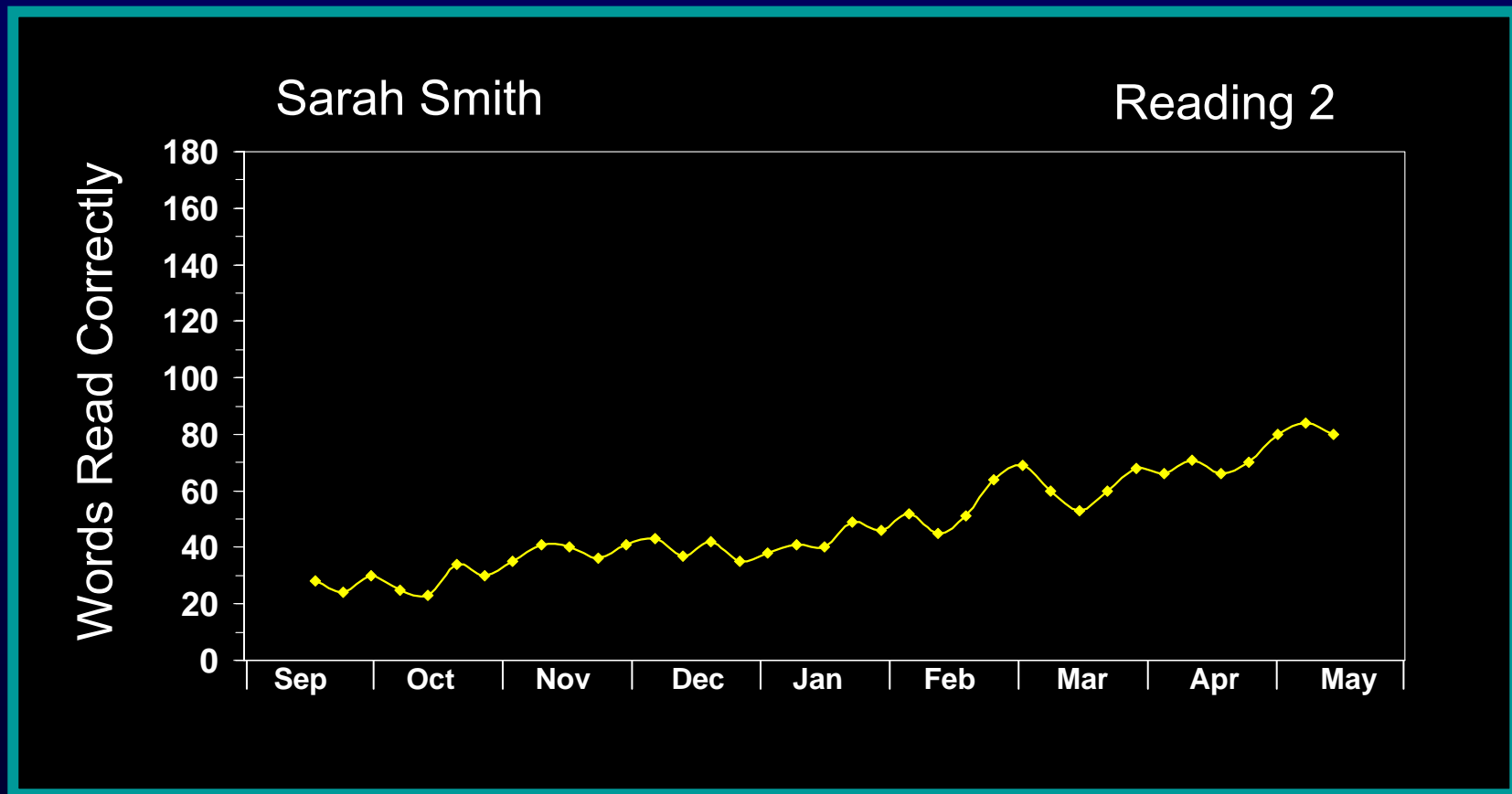
---

- Two types of errors challenge the accuracy of procedures for determining risk.
  - False positives
    - Children who eventually become good readers score below the screening cut score and are falsely identified as at risk.
    - Undermine RTI's prevention purpose by increasing the number of children identified at risk and thereby stressing school resources to provide intervention to an inflated percentage of the population.
  - False negatives
    - Children who later exhibit reading problems score above the cut score and are falsely identified as *not* at risk.
    - Diminish the utility of RTI prevention by failing to provide intervention to children who will eventually develop RD.

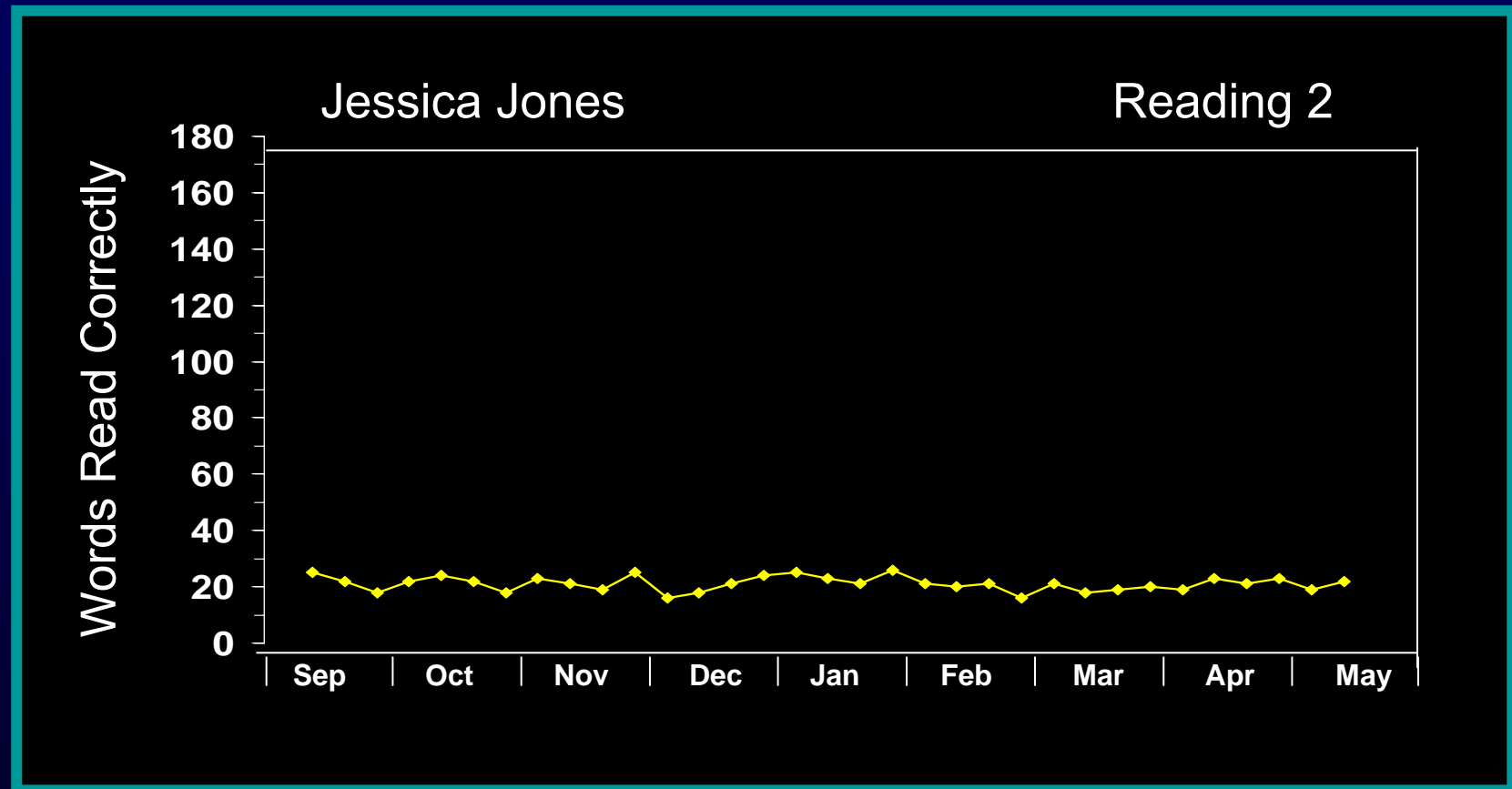
Identifying Children Who Are  
Responders (not at risk for LD)  
and Nonresponders (at risk for  
LD) to Tier 1 Instruction:

Hypothetical Case Studies

# Sarah's Progress on Words Read Correctly



# Jessica's Progress on Words Read Correctly



Do we have evidence that we  
can accurately identify children  
who are at-risk for becoming RD  
(i.e., Sarah vs. Jessica) ?



# NRCLD Study Purpose

---

- Explore issues affecting development of decision rules for selecting 1<sup>st</sup> graders for Tier 2 intervention within an RTI model of LD identification.
- Research questions:
  - What is the added predictive utility of including initial word identification fluency (WIF) or 5 weeks of WIF PM to a multivariate screening battery (that already includes phonemic awareness, rapid naming skill, and oral vocabulary)?
  - Are there advantages to using classification tree analysis over logistic regression in developing statistical prediction rules?

# Word Identification Fluency (WIF)

---

- CBM used to monitor the development of overall reading skill from beginning to end of 1<sup>st</sup> grade.
- In previous work, strong predictive validity for initial WIF and for year-long WIF slopes with respect to end-of-year decoding, word recognition, reading fluency, and reading comprehension performance.

# Example of a WIF Probe

---

## List 1

of	always	story
on	does	south
from	need	half
all	light	held
some	almost	table
them	kind	miles
him	better	that's
may	name	women
down	several	town
called	living	force
our	across	green
used	really	surface
come	means	coming
still	able	ask
life	book	books
between	inside	warm
few	anything	story

# Overview of Study Methods

---

In 42 classes in 16 middle-TN schools, identified low study entry 1<sup>st</sup> graders.

In October, administered a multivariate prediction battery: initial WIF, phonemic awareness, rapid naming, oral vocabulary.

Monitored progress with WIF, each week for 5 weeks; calculated 5-week slope and level.

At end of grade 2, administered standardized reading battery: untimed and timed measures of word identification and word attack and reading comprehension. Used the composite score across these measures to classify children as RD/non-RD.

Applied classification tree analysis and logistic regression to classify RD/non-RD at end of grade 2, using 1<sup>st</sup>-grade prediction battery and short-term PM as predictors.

Evaluated differences in classification accuracy using sensitivity, specificity, and area under ROC curve.

# Districts, Schools, and Teachers, and Sample

---

- 2 school districts in Tennessee (urban Metro-Nashville and suburban Williamson County)
- 8 Title 1 and 8 non-Title 1 elementary schools
- 42 first-grade teachers
- Assessed all students with consent on CTOPP-RLN and WIF
- The 6 lowest students per class on one or both measures, also judged as such by the teacher, were designated “low study entry” ( $n = 252$ ).
- At end of grade 2, 206 were found/tested (attrition rate of 18% over 2 years).

# Following Subject Selection

---

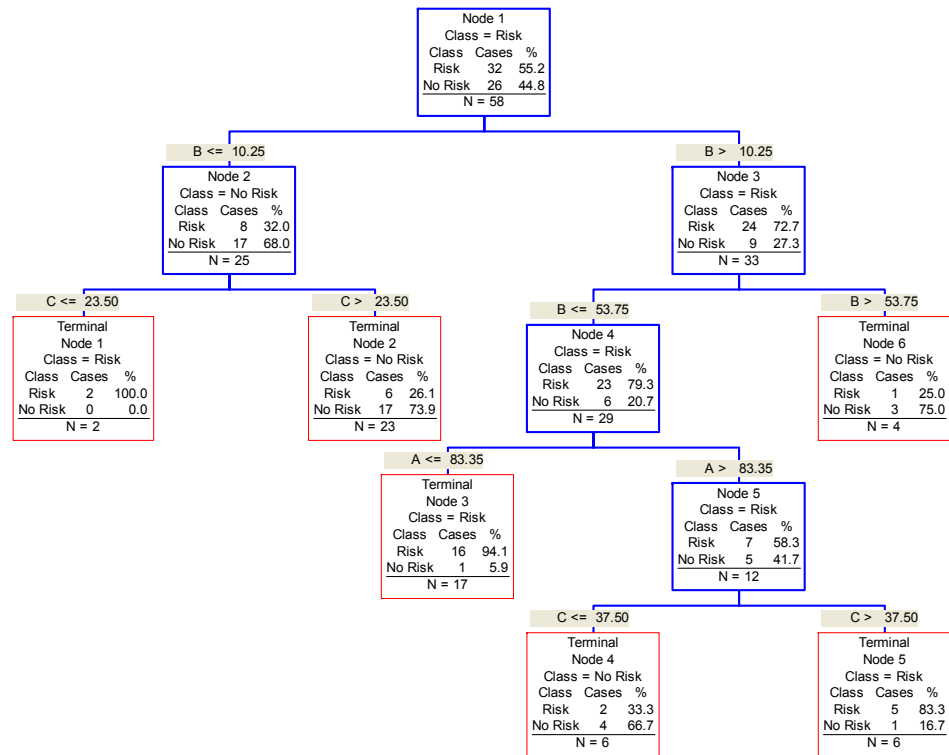
- First-grade (October) prediction battery
  - Phonemic awareness (CTOPP Sound Matching)
  - Rapid digit naming (CTOPP)
  - Oral vocabulary (WJ: Oral Vocabulary)
  - WIF
- Short-term PM on WIF for 5 weeks
- Outcome assessment in April of 2<sup>nd</sup> grade
  - Untimed decoding (WRMT Word Attack)
  - Untimed word identification (WRMT WID)
  - Timed decoding (TOWRE Phonemic Decoding Eff.)
  - Timed word identification (TOWRE Sight Word Eff.)
  - Reading comprehension (WRMT Passage Comprehension)

# Data Analysis

---

## *Classification Tree Analysis*

- Produces set of if-then conditions that permit classification
- At each split, a cut-point is set to provide the greatest improvement in predictive accuracy.
- Terminal nodes are designated “risk” or “no risk” based on whether the node has a higher concentration of risk or no risk children.
- Splitting continues until the predictors can no longer effectively split the parent nodes.
- Tree building is repeated many times with different randomly drawn samples from the data. The tree with the best average accuracy of cross-validated predicted classifications is selected.





# Data Analysis: Evaluating the various classification models

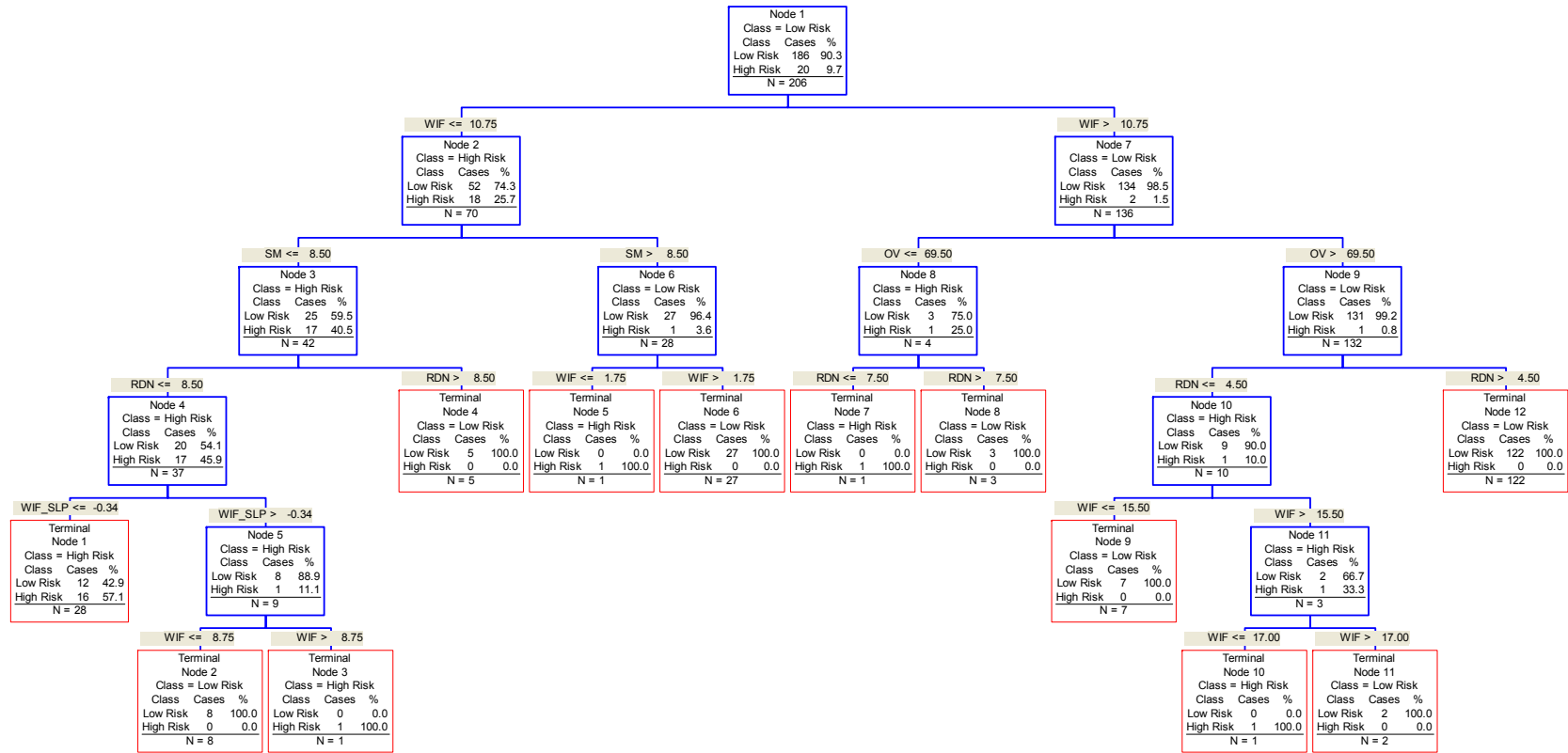
---

- Sensitivity
  - Correct classification of TP ( $TP/(TP + FN)$ )
- Specificity
  - Correct classification of TN ( $TN/(TN + FP)$ )
- Area under receiver operating characteristic (ROC) curves
  - To assess differences in predictive accuracy across models, we used “area under the ROC curve” (AUC), a measure of discrimination – or the ability to correctly classify RD vs. non-RD.
  - A ROC curve is a graphical representation of the trade off between the false negative and false positive rates for every possible cut off.
  - Tested AUC differences by calculating a critical ratio  $z$  value between the two AUCs.

# Results

---

	TN/FN/TP/FP	HR	Sens	Spec	AUC
<i>Initial Screen</i>	145/5/15/41	77.7	75.0	80.0	.863
Sound Matching					
Rapid Digits					
Vocabulary					
<i>Add Initial WIF</i>	150/3/17/36	81.1	85.0	80.6	.904
<i>Add 5-Week PM</i>	154/2/18/32	83.4	90.0	82.7	.912
<i>Classification</i>					
<i>Tree Analysis</i>	174/0/20/12	96.8	100.0	93.5	.982



# Implications for Tier 1 Screening

---

- For RTI to work successfully, reliable procedures for entering children into Tier 2 are required.
- This means identifying TP rates approaching 100%, with identifying a manageable risk pool by limiting FP.
- Previous kindergarten and 1<sup>st</sup>-grade studies demonstrate inadequate decision utility, where
  - some kids who develop RD are not identified for Tier 2
  - schools are stressed to provide Tier 2 intervention to many children who would not otherwise develop RD.

# Implications for Tier 1 Screening

---

- The final model, which relied on classification tree analysis, which allows the same set of predictors to interact, yielded significantly improved classification rates compared to the same set of variables analyzed via logistic regression.
  - Both sensitivity and specificity > 90
  - Only 3.5%-4.0% of 1<sup>st</sup>-graders entering Tier 2
  - With no FN.
- So, combination of 1<sup>st</sup>-grade screening battery of phonemic awareness, rapid naming, oral language, initial WIF, 5-week WIF Level, and 5-week WIF Slope, with decision rules based on classification tree analysis, may have the potential to push RD risk designation to a level of accuracy sufficient for RTI.

# Implications for Tier 1 Screening

---

- Results suggest that the potential exists to develop decision rules that allow identification of the “right” children to enter Tier 2 early in 1<sup>st</sup> grade.
- Additional work is needed to replicate and extend findings.
- *Schools planning to implement an RTI approach to LD identification should put considerable thought into designing an effective system for designating a risk pool that enters Tier 2 intervention that maximizes true positives and minimizes false negatives.*

# RTI Tier 2: Standardized Research-Based Preventative Treatment

---

## *Tutoring*

Small groups (2-4)

3-4 sessions per week (30-45 min per session)

Conducted by trained and supervised personnel (not the classroom teacher)

In or out of classroom

10-20 weeks

# Tutoring

---

## **Small Groups (1:1, 1:3, 1:5, 1:10)**

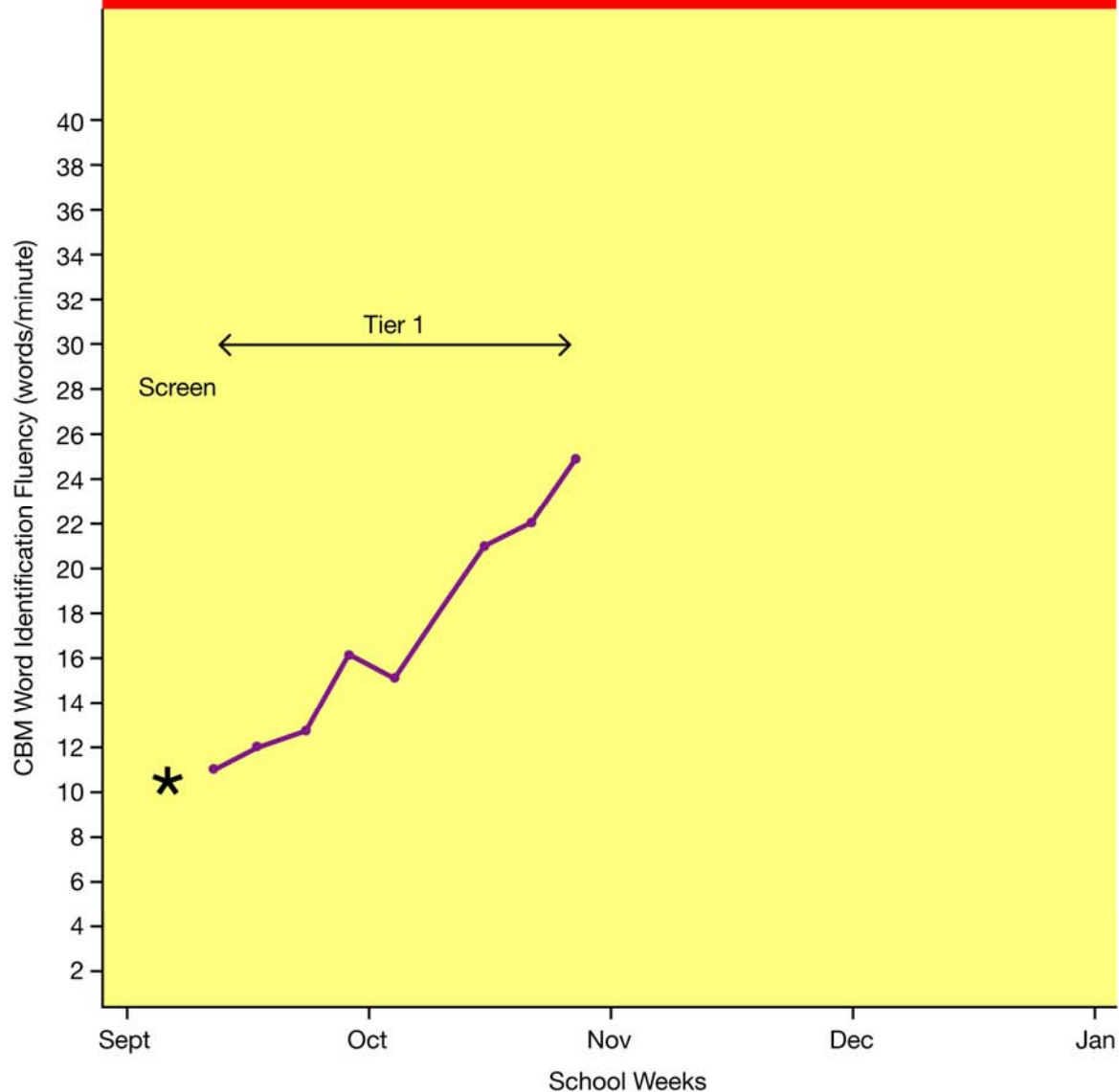
- 10-12 wks, 3-4x per wk, 35-45 min per session
- Point system for motivation
- Immediate corrective feedback
- Mastery of content before moving on
- More time on difficult activities
- More opportunities to respond
- Fewer transitions
- Setting goals and self monitoring
- Special relationship with tutor



What does Tier 2 look like?

Hypothetical Case Studies

## Case B



Student Does Not Have a Disability

### Step 1: Screening

Is this student at risk?

Word Identification Fluency = 10.5

No Yes

### Step 2: Assessing Tier 1 Response

Is this student responsive to general education?

Word Identification Fluency Slope = 1.8

Yes No

### Step 3: Assessing Tier 2 Response

Is this student responsive to diagnostic instructional trial?

Word Identification Fluency = NA

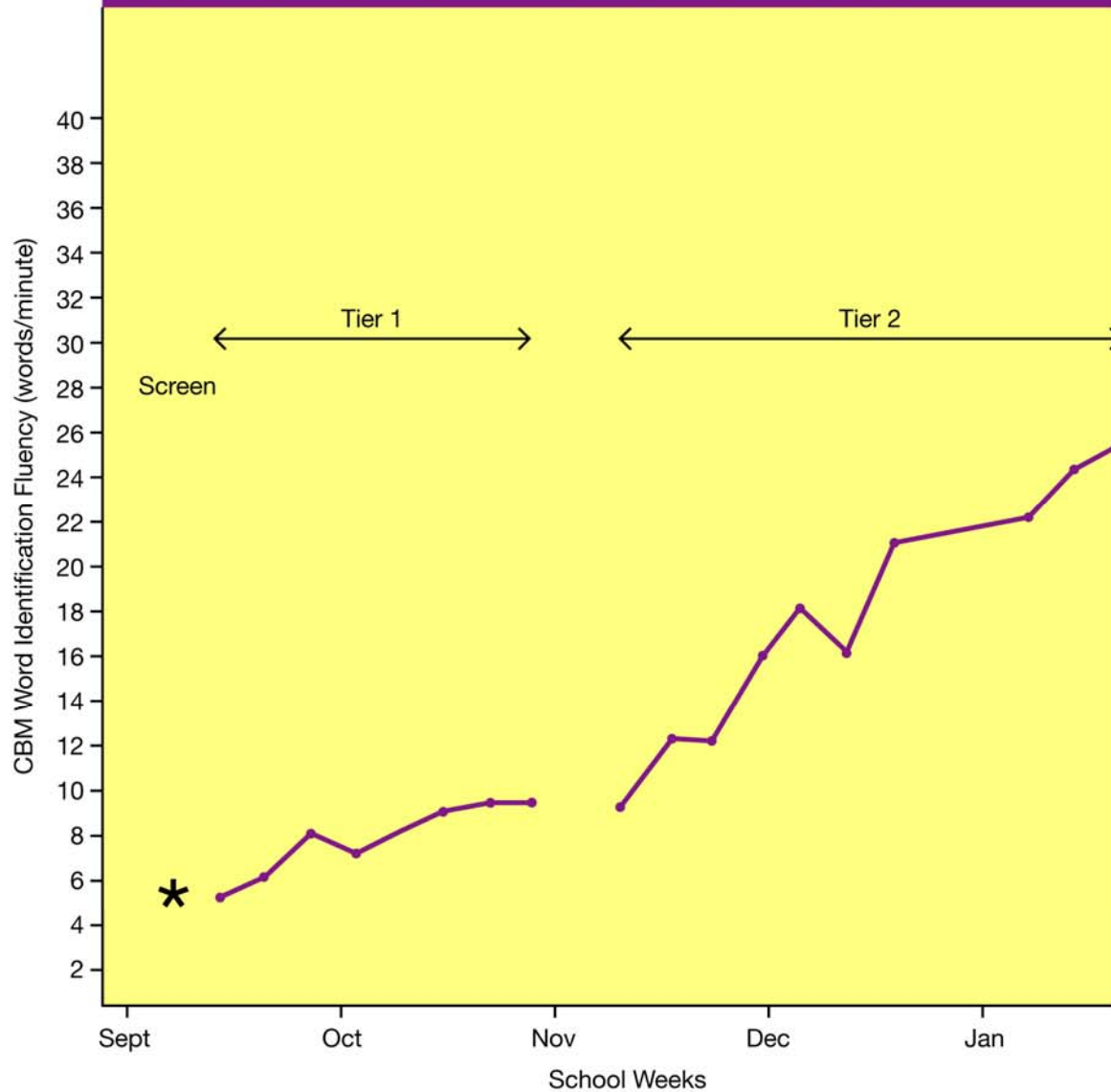
Yes No

### Step 4: Disability Classification/ Special Education Placement

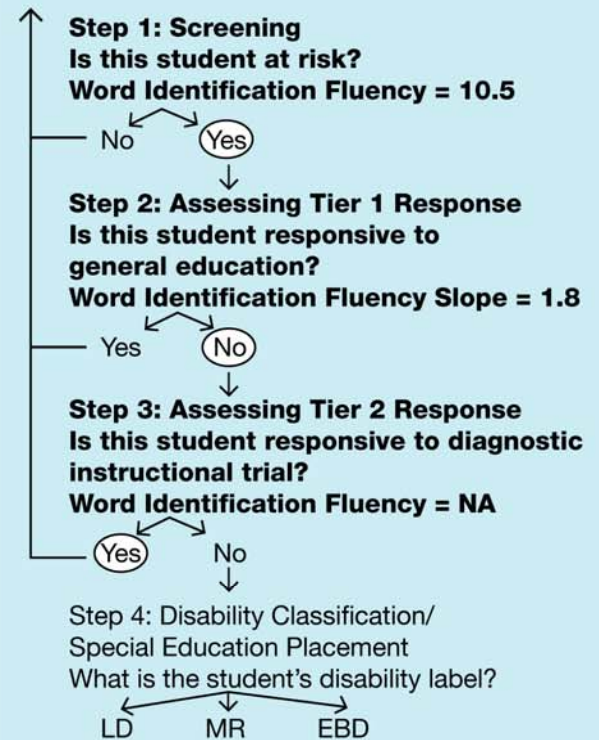
What is the student's disability label?

LD MR EBD

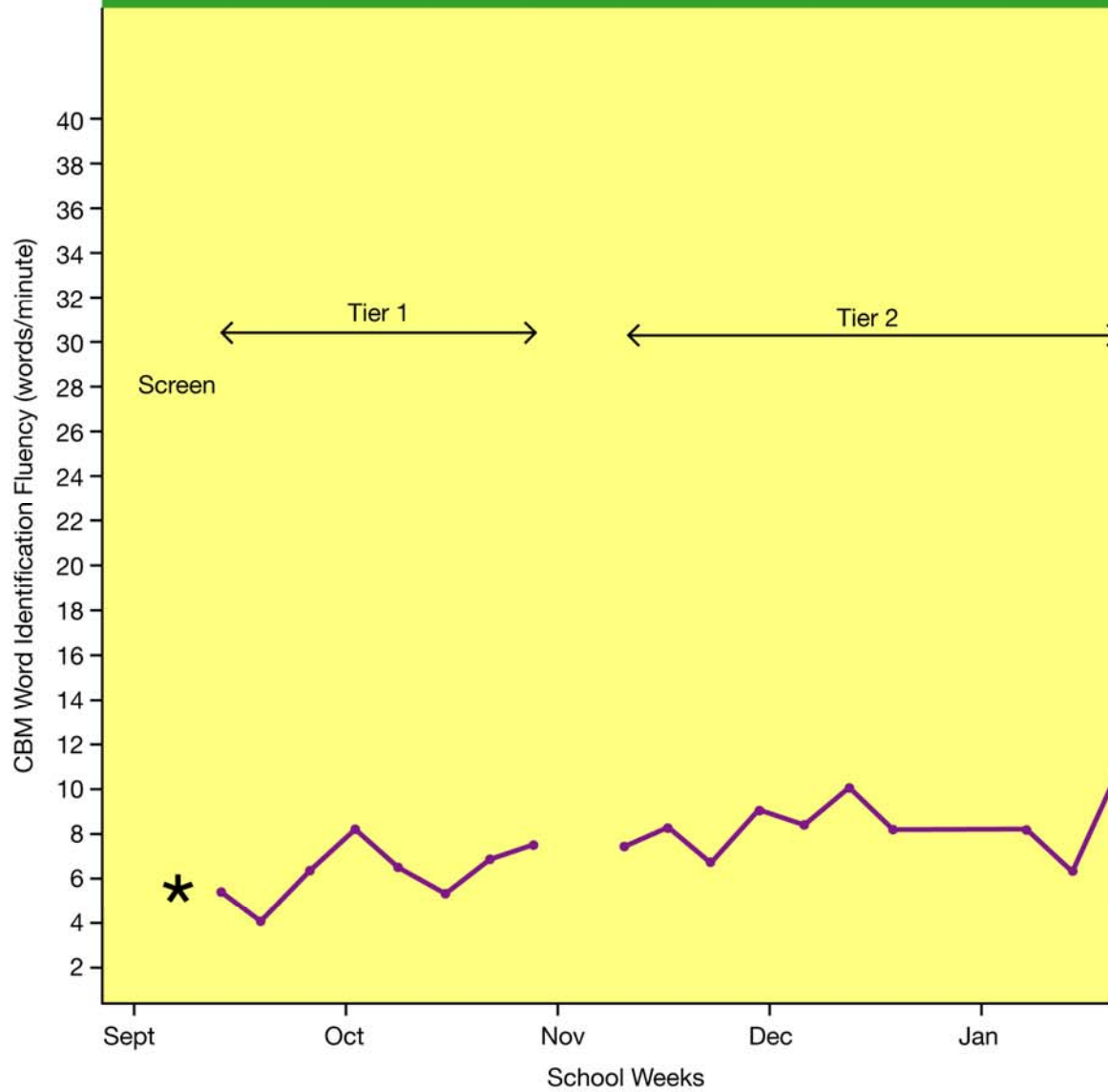
## Case C



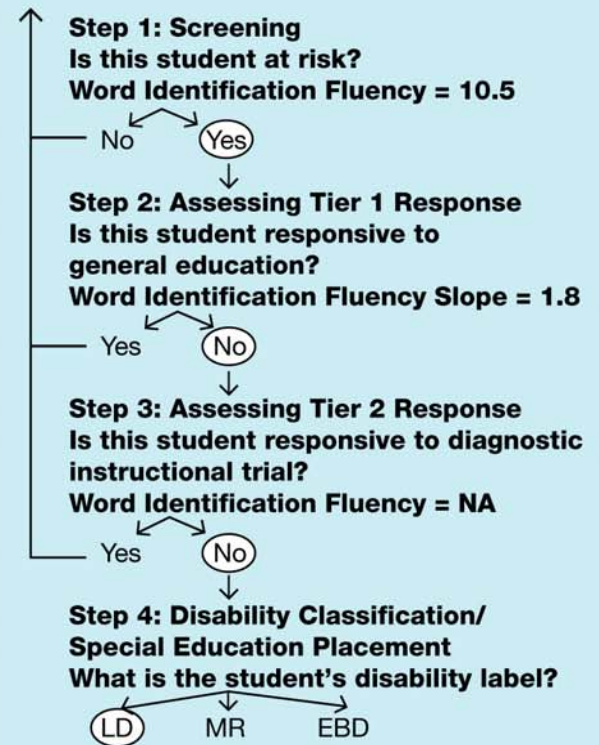
Student Does Not Have a Disability



## Case D



Student Does Not Have a Disability



# Data to Support Efficacy of Tier 2: First-Grade Study

---

- RTI is typically described as having two major functions:
  - To provide early intervention
  - To be a more valid means than IQ-achievement discrepancy of identifying children requiring more intensive services and maybe special education
- Some have researched #1 (e.g., McMaster et al., O'Connor et al., Torgesen et al., Vaughn et al., Vellutino et al.). Few have explored #2 (e.g., Fuchs et al.; Speece et al.).
- The validity of RTI as a method (or as methods) of identification is largely unknown.

# Purpose

---

- To explore:
  - Effects of multiple Tier 1 (classroom) and Tier 2 (pullout) instructional approaches on at-risk children's reading growth in a 9-wk treatment period in fall of 1<sup>st</sup> grade.
  - How responsiveness to the instructional approaches can be used to identify children as LD at the end of 1<sup>st</sup> grade.
  - Effects of alternative methods of LD classification on prevalence and severity.
  - Can characteristic growth patterns of children who are LD and not LD be identified for Tier 1 and Tier 2 instruction?

# Districts, Schools, and Teachers

---

- 2 school districts in Tennessee (urban Metro-Nashville and suburban Williamson County)
- 16 elementary schools serving mostly middle-class families
- 42 first-grade teachers assigned randomly within schools to PALS ( $n = 21$ ) and No-PALS ( $n = 21$ )

# Identifying “At-Risk” Students

---

- All first-grade students in each of the 42 classes were screened using:
  - RLN (CTOPP)
  - CBM Word Identification Fluency
  - Teacher judgment
- The 6 lowest students per class on one or both measures, and also judged as such by the teacher, were designated “at risk.”



# Instructional Groups and Outcome Measures

---

- 1 from the “high” group and 1 from the “low” group assigned randomly to *Fall Tutoring* ( $n = 84$ ); 1 and 1 to *Spring Tutoring—Maybe* ( $n = 84$ ); 1 and 1 to *No Tutoring* ( $n = 84$ ). Total  $N = 252$ .
- Short-term PM on WIF for 18 weeks of 1<sup>st</sup> grade
- Outcome assessment in April of 3<sup>rd</sup> grade
  - Untimed decoding (WRMT Word Attack)
  - Untimed word identification (WRMT WID)
  - Reading comprehension (WRMT Passage Comprehension)
- A variable for RD at the end of 3<sup>rd</sup> grade was created based on performance below a standard score of 85 on the WRMT measures. Complete records for 180 children.

# Example of a WIF Probe

---

## List 1

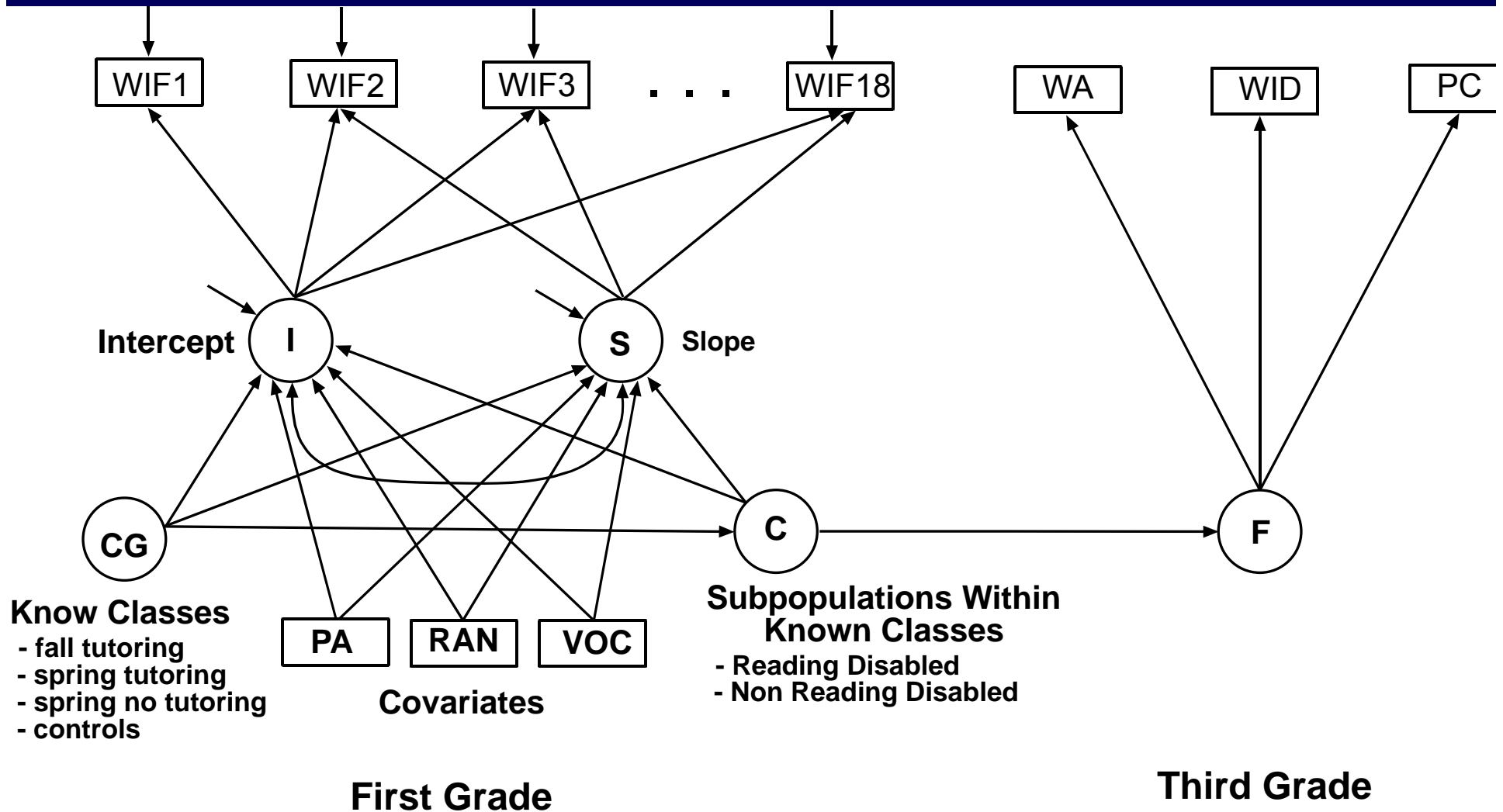
of	always	story
on	does	south
from	need	half
all	light	held
some	almost	table
them	kind	miles
him	better	that's
may	name	women
down	several	town
called	living	force
our	across	green
used	really	surface
come	means	coming
still	able	ask
life	book	books
between	inside	warm
few	anything	story

# Evidence-Based Tutoring

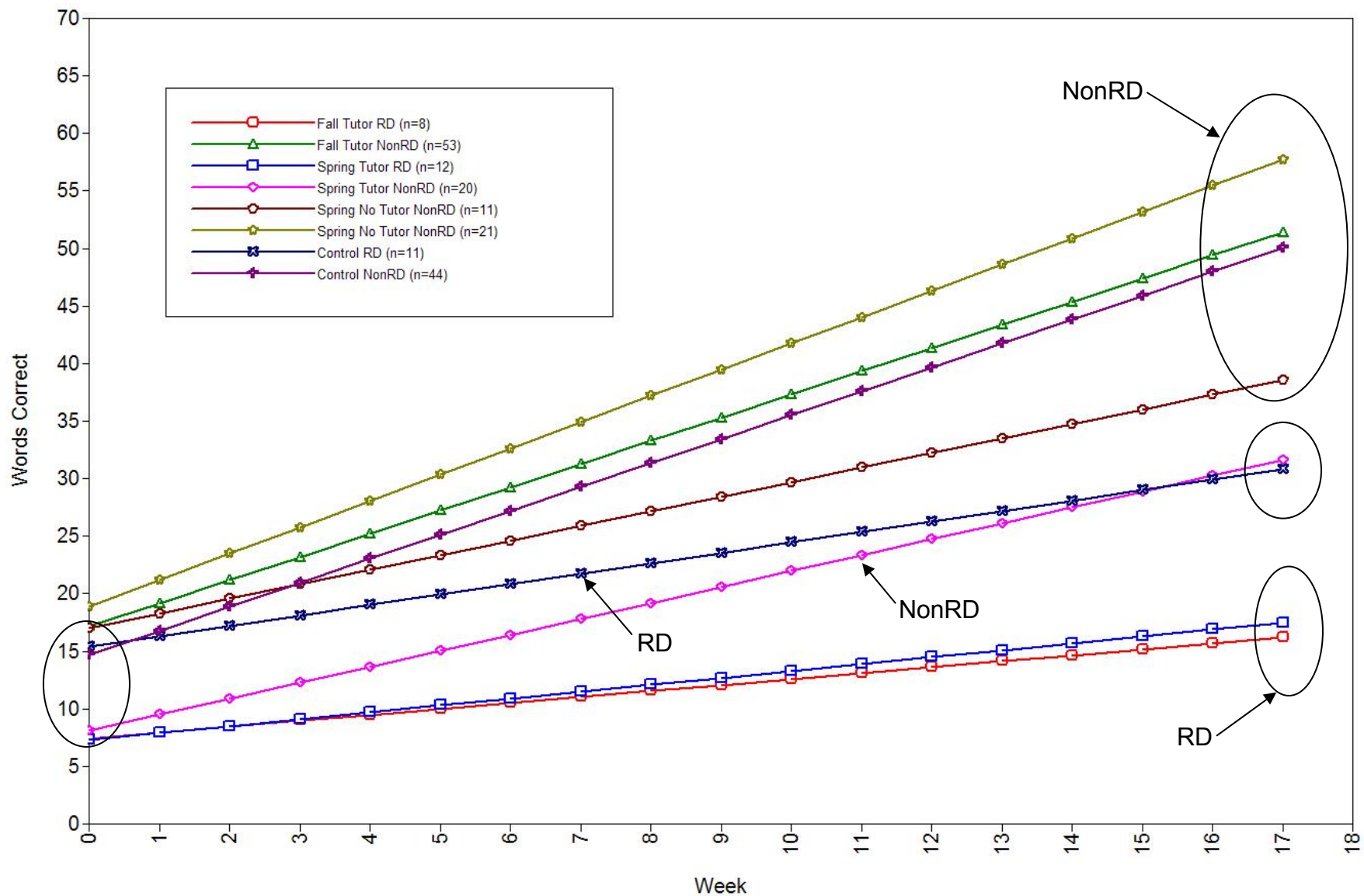
---

- Tutoring
  - Letter-Sound Recognition
  - Phonological awareness and decoding
  - Sight Words
  - Fluency
- Four Groups
  - Fall Tutoring (n=61)
  - Spring Tutoring for Nonresponsive Children (n=32)
  - Spring No Tutoring for Responsive Children (n=32)
  - Controls (No Tutoring, n=55)
- Sessions
  - Conducted by research assistants
  - 2-4 students per group
  - 4 sessions/week
  - 45 minutes/session
  - For a total of 36 sessions of tutoring

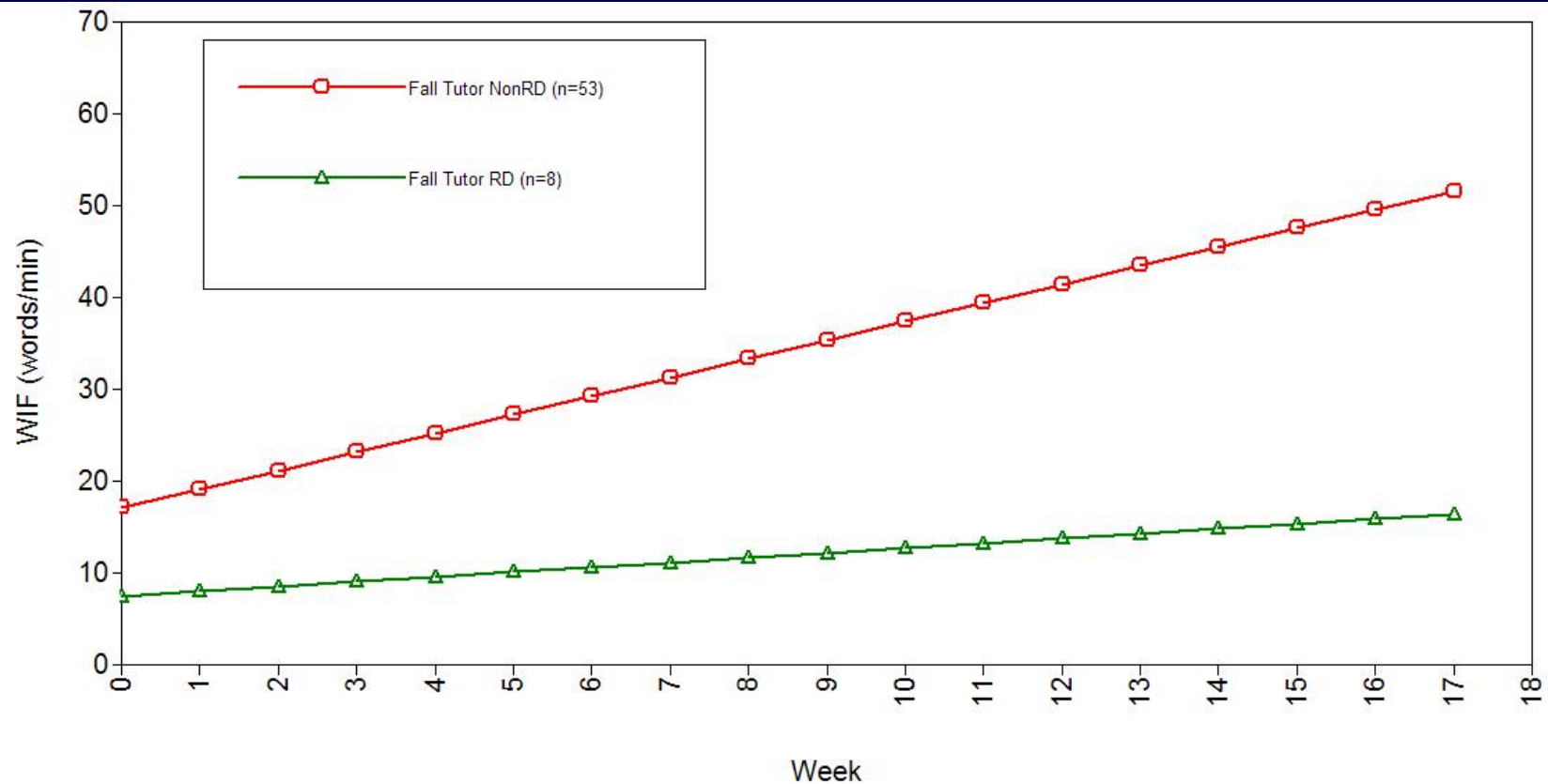
# Growth Mixture Model



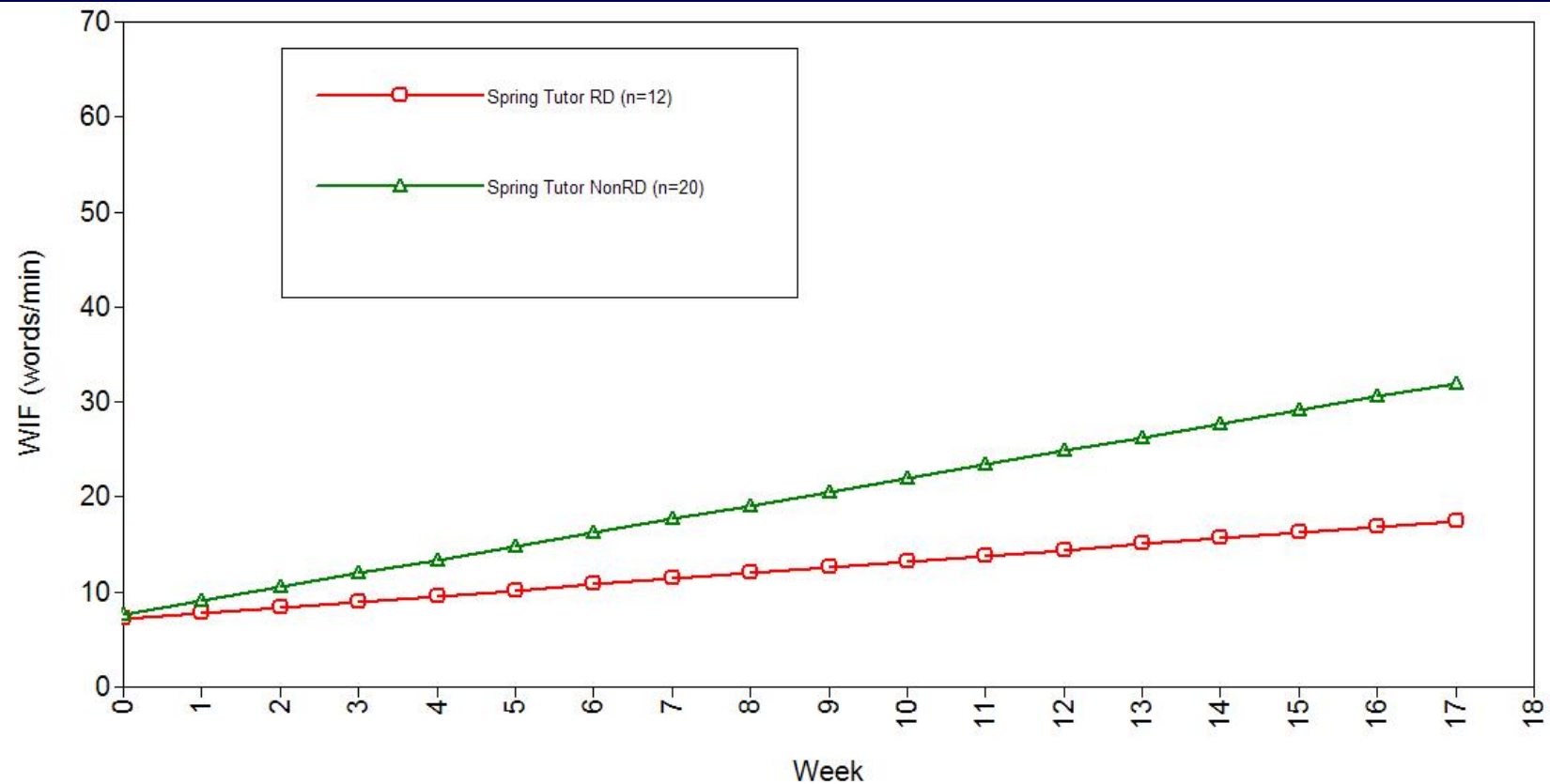
# Growth Patterns Associated with LD



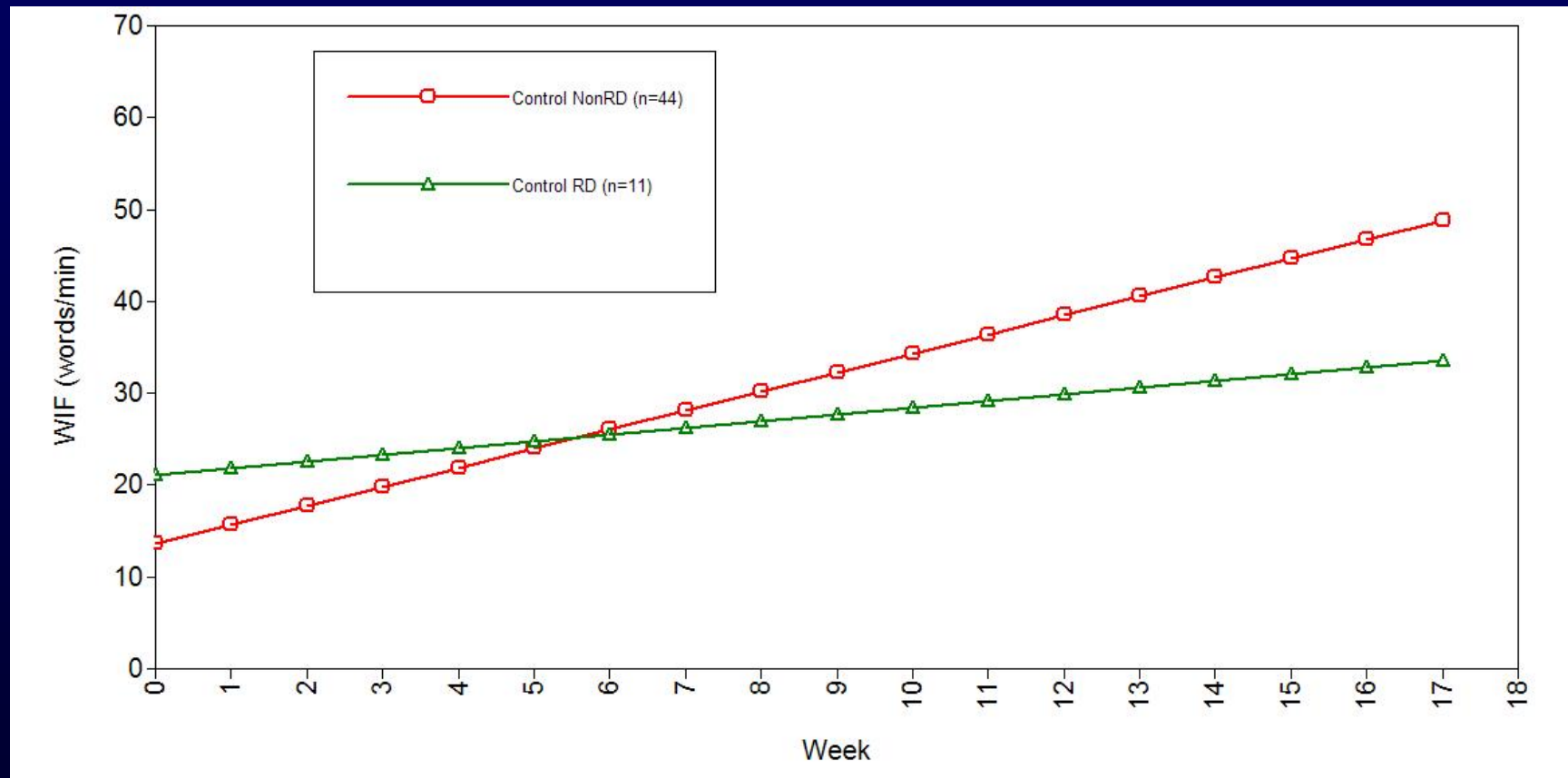
# Growth Patterns Associated with Fall Tutoring



# Growth Patterns Associated with Spring Tutoring



# Growth Patterns Associated with Control (No Tutoring)





# LD Rates by Group

---

Group	Within Group LD Rate (percentage)
Fall Tutoring	13% 8/61
Spring Tutoring Yes	38% 12/32
Spring Tutoring No	0% 0/32
Control	20% 11/55

# Can characteristic growth patterns of children who are either LD and not LD be identified for Tier 2 instruction?

---

- Characteristic LD and nonLD growth curves associated with third grade RD were identified for first graders using the progress monitoring measure.
- The percentage of children identified as LD varied as a function of group in a predictable fashion.
- Identification of children unresponsive to Tier 1 instruction (spring tutoring) yielded a more severe group of children entering into Tier 2 instruction and as expected a higher percentage of children who were unresponsive to Tier 2 instruction. However the number of children receiving Tier 2 tutoring was roughly half that receiving fall tutoring.
- There is reason to predict that response to Tier 1 and Tier 2 instruction can accurately identify children unresponsive to quality instruction while still identifying a manageable group of children entering Tier 2 instruction.

# Questions