

Understanding *Lexiles*: an example of a measure

A recent conversation with a former colleague (it was more of a lecture) about what psychometricians don't understand about students and education led me to resurrect an article that I wrote for the *Rasch Measurement Transactions* four or five years ago. It deals specifically with *Lexiles*[®] but it is really about how one defines and uses measures in education and science.

The antagonism toward *Lexiles* in particular and Rasch measures in general is an opportunity to highlight some distinctions between *measurement* and *analysis* and between a *measure* and an *assessment*. Often when trying to discuss the development of reading proficiency, specialists in measurement and reading seem to be talking at cross-purposes. Reverting to argument by metaphor, measurement specialists are talking about *measuring weight*; and reading specialists, about *providing proper nutrition*.

There is a great deal involved in physical development that is not captured when we measure a child's weight and the process of measuring weight tells us nothing about whether the result is good, bad, or *normal*; if you should continue on as you are, schedule a doctor's appointment, or go to the emergency room without changing your underwear. *Evaluation* of the result is an analysis that comes after the measurement and depends on the result being a measure. No one would suggest that, because it doesn't define health, weight is not worth measuring or that it is too politically sensitive to talk about in front of nutritionists. A high number does not imply good nutrition nor does a low number imply poor nutrition. Nonetheless, the measurement of weight is always a part of an *assessment* of well-being.

A *Lexile* score, applied to a person, is a measure of reading *ability*ⁱ, which I use to mean the *capability to decode words, sentences, paragraphs, and Supreme Court decisions*. *Lexiles*, as applied to a text, is a measure of how difficult the text is to decode. Hemingway's "For Whom the Bell Tolls" (840 *Lexile* score) has been cited as an instance where *Lexiles* do not work. Because a 50th percentile sixth-grade reader could engage with this text, something must be wrong because the book was written for adults. This counter-example, if true, is an interesting case. I have two counter-counter-arguments: first, all measuring instruments have limitations to their use and, second, *Lexiles* may be describing Hemingway appropriately.

First, outside the context of *Lexiles*, there is always difficulty for either humans or computer algorithms in scoring exceptional, highly creative writing. (I would venture to guess that many publishers, who make their livings recognizing good writingⁱⁱ, would reject Hemingway, Joyce, or Faulkner-like manuscripts if they received them from unknown authors.) I don't think it follows that we should avoid trying to evaluate exceptional writing. But we do need to know the limits of our instruments.

I rely, on a daily basis, on a bathroom scale. I rely on it even though I believe I shouldn't use it on the moon, under water, or for elephants or measuring height. It does not undermine the validity of *Lexiles* in general to discover an extraordinary case for which it does not apply. We need to know the limits of our instrument; when does it produce *valid* measures and when does it not.

Second, given that we have defined the *Lexile* for a text as the difficulty of decoding the words and sentences, the *Lexile* analyzer may be doing exactly what it should with a Hemingway text. Decoding the words and sentences in Hemingway is not that hard: the vocabulary is simple, the sentences short. That's pretty much what a *Lexile* score reflects.

Understanding or appreciating Hemingway is something else again. This may be getting into the distinction between reading *ability*, as I defined it, and reading *comprehension*, as the specialists define that. You must be able to read (i.e., decode) before you can comprehend. Analogously, you have to be able to do arithmetic before you can solve math word problemsⁱⁱⁱ. The latter requires the former; the former does not guarantee the latter.

The *Lexile* metric is a true developmental scale that is not related to instructional method or materials, or to grade-level content standards. The metric reflects increasing ability to read, in the narrow sense of *decode*, increasingly advanced text. As students advance through the reading/language arts curriculum, they should progress up the *Lexile* scale. Effective, even standards-based, instruction in ELA^{iv} should cause them to progress on the *Lexile* scale; analogously good nutrition should cause children to progress on the weight scale^v.

One could *coach* children to progress on the weight scale in ways counter to good nutrition^{vi}. One might subvert *Lexile* measurements by coaching students to write with big words and long sentences. This does not invalidate either weight or reading ability as useful things to measure. There do need to be checks to ensure we are effecting what we set out to effect.

The role of *standards-based assessment* is to identify which constituents of reading ability and reading comprehension are present and which absent. Understanding imagery and literary devices, locating topic sentences, identifying main ideas, recognizing sarcasm or satire, comparing authors' purposes in two passages are within its purview but are not considered in the *Lexile* score. Its analyzer relies on rather simple surrogates for semantic and syntactic complexity.

The role of measurement on the *Lexile* scale is to provide a narrowly defined measure of the student's status on an interval scale that extends over a broad range of reading from *Dick and Jane* to Scalia and Sotomayor. The *Lexile* scale does not define reading, recognize the breadth of the ELA curriculum, or replace grade-level content standards-based assessment, but it can help us design instruction and target assessment to be appropriate to the student. We do not expect students to say anything intelligent about text they cannot decode, nor should we attempt to assess their analytic skills using such text.

Jack Stenner (aka, *Dr. Lexile*) uses as one of his parables, you don't buy shoes for a child based on grade level but we don't think twice about assigning textbooks with the formula ($age - 5$). It's not one-size-fits-all in physical development. Cognitive development is probably no simpler if we were able to measure all its components. To paraphrase Ben Wright, how we measure weight has nothing to do with how skillful you are at football, but you better have some measures before you attempt the analysis.

ⁱ *Ability* may not be the best choice of a word. As used in psychometrics, *ability* is a generic placeholder for whatever we are trying to measure about a person. It implies nothing about where it came from, what it is good for, or how much is enough. In this case, we are using reading ability to refer to a very specific skill that must be taught, learned, and practiced.

ⁱⁱ It may be more realistic to say they make their livings recognizing *marketable* writing, but my cynicism may be showing.

ⁱⁱⁱ You also have to decode the word problem but that's not the point of this sentence. We assume, often erroneously, that the difficulty of decoding the text is not an impediment to anyone doing the math.

^{iv} Effective instruction in science, social studies, or basketball strategy should cause progress on the *Lexile* measure as well; perhaps not so directly. Anything that adds to the student's repertoire of words and ideas should contribute.

^v For weight, *progress* often does not equal *gain*.

^{vi} Metaphors, like measuring instruments, have their limits and I may have exceeded one. However, one might consider the extraordinary measures amateur wrestlers or professional models employ to achieve a target weight.