
OUTLINE OF A SAMPLE REQUEST FOR APPLICATION

The analyses and judgments of the RRSg that constitute the body of this report are intended to provide broad guidance for creating a program of research and development that will support the improvement of reading comprehension instruction in U.S. schools. However, this broad guidance is not sufficient for actually providing instructions to potential applicants for funding. Funding agencies that intend to support such R&D will need to write specific requests for applications (RFAs) that delineate narrower areas of R&D that the funding is intended to support and that provide more explicit guidance to the applicants.

In this appendix, the RRSg has illustrated its own sense of what would constitute the core of an effective RFA in a single area—assessment. We chose assessment because advances in assessment are crucial to effective R&D related to reading comprehension. If time had permitted, we would have also developed similar core RFAs in several other areas identified in the body of the report including (1) improving instruction using currently available knowledge; (2) developing new knowledge to inform more radical changes in instruction; (3) enhancing teacher preparation and professional development; (4) gaining a better understanding of the properties of reader-text interactions through studies using various genres of text as well as comparisons of electronic and linear texts; and (5) exploring the impact of technology on reading comprehension, reading activities being undertaken by learners, and opportunities for instruction.

The example that follows¹ should probably be viewed as a starting point. Any funding agency would want to elaborate on this RFA, perhaps by including a review of the literature, a more detailed specification of research goals, examples of possible R&D activities, funding levels, and so on. The proposal includes examples of short-, mid-, and long-term projects, and a funding agency, such as OERI, might choose to address these in separate RFAs issued at different points in time.

¹In this example, we have drawn heavily on material in earlier chapters of this report.

However, the core content we have provided here seems to us to convey the key points and spirit of what we view as being important in guiding potential applicants for work on assessment of reading comprehension.

REQUEST FOR APPLICATIONS IN THE DOMAIN OF ASSESSMENT

Statement of the Problem

Currently available assessments of student performance in the field of reading comprehension are a persistent source of complaints from both practitioners and researchers. These complaints claim that the assessments

- inadequately represent the complexity of the target domain
- conflate comprehension with vocabulary, domain-specific knowledge, word-reading ability, and other reader capacities involved in comprehension
- do not reflect an understanding of reading comprehension as a developmental process or as a product of instruction
- do not examine the assumptions underlying the relationship of successful performance to the dominant group's interests and values
- are not useful for teachers
- tend to narrow the curriculum
- are one-dimensional and method-dependent, and often fail to address even minimal criteria for reliability and validity.

Indeed, most currently used comprehension assessments reflect the purpose for which such assessments were originally developed—to sort children on a single dimension by using a single method. More important, though, is that none of the currently available comprehension assessments is based in a viable or articulated theory of comprehension. Because currently used comprehension instruments are all unsatisfactory in various ways, the field has not selected a common standard for assessing comprehension. Until instrumentation and operationalization of comprehension is widely agreed on, mounting a truly programmatic research effort will be difficult.

These considerations, as well as current thinking about the nature of reading comprehension (see the body of this report), create a demand for new kinds of assessment strategies and instruments that more robustly reflect the dynamic, developmental nature of comprehension and represent adequately the interactions among the dimensions of reader, activity, text, and context.

Currently, widely used comprehension assessments focus heavily on only a few tasks: reading for immediate recall, reading for gist, and reading to infer or disambiguate word meaning. Assessment procedures to evaluate learners' capacities to modify old or build new knowledge structures, to use information acquired while reading in the interest of problem solving, to evaluate texts on particular criteria, or to become absorbed in reading and develop affective or aesthetic responses to text have occasionally been developed in the pursuit of particular research programs, but they have not influenced standard assessment practices. Because knowledge, application, and engagement are the crucial consequences of reading with comprehension, assessments that reflect all three are needed. Further, the absence of attention to these consequences in widely used reading assessments diminishes the emphasis on them in instructional practices as well.

Because existing measures of comprehension fail to reflect adequately the inherent complexities of the comprehension process, they limit the kinds of research that can be done and undermine optimal instruction. Good assessment tools are crucial to conducting sensible research. In addition, though, as long as we define comprehension by default as reading a passage and answering a few multiple-choice questions about it, teachers will have little incentive to instruct children in ways that reflect a deeper and more accurate conceptualization of the construct.

Efforts have been made to develop measures of comprehension that are referenced to the characteristics of text, that is, a way of relating an assessment of comprehension to the difficulty of the text. Although such measures (e.g., Lexile) provide information that supports instruction by indicating a child's level of comprehension, different estimates of readability do not correspond well with one another. Moreover, they provide no diagnostic information for individualizing instruction, nor do efforts such as the Lexile text system fit easily into a larger assessment system.

Requirements for the Assessment System to Be Developed

A comprehensive assessment program that reflects the current thinking about reading comprehension must satisfy many requirements that have not been addressed by any assessment instruments, while also satisfying the standard psychometric criteria (e.g., reliability and validity). A list of requirements for such a system includes, at a minimum, the following:

- *Capacity to reflect authentic outcomes.* Although any particular assessment may not reflect the full array of consequences, the inclusion of a wider array than that currently being tested is crucial. For example, students' beliefs

about reading and about themselves as readers may constitute supports or obstacles to their optimal development as comprehenders; teachers may benefit enormously from having ways to elicit and assess such beliefs.

- *Congruence between assessments and the processes involved in comprehension.* Assessments must be available that target particular operations involved in comprehension, in the interest of revealing inter- and intra-individual differences that might inform our understanding of the comprehension process and of outcome differences. The dimensionality of the instruments in relation to theory should be clearly apparent.
- *Developmental sensitivity.* Any assessment system needs to be sensitive across the full developmental range of interest and to reflect developmentally central phenomena related to comprehension. Assessments of young children's reading tend to depend on the child's level of word reading and must control for the level of decoding to assess comprehension adequately. The available listening comprehension assessments for young children do not reflect their rich oral language processing capacities, discourse skills, or even the full complexity of their sentence processing.
- *Capacity to identify individual children as poor comprehenders.* An effective assessment system should be able to identify individual children as poor comprehenders, not only in terms of prerequisite skills such as fluency in word identification and decoding, but also in terms of cognitive deficits and gaps in relevant knowledge (e.g., background and domain specific) that might adversely affect reading and comprehension, even in children who have adequate word-level skills. It is also critically important that such a system provide for the early identification of children who are apt to encounter difficulties in reading comprehension because of limited resources to carry out one or another operation involved in comprehension.
- *Capacity to identify subtypes of poor comprehenders.* Reading comprehension is a complex process. It therefore follows that comprehension difficulties could come about because of deficiencies in one or another of the components of comprehension specified in the model. Thus, an effective assessment system should have the means to identify subtypes of poor comprehenders in terms of the components and the desired outcomes of comprehension and in terms of both intra- and inter-individual differences in acquiring the knowledge and skills necessary for becoming a good comprehender.
- *Instructional sensitivity.* The major purposes for assessments are to inform instruction and to reflect the effect of instruction or intervention. Thus, an effective assessment system should provide not only important information about a child's relative standing in appropriate normative populations

(school, state, or national norms groups), but also important information about a child's relative strengths and weaknesses for purposes of educational planning.

- *Openness to intra-individual differences.* Understanding the performance of an individual often requires attending to differences in performance across activities with varying purposes and with a variety of texts and types of text.
- *Utility for instructional decisionmaking.* Assessments can inform instructional practice if they are designed to identify domains that instruction might target, rather than to provide summary scores useful only for comparison with other learners' scores. Another aspect of utility for instructional decisionmaking is the transparency of the information provided by the test given to teachers who are not technically trained.
- *Adaptability with respect to individual, social, linguistic, and cultural variation.* Good tests of reading comprehension, of listening comprehension, and of oral language production target authentic outcomes and reflect key component processes. If performance on the task reflects differences owing to individual, social, linguistic, or cultural variation that are not directly related to reading comprehension performance, the tests are inadequate for the purposes of the research agenda proposed here.
- *A basis in measurement theory and psychometrics.* This aspect should address reliability within scales and over time, as well as multiple components of validity at the item level, concurrently with other measures, and predictively relative to the longer-term development of reading proficiency. Studies of the dimensionality of the instruments in relationship to the theory underlying their construction are particularly important. Test construction and the evaluation of instruments are important areas of investigation and highly relevant to the proposed research agenda.

Clearly, no single assessment will meet all these criteria. Instead, this RFA seeks work that will build toward an integrated system of assessments, some of which may be particularly appropriate for particular groups (e.g., emergent or beginning readers, older struggling readers, second-language readers, or readers with a particular interest in dinosaurs). Furthermore, the various assessments included in the system will address different purposes (e.g., a portmanteau assessment for accountability or screening purposes, diagnostic assessments for guiding intervention, curriculum-linked assessments for guiding instruction, and so forth). Given that multiple assessments are proposed, studies of their dimensionality and the interrelations of these dimensions across measures are especially critical.

Research Plans

Responses to this RFA should take the following issues into account. We seek a mix of large- and small-scale efforts and a mix of short- and long-term efforts. We will consider an application more positively if it builds into the work plan mechanisms for developing the expertise of doctoral and young postdoctoral scholars in the psychometric and content areas relevant to assessing reading comprehension. Applications that have an assessment agenda embedded within another research undertaking (e.g., evaluating the effectiveness of an intervention or developing model teacher education efforts) will be considered for co-funding.

A variety of activities will be considered for funding under this initiative. A few examples for short-, medium-, and long-term research efforts are provided below. These examples are meant to stimulate thinking and are not an exhaustive list of the possible relevant kinds of activities. Consideration of the reliability, validity, and dimensionality of different assessment instruments and approaches is essential to all these endeavors.

Examples of Short-Term Activities

- Generate an inventory of available tests purporting to assess comprehension and evaluate them on the requirements indicated above. Mechanisms for carrying out this activity might include a panel who would provide a consensus or an empirically driven evaluation, or a research effort to review the available norms, which would then generate cross-test reliability data.
- Generate an inventory of and evaluate the assessment strategies, tools, and instruments that teachers, schools, districts, and states are using and the ways that teachers use the information collected within classrooms or schools. Again, the specific mechanism for carrying out this activity might include a panel of assessment officers from states and large districts, a teacher survey, or observational research in a stratified sample of schools.
- Generate an inventory of and evaluate what pre-service teachers learn about the assessment of reading comprehension. A survey of teacher education institutions, an analysis of syllabi and texts used widely in teacher education programs, interview studies with teacher educators, or other mechanisms might be used to address this need.
- Use established databases to evaluate the reliability, validity, and dimensionality of existing assessments of reading comprehension.

Examples of Medium-Term Activities

- Use the information gathered on which instruments are good for what purpose to develop assessments of lesser-studied domains. Such assessments will be useful in instructional settings and will differentiate the relative contribution that variations among readers make to an overall level of performance. For example:
 - discourse structure, including genre
 - understanding written syntax
 - mental model construction—content segmentation
 - metacognitive strategy use
 - vocabulary.

Developing these measures ultimately will enable researchers to look at the similarity of comprehension operations across a variety of text types and of content.

- Develop measures that reflect engagement and the application of knowledge as consequences of comprehension in order to relate those consequences to the more commonly studied ones of (often temporary) knowledge accumulation. Such measures could then also serve as a research agenda for seeking to understand how interest or motivation affects the reading comprehension process and could be used by teachers to select optimally engaging texts for instructing their struggling students.
- Assess systematically the effect of various accommodations for second-language readers on comprehension outcome measures. Accommodations to be tested might include various manipulations of the text (simplified syntax, modified rhetorical structures, access to translations of key words), manipulations of preparation for reading (providing background knowledge in first-language reading or pre-teaching key vocabulary with translations), or manipulations of response modes (responding in the first language or responding with support from a first-language dictionary).
- Evaluate systematically the dimensions measured by different assessments in relation to more traditional assessments and the proposed new approaches to assessment. How well does the dimensionality map onto the theories underlying the development of the assessments?

Examples of Long-Term Activities

- Use the accumulated information about what assessments are available and how they might be well used to develop professional training for teachers and other decisionmakers in how to interpret and use assessment data optimally.
- Collate information from several states or large districts that are using comparable and adequate assessments of reading comprehension to establish benchmarks for appropriate progress in reading comprehension and determine scores that reflect those benchmarks on a variety of measures. This would be a first step in formulating a detailed picture of how serious the problem of comprehension achievement in the United States really is.