



# ***Assessing Reading Skills in Young Children: The TBALL Project***

***(Technology Based Assessment of Language and Literacy)\****

**Elaine Andersen**

**USC**

**Patti Price**

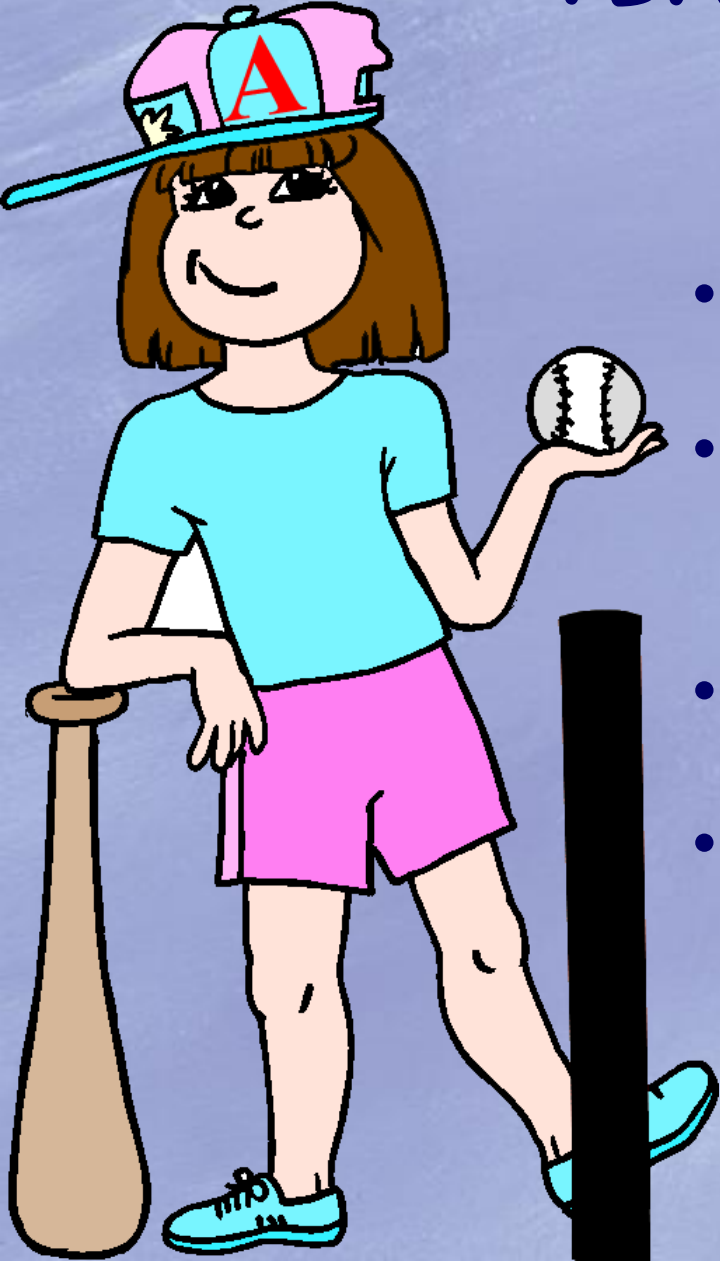
**PPRICE Speech and  
Language Technology**

***CRESST Conference, UCLA***

***Sept 9th, 2004***

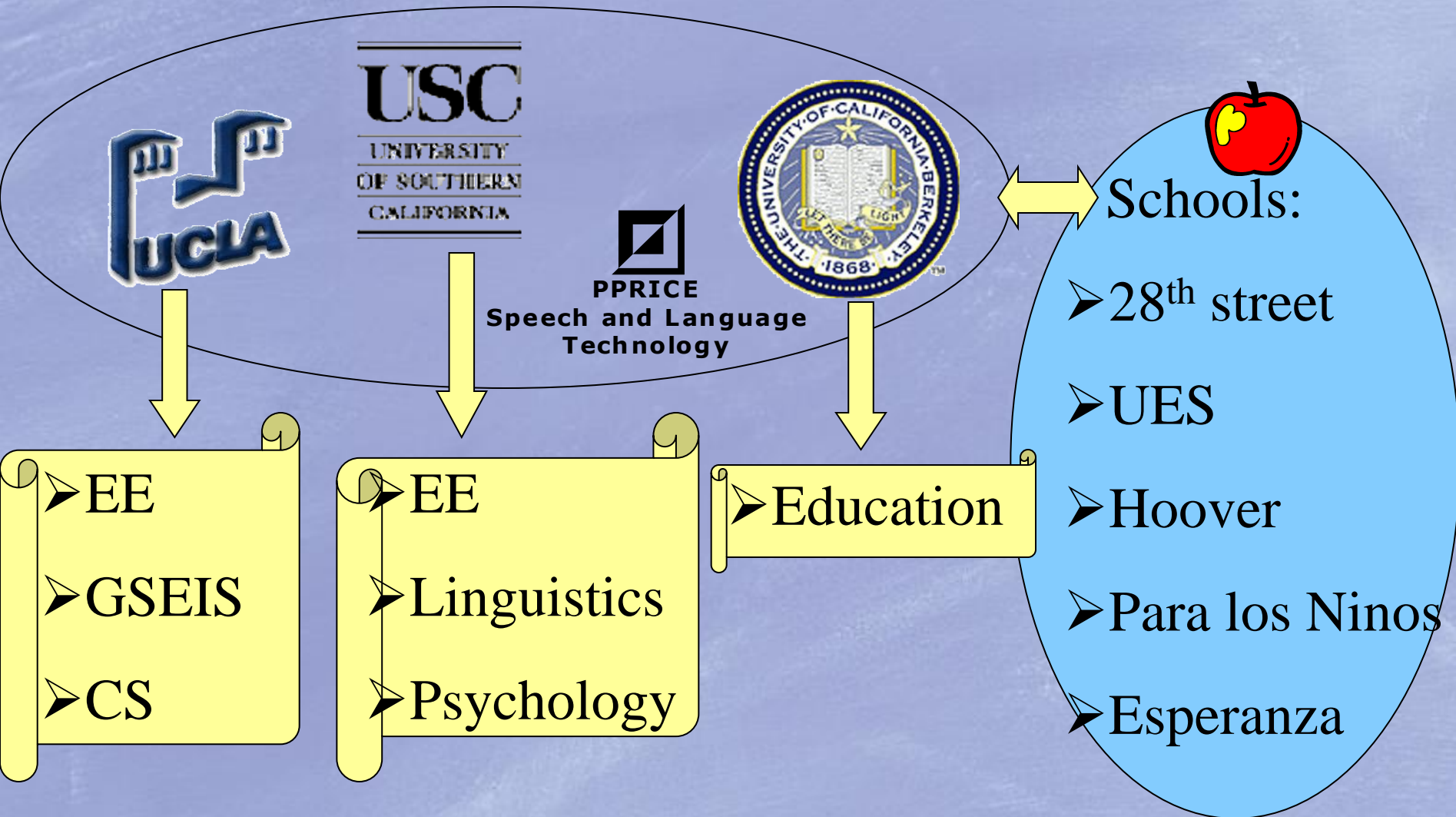
***\*Funding by NSF is gratefully acknowledged.***

# TBALL Overview



- TBALL people and goals
- Challenges (with examples)
- Approach and progress
- Plans

# TBALL Team





# TBALL Team

- **UCLA:** Alwan (PI), Baker (co-PI), Bailey, Boscardin, Heritage, Muntz, Zaniolo
- **Berkeley:** Pearson
- **USC:** Andersen, Narayanan (co-PI)
- **Consultant:** Patti Price
- **Students:** from all three sites
- **Teachers:** RETs
- **Advisory Board:** Neumeyer, Picheny, Rueda, Seda





# **TBALL Specific Aims**

## **Develop assessment system and tools**

- **Helpful for teachers**
- **Test mono and multi-lingual students consistently**
- **Automatically score, analyze K-2 children**

**Investigate emerging literacy measures that are reliable indicators of later academic performance**



# **Why Technology-Based Assessment?**

- **Teacher time constraints**
- **Teacher knowledge constraints**
- **Attractive activity for children**
- **Assessment tailored to individual students needs**
- **Valid, reliable information about students' progress and needs**



# **Components of Assessment**

**Present selected test materials**

**Measure/score collected responses**

**Analyze and adapt to responses**

**Monitor progress, compare, experiment**

**Displays to help teachers make decisions**

**Resources for teacher development**



# **Sample Challenges**

- \* What materials to present and how?**
- \* How to adapt speech recognition to children's speech?**
- \* How to diagnosis discrepancies arising from**
  - Pronunciation differences**
  - Language exposure differences**
- \* How to detect distinct learner profiles?**  
**(Displaying data for different groups and needs)**





# **What Material to Present?**

**Many different aspects of reading skills**

Phonemic Awareness

Letter-sound knowledge, Blending, Spelling

Word Recognition, Rate and Accuracy

Morphology, Syntax, Comprehension

**How to diagnostically assess all aspects within the focus span of a young child?**



# **.. And how to present it??**

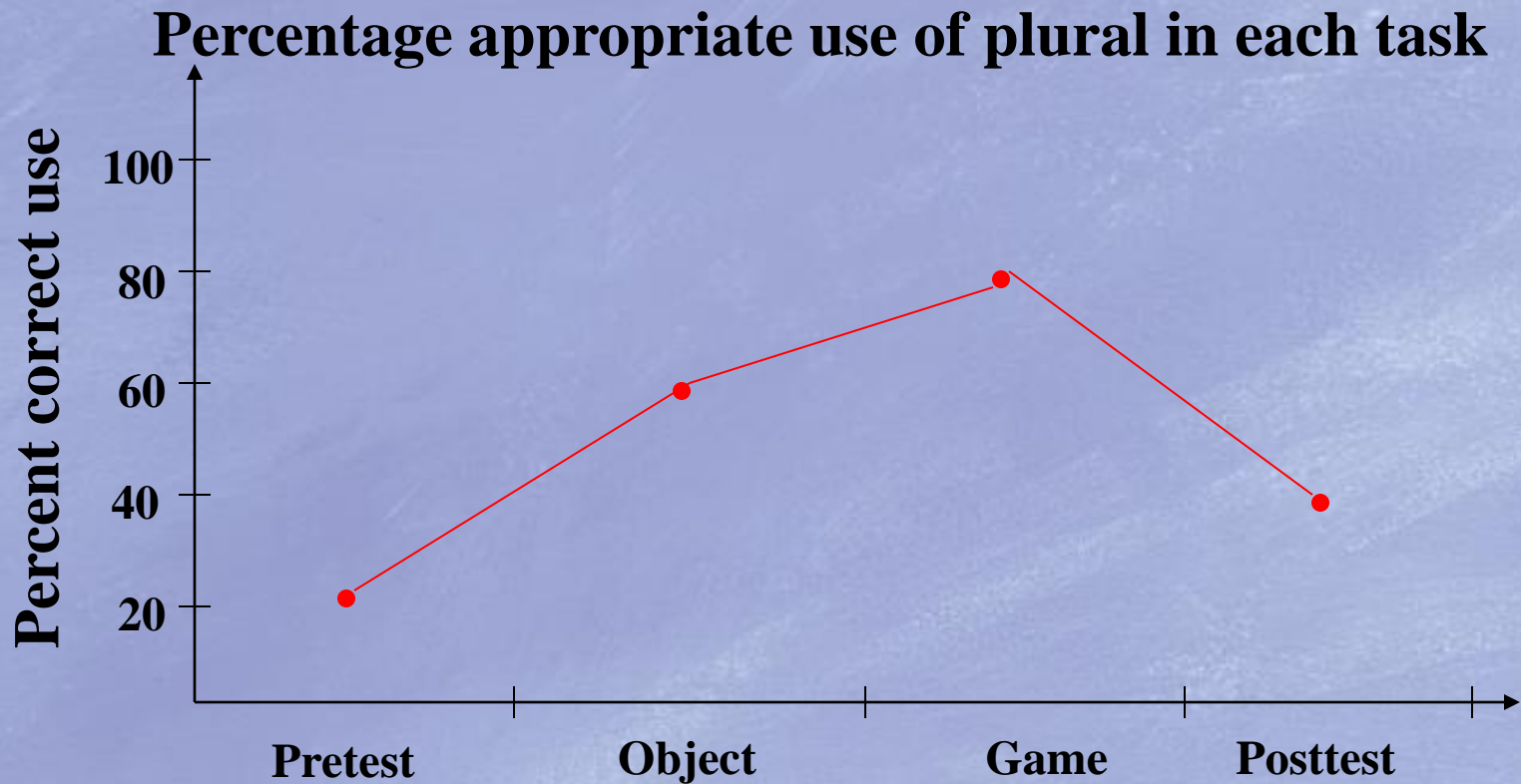
**Children's demonstration of language & cognitive skills is highly variable across contexts**

**Researchers need to be sensitive to ecological validity of procedures**

**How will our collection technique affect the data?**

**Will it disadvantage some children in the measures?**

# Example of Presentation Differences: Hecht's Results

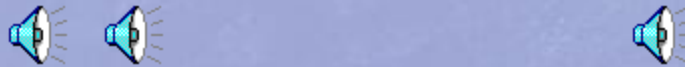


# Speech Recognition Challenges

**Shorter vocal tract lengths,  
higher pitch**



**Significant intra- and inter-speaker  
variability**



**Significant variability**

- **Different linguistic backgrounds**
- **Misarticulations**
- **Signal to noise ratio**







# Reading Error or Pronunciation Difference?

How do we know that reading is correct? /k aw/

- A misreading of 'car' (saw first letter and guessed)
- Or, a misarticulation/idiolect (can't say 'r')
- Or, possibly a dialect/accent issue (/jh eh s/ for 'yes')

**We don't know what the word is**  
**unless we know something about the system**

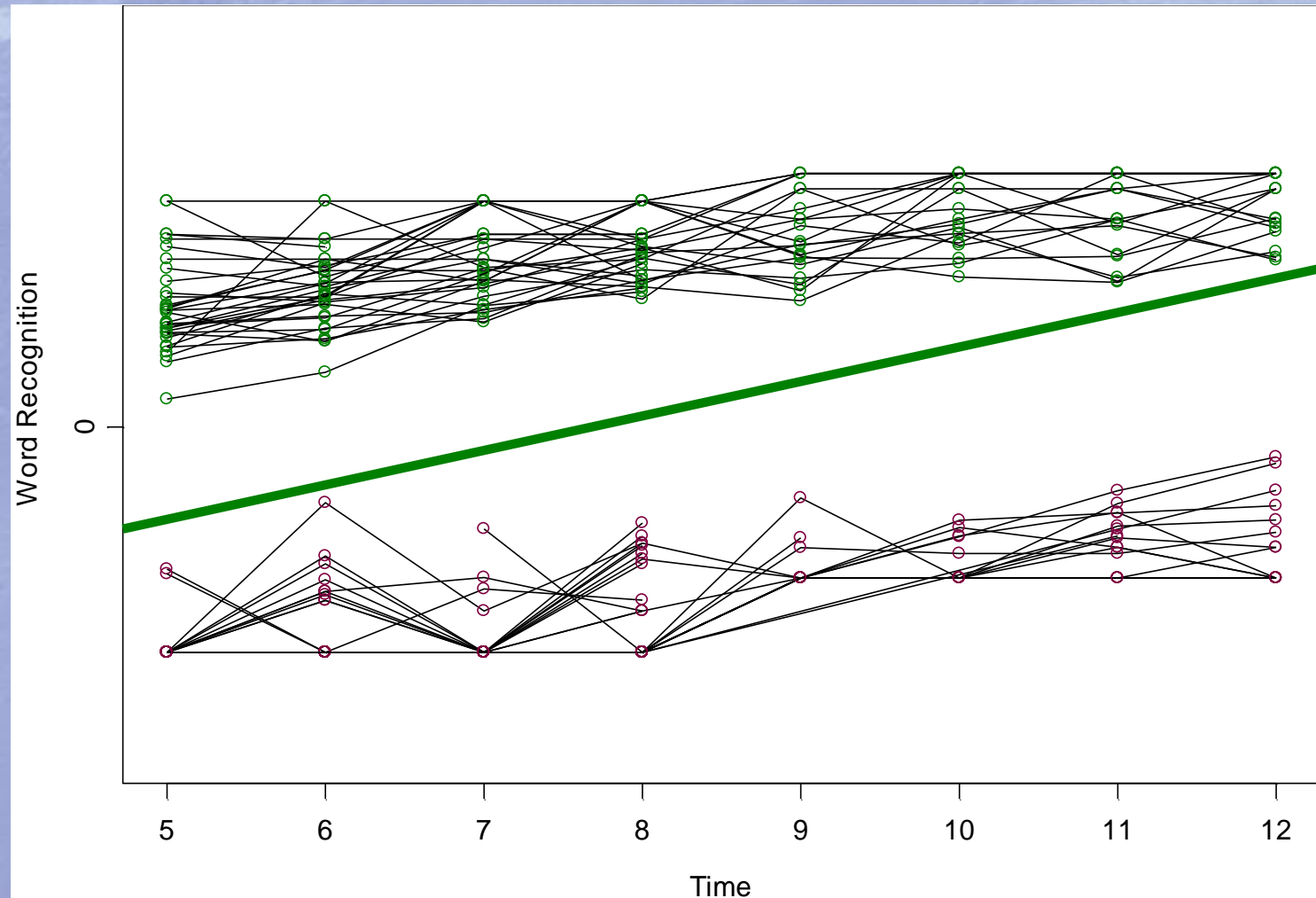


# What marks “Hispanic accent” in English?

	In Spanish, compared to English
<b>Phonetics</b>	<p>p t k closer to Eng b d g than to p t k</p> <p>s z n t d: tongue on teeth, not behind them</p> <p>Sounds missing: th, oy, etc.</p>
<b>Phonology</b>	<p>s+ptkbg only across syllables</p> <p>Distinctions like ‘bit-beat’ not made</p>
<b>Literacy</b>	<p>Words spelled ‘y’ pronounced ‘j’, (by some)</p> <p>Words spelled ‘i’ pronounced ‘ee’, etc.</p>
<b>Exposure</b>	<p>May be more likely to hear much BEV</p>



# Learner Profiles



**Individual data is 'messy', but the 'average' line hides the two distinct groups of learners.**





# **The Biggest Challenge**

**Multidisciplinary collaboration**

**To solve these challenges requires**





- **Engineering**
- **Psychology, linguistics, psycholinguistics**
- **Experts in reading, assessment, datamining**

**Starting from such different points of view**

- **Difficult to integrate into one coherent view**
- **Also the biggest opportunity**
- **And probably essential**



# Samples For You To Rate!

Target	Rating, Explanation
put 	Wrong, confuses letter b and letter p Wrong, not paying attention Right, Hispanic accent
watch 	Wrong, doesn't know -tch Right, Hispanic accent
cold 	Wrong, confuses short and long vowels Right, just child's way of pronouncing word
full 	Wrong, confuses short and long vowels Right, just child's way of pronouncing word

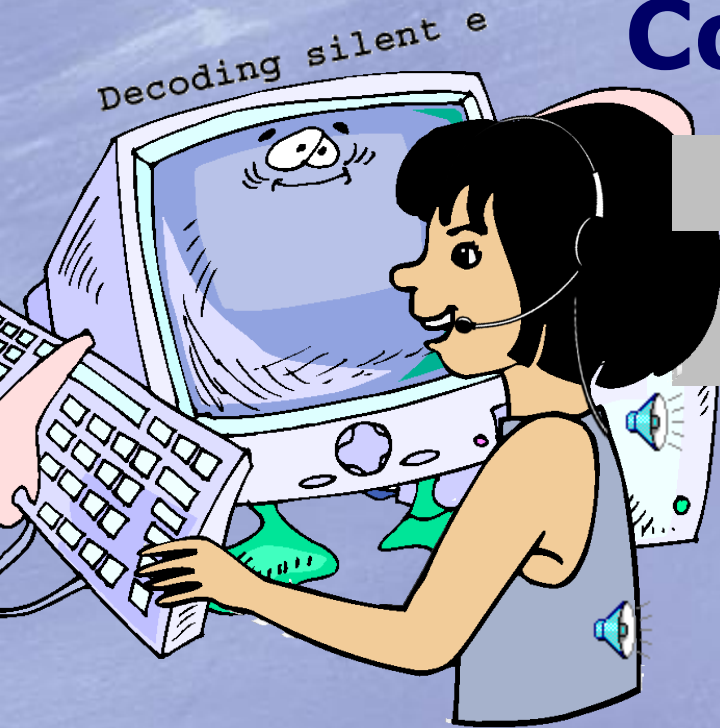
# Components

**\* Present auditory, text, graphical stimuli**

A cartoon illustration of a purple dinosaur with a pink and yellow star pattern on its back, standing on a green grassy bank next to a blue body of water. The dinosaur is holding a white rectangular sign with a green border. On the sign, the word "eighteen" is written in red. The dinosaur has a surprised expression on its face.

**eighteen**

# Components



- \* **Present** auditory, text, graphical stimuli

- \* **Measure** decoding, comprehension skills



# Components



- \* Present auditory, text, graphical stimuli

- \* Measure decoding, comprehension skills

- \* Score, analyze, and adapt to responses

(Query-based datamining: monitor progress, compare, experiment)

Which improved most?

Which data set performs best?

Who is teacher C?

# Components

- \* **Present** auditory, text, graphical stimuli

- \* **Measure** decoding, comprehension skills

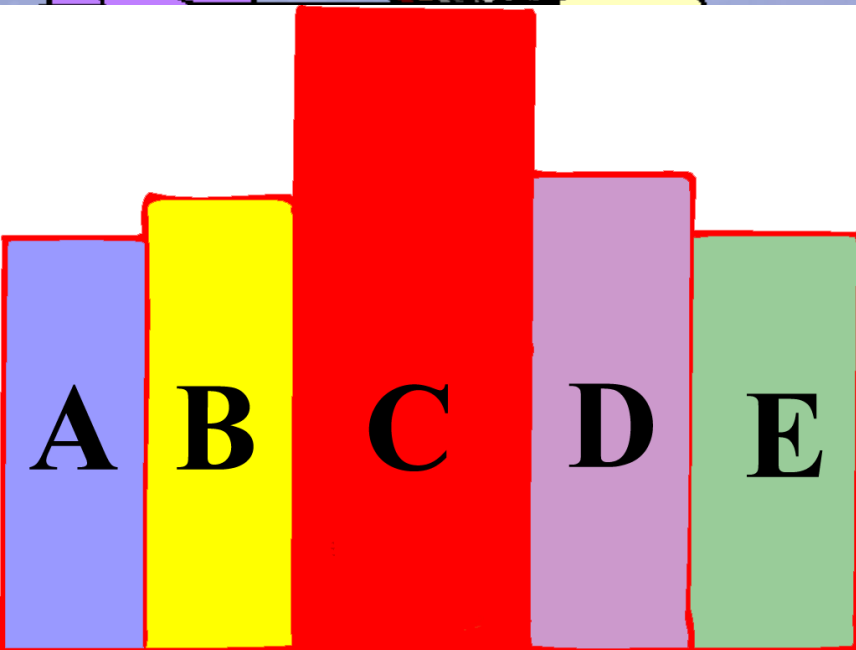
- \* **Score, analyze, and adapt** to responses



**(Query-based datamining: monitor progress, compare, experiment)**

**(Displays for teachers to combine data to help make decisions)**

**(Resources for teacher development)**





# Development Process

- **Task specifications**
- **Write items**
- **Teacher review**
- **Try with students ( Instructional utility)**
- **Design interface**
- **Try interface**
- **Teacher review**
- **Displaying results**



# Sampling Domain

**Core that all do, sampling of rest**

**Focus on high frequency items**

**Oral Language**

**Name letters**

**Say Sound of letters**

**Decoding**

**Hear sound, point to letter**

**Rhyming, blending**

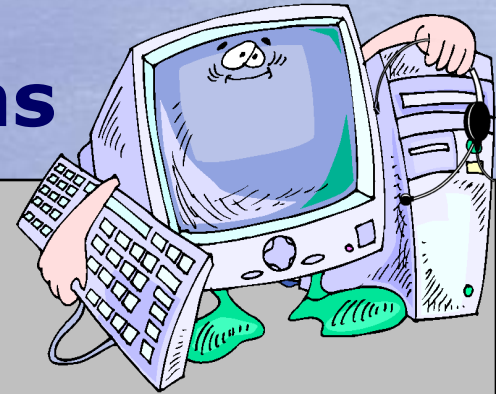
**Fluency**

**Reading words, timed and not**

**Naming images, timed and not**

**Comprehension**

**Reading sentences, and pointing  
to image matching word**







# Fall Battery (K example)

<b>Letter Names</b>	<b>b, k, y, s, j, z + 6 random</b>
<b>Reading (LTS)</b>	<b>d, a, i, s, j + 5 random</b>
<b>Spelling (STL)</b>	<b>p ih v iy z + 5 random</b>
<b>Blending</b>	<b>z+oo, t+ub, s+ix, ch+ick, thr+ee + 5 random</b>
<b>Reading Words</b>	<b>5 fixed, 5-35 random, hi freq words sorted by decodability</b>
<b>Naming Pictures</b>	<b>As above, but with images</b>
<b>Rapid Naming</b>	<b>Words and images, timed</b>



# **Speech Recognition Approach**

**Speaker adaptation techniques**

**Pronunciation modeling**

**Noise robust (front end  
and/or back end)**

**Source and vocal tract  
parameter estimation**



# Sorting Data

**Database design allocates a place to 'put' the collected data and its context, e.g.,**

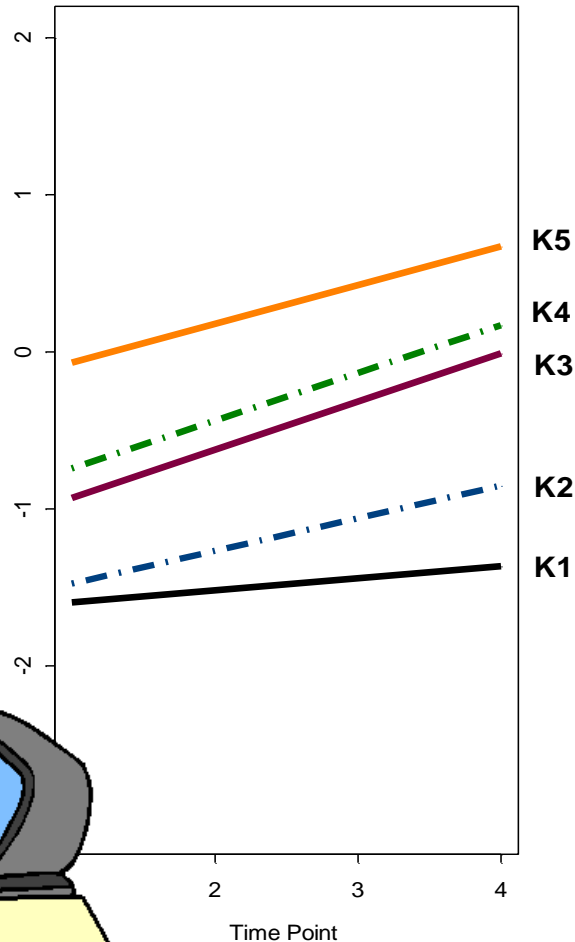
- **Demographic info from parent, date, time, type of test**
- **Data from test**

**Later the data can used for computations, e.g.,**

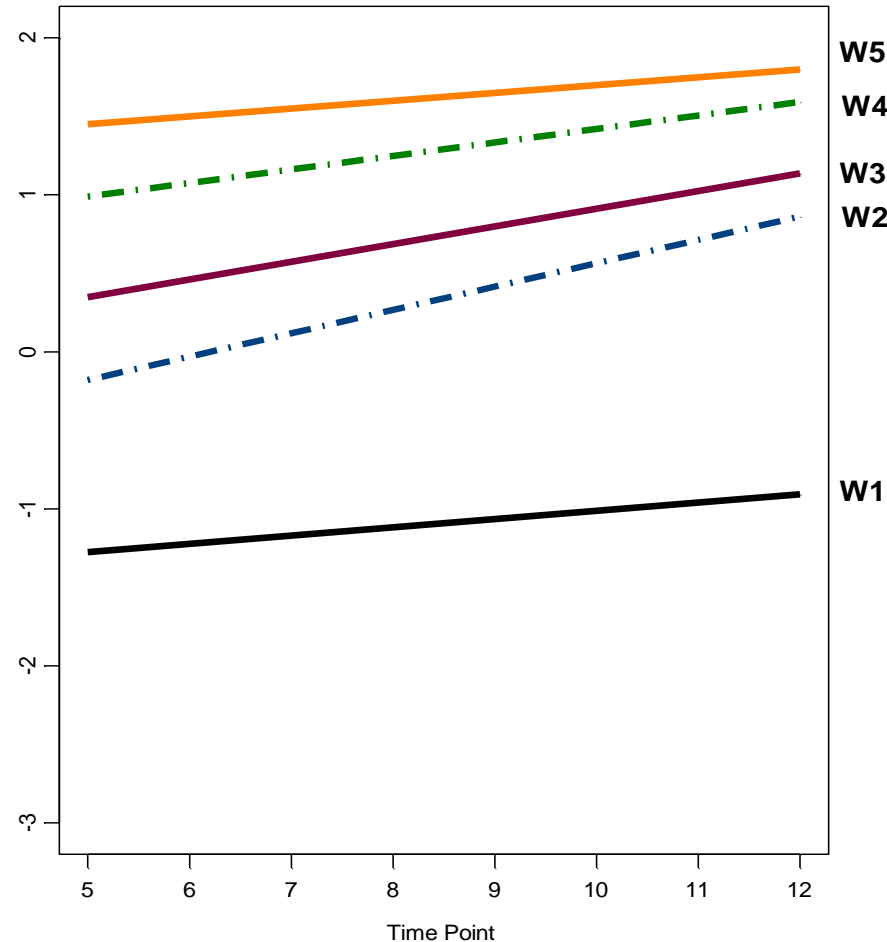
- **Words in isolation correct:  $21/51 = 41\%$**
- **Words in connected text:  $20/36 = 55\%$**
- **75% of native speakers do better in connected text..**
- **Level of accentedness: 70%**

# Example of Data Modeling

Kindergarten Growth (Five Classes)  
full7wtrnl.out

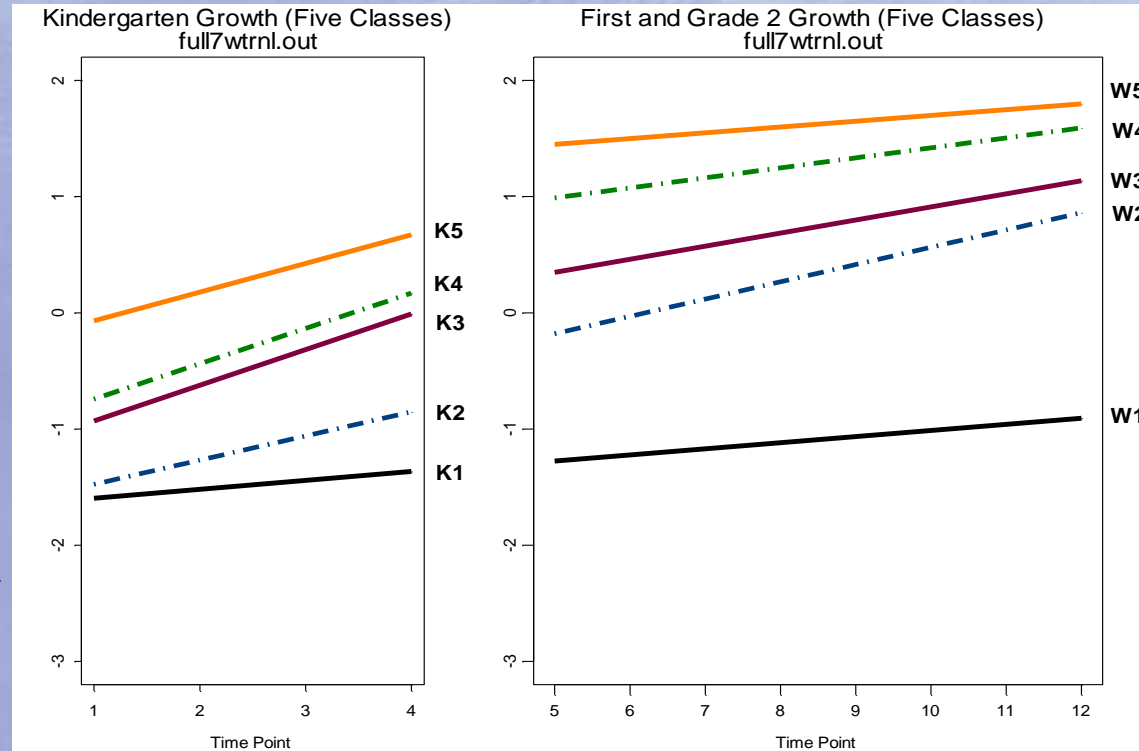


First and Grade 2 Growth (Five Classes)  
full7wtrnl.out





**Growth Mixture Modeling can reveal unobserved heterogeneity in the model**



**Different developmental trajectories are accurately estimated**

**Students who are most at-risk for reading problems can be identified**



# **Plans**

**Content selection plans**

**Data collection plans**

**Database plans**

**Datamining/longitudinal plans**

**Feedback plans**

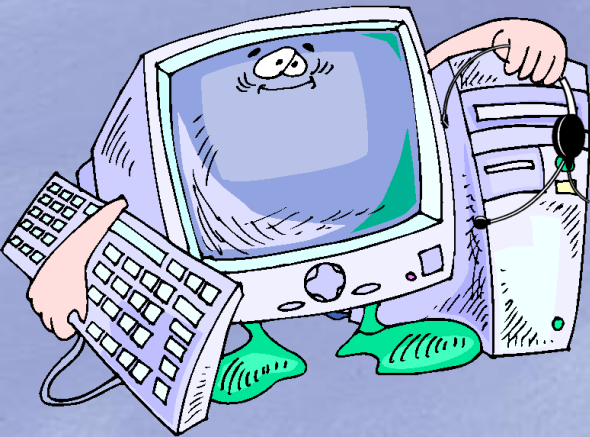
**Longer term plans**

# **Content Selection Plans**

**Refine assessment tasks, materials, and automated techniques based on feedback**

**Address validity, utility, and impact for native and non-native speakers**

**Pilot studies on comprehension and reading in context tasks**





# **Data Collection Plans**

**Train teachers to use the system**

**Deploy in more classrooms each year**

**Further evaluate and refine the ASR system**

**Try assessment with children (native speakers as well as various ELL levels)**

**Get information on teachers' interpretations and evaluation of instructional use**







# **TBALL Specific Aims**

**Develop assessment system and tools**

- **Helpful for teachers**
- **Test students consistently**
- **Automatically score, analyze K-2 children**

**Investigate emerging literacy measures that are reliable indicators of later academic performance**