TEXT COMPLEXITY RESEARCH PROJECT AND EDUCATIONAL IMPLICATIONS

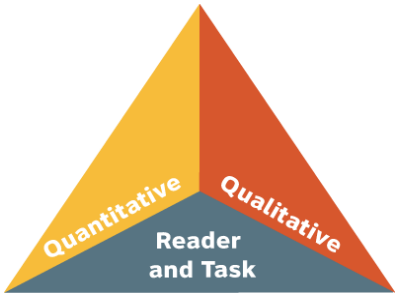
-David and Meredith Liben

The Common Core State Standards and Text Complexity

The Common Core State Standards (CCSS) are the first standards to require text complexity as a standard (Reading Standard Ten). Appendix A makes clear the reasons for this inclusion and presents a comprehensive discussion of text complexity. The fact of text complexity as a standard requires educators, publishers and assessment designers to pay careful attention to complexity in choosing texts for instruction and assessment at each grade level.

The goal of the Text Complexity Research Project was to discover valid, reliable, transparent, and user friendly tools for text measurement and test their ability to analyze complexity through the K-12 progression. Another part of this work was to ensure the measures would provide a smooth trajectory of text complexity from the early grades through to college and career readiness. This paper presents the results of these efforts: the research results, the tools developed for measuring text complexity, and the common table of the quantitative measures that eliminates the gap in reading complexity that has existed between high school texts and college level reading materials. It also examines the implications of this research for education.

The Standards approached measuring text complexity in three ways. First are qualitative considerations, those features of text that cannot be measured by computers at this time but must be evaluated by educators; second, quantitative measures, for the features computers can analyze; third, professional judgment, to look at how suited a text is for a specific instructional purpose with a particular set of students.



Text Complexity Research Project

The goal of the Text Complexity Research project was to further develop and refine each leg of the triangle from Appendix A (above). It is important to understand that the primary goal was not to find a better tool for *matching* specific texts to individual readers but to determine

appropriate levels of complexity for instructional purposes at each grade or band level and to ensure that the complexity level of instructional materials provided a smooth pathway to college and career readiness by the end of high school rather than leaving graduates unprepared.

Qualitative Leg

The CCSS approaches text complexity by grade bands (K-1, 2-3, 4-5, 6-8, 9-10, 11-12) rather than single grades. Appendix A describes four features of text that inform complexity and have long been considered impossible to determine by automated measuring: levels of meaning (literary texts) or purpose (informational text); text structure; language conventionality and clarity; and knowledge demands. During the 2010-2011 school year, systematic work was done using these features to develop scales for helping educators determine qualitative features of text complexity. Teachers need to cultivate confidence in their ability to evaluate texts for complexity and for determining the purpose and task for each text. The qualitative scale was conceived and refined by focus groups of teachers who represented many adopting states and a wide variety of teaching backgrounds. A scale exists for each of the Common Core grade bands, and a sample scale is included in in the resources section that follows this paper.

This qualitative tool is designed to be used by educators in conjunction with one of the quantitative measurement tools described below.

Quantitative Leg:

For the quantitative portion of the research project, seven partners, each of which had an existing computerized tool to determine text complexity or were in the early stages of developing one[1], agreed to participate. There were several goals.

One was to determine which tool – if any – best correlated with educators' determination of grade level texts. For example, given a group of texts that educators had determined to be 6th grade level, each organization would use their tool to analyze the texts without knowing the grade level, and provide a grade number to the lead researcher. In this hypothetical example, the tool that had the scores which averaged closest to 6th grade would be deemed the most accurate.

A second goal was to evaluate whether the features that make text seem complex are the same features that make text difficult for readers.

[1] [1] The measurements and the groups that developed them are as follows: ATOS® (Renaissance Learning), DRP® (Questar), Flesch Kincaid (public domain, analysis run courtesy of Coh-Metrix/University of Memphis), Lexile® (MetaMetrics), Source Rater© (Educational Testing Service), and Pearson Reading Maturity Metric© (Pearson Knowledge Technologies/Pearson Education).

A third goal was to ensure that any tool for measuring text complexity had as the high end of its trajectory an anchor in college and career level complexity. It also needed to provide a smooth pathway from the earlier grades through to that endpoint. In other words, by using a given tool to measure appropriate complexity at each grade level, educators at each level could feel confident that their students would be on track to handle the demands of college and career level text by the time they completed high school.

The Tools in the Study

The first four of these tools: ATOS, DRP, Flesch-Kincaid and Lexile are readability measures. Readability determines the complexity of a text based on two features: a word frequency measure and a syntactic complexity measure, both based on the averages the computer finds in the scanned text. The common sense notion behind word frequency is that the greater the number of less common words a text has the more complex it is, and the greater the number of more common words the less complex it is. The syntactic component of readability is usually a measure of average sentence length; the idea being that texts with greater percentages of longer sentences are more complex than texts with shorter sentences. With both parts of readability, the use of averages can mean that sections of the text that might be much harder than average could be glossed over, an issue when those might be the very parts that readers could find challenging.

Three of the other tools Coh-Metrix, Text Maturity, and Source Rater are what we have called "broad spectrum" measures. These measures include word frequency and syntax too, but they contain a number of other text features as well. The most important of these other features are academic orientation/narrativity, cohesion, and a newer measure called word maturity that is central to the Text Maturity tool.

Academic orientation and narrativity are opposite sides of the same coin; generally the more of one, the less of the other. Academic orientation, as the name implies, measures how academic the text is. Academic text usually has more uncommon words, more nouns and nominalizations (words ending in "tion"), fewer verbs, fewer words connoting emotions and fewer negations. These language patterns can all be picked up by quantitative tools. Academic texts are generally informational texts – text intended to inform. Informational text, as has long been known by teachers, is more difficult for students to read. Formal research has found the same phenomenon (cite). Narrativity measures how story- like a text is. Texts with more narrativity tend to have more ordinary words. They have less nominalization, more verbs, and more words communicating emotions. As we've seen, these can all be picked up by the broad spectrum quantitative tools. Texts with clusters of these features usually have less information. Texts

that are more story- like are considered easier to read, a finding well documented by research (cite).

Cohesion is the collection of features that authors include- whether by design or not - that glue a text together for the reader. There are many types of cohesion but the text complexity research revealed that repetition of the same words or phrases between adjacent sentences or within paragraphs added to or eased the complexity of texts demonstrably. The broad spectrum measures can also determine the relative proportion of words in the text that are abstract. Abstract words are often more difficult to determine from context (cite) and harder to quickly grasp the meaning of. Because of these features, the presence of many abstract words can make a text more complex (cite).

Word maturity is the newest of the broad spectrum measures. Word maturity is based partly on the fact that many words have related but subtly different meanings depending on the context in which they appear. Ground, for example, is the earth or the soil beneath our feet. As a verb it is the past participle of "to grind" as in to crush something ("He ground his teeth in annoyance"). But it can also be used as "He has no grounds for his argument," or "She is a well-grounded person." Texts which use ground in one of these latter senses are likely to be more complex than texts using it in the first sense. Word maturity can determine how ground is being used by looking at the other words around it that form its context. A text using ground in the sense of, "he has no grounds for his argument" would likely be a text requiring greater word and world knowledge to comprehend. Word maturity can measure this and estimate complexity from the presence or absence of these words.

In sum, the readability tools tell us about complexity formed by two dimensions of text: how common the words are and how long the sentences are. The broad spectrum tools tell us about complexity formed by these as well as features such as the examples described above. In all cases, each feature is reported as an average over the full text selection being analyzed.

<u>How the Study was Conducted</u>

During the past year, a team led by Jessica Nelson, PhD (Carnegie Mellon University) and Charles Perfetti PhD (University of Pittsburgh) and coordinated by the authors has been working to test and validate these seven quantitative measures of text complexity.

To do this, the research team performed a detailed analysis of each tool: its ability to predict student performance on standardized high stakes tests, its ability to predict educator judgment of text complexity, and an examination of how each tool performs measuring complexity across a broad array of types of texts selected by different groups of educators for widely varying purposes.

The texts we used for analysis included 683 retired state test passages, 168 exemplar texts from Appendix B of the CCSS, 97 secure passages from the Gates MacGinitie Reading test and 98 passages from the Stanford Achievement Test version 9, the latter two accompanied with student performance data. The passages in each of these sets of text had a designated grade level determined by educators.

When we focused on the question of the tools' abilities to predict educator designations, we found the correlation between all the measures and educators' designated grade level was quite high for every grade, for narrative as well as for informational text. However, the broad spectrum measures produced stronger correlations for all texts, especially for texts at higher grade levels.

This finding would indicate that the features of text that educators rely on – consciously or not - when determining complexity match up more tightly with the additional features plus readability measures that are all contained in the broad spectrum measures than they do with the readability measures alone. So the broad spectrum measures can 'see' much of what the educators evaluating text see.

Complexity, however, is not necessarily the same as difficulty. It is possible that what educators see as the features that make a text more or less complex are not those that present most difficulty to the students trying to understand the text. For this reason, the second part of the research addressed the question of which tools correlated most with students' scores on standardized reading tests.

In this case, the researchers focused on the tools' abilities to produce measures that correlated with the student performance data. If the tools could predict where students had higher or lower scores on specific passages, then they could be said to measure text *difficulty* as well as text complexity. Here, unlike the results for educator grade level designations, where the broad spectrum measures outperformed the readability measures, *all* the tools performed comparably - and well - in correlating with student scores. Though the broad spectrum measures had *slightly* higher correlations in the upper grades, these differences were not large enough to be statistically significant.

Given how well each group's tool performed in measuring student performance, it was clear to the research team that any of them could be used with confidence for measuring text to determine difficulty.

The third task of the research team was to align each of these measures on a common scale. That scale had to provide for a smooth trajectory from the early grades to college and career level complexity. The text complexity concordance table that resulted is reproduced below along with a guide to locating the public access portals of each tool.

Figure A

Common Scale for Band Level Text Difficulty Ranges

Common Core Bands:	Text Analyzer Tools					
	ATOS	DRP	FK	Lexile	SR	RM
2nd-3rd	2.75-5.14	42-54	1.98-5.34	420-820	0.05 – 2.48	3.53 - 6.13
4th-5th	4.97-7.03	52-60	4.51-7.73	740-1010	0.84 – 5.75	5.42 - 7.92
6th-8th	7.00-9.98	57-67	6.51-10.34	925-1185	4.11 – 10.66	7.04 - 9.57
9th-10th	9.67-12.01	62-72	8.32-12.12	1050-1335	9.02 – 13.93	8.41 - 10.81
11th-CCR	11.20-14.10	67-74	10.34-14.2	1185-1385	12.30 – 14.50	9.57 - 12

Key:
ATOS - ATOS® (Renaissance Learning)
DRP - Degrees of Reading Power ®(Questar)
FK - Flesch Kincaid ®(public domain, no mass analyzer tool available)
Lexile - Lexile Framework® (MetaMetrics)
SR - Soure Rater ©(Educational Testing Service)
RM- Pearson Reading Maturity Metric© (Pearson Education)

Measures not in concordance table:
Coh-Metrix (Univ of Memphis)
REAP® (Carnegie Mellon University)

Figure B

How to access each of the text measurement tools:

1. ATOS analyzer: Renaissance Learning. http://www.renlearn.com/ar/overview/atos/
2. Coh-Metrix Easability Tool. University of Memphis. Beta site available at: http://141.225.42.101/cohmetrixgates/Home.aspx?Login=1
3. DRP – Questar www.questarai.com Contact Quester with requests for text analysis.
4. Lexiles – Metrametrics www.lexile.com/analyzer/
5. Pearson Reading Maturity – Pearson Knowledge Technologies Beta site available at: www.readingmaturity.com

6. REAP – Carnegie Mellon University www.reap.cs.cmu.edu/
7. Source Rater Educators Testing Service *Beta site available at:* http://naeptba.ets.org/SourceRater3/

Reader and Task: Applying Professional Judgment

The third leg of the text complexity triangle is equal in importance to the qualitative and quantitative legs. It is the exercise of professional judgment by the individual educator who knows his or her students, has a firm sense of desired learning outcomes, and is in the best position to analyze how best to accomplish those goals. Since the added variable of text complexity is a new consideration for most educators, the research team developed the systematic checklist below to assist professional judgment.

Figure C

✓	Reader and Task Considerations: Professional Judgment			
	Quality of text and/or information it conveys: Is this text worth student and instructor time, attention and effort?			
	Suitable for purpose: Is this the best text available for my purpose?			
	Level of knowledge needed: Do the students have the knowledge needed to learn from this text? If not, can that learning happen with careful reading? Have I planned adequately for that learning?			
		Background in Topic		**Syntax and language features**
		Vocabulary		**Strategies for comprehending challenging sections**
	Appropriate: Is the content, style and tone suited to this age student?			
	Aligning the task(s) between the text, the desired outcome and the readers: Does the task align to the text? Will it elicit the learning I want for my students? What are my learning goals? What supports do I need to plan to include?			

EDUCATIONAL IMPLICATIONS

Choosing Texts

Figure A shows the concordance table for all of the tools used in the study. Each of the tools has a different numbering system, and the concordance table shows each tool's numbers for each band level. This allows publishers and educators to determine if a text fits into one of the Common Core band complexity levels based on any – or any combination – of the different measures. For example, a text for placement at the 4-5 band level would need to have a Lexile score between 740 and 1010 or a Reading Maturity score between 5.42 and 7.92.

Figure B shows the access points for each of the tool's public portals. While large scale use will be contracted formally through whichever group a publisher or school system choses, an individual user who wants to see the complexity of an article or excerpt may upload a sample (up to a thousand words in length) into any of the analyzer tools s/he wishes to access.

The CCSS place texts by bands and call for students who are toward the end of a given band to be able to read any texts in the complexity range of the band "independently and proficiently." Students in the earlier grade of a band range need to read texts that measure at the end of the band level "proficiently, with scaffolding as needed at the high end of the range." The concordance table combined with qualitative analysis will allow publishers and educators to determine texts across and within band levels.

Curriculum Considerations

The research results highlight the essential and central role of both vocabulary and syntax in determining students' success in reading comprehension. As you'll recall, all the measures correlated equally well with students' scores. Vocabulary and syntax are the sole features all the measures have in common. They are the *only* features measured by the readability tools. The broad spectrum tools measure many other features: the academic orientation of a text; how much the language of the text ties together the events, ideas and information, and the relative density of abstract words Word maturity can additionally measure the relative sophistication of a given text by examining the context words appear in. The results of our research would strongly suggest – especially for the elementary years – that these features play less of a role in determining student success than do the two factors of word knowledge and comfort level with more complicated syntax.

While more research remains to be done on the question of precisely what role if any these additional features play in affecting readers' understanding, it is clear at this point that students need a secure and solid foundation in word knowledge and the ability to handle diverse sentence structures as a starting point to accessing complex text. Another way of understanding this is that when the features measured by the broad spectrum tools add complexity to the text, students will be more prepared for handling them if they have sufficient word knowledge and syntactical abilities. These are the implications of the research results, and they point squarely at an educational challenge.

Vocabulary is the least systematic and least intensive part of ELA instruction (cite). Aside from the formal studies, this fact can be confirmed by asking any school or district ELA leader to describe their ELA program. They will likely mention some combination of the following: guided reading; shared reading, balanced literacy, leveled reading, discipline based reading... and then go on to describe how these are implemented. If vocabulary is mentioned at all, rarely will details be provided as to how words are chosen; how they are taught, what tools are used, what assessment looks like, or what changes in vocabulary instruction as students progress through the grades.

The relation between vocabulary and comprehension is well documented (cite). The research on syntax, though far less extensive, also shows a relation to comprehension (cite). As haphazard as vocabulary instruction has been in the ELA curriculum, formal instruction in syntax and varied sentence structure has been consistent. Consistently absent. Work with syntax has been virtually absent for almost half a century from American K-12 education. We have not even been able to find research about the frequency of instruction with syntax (other than references to language instruction with English Language Learners and in foreign language instruction). This has been anecdotally confirmed by every educator we have discussed this question with since the publication of the CCSS, and probably rings true in your experience.

CONCLUSION

The Common Core State Standards call for raising the complexity level of the texts students read. No matter what tool or combination of tools is chosen to determine complexity, the texts selected will increase demands on students at every grade after first grade. This will in turn raise the level of vocabulary and syntactic demands placed on students at every grade level.

Students whose vocabulary levels were already behind will fall further behind; students unable to grasp syntax at previous levels of complexity will now confront substantially greater obstacles. Many students who were previously proficient with both these features of text will suddenly find themselves stretched beyond their capacity when faced with the more complex texts called for by the standard.

If instruction and curriculum aligned with the standards does not directly address these concerns, our current educational gaps will expand rather than contract. Students with robust vocabularies who are experienced and resilient in the face of varied syntax will be in a position to soar as they become increasingly able to *independently* meet the challenge of more complex text and rigorous tasks; those not so prepared will falter, unprepared and unsupported.
