Considerations for Revisiting the NAEP Reading Framework

Margaret G. McKeown

Background: The NAEP reading assessment is taken as the gold standard on reading ability in the United States. We should be cognizant of the assessment's reputation and adhere to principles that will be faithful to that status. This special status provides the imperative to attend to the relevance and robustness of NAEP's Reading Framework. The impetus for keeping the framework up to date and relevant is to assure that the framework remains a productive tool for guiding the development of the NAEP reading assessment.

Key to keeping the NAEP reading assessment true to its status and purpose is that any changes to the framework, and the assessment itself, should be undertaken with care to assure that the passages and items measure the intended construct; that is, they elicit reading behaviors and assesses reading ability. This may seem an obvious statement for a reading test. However, when test items are being created, what is truly being measured can drift. For example, the way tasks are structured can lead students to respond based on information gleaned from images or videos rather than on interactions with written language. Alternatively, the items might give precedence to assessing skills other than comprehending text information. Against the backdrop of maintaining reading comprehension as the key construct, I address three aspects of the framework and assessment:

1. Definition of reading: What reading is and what we are measuring

2. Text and task selection: How the framework is used to construct the assessment

3. Assessing vocabulary: How word meaning is used to comprehend text

(1) Definition of Reading: In the current NAEP Reading Framework, developed for the 2009 assessment, great care was taken to be clear about what reading is and to assure that the assessment would indeed measure reading. The definition of reading was developed to reflect the understanding that reading includes the ability to retell text information and to integrate ideas within a text in order to interpret meaning. Essential here is the recognition that reading involves more than a literal lifting of information from a text, yet the process is driven by dealing with text information.

The care in developing this definition was in part motivated to address a concern with the earlier framework (2003) in which making reader/text connections was a focus. Although a reader's making connections to text is important to attaining high literacy, attempting to measure that aspect tends to overemphasize knowledge that a reader brings to a text. Thus prior knowledge, rather than the ability to read and make sense of text, may be the phenomenon that is measured. This situation can render the measure of reading inaccurate; the reading ability of high knowledge students can easily be overestimated while the ability of students with lower levels of knowledge is underestimated.

Of course it is not possible to remove prior knowledge from a measure of reading, but in the current framework efforts were made to prevent knowledge from driving responses. This was done through the definition of reading, which drove the item types, through selecting texts whose topics seemed generally accessible and using a broad array of topics.

The above is intended to illustrate why the current framework's articulation of reading comprehension is still relevant and appropriate. The current Programme for International Assessment (PISA [2015]) and Progress in International Reading Literacy Study (PIRLS [2016])

definitions, which remain largely the same as those used to inform the NAEP definition of reading for the 2009 Reading Framework, confirm the relevance of NAEP's definition.

(2) Text and task selection: The texts and tasks prescribed in the framework instantiate the definition of reading. Some further specification of texts and tasks to be used in the assessment may be necessary, based on the expansion of the kinds of texts and tasks vying for inclusion.

In the case of texts, digitally-native texts, which differ from printed material in their nonlinearity and opportunities for navigation through menus and embedded links, are available. These characteristics present affordances but also challenges for readers. Digitally-native texts add layers of complexity and decision-making that do not tap reading ability but rather the ability to utilize features specific to online material. Thus if online texts are used in the assessment, it needs to be done with attention to assuring that measuring reading remains at the forefront.

Dealing with online material is increasingly part of most students' lives, and may become more necessary for academic success. But the fact remains that the ability to deal with these materials goes well beyond reading. The implication is that perhaps creation of an additional type of assessment should be considered, to assess *reading online* (similar to ePIRLS). But that purpose should not be mixed with the reading assessment. If it is, those two purposes will be confounded, and we will learn less, not more, about students' abilities.

The types of tasks and response formats available to the assessment in electronic presentation have similarly expanded. Of particular focus in this regard are scenario-based assessments (SBA). These represent an intriguing new approach that is excellent for some purposes, such as

assessing argument development (Deane et al, 2018). However its use in a reading assessment requires caution to assure that the measured construct is indeed reading comprehension.

SBAs require setting up a scenario, including multiple characters ("avatars"), and specifying a task, which presents students with an extra reading burden as well as the conceptual burden of working through a separate task. These can divert attention from the measured construct, reading comprehension, especially for students at risk.

The rationale underlying SBAs involves the concept of using information gained from reading to accomplish some purpose. The concept is common in the reading field, and is included in the NAEP framework as well as in PISA and PIRLS. But boundaries need to be set on the extent to which assessing how a reader uses information from reading is still measuring reading. To illustrate, imagine providing students a recipe for blueberry pie; evaluating how good the pie tasted would not be an appropriate measure of their reading ability. And indeed, in the NAEP, PISA, and PIRLS frameworks, the concept of using reading remains directly related to interactions with the text rather than performing some related task (NAEP framework, p. 3).

The use of new task types in SBAs should depend on a deliberate decision process that includes exploring potential tasks and explicitly describing their use in a new or revised framework. Making the use of such tasks explicit in the framework will promote careful consideration of their construction in a way that exploits their affordances but keeps the focus on reading.

(3) Assessing Vocabulary: The current framework is unique in its attention to vocabulary. This component was included based on the understanding that vocabulary plays a key role in reading comprehension and that shallow knowledge of words, such as memorizing definitions, is unlikely to assist comprehension (McKeown et al, 1985; 2018). NAEP

vocabulary items measure how well students can use word meanings to understand the
passages they read. Essentially, each item assesses how knowledge of a particular word is
used to develop understanding of a portion of the passage.

Since the development of the framework, understanding of the centrality of vocabulary to
comprehension has only strengthened (Perfetti & Stafura, 2014). Contemporary research
confirms that comprehension develops as readers integrate word meanings and context, and has
emphasized that a reader's facility with multiple dimensions of vocabulary is key to that
integration (Proctor et al, 2009; McKeown, et al, 2017). Thus it seems important to maintain
vocabulary as an aspect of the framework and assessment, and to consider expanding the
framework to allow items that assess or embody the multidimensionality of words that underlies
effective word knowledge.

For example, items could capture semantic dimensions of words by tapping into the concept that
words have various senses in different contexts, such that the *foundation* of someone's beliefs is
not the base that a building stands on; or items could capture morphological knowledge, such as
understanding the relationship of *wilderness* to *wild*.

The framework restricts item types for vocabulary to multiple choice. The assessment might
benefit from expanding to include innovative item types to capture students' vocabulary
knowledge in more distinct ways as well as items that could take advantage of response modes in
the electronic format.

Most important in any rethinking of the vocabulary portion of the framework is to adhere to the
concept that vocabulary is measured as part of comprehension, that is, to learn how well students
can use word knowledge to make sense of text.

Conclusion: The focus of revisiting the NAEP Reading Framework should be to invigorate its role as the conceptual basis of the reading assessment, to assure that the framework represents our best judgment of how to measure reading, and that the assessment faithfully instantiates the framework. This includes revising and maintaining aspects of the framework.

1. Review and potentially revise, clarify, or expand:

   - Text types – review based on electronic formats

   - Task types – review based on recent new approaches

   - Cognitive targets – review to assure alignment of new formats

   - Response formats – revise based on electronic test format

   - Vocabulary assessment – expand based on recent work

2. Maintain:

   - Definition of reading

   - Reading comprehension as the construct of interest

   - Goal of assessing vocabulary as comprehension

References

Deane, P., Song, Y., van Rijn, P., O'Reilly, T., Fowles, M., Bennett, R., Sabatini, J., & Zhang, M. (2018). The case for scenario-based assessment of written argumentation. *Reading and Writing*. https://doi.org/10.1007/s11145-018-9852-7.

McKeown, M. G., Deane, P. D., Scott, J. A., Krovetz, R., & Lawless, R. R. (2017). *Vocabulary assessment to support instruction: Building rich word-learning experiences*. New York: Guilford Press.

McKeown, M.G., Crosson, A.C., Moore, D., & Beck, I. L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal 55*(3), 572-616. doi.org/10.3102/0002831217744181

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20,* 522-535.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687

Proctor, C. P., Uccelli, P., Dalton, B., & Snow, C. E. (2009). Understanding depth of vocabulary online with bilingual and monolingual children. *Reading and Writing Quarterly*, *25*(4), 311-333. https://doi.org/10.1080/10573560903123502