**Mary DeKonty Applegate**
**Kathleen Benson Quinn**
**Anthony J. Applegate**

# Levels of thinking required by comprehension questions in informal reading inventories

*Informal reading inventories may not be the best tool for assessing higher level thinking skills*

Despite widely publicized controversies about how reading should be taught, it is probably fair to say that once we are past the point of sheer word recognition, there is a remarkable amount of agreement about what a good reader is and does (Flippo, 1998). Most reading theorists would agree that when readers encounter printed text, they comprehend by retrieving prior experiences and concepts rooted in their culture and their language. Through a complex interaction of systems, readers construct a plausible interpretation of the text in order to understand, respond to, and react to the meaning intended by the writer (Anderson, Osborn, & Tierney, 1984; Pearson, 1992; Tierney & Shanahan, 1991). Good readers connect past experiences with the text, interpreting, evaluating, and considering alternative responses and interpretations. Rosenblatt (1983) viewed the experience of literature as a uniquely personal interpretation of text that offers an opportunity for readers to expand self-knowledge and to develop their understanding of others. As readers reflect on the experiences of others in light of their own beliefs, those who make effective emotional connections are more likely to develop a lifelong love of reading and to read regularly.

Despite what theorists have said about the nature of reading, many teachers and assessment specialists in the field still measure comprehension by how well children recall the details of what they have read (Allington, 2001). Thus many children are judged as proficient readers because they can answer questions related to the factual information included in the text. However, many test constructors see the issue of comprehension quite differently. Well-publicized changes in the National Assessment of Education Progress (NAEP), Stanford Achievement Tests (SAT-9), and numerous statewide assessments suggest a shift, albeit a gradual one, from objective to more open-ended responses to text (Sarroub & Pearson, 1998). Open-ended items better measure children's ability to think about a story and to use the information in a story to explain their thinking. Thus a combination of objective and open-ended questions in any given assessment may make it possible to gather more specific information about readers. For example, the NAEP results from 1998 suggest that readers in the United States are performing at historically high levels in overall reading achievement (Donahue, Voelkl, Campbell, & Mazzeo, 1999). However, when the assessment focuses on critical reading and responding to text, only a few children demonstrate even minimal proficiency (Allington, 2001).

This shift in the emphasis of reading assessment does not bode well for many children whom the three of us have encountered in our experiences as teacher educators. The children to whom we refer are quite proficient at extracting meaning from text, recalling, and even answering most


 (pp. 174–180)

text-based questions. Many of these same children, however, see reading as an insignificant, boring, annoying, or even painful chore. When these children are asked to discuss the key issues or the significance of what they have read and to support their responses with details and arguments, they are often unable to do so (Foertsch, 1992). They do not spontaneously identify with characters or imagine themselves in a story. They do not make connections between what they know and new information in the text. They seem to have fixated at the point of literal reading comprehension. Consequently, these children often view reading as a mechanical and laborious task, far removed from the type of critical analysis and response in which they regularly engage in their lives outside the classroom.

If these children are in classrooms where the major emphasis of instruction or assessment is on literal recall, their opportunities to discuss ideas related to the text may be limited. Unfortunately, research suggests that classroom questioning is largely literal (Brown, 1991; Elmore, Peterson, & McCarthey, 1996; Johnston, 1997; Knapp, 1995; Tharp & Gallimore, 1989). Literal comprehenders may function effectively and may even be judged as competent readers while they cultivate a negative attitude toward reading. If an insightful teacher or parent suspects a problem, then a more careful and specific diagnosis may be called for. We wondered, however, to what extent commercial informal reading inventories (IRIs), the instruments most likely to be used in the event of a more extensive diagnosis, are effective tools to help teachers and reading specialists break the cycle of low-level questioning. That is, if they consist of items that call for readers to think about text, they may be instruments that reading specialists can use to help teachers become more sensitive to the needs of readers who have difficulty responding to text. Even more to the point, they may provide evidence of the need for teachers to promote responding to literature and reading, a process that is rewarding and effective at improving test performance (Guthrie, Schafer, & Huang, 2001). If, however, they consist of primarily literal and low-level inference items, they are unlikely to be sensitive enough to detect the problems experienced by a great many readers. And while open-ended comprehension questions are by no means the only diagnostic technique used by IRIs, they are always central to estimations of reading levels, one of the key uses of the inventories.

The purpose of this study was to determine whether the types of open-ended questions and the levels of thinking that the questions called for were representative enough to allow users, whether teachers or reading specialists, to determine the extent to which students remember, think about, or respond to what they are reading. To investigate this issue, we defined the following four types of open-ended reading comprehension questions that may appear on the typical IRI: literal, low-level inference, high-level inference, and response items (see Figure). We did not include in our analysis items that asked readers to define vocabulary words, primarily because we could not be certain whether readers used context clues to arrive at their definitions or already knew the definition from prior exposure.

Much of the previous research was characterized by difficulties with item type taxonomies. As Pearson (1983) pointed out, it is hard to classify reading comprehension items in the absence of the information available in the text and the apparent source of a reader's response. Some readers, for example, may interpret a literal question as requiring more than just recall and make an inference. Pearson and Johnson (1978) went so far as to identify a question-answer relationship (QAR) as a more effective construct for the classification of comprehension questions than the analysis of questions alone. We, however, were less concerned about a precise taxonomy of item types and more concerned with the orientation toward reading that sets of comprehension questions would convey to the reader. That is, if examiners asked sets of questions that called primarily for literal recall and low-level inferences, they would, in effect, be conveying a message to examinees about their expectations and beliefs about the nature of reading. If, on the other hand, their questions also called for conclusions and responses on the part of readers, it would be possible for examiners to send a more theoretically sound message about the nature of reading comprehension. Thus for our purposes, the precise classification of an item was less important than the potential message the item conveyed. That is, we asked whether reading comprehension appeared to be defined by the authors of various IRIs as largely text based (literal

**Types and examples of open-ended comprehension questions**

1. **Literal items**: Answers to these items are stated explicitly (verbatim) in the text. They simply require that the readers recall what they have read.

   **Literal questions**: The text states that Mary, a character in the story, is in fourth grade. The literal question asks, "What grade was Mary in?"

2. **Low-level inference items**: The answers to low-level inferences are not stated verbatim in the text but may be so close to literal as to be obvious. All inference items require that the reader draw a conclusion on the basis of the text and use their background experiences to some extent as well. However, low-level inferences require very little in the way of drawing conclusions. For example, we classified the following as low-level inferences.
   - Those that involve the recognition of information in different words from those used in the original text. Such items require of the reader only a translation of the printed text;
   - Those that require the reader to identify relationships that exist between ideas in the text. Such items as these are not literal only because the writer has not made the relationship explicit by using a grammatical marker (e.g., *because*). This in not to say that the skill of making such connections is unimportant. Classification of an item as low level is merely reflective of the fact that the writer assumes that at a given grade level, the reader can and will make the connection;
   - Those that deal with details largely irrelevant to the central message of the text; or
   - Those that require that the reader draw solely on background knowledge or speculate about the actions of characters without the benefit of information in the text that may transform speculation into a logical prediction.

   **Low-level inference questions**: The text states that "Mr. Wilson's car would not start. Mr. Wilson was late for work." The low-level inference question asks, "Why was Mr. Wilson late for work?"

3. **High-level inference items**: These items call for the reader to link experience with the text and to draw a logical conclusion. Answers to these items require significantly more complex thinking than low-level inferences. Examples include those items that require the reader to do the following:
   - Devise an alternative solution to a specific problem described in the text.
   - Describe a plausible motivation that explains a character's actions.
   - Provide a plausible explanation for a situation, problem, or action.
   - Predict a past or future action based on characteristics or qualities developed in the text.
   - Describe a character or action based on the events in a story.

   **High-level inference questions**: The text describes two characters and several circumstances in their lives. The high-level inference question asks, "Why do you think that the two characters in the story became friends?"

4. **Response items**: These items call for a reader to express and defend an idea related to the actions of characters or the outcome of events. Response items differ from high-level inference items in that they are usually directed toward broader ideas or underlying themes that relate to the significance of the passage. While high-level inference items are directed toward a specific element or problem in the passage, response items require a reader to discuss and react to the underlying meaning of the passage as a whole. Examples include items that ask the reader to do the following:
   - Describe the lesson(s) a character may have learned from experience.
   - Judge the efficacy of the action or decisions of a character and defend the judgment.
   - Devise and defend alternative solutions to a complex problem described in a story.
   - Respond positively or negatively to a character based on a logical assessment of the actions or traits of that character.

   **Response questions**: The story describes characteristics of two young children on a field trip. The response question asks, "If you were a teacher, which of the two children would you rather have in your class and why?"

and low-level inferences) or if it also involved conclusions about and responses to ideas in the text (high-level inference and response).

To address that question, we identified, via sales data and references in the professional literature, eight of the most widely used and cited IRIs. We then randomly chose two passages and their accompanying comprehension questions at each elementary grade level (pre-primer through sixth grade). We selected only those IRI forms identified by the author(s) as narrative because higher level items are more likely to be found in response to narrative text. We found wide variations in the narrative passage format, with biographies and content text included along with passages that would fit the more classic definition of a story. We met briefly to review and clarify the descriptions associated with each of the item types as well as to resolve any questions. Each of us then classified the items independently. Any disagreements were resolved through subsequent discussion.

Interrater reliability was calculated by comparing item classifications among all three of us and then calculating a percentage of agreement. The overall level of agreement was 92.2%, which we regarded as exceptionally high. Much of that level of agreement could be accounted for by the ease with which we were able to classify literal items, those whose answers were stated verbatim in the text. We also found it rather easy to distinguish between items that required higher levels of thinking (high-level inference and response items) and those that required low levels (low-level inference and literal). Among those items where there was disagreement, there were remarkably few incidences (less than 4%) where one researcher viewed an item as requiring high-level thinking and another viewed it as requiring low-level thinking. Thus we felt that the classification system we constructed could be used reliably by others, particularly those interested in distinguishing between items that called for response-based thinking and text-based thinking as we have defined these. The results of the item classification are presented in the Table.

**Percentage of item types and number of items in informal reading inventories**

| | Literal | Low-level inference | Total text-based | High-level inference | Response | Total response-based |
|---|---|---|---|---|---|---|
| Johns (1994) | 55.2 (*N* = 74) | 36.6 (*N* =49) | 91.8 (*N* = 123) | 7.5 (*N* = 10) | 0.7 (*N* = 1) | 8.2 (*N* = 11) |
| Bader (1998) | 78.9 (*N* = 116) | 13.6 (*N* = 20) | 92.5 (*N* = 136) | 7.5 (*N* = 11) | 0.0 (*N* = 0) | 7.5 (*N* =1 1) |
| Woods & Moe (1999) | 54.3 (*N* = 44) | 35.8 (*N* = 29) | 90.1 (*N* = 73) | 7.4 (*N* = 6) | 2.5 (*N* = 2) | 9.9 (*N* = 8) |
| Silvaroli & Wheelock (2001) | 65.7 (*N* = 44) | 32.8 (*N* = 22) | 98.5 (*N* = 66) | 1.5 (*N* = 1) | 0.0 (*N* = 0) | 1.5 (*N* = 1) |
| Burns & Roe (1993) | 54.8 (*N* = 69) | 33.3 (*N* = 42) | 88.1 (*N* = 111) | 11.9 (*N* = 15) | 0.0 (*N* = 0) | 11.9 (*N* =1 5) |
| Flynt & Cooter (2001) | 61.5 (*N* = 59) | 28.1 (*N* = 27) | 89.6 (*N* = 86) | 7.3 (*N* = 7) | 3.1 (*N* = 3) | 10.4 (*N* = 10) |
| Leslie & Caldwell (2001) | 57.9 (*N* = 66) | 23.7 (*N* = 27) | 81.6 (*N* = 93) | 17.5 (*N* = 20) | 0.9 (*N* = 1) | 18.4 (*N* = 21) |
| Shanker & Ekwall (2000) | 92.6 (*N* = 113) | 6.6 (*N* = 8) | 99.2 (*N* = 121) | 0.8 (*N* = 1) | 0.0 (*N* = 0) | 0.8 (*N* = 1) |
| **Total** | **66.0 (*N* = 585)** | **25.2 (*N* = 224)** | **91.2 (*N* = 809)** | **8.0 (*N* = 71)** | **0.8 (*N* = 7)** | **8.8 (*N* = 78)** |

## Discussion of results

As can be seen from the data presented in the Table, more than 91% of the nearly 900 items that we classified required text-based thinking; that is, either pure recall or low-level inferences. Nearly two thirds of the items we classified fell into the purely literal category, requiring only that the reader remember information stated directly in the text. Items that called for readers to draw logical conclusions accounted for less than 10% of all items, while items that called for a response to the text were rarer still, occurring at a rate of less than 1%. As might be expected, not all of the inventories were equal in terms of the thinking demands that they placed upon readers, but even in the most demanding of the IRIs that we analyzed, less than one fifth of the items required more than text-based recall or low-level thinking. We thought we would find more text-based items accompanying passages from the lower grades (pre-primer through second grade), but this was not the case. The balance of text-based and response-based items remained consistent throughout the different grade levels included in this sample.

Thus it appears that widely used IRIs are overwhelmingly text based, emphasizing the reader's ability to reproduce the ideas of the author rather than to integrate those ideas with their own knowledge, question their or the author's beliefs, or actually use ideas from the text (Tierney & Pearson, 1992). We found as well that a number of items that the authors had intended to measure the ability to infer about text actually centered solely upon the background knowledge of the reader. For example, in a story that ends with a child successfully catching a fish, the question, "What do you think he will do with the fish?" in combination with the instruction to accept any plausible response, suggest that pure speculation is called for in the item. The essence of reading, however, is the ability to link past experience with the text. Items that would require the reader to predict, based on information or character development in the text, the fate of the fish in the story would meet the criteria for response-based comprehension questions and high-level thinking about text.

## Implications

Open-ended questions can take the reading teacher where multiple-choice items cannot: to the children's ability to use their experiences to construct meaning in response to text. But the overwhelming emphasis on low-level, text-based thinking reflected in our analysis represents a missed opportunity to use the structure and potential of the IRI to tap into the reader's higher level thinking and response skills. Without a balance of items that require reader response and those that require only memory, the IRI may not be sensitive enough to distinguish between those children who can remember text and those who can think about it. Struggling readers with cultural, linguistic, economic, or educational disadvantages are those most likely to be diagnosed by means of the IRI. Ironically, these children have often experienced a curriculum that focuses on low-level thinking tasks and memory for details (Allington & Walmsley, 1995). Consequently, they may experience more success on text-based items and thus mask their need for instruction in higher level thinking.

Perhaps even more important, however, is that an IRI may ultimately be unable to help us distinguish between children who can read and those who are likely to. That is, if children, through their experiences in literacy classrooms, arrive at the internal view that reading involves no more than the storage and retrieval of the details of text, there would appear to be very limited motivation for them to engage in reading.

In the broader educational community, the implications are even more disturbing. Reading specialists and reading teachers often serve as catalysts for change in the elementary school classroom. But if they are unable to assess children's ability to think about and respond to text, they will lose one of their most powerful tools for increasing the awareness of their colleagues about the importance of a child's thoughtful response to text. Notwithstanding the debate on the wisdom of assessment driving the curriculum (see Calfee, 1987; Popham, Cruse, Rankin, Sandifer, & Williams, 1985; Valencia & Pearson, 1987), it would be naïve of reading teachers and specialists to conclude that assessment does not profoundly affect the way that reading (or any other subject) is taught. If we wish to have an impact upon that teaching, we need to be able to

produce evidence that important dimensions of the act of comprehension are not part of the view of reading held by a significant number of children. Otherwise, many children, parents, and teachers will have little or no reason to question the widespread assumption that literal comprehension is the final word in reading assessment.

More specifically, we must be able to demonstrate that some children taught in classrooms where teachers emphasize the recall of details cannot think about what they read. Otherwise, there would be no reason for literal-minded teachers to reflect upon their teaching or to try techniques that promote a higher level of thoughtful literacy. If the IRIs that we use to assess children are insensitive to the differences between recalling and thinking about text, our ability to provide evidence of any given child's instructional needs, let alone to have an impact upon instruction, is severely limited.

## Recommendations

We believe that the results of our study of IRIs warrant several specific recommendations to the reading community.

1. When IRIs are needed, teachers and reading specialists must choose them judiciously. Our results suggest that not all IRIs are created equal in terms of their ability to assess thinking and responding to text. Reading teachers and specialists must take care to assess the IRI in order to ensure a level of congruence between their assessment and the objectives that they wish to promote.

2. We recommend that publishers of IRIs seek out opportunities to encourage and promote the assessment of thoughtful literacy within the structure of current and future IRIs. Because IRIs are not designed to compare individuals or groups, some teachers and specialists may even choose to include their own items that call for higher levels of thought and response.

3. We recommend that teachers and reading specialists emphasize the assessment and teaching of thoughtful responses to text in their classrooms. Incorporating questioning techniques that encourage a response from children, as well as establishing the expectation among children that literal comprehension is not sufficient in their classrooms, can help to identify reluctant or struggling readers as early as possible. Techniques such as literature circles and response journals, when accompanied by appropriate modeling, can promote not only thoughtful literacy but also provide diagnostic assessment before the more formal assessment of the IRI.

4. We recommend that teacher education programs assess the extent to which they promote thoughtful literacy on the part of their teacher candidates. It is unrealistic to expect teachers to help students develop thinking and responding abilities that they do not themselves have and use. Teacher educators must also assess the extent to which their candidates are learning to become learners and critical thinkers. It is essential that teacher candidates demonstrate the ability to integrate, critique, and use insights from all of the segments of a sometimes fragmented curriculum and to bring these insights to bear on the education and preparation of their future students.

5. We need to promote professional dialogue within the school and community about the nature and the necessity of thoughtful literacy. Teachers, administrators, parents, and children should be aware of the central importance of thinking about and responding to text, as well as the ever-increasing role that thinking and responding play in the assessment of reading.

6. Finally, we recommend that teachers and reading specialists position themselves at the cutting edge of the assessment revolution in reading. Diagnostic instruments such as the IRI have long been entrenched in the domain of reading teachers and specialists. We should not relegate ourselves to the position of waiting for standardized assessment measures to lead the way toward a fundamental change in instruction. In our diagnostic instruments and in the approach we take to teaching, the ability to think and respond to text must be at the core of our philosophy of reading.

---

*Mary Applegate teaches at St. Joseph's University in Philadelphia. She may be contacted at 401 Santillo Way, Downingtown, PA 19335, USA or by e-mail at Tapple1492@aol.com. Quinn and Anthony Applegate teach at Holy Family College in Philadelphia.*

### References

Allington, R.L. (2001). *What really matters for struggling readers: Designing research-based programs*. New York: Addison Wesley Longman.

Allington, R.L., & Walmsley, S.A. (1995). *No quick fix: Rethinking literacy programs in America's elementary schools*. Newark, DE: International Reading Association.

Anderson, R.C., Osborn, J., & Tierney, R.J. (1984). *Learning to read in American schools: Basal readers and content texts*. Hillsdale, NJ: Erlbaum.

Bader, L.A. (1998). *Reading and language inventory* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Brown, R.G. (1991). *Schools of thought: How the politics of literacy shape thinking in the classroom.* San Francisco: Jossey-Bass.

Burns, P.C., & Roe, B. (1993). *Informal reading inventory*. Boston: Houghton-Mifflin.

Calfee, R. (1987). The school as a context for the assessment of literacy. *The Reading Teacher, 40*, 738–743.

Donahue, P.L., Voelkl, K.E., Campbell, J.R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states* (NCES 1999-500). Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education.

Elmore, R.S., Peterson, P.L., & McCarthey, S.J. (1996). *Restructuring in the classroom: Teaching, learning and school organization.* San Francisco: Jossey-Bass.

Flippo, R.F. (1998). Points of agreement: A display of professional unity in our field. *The Reading Teacher, 52*, 30–40.

Flynt, E.S., & Cooter, R.B. (2001). *Reading inventory for the classroom* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Foertsch, M.A. (1992). *Reading in and out of school: Achievement of American students in grades 4, 8 and 12 in 1989–90*. Washington, DC: National Center for Educational Statistics: U.S. Government Printing Office.

Guthrie, J.T., Schafer, W.D., & Huang, C. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94*, 145–162.

Johns, J. (1994). *Basic reading inventory*. Dubuque, IA: Kendall-Hunt.

Johnston, P.H. (1997). *Knowing literacy: Constructive literacy assessment.* York, ME: Stenhouse.

Knapp, M.S. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.

Leslie, L., & Caldwell, J. (2001). *Qualitative reading inventory* (3rd ed.). New York: Longman.

Pearson, P.D. (1983). A critique of F.J. Guszak's study: Teacher questioning and reading. In L. Gentile, M. Kamil, & J. Blanchard (Eds.), *Reading research revisited* (pp. 271–281). Columbus, OH: Merrill.

Pearson, P.D., (1992). Reading. In *The encyclopedia of educational research* (pp. 1075–1085).

Pearson, P.D. & Johnson, D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart & Winston.

Popham, W.J., Cruse, K.L., Rankin, S.C., Sandifer, P.D., & Williams, P.L. (1985). Measurement driven instruction: It's on the road. *Phi Delta Kappan, 55*, 628–634.

Rosenblatt, L.M. (1983). *Literature as exploration* (4th ed.). New York: Modern Language Association of America.

Sarroub, L., & Pearson, P.D. (1998). Two steps forward, three steps back: The stormy history of reading comprehension assessment. *The Clearing House, 72*, 97–106.

Shanker, J.L., & Ekwall, E.E. (2000). *Ekwall-Shanker reading inventory* (4th ed.). Boston: Allyn & Bacon.

Silvaroli, N.J., & Wheelock, W.H. (2001). *Classroom reading inventory* (9th ed.). New York: McGraw-Hill.

Tharp, R.G., & Gallimore, R. (1989). Rousing schools to life. *American Educator*, 13(2), 20–25, 46–52.

Tierney, R.J., & Pearson, P.D. (1992). A revisionist perspective on "Learning to learn from text: A framework for improving classroom practice." In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 514–519). Newark, DE: International Reading Association.

Tierney, R.J., & Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions and outcomes. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 246–280). White Plains, NY: Longman.

Valencia, S., & Pearson, P.D. (1987). Reading assessment: Time for a change. *The Reading Teacher, 40*, 726–732.

Woods, M.L., & Moe, A.J. (1999). *Analytical reading inventory* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

## Erratum

The discussion of recommender systems that appeared on p. 620 in "Where Do You Want to Go Today? Inquiry-based Learning and Technology Integration" by Roxanne Farwick Owens, Jennifer L. Hester, and William H. Teale (April 2002 *RT*) included a reference to http://it-edtech.ed.usu.edu/alteredvista as the source of the work on recommender systems. However, the article did not make clear that these ideas were attributable solely to work conducted by Mimi Recker, Kimberly Lawless, and Andrew Walker. Teale, the author of this section of the article, apologizes for inadvertently not providing proper attribution to these researchers and wishes to make clear that the Recker et al. sources were the basis for the entire discussion. In addition, the following reference should have been included:

Recker, M.M., Walker, A., & Wiley, D. (2001). Collaboratively filtering learning resources. In D. Wiley (Ed.), *Designing instruction with learning objects* (pp. 243–259). Bloomington, IN: Association for Educational Communications and Technology.

Finally, it should be noted that the new URL for the website related to this work on systems is http://alteredvista.usu.edu.