Assessment of Reading Comprehension

Written by: Scott G. Paris, Department of Psychology, University of Michigan

Introduction

Current definitions acknowledge that reading comprehension involves the construction of meaning from text using a wide variety of skills and knowledge (e.g., National Reading Panel, 2002; Snow, Burns, & Griffin, 1998). The National Assessment of Educational Progress (NAEP) 2009 Reading Framework Committee defines reading comprehension as …"an active and complex process that involves understanding written text, developing and interpreting meaning, and using meaning as appropriate to type of text, purpose and situation (National Center for Educational Statistics, 2005, p. 2). To construct meaning, readers must decode words fluently, understand vocabulary, make inferences, and relate the ideas in text to their prior knowledge and experiences. These skills vary with age, experience, instruction, context, and motivation so both the processes and the products of reading comprehension are constructive, multidimensional, developmental, and variable. Thus, reading comprehension is difficult to define simply and measure neatly.

Assessment of reading comprehension has been controversial because (a) summative measures of reading have been used in high-stakes tests to make comparisons about proficiency levels of students and (b) researchers have shown how the complex interaction of many factors can influence the assessment of comprehension across texts, instruction, and response formats. Sweet (2005) summarized the findings of the 2002 RAND Reading Study Group by noting that:

> Current available assessments in the field of reading comprehension generate persistent complaints that these instruments:
> • 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 rest on an understanding of reading comprehension as a developmental process or as a product of instruction,
> • do not examine the assumptions underlying the relation of successful performance to the dominant group's interests and values,
> • are not useful for teachers,
> • tend to narrow the curriculum,
> • are unidimensional and method-dependent, often failing to address even minimal criteria for reliability and validity. (pp. 4-5)

From the first reading tests at the turn of the 20th century to the "cognitive revolution" in the 1970s, the dominant method of assessing reading comprehension required students to read passages silently and to respond to short-answer or multiple-choice questions

(Pearson & Hamm, 2005). The traditional measurement of reading comprehension remains popular today because the quantitative scores on the same scales provide summative measures of reading that can be used to sort and compare students. In contrast, cognitive approaches acknowledge that measures of reading comprehension are variable and indirect indicators, such as oral and written responses to text, that serve as formative measures that are useful for instruction and remediation (Fletcher, 2006; Johnston, 1984). Choosing a measure of reading comprehension therefore depends on the purpose of assessment.

Research Questions

1. What are the purposes for assessing reading comprehension?
2. How is reading comprehension usually assessed?
3. How is comprehension assessed among beginning readers?
4. Can informal reading inventories and curriculum-based measurements assess reading comprehension?

Recent Research Results

What are the purposes for assessing reading comprehension?

Kameenui et al. (2006) identified four decision-making purposes of early reading assessment: screening, diagnosis, progress monitoring, and outcome evaluation. Carlisle and Rice (2004) identified four similar purposes for assessing reading comprehension in school settings:

1. state and district evaluation of programs and curricula;
2. identification of children at risk for reading problems;
3. diagnoses of children's reading problems; and
4. measurement of student progress during instruction or intervention.

The first purpose includes testing for accountability, and it has been the main focus historically. Reading comprehension has been assessed as an indicator of both literacy skill and academic achievement, so it has been used to measure the effectiveness of teachers, curricula, instruction, and new programs. Standardized tests with scaled scores are also used as summative measures to sort students by ability (purpose # 2) and to monitor progress of students and schools (purpose #4). Around the world, the test scores are used as proxy measures of the quality of instruction provided by teachers and schools so the reading scores are often reported in media and used in comparisons of districts, schools, states, and nations (purpose #1). High-scoring students are designated for awards and academic tracks in future schooling. Low-scoring students are identified for remedial services or vocational tracks of study in many countries. Fewer assessments of reading comprehension are designed for diagnostic purposes (purpose #3). These tests focus on specific cognitive processes, such as differences in memory, monitoring, inference generation, or strategy use among students, and they are often used with beginning or struggling readers. Diagnostic assessments are designed to be aligned with and inform instruction in classrooms, and

they are becoming more numerous and often embedded in new technologies (Fletcher, 2006).

How is reading comprehension usually assessed?

Most assessments of reading comprehension have been designed for students at or above grades 3-4 when decoding skills have become independent, fluent, and silent (Rathvon, 2004). The assessments usually require students to read (silently and without assistance) many short passages and to answer a variety of questions, multiple-choice or open-ended, about implicit and explicit information in the text. The method is based on (a) quantitative models of normative skills used to sort students on the basis of uniform scores, (b) psychometric models with samples of text and questions in item pools that yield high reliability and validity across items and test forms, and (c) economically efficient models of standardized testing with group administration and computerized scoring. The popularity of the test format and method can be traced to the appeal of the psychometric approach to measuring human abilities that emerged in psychology in the first half of the 20th century, the invention of scanners and computers to score tests mechanically, and the rise of mastery learning and criterion-referenced testing in the 1970s (Pearson & Hamm, 2005). Traditional methods of assessing reading comprehension with standardized tests and multiple-choice questions are the most frequent type of assessment used in commercial reading tests, state-mandated achievement tests, and tests used to compare reading proficiency among countries.

Although standardized tests of reading comprehension are used with students in grades 4-12, teachers in grades K-3 are more likely to use formative and informal assessments of reading comprehension. Surveys have revealed a wide variety of commercial tests that teachers can use to assess hundreds of skills in young readers (Pearson, Sensale, Vyas, & Kim, 1999; Stallman & Pearson, 1990). Most early reading tests are designed for individual administration and focus on decoding skills, word recognition, and vocabulary; few commercial tests assess comprehension. Kameenui et al. (2006) reviewed the adequacy of reading assessments in K-3 and found great variation among assessments. They created a set of evidence-based criteria to evaluate reading assessments, and they concluded, “many measures do not provide enough evidence of trustworthiness to warrant use” (p.9).

Teachers in K-3 use anecdotal records, daily performance in reading groups, and observations to assess their students’ comprehension, and they prefer informal measures to standardized tests (Paris & Hoffman, 2004). In order to provide more formal and uniform measures, some states have created their own diagnostic reading assessments. For example, the 2002 Texas Primary Reading Inventory (TPRI), the 2002 Virginia Phonological Awareness Literacy Screening (PALS), the 2002 Michigan Literacy Progress Profile (MLPP), and the 2003 Illinois Snapshots of Early Literacy (ISEL) all include measures of the five essential components of reading identified by the National Reading Panel (2000): alphabetic knowledge, phonological awareness, oral reading fluency, vocabulary, and comprehension. However, comprehension is a minor focus of the tests and usually is measured with retellings and multiple-choice questions after children read or hear a short passage.

How is comprehension assessed among beginning readers?

Reading comprehension of students in grades K-3 is usually assessed with formative (i.e., informal diagnostic tasks that inform instructional decision-making) rather than summative (i.e., scores that summarize performance and allow comparisons among test-takers) measures because the main purpose of assessment with beginning readers is to identify children who need additional instruction. This "low-stakes" approach may be partly responsible for the lack of rigorous evidence about the validity, reliability, and utility of early assessments. Teachers use three types of informal comprehension assessments most frequently; oral retellings, answering questions, and cloze tasks. Each format can be used as a comprehension assessment of listening, viewing, or reading so they provide developmental bridges from listening and viewing tasks to reading comprehension. First, oral retellings of text information that children hear, view, or read can assess children's understanding of main ideas, sequences of events, and narrative elements such as characters, settings, and plots (Yussen & Ozcan, 1996). Retelling stories has been shown to facilitate comprehension and oral language in young children, and assessments of retellings are correlated with reading comprehension scores (Morrow, 1985; 1990).

Second, children's answers to questions can occur after viewing, hearing, or reading text. The memory and language demands are present in each modality, but confounding due to differences in decoding skills is removed when children are questioned about information they see or hear. For example, children's comprehension of televised narratives is correlated significantly with their reading comprehension (van den Broek., Kendeou, Kremer, Lynch, Butler, White, & Lorch, 2005). Likewise, children's comprehension of narrative elements and relations depicted in wordless picture books during grades K-2 is correlated significantly with their reading comprehension assessed 1-2 years later (Paris & Paris, 2003; van Kraayenoord & Paris, 1996). Regardless of the modality, children usually have more difficulty answering questions based on implicit information, such as inferences in the text and the author's purpose, as opposed to explicit text information, such as facts and details.

The third format used to assess both beginning and skilled readers involves supplying missing words. Cloze tasks require children to fill in missing words in text, whereas maze tasks require children to choose the missing word from several alternatives, i.e., a multiple-choice task. Cloze tasks have been criticized because they may measure comprehension only within sentences based on word associations as opposed to comprehension of meaning across sentences (Shanahan, Kamil, & Tobin, 1982), yet they remain popular assessment tasks in commercial as well as informal assessments.

Can informal reading inventories and curriculum-based measurements assess reading comprehension?

Oral reading has been a focus for the assessment of early reading development throughout the 20th century, and informal reading inventories remain popular diagnostic assessments (Rasinski & Hoffman, 2003). Most informal reading inventories include a variety of graded passages that children read alone and aloud while a teacher records

the rate, accuracy, and intonation of the oral reading. Inventories often include a retelling task and a set of questions to assess comprehension. When children can read text passages at their own grade level with at least 98% correct word recognition and 90% correct comprehension, they are considered to be independent readers at their grade level (e.g., Leslie & Caldwell, 2001). However, these criteria may vary with the difficulty of the text and the purpose of the test.

Some researchers have suggested that oral reading fluency in informal reading inventories is the best predictor of reading achievement in elementary grades (Fuchs, Fuchs, & Maxwell, 1988). Researchers in special education have argued that brief assessments of oral reading rate are good measures of reading competence, and by inference, oral reading rates may indirectly assess reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Hosp & Fuchs, 2005). Because the text passages can be drawn from the students' curriculum, this approach was labeled "curriculum-based measurement" (CBM), but researchers have shown that one-minute samples of oral reading from grade-appropriate texts yield similar results. CBM has been used for screening, diagnostic, and progress-monitoring purposes. The brief tests are appealing for their efficiency and quantitative scores as well as their usefulness for identifying struggling readers. Because oral reading fluency is necessary for decoding and comprehension, CBM is strongly correlated with many measures of reading competence, including comprehension. Some evidence, however, suggests that oral reading fluency becomes more dissociated from comprehension after grade 3 (Kranzler, Miller, & Jordan, 1999; Paris, Carpenter, Paris, & Hamilton, 2005; Stahl & Hiebert, 2006).

Conclusions, Implications, and Future Directions

There are many ways to assess reading comprehension so care must be exercised to match the level of proficiency of the readers and the purpose of testing with the format, content, and method of assessment. Responses based on retellings, constructed responses, selection among multiple-choice answers, and filling in the missing words can all yield useful measures of reading comprehension. The reliability and validity of the measures usually increase as decoding skills, topic familiarity, and test-taking skills increase because these, and other factors, may confound measures of comprehension.

Future research will lead to better and earlier identification of children who are at risk for reading comprehension problems due to factors such as impoverished literacy environments (Scarborough & Dobrich, 1994), inadequate vocabularies (Qian, 1999), and language impairments (Nation, 2005). Future research will also provide technological tools to administer, score, and interpret test results so that teachers can spend less time testing students and more time providing individualized instruction.

Date Posted Online: 2007-09-21 10:45:42

References

Carlisle, J., & Rice, M. (2004). Assessment of reading comprehension. In A. Stone, E. Silliman, B. Ehren, & K. Apel (Eds*.), Handbook of language and literacy* (pp. 521-555). New York: Guilford Press.

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*(3), 323-330.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 241-258.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20-28.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

Johnston, P. H. (1984) Assessment in reading. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 147-182). New York: Longman.

Kameenui, E. J., Fuchs, L., Francis, D. J., Good, R., O'Connor, R. E., Simmons, D. C., et al. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35*(4), 3-11.

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading*. School Psychology Quarterly, 14,* 327-342.

Leslie, L., & Caldwell, J. (2001). *Qualitative reading inventory – 3*. New York: Addison Wesley Longman.

Morrow, L. M. (1985). Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity. *Elementary School Journal, 85(*5), 646-661.

Morrow, L. M. (1990). Assessing children's understanding of story through their construction and reconstruction of narrative. In L. M. Morrow & J. K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 110-134). Englewood Cliffs, NJ: Prentice-Hall.

Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: a handbook* (pp. 248-265). Oxford: Blackwell.

National Center for Educational Statistics. (2005*). 2009 NAEP reading framework*. Washington DC: Author.

NAEP Reading Consensus Project. (1992). *Reading framework for the 1992 national assessment of educational progress*. Washington, D.C.: US Printing Office.

National Reading Panel (NRP) (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups.* Bethesda, MD: NICHD.

Paris, S. G., Carpenter, R. D., Paris, A. H., & Hamilton, E. E. (2005). Spurious and genuine correlates of children's reading comprehension. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 131-160). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Paris, S. G., & Hoffman, J. V. (2004). Reading assessments in kindergarten through third grade: Findings from the center for the improvement of early reading achievement. *Elementary School Journal, 105*(2), 199-217.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*(1), 36-76.

Pearson, P. D. & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices – past, present, and future (pp. 13-69). In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment.* (pp. 131-160). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Pearson, P. D., Sensale, L., Vyas, S., & Kim, Y. (1999, June). *Early literacy assessment: A marketplace analysis.* Paper presented at the National Conference on Large-Scale Assessment, Snowbird, UT, 1999.

Rasinski, T. V., & Hoffman, J. V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly, 38*(4), 510-522.

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. New York: Guilford Press.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review/ La revue canadienne des langues vivantes. 56*(2), 282-308.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, *14*, 245-302.

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 17*, 229-255.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stahl, S. A., & Hiebert, E. H. (2006). The "word factors". In K. A. Stahl & M. McKenna (Eds*.), Reading research at work* (pp. 403-424). New York: Guilford.

Stallman, A. C., & Pearson, P. D. (1990). Formal measures of early literacy. In L. M. Morrow & J. K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 7 - 44). Englewood Cliffs, NJ: Prentice Hall.

Sweet, A. P. (2005). Assessment of reading comprehension: The RAND reading study group vision. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment.* (pp. 3-12). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. G. Paris & S. A. Stahl (Eds.), *Current issues in reading comprehension and assessment.* Mahwah, NJ: Erlbaum.

van Kraayenoord, C. E., & Paris, S. G. (1996). Story construction from a picture book: An assessment activity for young learners. *Early Childhood Research Quarterly*, *11*, 41-61.

Yussen, S., & Ozcan, N. M. (1996). The development of knowledge about narratives. *Issues in Education*, *2*(1), 1-68.

To cite this document:

Paris, S.G. (2007). Assessment of Reading Comprehension. *Encyclopedia of Language and Literacy Development* (pp. 1-8). London, ON: Canadian Language and Literacy Research Network. Retrieved [insert date] from http://www.literacyencyclopedia.ca/pdfs/topic.php?topId=226