Advances in Large Scale Oral Reading Fluency Assessment

March 22, 2008

Ryan Downey, David Rubin, & Alistair Van Moere
Ordinate Corporation, Palo Alto, California
Oral Reading Fluency

• National Reading Panel (2000):
  – “The ability to read a text quickly, accurately, and with proper expression”.
  – “A critical component of skilled reading.”

• “Thermometer analogy”
  – Oral Reading Fluency scores can be used by educators and clinicians as a dependable indicator of a student’s academic “health” or “illness” (Hasbrouck & Tindal, 2006).
Current Approach

Scoring
Mark reading errors
Reading Rate (WCPM)
DIBELS, AIMSweb, etc.

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.
Tracking ORF

- Benchmark tests for all students to identify struggling or at-risk students

Student: Student, Johnny – Oral Reading Fluency Scores (WCPM) - Compared to Class
Rincanada 2nd Grade – 2007 Academic Year
Tracking ORF

- Progress Monitoring tests for tracking students who require special instruction/attention

Student: Student, Johnny – Oral Reading Fluency Scores (WCPM) - Compared to Class
Rincanada 2nd Grade – 2007 Academic Year

<table>
<thead>
<tr>
<th></th>
<th>Fall Benchmark</th>
<th>Winter Benchmark</th>
<th>Spring Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
</tbody>
</table>

Words Correct Per Minute (WCPM)
Challenges

• **Important**
  – Recognized by the National Reading Panel (2000) as one of five critical components of skilled reading
  – Reporting required by NCLB (2001) for demonstrating “adequate yearly progress”

• **Time Consuming**
  – Individually administered
  – Teachers need training to score accurately
  – Teachers spend time scoring, annotating, and entering data
Oral Reading Assessment

Please read this passage.

Teacher  Student
Oral Reading Assessment

Here is my reading…

Teacher  Student
Oral Reading Assessment

Here I am reading…

Teacher

Student
Goals of the Study

Freeing up teachers’ time with:

• Automatic administration and scoring
  – Telephone with booklet
  – Reading Rate: words correct per minute (WCPM)

• Accurate scoring
  – How well can machine scores predict human scores?

• Data management and reporting program
The web application
Web delivery
Web delivery

![San Jose Summer Program](image)

<table>
<thead>
<tr>
<th>Student</th>
<th>Test</th>
<th>Call</th>
<th>Score</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Smith</td>
<td>4th Grade - Week 2</td>
<td><img src="image" alt="call" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bob Jones</td>
<td>4th Grade - Week 3</td>
<td><img src="image" alt="dialing" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>James Dean</td>
<td>4th Grade - Week 4</td>
<td><img src="image" alt="call" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Web delivery

![San Jose Summer Program Test Delivery](image)

<table>
<thead>
<tr>
<th>Student</th>
<th>Test</th>
<th>Call</th>
<th>Score</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Smith</td>
<td>4th Grade - Week 2</td>
<td>📞</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bob Jones</td>
<td>4th Grade - Week 3</td>
<td>📞</td>
<td>📞</td>
<td>dialing...</td>
</tr>
<tr>
<td>James Dean</td>
<td>4th Grade - Week 4</td>
<td>📞</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Web delivery

Albert was a goldfish in a bowl. He ate a breakfast of green and brown flakes each morning. Then he watched the children go off to school.

Albert hated being stuck in his bowl because he could only swim around in circles. He'd rather go to school. Poor Albert couldn't even read a book. The pages would get soaked!

Albert was quite a smart fish. He could do flips under water. He could spell his name in the pebbles on the bottom of his bowl. No matter how brilliant Albert was, though, he still had a problem. Only the cat spoke to him. And the cat was not particularly nice to him.
Web delivery
Web delivery

San Jose Summer Program
Bill's Class

Callback Number
4152005502

This number will be called by the Ordinate System to deliver a test. Click "Test" to verify the phone number by initiating a callback.

<table>
<thead>
<tr>
<th>Student</th>
<th>Test</th>
<th>Call</th>
<th>Score</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Smith</td>
<td>4th Grade - Week 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bob Jones</td>
<td>4th Grade - Week 3</td>
<td>completed</td>
<td>Accuracy 81%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WCPM 85</td>
<td>re-test</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Expressiveness 3</td>
<td></td>
</tr>
<tr>
<td>James Dean</td>
<td>4th Grade - Week 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Goals of the Study

Freeing up teachers’ time with:

• Automatic administration and scoring
  – Telephone with booklet
  – Reading Rate: words correct per minute (WCPM)

• Accurate scoring
  – How well can machine scores predict human scores?

• Data management and reporting program
"It’s supposed to rain tomorrow isn’t it"
Alignment

waveform

spectrum

words

segmentation

75-90 Words/Min

5.8 Phones/Sec
Development

**PLAN**
- Define Test
- Construct
- Design Test
- Create Item Specification

**DEVELOP**
- Write Items
- Review Items
- Record Items
- Create Data Collection System

**SCORE**
- Collect Data
- Train Speech Recognizer
- Create Models
- Generate Scores
Optimization

Reading experts → Human WCPM

Machine scoring → Machine WCPM

N = 183,000 individual reading samples
Validation

Reading experts → Human WCPM

Machine scoring → Machine WCPM

r = .98

N = 475 individual reading samples
Correlation of WCPM

- $n = 475$
- $r = 0.98$

Grades 1, 3, and 5 together
Conclusions

Automating administration and scoring of oral reading fluency performance using ASR:

- Reduces time spent scoring and administering (e.g., teachers can use proctors or aides)

- Produces scores as reliable as trained human raters (0.98)

- Preserves digitized readings for replaying later from web page testing record
Conclusions

Future directions

– Broader profile of reading ability:
  • Percent Accuracy
  • Expressiveness (phrasing, pausing, prosody)
  • Diagnostic measures

– Homework assignments/learning modules
Thank You

rdowney@ordinate.com
avanmoere@ordinate.com

More information at: www.ordinate.com