Usability testing with stakeholders in the development of a formative assessment of oral reading fluency

Alistair Van Moere & Ryan Downey
Pearson, Knowledge Technologies
Overview

- Why Oral Reading Fluency?
- Demonstration
- Pilot & focus groups with stakeholders
Oral Reading Fluency

“The ability to read a text quickly, accurately, and with proper expression.”

National Reading Panel (2000)
Word recognition: is the initial and most simple step in the reading process, yet it is this that differentiates native readers from even highly skilled L2 readers.

(Segalowitz et al, 1991; Carver, 1982)
5 Critical Components of Reading

- Phonemic awareness
- Phonics
- Oral reading fluency
- Vocabulary
- Comprehension

National Reading Panel (2000)
The Thermometer

A dependable indicator of a student’s academic “health” or “illness”

(Hasbrouck & Tindal, 2006)
“No Child Left Behind”

- School districts need to show **Adequate Yearly Progress**

- Oral Reading Fluency may be used as part of the Reading First initiative
National Norms

<table>
<thead>
<tr>
<th>Grade</th>
<th>Autumn Norms</th>
<th>Spring Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Grade</td>
<td>30 WCPM</td>
<td>90 WCPM</td>
</tr>
<tr>
<td>2nd Grade</td>
<td>80 WCPM</td>
<td>120 WCPM</td>
</tr>
<tr>
<td>3rd Grade</td>
<td>110 WCPM</td>
<td>140 WCPM</td>
</tr>
</tbody>
</table>

“At risk” = 1 grade level below norms

“Severe” = 2 grade levels below norms
Pedro had just moved from Mexico when he saw an accident. A little boy had fallen into an open manhole, and now his leg was caught between two pipes. Pedro was just ten. He didn’t think he could rescue the boy alone.
Sample school year

AUTUMN  WINTER  SPRING

Summative Test

Formative Tests

Summative Test

Formative Tests

Summative Test

Formative Tests
Challenges to this approach

- Teachers need training
- One-on-one testing
- Time-wasted: scoring, annotating, and entering data
- Expressiveness not included
- Standardization
Goals of our study

1. Implement automated testing
2. Validate accuracy of automated scoring
3. Involve stakeholders
Demonstration
How It Works

1. Browser initiates the call
2. Server calls the number
3. Student reads for a minute
4. Server displays the score

Callback Number

<table>
<thead>
<tr>
<th>Student</th>
<th>Test</th>
<th>Call</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Doe</td>
<td>4th Grade - Week 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jane Smith</td>
<td>4th Grade - Week 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bob Jones</td>
<td>4th Grade - Week 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>James Dean</td>
<td>4th Grade - Week 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Involving Stakeholders

• Pilots:
  - Two local school districts

• Focus group:
  - San Jose School District office, June 2008
Pilot

Participants:
• 87 students, grades 1 to 3
• 3 reading passages per student (261 passages)

Data:
• Machine scores
• A score from the regular class teacher
• Scores from 3 expert raters (r=.998)
Expert scores vs. Machine scores

Machine Words Correct vs. Expert Words Correct
Expert scores vs. Machine scores

Machine Words Correct

Expert Words Correct
Expert scores vs. Machine scores

- Teacher vs machine, $r=.99$
- Expert vs machine, $r=.99$
Focus Sessions (teachers)

- Immediate feedback
- Saves teaching time
- Email recording to parents

“When can we start using it?”
Focus Sessions (students)

- Enjoyable
- Self-realization

“That’s not how you say that word.”
Conclusions

• Tracks performance against national norms
• Saves teacher time
• Reliable scores
• Recordings involve parents
Future directions

Broader profile of reading ability:

- Percent Accuracy
- Expressiveness (phrasing, pausing, prosody)
- Diagnostic measures, e.g. pronunciation
- Comprehension tasks, e.g. summarizing

Can be adapted to other languages
Thank You!

Questions?

Alistair.VanMoere@pearsonkt.com
Ryan.Downey@pearsonkt.com