Relationship Between Theoretical and Empirical Estimates of Text Complexity for Two Different Reading Item Types

A. Jackson Stenner, Carl W. Swartz, Harold Burdick, Donald S. Burdick, Sean T. Hanlon

MetaMetrics research on task types has been ongoing for several years. Early criticism of the native Lexile item Type centered around the length of the source text (see Appendix A). Typically a native item includes a paragraph or less of text followed by an embedded completion sentence that calls on the reader to choose one of four words to complete the embedded sentence. This item type clearly requires understanding of inter-sentential relationships but may involve too little text to make coherence and cohesion relations relevant to the response process (Appendix B).

In response to this criticism MetaMetrics conducted a study involving 475 article length passages. Empirical text complexities were computed using machine generated cloze items (Appendix C). Then empirical text complexity measures were correlated with the theoretical text complexity measures provided by the Lexile Analyzer. After correction for reliability 92% of the variance in empirical text complexity measures was explained by the Lexile analyzer measures. Attempts to find other variables to account for the 8% unexplained variance were unsuccessful. Collaborations with Coh-Metrix and ETS are under way to find explanatory variables beyond the word level and sentence level variables.

One explanation for the failure to find paragraph or discourse variables that account for the unexplained variance is that the task type (machine generated cloze) used to estimate empirical text complexity is decidedly intra-sentential, in that, the information needed to choose among four word choices to complete a sentence rarely extends beyond the sentence boundary.

In response to this criticism MetaMetrics conducted a recent study to look at how differently an intra-sentential task type (machine generated cloze) and an inter-sentential task type ordered 63 passages from Pearson Educational Measurement’s PASeries reading test (Grades 2-9). Please see Table 1 for the raw correlations (upper right) and the error corrected correlations (lower left) among three text complexity variables: (1) theoretical text complexity, (2) empirical text complexity based auto-generated semantic cloze, and (3) empirical text complexity based human generated embedded extended passage sentence completion items. The diagonal includes the reliabilities for each of the variables. The correlation between the two item types is $r = .87$ and corrected for measurement error is $r = .95$. Thus, the two different task types, one human generated and the other machine generated, order the 63 passages similarly. A further analysis showed no interaction with text type (narrative, expository or hybrid).

In conclusion, MetaMetrics research shows, (1) no evidence for a third variable that improves prediction of empirical text complexity beyond the word and sentence variables, (2) no evidence that genre (narrative or informational) is an important distinction to make when using a text complexity equation, and (3) no evidence that intra-sentential tasks order texts as to empirical complexity wildly differently than inter-sentential tasks.

We emphasize that the failure of any variable to account for variance beyond word and sentence level variables does not mean that those variables are unimportant. For many uses it is
important to know whether a text is fiction or non-fiction and, thus, a “code” that makes this
designation may prove useful, even though, this designation adds nothing to our ability to predict
how easy or hard a text will be for an individual reader.

Finally, Figure 1 summarizes the fact that the human generated inter-sentential task type is
102L harder than the intra-sentential machine generated task type. This is a main effect
difference not an interaction and thus does not interfere with the objective of ordering texts as to
their complexity.
Table 1

Estimates of Text Complexity Measured with Two Item Types Using PASeries Reading Passages (n=63)

<table>
<thead>
<tr>
<th></th>
<th>Theory</th>
<th>Auto-Generated Semantic Cloze</th>
<th>Extended Passage Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory</td>
<td>1.0</td>
<td>.89</td>
<td>.93</td>
</tr>
<tr>
<td>Auto-Generated Semantic Cloze</td>
<td>.94</td>
<td>.90</td>
<td>.87</td>
</tr>
<tr>
<td>Extended Passage Native</td>
<td>.96</td>
<td>.95</td>
<td>.94</td>
</tr>
</tbody>
</table>

Research Notes: (1) Upper triangle: Pearson-product moment correlations, (2) Lower triangle: Error corrected correlations (bold), (3) Diagonal: Reliabilities (bold, italics)
Appendix A

An Example of a Native Lexile Reading Item Type

Mom and Dad expected their children to do well in school. Stephen did well because he wanted to get into a distant college. Jodie did well because she liked to come in first. Brendan did well because the school imposed standards on athletes. They had different _______.

a) motivations  
b) families  
c) memories  
d) questions
Megan's First Ride

Megan found a seat at the front of the plane. She tightened her seat belt. She hugged her pillow. She ____ (held, passed, threw, dropped) it close. It was the first time she had been on an airplane. She was very excited.

Soon the plane started going down the runway. It went faster and faster. Then Megan looked out the window. She wanted to _____ (see, hide, sleep, hurry). The plane was in the air! Up, up, up, they flew. It was fun to go so fast.

After a few minutes, the plane was very high. The _____ (ground, plane, wind, machine) was far below. Megan looked out the window again. She could see big, puffy clouds. They looked like giant marshmallows. "What do clouds taste like?" wondered Megan.

The flight attendant came by. She asked for Megan's help. She wanted her to help pass out the peanuts. Megan grinned. She felt _____ (important, sorry, lonely, cold). She could do that! She had to work hard to keep her balance. The floor of the plane wouldn't stay still. It was _____ (moving, closing, leaving, waiting).

Before long, it was time to land. Megan stared out the window to watch. The plane zoomed down. Bump! The plane skidded to a stop. The flight was over. It had been fast but fun. Megan wanted to fly again soon. She had _____ (liked, found, used, picked) it.
Appendix C

An Example of an Auto-Generated Semantic Cloze Item Type

Pizza Problems (Weekly Reader)

Some experts question rewarding reading with pizza.

Should kids be reading for pizza? A nationwide reading program is drawing criticism for putting pizza front and center. Since 1985, the restaurant chain Pizza Hut has been running the Book It program, which rewards students for reading with personal pan pizzas. Now some people are taking a slice at Book It. They say the program promotes poor eating habits and gives Pizza Hut free advertising in public schools.

“It’s clear that Pizza Hut’s Book It has no place in public schools. …It promotes junk food,” says Susan Linn, a cofounder of the Campaign for a Commercial-Free Childhood. The group has called for schools to stop participating in the Book It program.

Nearly 22 million children in 50,000 U.S. schools take part in Book It. Since the program began, more than 200 million pizzas have been given away. Each year, from October 1 through March 31, students read books to meet monthly reading goals set by teachers. When students meet their goal for the month, they get a certificate for a personal pan pizza.

“We’re really proud of the program,” says Leslie Tubbs, the 5th grade teacher of Book It. “We get hundreds of e-mails from [past participants] who praise it and say it helped them get started with reading.” In 1988, President Ronald Reagan honored the program for literacy, or reading, skills.

comfrey	laborer
cascade
director
Figure 1. Reading Task Complexity Continuum for Three Types of Dichotomously Scored Items