A Critical Analysis of Eight Informal Reading Inventories

By: International Reading Association

There are a number of current informal reading inventories — each has its strengths, limitations, and unique characteristics, which should be considered in order to best fit a teacher's needs.

Comment

As a classroom teacher, reading specialist, and university professor, I have always found helpful published summaries or syntheses of professional-related information relevant to my work. In this article, I review the current editions of eight informal reading inventories (IRIs) published since 2002 that are available at the time of this writing. Specifically, I identify key issues surrounding the use of IRIs and examine ways in which the various IRIs reviewed approach them. A goal of this undertaking is to guide teachers, reading specialists, reading coaches, administrators, professionals in higher education, and others charged with the education or professional development of preservice or inservice teachers in their quest to find IRIs best suited to their specific needs. I hope the findings point to new ways in which IRIs can be made even more effective in the near future.

What are informal reading inventories (IRIs)?
IRIs are individually administered diagnostic assessments designed to evaluate a number of different aspects of students' reading performance. Typically, IRIs consist of graded word lists and passages ranging from preprimer level to middle or high school levels (Paris & Carpenter, 2003). After reading each leveled passage, a student responds orally to follow-up questions assessing comprehension and recall. Using comprehension and word recognition scores for students who read the passages orally, along with additional factors taken into consideration (e.g., prior knowledge, fluency, emotional status, among other possible factors), teachers or other education-related professionals determine students' reading levels.

They also use this information to match students with appropriate reading materials, place children in guided reading groups, design instruction to address students' noted strengths and needs, and document reading progress over time. While IRIs serve a variety of

purposes, perhaps their greatest value is linked to the important role they play in helping educators to diagnose the gaps in the abilities of readers who struggle the most.

Based on notions implicit in developmental (Chall, 1983; Spear-Swerling & Sternberg, 1996) and interactive models of reading (Rumelhart, 1977; Stanovich, 1980), IRIs provide information about students' reading stages and knowledge sources. For example, by charting and analyzing patterns in oral reading error types, educators identify whether students rely on one cueing system (i.e., graphophonic, syntactic, or semantic cueing system) to the exclusion of the others, as beginning readers typically do, or if they use a balance of strategies, as mature readers at more advanced stages do in their reading development when they encounter challenges while processing text. Supplemented by other measures of literacy-related knowledge and abilities, as needed, IRIs contribute valuable information to the school's instructional literacy program.

Rationale for selecting IRIs to evaluate

Given the sweeping, education-related policy changes associated with the No Child Left Behind Act signed into U.S. law in 2002, the IRIs included in this analysis were limited to those published since 2002 because it was felt that they would be more likely to reflect features relevant to the policy changes than IRIs published earlier. For example, federal guidelines specify that the screening, diagnostic, and classroom-based, instructional assessments used by schools receiving Reading First grants to evaluate K-3 student performance must have proven validity and reliability (U.S. Department of Education, 2002) — aspects noted as weak with regard to IRIs published earlier (Kinney & Harry, 1991; Klesius & Homan, 1985; Newcomer, 1985).

In addition, specifications in *Guidance for the Reading First Program* (U.S. Department of Education, 2002) require that educators in Reading First schools evaluate students in the five critical areas of reading instruction (i.e., comprehension, vocabulary, fluency, phonemic awareness, and phonics) as defined by the National Reading Panel (NRP; National Institute of Child Health and Human Development [NICHD], 2000) and screen, diagnose, and monitor students' progress over time. Given these federal mandates, it was assumed that IRIs published since 2002 would be more apt to exhibit the technical rigor and breadth in assessment options necessary to help reading professionals achieve these goals.

The names of specific IRI instruments identified were obtained from searches in the professional literature or recommended by professionals in the field of literacy. In all, eight IRIs were identified, examined, and cross-compared with regard to selected features of their most current editions. The following were the IRIs included in this analysis: Analytical Reading Inventory (ARI; Woods & Moe, 2007), Bader Reading and Language Inventory (BRLI; Bader, 2005), Basic Reading Inventory (BRI; Johns, 2005), Classroom Reading Inventory (CRI-SW; Silvaroli & Wheelock, 2004), Comprehensive Reading Inventory (CRI-CFC; Cooter, Flynt, & Cooter, 2007), Informal Reading Inventory (IRI-BR; Burns & Roe, 2007), Qualitative

Reading Inventory-4 (QRI-4; Leslie & Caldwell, 2006), and The Critical Reading Inventory (CRI-2; Applegate, Quinn, & Applegate, 2008).

Analyzing the IRIs

In order to cross-compare selected features of the current editions of all eight IRIs, a coding spreadsheet was prepared and used to assist in the systematic collection of data. The categories used were chosen because of their relevance to issues in the professional literature (e.g., length of passages, type of comprehension question scheme used) or to policy and other changes affecting the field today (e.g., assessment options related to the five critical areas of reading, reliability, and validity information).

Interrater reliability measures

To ensure the accuracy of the coded data, I enlisted the assistance of a graduate student who independently coded one of the IRIs. Afterward, our data charts were compared and the percentage of agreement was determined with differences resolved by discussion. Following this interrater reliability check, data from the separate coding sheets for each IRI were rearranged and compiled onto additional charts in various ways in order to facilitate comparisons and the detection of patterns among variables of interest.

Results

In all, eight IRIs published since 2002 were analyzed and compared in order to identify the variety of ways in which the instruments approach key issues relevant to their use. Based on the analysis, it is evident that the eight IRIs reviewed range in the assessment components they include and in which critical aspects of reading instruction identified by the NRP (NICHD, 2000) they assess. For example, measures for reading comprehension and vocabulary (i.e., sight word vocabulary) were more common than measures in the other areas. An analysis of the IRI features related to each of the five pillars of reading follows.

IRI	Forms or passage types (grade levels)	Passage word length	Question scheme/ retelling rubric focus
Applegate, Quinn, & Applegate (2008)	3 narrative passages (pp-12) 3 expository passages (pp-12)	**Varies:** 66-844	**Questions:** Text-based, inferential, and critical response questions
Bader	3 forms[a] (pp-8,	**Varies:**	**Questions:**

(2005)	9/10, 11/12) Form C (for children) Form C/A (for children, adolescents/adults) Form A (for adults)	31-278	Literal questions and one interpretive question per passage (not included in the total memory score)
Burns & Roe (2005)	4 forms[b] (pp-12)	**Varies: 60-195**	**Questions:** Main idea, detail, inference, sequence, cause/effect, vocabulary **Retelling rubrics:** Two options — a focus on story elements specific to narratives and another rubric option
Cooter, Flynt, & Cooter (2007)	Forms A and B, English, narrative (pp-9) Forms A and B, Spanish, narrative (pp-9) Forms C and D, English, expository (1-9)	**Varies: 25-760**	**Questions:** Literal, inferential, and evaluative questions about story grammar elements for narratives and expository grammar elements for

	Form C, Spanish, expository (1-9) Form D, Spanish, expository (10, 11, 12) Form E, English, expository (10, 11, 12) Emergent literacy assessments		expository passages **Retelling rubrics:** A variety with a focus on story elements for narratives, major points and supporting details for expository text, and other rubric option
Johns (2005)	7 forms (not entirely equivalent): Forms A, B: oral reading (pp-8) Form C: oral reading, expository (pp-8) Form D: silent reading (pp-8) Form E: oral reading, expository (pp-8) Form LN: longer narrative (3-12) Form LE: longer expository (3-12)	**Consistent:** pp = 25 and 50 Forms A-E = 100 Form LN = 250 words Form LE = 250 words	**Questions:** Topic, fact (lower-inference, and vocabulary (higher-level comprehension) questions

	Emergent literacy assessments		
Leslie & Caldwell (2006)	4 narrative, 1 expository (pp) 3 narrative, 2 expository (p-2) 3 narrative, 3 expository (3-5) 3 literature, 2 social studies, 2 science (6) 2 literature, 2 social studies, 2 science (UMS) 1 literature, 1 social studies, 1 science (HS)	**Varies:** 44-786 (pp-UMS) 354-1,224 (HS, passage sections	**Questions:** Explicit and implicit questions that focus on the most important information (e.g., the goal of the protagonist for narratives and the implicit main idea for expository passages and other important information) **Retelling rubrics:** A focus on the most important information for narratives and main idea/supporting details for expository materials
Silvaroli & Wheelock (2004)	3 forms[b] with pre-/posttests for each: Form A, Subskills Format (pp-8)	**Varies:** 38-268	**Questions (subskills format):** Factual, inferential, vocabulary questions

	38-268 Form B, Reader Response Format (pp-8) Form C, Subskills Format for high school and adult education students (1-8)		**Questions (response format):** A focus on story grammar elements
Woods & Moe (2007)	**3 equivalent narrative forms (pp-9):** Form A Form B Form C **2 expository forms (1-9):** Form S (science) Form SS (social studies)	**Varies:** 28-352	**Questions:** Defined by the reader-text relationship: "Retells in Fact" (RIF) "Puts Information Together" (PIT) "Connects Author and Reader" (CAR) "Evaluates and Substantiates" (EAS) **Retelling rubrics:** A focus on story elements for narratives and expository elements for factual text

Note: pp = preprimer level, p = primer level, UMS = upper middle school, HS = high school. [a]Narrative and expository text passages are distributed across levels as

follows: pp.2: narratives only; 3-5: 2 narrative, 1 expository; 6: all expository; 7: 1 narrative, 2 expository; 8: 2 narrative, 1expository; 9-10: 1 narrative, 2 expository; 11-12: all expository.[b]Forms include narrative and expository text passages that are not explicitly identified by genre.

Reading comprehension and recall

Evidence of content validity

According to *Standards for Educational and Psychological Testing* (1999), a fundamental concern in judging assessments is evidence of validity. Assessments should represent clearly the content domain they purport to measure. For example, if the intention is to learn more about a student's ability to read content area textbooks, then it is critical that the text passages used for assessment be structured similarly. Based on their study of eight widely used and cited IRIs, Applegate, Quinn, and Applegate (2002) concluded that there were great variations in the way IRI text passages were structured, including passages with factual content. They observed that biographies and content area text, in some cases, matched up better with the classic definition of a story.

In a similar manner, Kinney and Harry (1991) noted little resemblance between the type of text passages included in many IRIs and the text type typically read by students in middle and high school. As researchers have demonstrated through their studies and analyses, narrative and expository texts are structured differently (Mandler & Johnson, 1977; Meyer & Freedle, 1984), and readers of all sorts, including general education students and children with learning disabilities, process contrasting text types in different ways (Dickson, Simmons, & Kame'enui, 1995). Thus, it makes sense that if the goal of assessment is to gain insights on a student's reading of textbooks that are expository, then the text used for the assessment should also be expository.

Relative to the IRIs examined for this analysis, text passages varied by genre and length as well as by whether the text included illustrations, photos, maps, graphs, and diagrams. A discussion of the ways in which the various IRIs approach these issues follows.

Passage genre

With regard to the text types included in the IRIs under review here (aligned with the perspective that reading comprehension varies by text type), five of the eight IRIs provide separate sections, or forms, for narrative and expository passages for all levels, making it easy to evaluate reading comprehension and recall for narrative text apart from expository material (Applegate et al., 2008; Cooter et al., 2007; Johns, 2005; Leslie & Caldwell, 2006; Woods & Moe, 2007).

However, caution is advised. Despite the separation of genres, in some of the current IRIs, consistent with Applegate et al.'s (2002) observations, some passages classified as expository are actually more like narrative. For example, in BRI (Johns, 2005), the passage

"Have You Played This Game?" contains factual information about the board game Monopoly, but it is written in a narrative style. The passage is placed in the Expository Form LE section; however, the first comprehension question asks, "What is this story about?" Even for passages more expository-like in text structure, at times authors refer to them as "stories" (e.g., "Here is a story about driver's license requirements," Bader, 2005, p. 65; "Tell me about the story you just read," Cooter et al., 2007, p. 275, in reference to the factual passage "Bears").

Of all the IRIs considered, ARI (Woods & Moe, 2007) and QRI-4 (Leslie & Caldwell, 2006) provide expository text passages with features most like text found in science and social studies textbooks. In fact, the authors note most of the passages were drawn from textbooks.

A few of the IRIs appear to take a more holistic approach in their representation of the content domain. For three of the IRIs, the assessment includes a "mix" (Burns & Roe, 2007, p. 227) or "balance" (Bader, 2005, p. 4) of text types with greater emphasis on narratives and no expository passages at lower levels (Bader, 2005; Burns & Roe, 2007; Silvaroli & Wheelock, 2004). In these IRIs, there is no clear separation of narrative and expository text passages.

Passage length

While the passages generally become longer at the upper levels to align with the more demanding texts read by older students, across inventories passage lengths at the same levels vary; some cases, within the same inventory, authors offer passages of different lengths as options at the same levels (see Table 1). For example, finding that beginning readers sometimes struggled with the 50-word, pre-primer passage in earlier editions, Johns (2005) now includes in the ninth edition of BRI a second, shorter passage option of 25 words for each form that offers passages at the preprimer level. In a similar manner, he offers passages of two different lengths at levels 3-12.

Pictures and graphic supplements

Noting the benefits and drawbacks of including illustrations and other graphic supplements with the passages, IRI authors vary in their opinions on this matter. To eliminate the possibility of readers' relying on picture clues rather than their understanding of the text, Silvaroli and Wheelock (2004) and Burns and Roe (2007) exclude illustrations entirely. Bader (2005), Cooter et al. (2007), Johns (2005), and Woods and Moe (2007) limit illustrated passages to lower levels only. Providing examiners with options for comparing beginning readers' performance, Applegate et al. (2008) and Leslie and Caldwell (2006) provide passages with and without illustrations or photos. Moreover, Leslie and Caldwell provide a number of assessment choices at levels 5 through high school, allowing for in-depth and varied evaluations of students' abilities to use different types of graphic supplements

typically found in science and social studies textbooks, such as diagrams, maps, photos, and pie graphs.

Evidence of construct validity

According to *Standards for Educational and Psychological Testing* (1999), a valid test also captures all the important aspects of the construct (i.e., the characteristic or concept that the test is designed to measure), and it also provides evidence that processes irrelevant to the construct do not interfere or distort results. Across IRIs examined, comprehension question frameworks varied in terms of which aspects of narrative or expository text comprehension they centered on, as well as what dimensions, or levels, of comprehension they measured. In addition, across the IRIs reviewed, assorted measures were used to identify extraneous factors potentially affecting comprehension scores. A discussion of the various ways in which each IRI handles these issues follows.

Comprehension/Recall measures

For most of the IRIs reviewed, question schemes introduced alone or in conjunction with retelling rubrics or scoring guides serve to assess a reader's comprehension or recall in two areas: (1) the reader's grasp of narrative and expository text structure and (2) various dimensions or levels of reading comprehension (e.g., literal and inferential comprehension). All of the IRIs attempt to assess these areas either through their question schemes alone or in combination with a retelling and rubric assessment; however, in some cases, the authors use different terms for the dimensions of comprehension they measure.

For measuring narrative text comprehension and recall, six of the eight IRIs focus their question schemes and retelling rubrics on story elements (e.g., character, setting, problem or goal, resolution; Applegate et al., 2008; Burns & Roe, 2007; Cooter et al., 2007; Johns, 2005; Leslie & Caldwell, 2006; Woods & Moe, 2007) based on story grammar theory. It should be noted that the question schemes of Burns and Roe, Johns, and Woods and Moe are structured differently (see Table 1). Thus, if their question schemes are used to evaluate narrative comprehension independently without a retelling and the associated rubric with story elements criteria, then a student's grasp of narrative text structure will not be evaluated.

In the assessment of expository text comprehension and recall, there is greater variety across IRIs. Four IRIs use question schemes or rubrics based on the levels of importance of information (e.g., macro vs. micro concepts, main ideas vs. details; Applegate et al., 2008; Burns & Roe, 2007; Johns, 2005; Leslie & Caldwell, 2006). Taking a different approach, Woods and Moe (2007) and Cooter et al. (2007) provide checklists and question schemes, respectively, for evaluating student recall of expository elements (e.g., description, collection, causation, problem and solution, comparison). Johns includes a variety of rubric options specific to narrative and expository text passages but also more holistic rubrics that he suggests can be used with retellings of any text type.

In addition, in the QRI-4, Leslie and Caldwell provide a think-aloud assessment option useful for capturing information about the strategies readers use while they are in the process of constructing meaning based on the text. To facilitate the use of this assessment option, some of the expository text passages at the sixth, upper middle school, and high school levels are formatted in two different ways that allow for conducting assessments with or without student think-alouds. The authors also provide a coding system for categorizing the think-aloud types based on whether they indicate an understanding or lack of understanding of the text.

It should be noted that Bader (2005) and Silvaroli and Wheelock (2004) use similar criteria for assessing comprehension and recall of narrative versus expository text. For example, in using the BRLI (Bader, 2005) for the assessment of narrative and expository passages, readers are asked to retell the "story" (p. 59), and the idea units recalled are checked off from a list that does not categorize the idea units in any way (e.g., according to story grammar elements in the case of narratives or levels of importance for expository material). In addition, there is a place on the evaluation sheet for checking off whether a student's retelling is organized; however, criteria for making this judgment are lacking. Without a theoretical framework and clearly defined criteria to guide the examiner, it is difficult to determine if the assessment effectively captures the essential qualities of reading comprehension and recall.

The CRI-SW (Silvaroli & Wheelock, 2004) is similar in that there is little distinction in criteria used for judging comprehension or recall of contrasting text types. For example, in the Reader Response Format section of the IRI (the same scoring guide used to evaluate a student's recall of characters, problems, and outcome or solutions for the narrative) "It's My Ball" (p. 136) is provided as a tool for evaluating the factual selection "The World of Dinosaurs" (p. 143). Use of a scoring guide based on story grammar theory seems misplaced as a tool for judging comprehension of expository text.

As noted, in addition to assessing students' understanding of the structural features of narrative and expository text, IRI authors provide measures of various dimensions, or levels, of reading comprehension — most commonly literal and inferential comprehension (Applegate et al., 2008; Bader, 2005; Burns & Roe, 2007; Cooter et al., 2007; Johns, 2005; Leslie & Caldwell, 2006; Silvaroli & Wheelock, 2004; Woods & Moe, 2007). Although the terms for these constructs vary, and there may be subtle differences in meanings across inventories, the dimensions overlap. For example, Leslie and Caldwell refer to explicit and implicit comprehension. Woods and Moe, however, using a reader-text relationship question scheme stemming from Raphael's (1982, 1986) Question-Answer Relationships framework, provide questions measuring fact-based, literal comprehension that call for responses "from the text" as well as questions that measure inferential comprehension or responses "from head to text" (Woods & Moe, 2007, pp. 28-29).

Taking a different approach, Applegate et al. include questions to measure critical response (i.e., a response requiring analysis, reaction, and response to text based on personal experiences and values and usually allowing for more than one possible answer). Cooter et al. (2007) provide questions as measures of evaluative comprehension. Johns's questions measure comprehension dimensions called "lower-level" (i.e., assessed by fact questions) and "higher-level" (i.e., assessed by topic, evaluation, inference, and vocabulary questions; Johns, 2005, p. 76). It should be noted that Silvaroli and Wheelock include assessment of different levels of comprehension (i.e., inferential vs. factual questions) as part of the question taxonomy in the Subskills Format section of their IRI, but this aspect of comprehension is not assessed by the question scheme in the Reader Response Format.

Despite concerns (Applegate et al., 2002; Duffelmeyer & Duffelmeyer, 1987, 1989; Johns, 2005; Schell & Hanna, 1981), a few of the IRIs reviewed continue to use question taxonomies with main idea, fact and detail, inference, and vocabulary questions, among other question types (Burns & Roe, 2007; Johns, 2005; Silvaroli & Wheelock, 2004). In the past, criticisms targeting these question schemes arose out of concern due to lacking empirical support and confusion over what main idea questions in some of the IRIs actually measured.

In the ninth edition of BRI reviewed for this study, citing Schell and Hanna (1981) as his information source, even Johns himself cautions readers, "Lest teachers glibly use the classification scheme suggested, it must be emphasized that these categories of comprehension questions, although widely used, have little or no empirical support" (Johns, 2005, p. 72). For this reason, Johns advises using his own question classification scheme informally and with discretion.

Other scholars in the field of literacy, as well, have suggested that main idea question types included in some IRIs were actually no more than "topic" questions that could be answered in one-word or simple phrase responses rather than full statements of the moral or underlying theme of a story, requiring the integration of selection content (Applegate et al., 2002; Duffelmeyer & Duffelmeyer, 1987, 1989; Schell & Hanna, 1981).

As Applegate et al. pointed out, the ramifications of confusions over question types can be serious in that children who are proficient in responding to questions of one sort, such as questions requiring literal recall and low-level inferences that are largely text based, sometimes experience great difficulty in answering questions of other types, such as those that require more critical thinking. The confusion over question types and just what the questions actually measure restricts the usefulness of the assessment data they yield in terms of helping teachers pinpoint and address children's instructional needs.

While IRI-BR (Burns & Roe, 2007) continues to use a question classification system with main idea questions vulnerable to these criticisms, it is evident that Johns (2005) has made changes to address the terminology issue in BRI. Items that he previously called main idea

questions are now labeled "topic" questions. Otherwise, his classification system remains similar to that in earlier editions. As a result, some of the confusion over question type is eliminated, but if a teacher relies strictly on Johns's question scheme to assess comprehension, a reader's ability to synthesize the content and come up with the main or "big idea" (Walmsley, 2006) of a passage (an important aspect of reading) will not be evaluated.

Silvaroli and Wheelock (2004) include not only the traditional question scheme from earlier editions of CRI (Silvaroli, 1990), but also the authors have added a whole new question framework that supplements, or serves as another option, to the question scheme of their earlier editions. Those who use the newest edition of CRI-SW have a choice as to whether to administer the passages and follow-up questions that fall into the Subskills Format or an alternative set of questions included in what the authors call the Reader Response Format. Accordingly, the five questions accompanying the passages in the Subskills Format, as in the earlier editions, include factual, inferential, and vocabulary question types.

The question types for the retelling portion of the newer Reader Response Format, however, include a prediction question followed by three questions pertaining to the characters (i.e., "Who was the main person in the story?"), the problems (i.e., "What was the problem?"), and the outcomes or solutions (i.e., "How was the problem solved?") of the passage. The authors explain that the rationale for adding the Reader Response Format was to accommodate literacy programs that have shifted from a "subskills instructional emphasis" to a "literacy emphasis" (Silvaroli & Wheelock, 2004, pp. 1-2). They suggest the passages and questions included in each format can be used separately or in some combination, as desired.

Measures of extraneous variables
In order to control for extraneous variables that can affect comprehension and recall, some of the IRI authors include measures of prior knowledge (Bader, 2005; Johns, 2005; Leslie & Caldwell, 2006; Silvaroli & Wheelock, 2004; Woods & Moe, 2007), emotional status (Burns & Roe, 2007; Woods & Moe, 2007), and level of engagement (Johns, 2005). Other authors suggest the administrator informally note observations and student comments in related areas (Burns & Roe, 2007).

Form equivalence/Reliability
Because federal guidelines for Reading First schools require educators to monitor student progress over time (U.S. Department of Education, 2002), it can be valuable to know if the parallel forms within each IRI can be used interchangeably. In order to know how consistent the scores are across forms, it is necessary to obtain the alternate form reliability coefficient. Generally, a correlation of .85 or higher is desirable, with the maximum a correlation can be at +1.00. It is also necessary to have information about the sample

population on which the reliability figure was based in order to generalize to a different student population (Bracey, 2000).

Although the *Standards for Educational and Psychological Testing* (1999) suggests a need to report critical information indicating the degree of generalizability of scores across alternate-forms, few of the IRI authors do. Only one IRI (Leslie & Caldwell, 2006) provides data suggesting the forms for determining that reading comprehension levels may be used interchangeably, although the specific IRI edition used for that reliability study is not reported.

With respect to the alternate forms of the QRI text passages, Leslie and Caldwell found the reliabilities based on comprehension scores were all above 0.80, and 75% of the scores were greater than or equal to .90. In addition, the authors examined whether the same instructional level would be determined based on the comprehension scores of each passage and report that 71% to 84% of the time the instructional level was the same on both. The individual reliability levels for each grade-level text from primer level through upper middle school are reported.

In some IRIs, the authors infer that alternate-form reliability levels are acceptable; however, information is lacking to confirm that. For example, based on the similar content that occurs across all three narrative forms in ARI (e.g., all three passages at level 6 are written about famous African American scientists or inventors), Woods and Moe (2007) suggest, "This consistency enables the examiner to change forms when determined necessary" (p. 257); however, because no correlation coefficient indicating degree of equivalence is reported, this inference cannot be made with confidence.

With respect to IRI-BR, Burns and Roe (2007) state, "Alternate forms testing revealed that the levels indicated by different forms administered to the same students were consistent" (p. 229); however, without reliability figures reported, the examiner cannot make a judgment about the degree of reliability. Also, without a sample description, even if the forms are equivalent for one sample population, given the possible differences across groups, it may not be possible to generalize those results to another student population.

In addition, in John's (2005) ninth edition of BRI, he refers to an alternate-form reliability study (Helgren-Lempesis & Mangrum, 1986) of BRI (Johns, 1981) and two other IRIs, which indicated the Pearson *r* coefficients for BRI were .64 for independent level, .72 for instructional level, and .73 for the frustration level. However, these results pertain to a 1986 study with fourth-grade students who orally read passages from Forms A and B of the second edition of BRI. New reliability information pertaining to all forms in the current edition and for passages at all levels read orally and silently is needed in order to use parallel forms interchangeably without question.

Some of the alternate-form reliability figures reported are lower than is desirable. For example, based on the figures reported by Cooter et al. (2007) for grades 1, 2, and 3 (i.e., 0.58, 0.63, and 0.70), the authors caution that Forms A and B may not be equivalent. The authors also report that due to small sample sizes, they were not able to obtain reliability figures for other grade levels. Of note, CRI-CFC was published in its first edition in 2007.

In some cases, there are not enough data reported for interpreting the degree of reliability. For example, it is not clear just what variables the reliability coefficients (i.e., 0.80 for oral, elementary; 0.78 for silent, elementary; 0.83 for oral high school and adult; 0.79 for silent, high school and adult) reported by Bader (2005) apply to, such as word recognition, comprehension, or both.

Vocabulary

Meaning vocabulary

Although norm-referenced tests typically report scores for vocabulary knowledge both as a separate and combined reading score (Pearson, Hiebert, & Kamil, 2007), none of the IRIs reviewed include enough vocabulary items accompanying the text passages to make this feasible. For example, Burns and Roe (2007), Johns (2005), and Silvaroli and Wheelock (2004) treat vocabulary as an embedded construct contributing to reading comprehension; however, out of five to eight questions, only one or two items are vocabulary related.

Sight word vocabulary and word recognition strategies

While Cooter et al. (2007) treat vocabulary as a separate construct with its own set of test items and score in CRI-CFC, this section is more a measure of high-frequency or sight words recognized than meaning vocabulary knowledge. It should be noted that the word list components of the other IRIs reviewed also provide information related more to word recognition than to knowledge of word meanings.

Each of the other inventories takes a different approach to the assessment of sight word recognition, as well as general word identification strategies, by including a series of word lists administered at the beginning of the IRI assessment in order to gain insights on a student's word recognition strategies as well as to determine a reading passage starting point. Across inventories, although the specific sources for the word lists are not always identified (Bader, 2005; Burns & Roe, 2007; Silvaroli & Wheelock, 2004; Woods & Moe, 2007), two of the authors report some or all of the word list words are drawn from the reading passages (Applegate et al., 2008; Leslie & Caldwell, 2006) or various named, high-frequency word lists (e.g., Fry's Instant Words; Applegate et al., 2008; Johns, 2005).

With regard to CRI-2 and QRI-4, because some of the words were drawn from the reading passages, evaluators can compare word identification abilities in context versus out of context. These two inventories also allow for making distinctions between words recognized instantly (i.e., sight words) versus words that are decoded when readers are allowed more time.

BRLI (Bader, 2005) includes separate lists of "experiential" words (i.e., words commonly found in instructional materials and on tests), as well as lists of "adult thematic" words (i.e.,

office-related vocabulary, words related to health and safety, vehicle-related words), which could be useful with English-language learners and adult literacy students. Because students are asked to read each item but to explain the meanings only as needed, this assessment appears to provide more information related to sight word vocabulary and word recognition strategies than meaning vocabulary, similar to the other IRI word lists. Information about the development of these word lists, however, or pilot testing of items is lacking.

Phonemic awareness

Three of the IRI authors include phonemic awareness assessments (Bader, 2005; Cooter et al., 2007; Johns, 2005) in their manuals. It should be noted that these assessments are not integral parts of the inventories; instead, they are provided as supplements for optional use. Given the fact that there are other instruments available that are more developmental, systematic, and comprehensive for assessing phonemic awareness, these IRI assessment supplements are not recommended for evaluating children's knowledge in this area.

Phonics

As with phonemic awareness, an IRI is not intended to provide a thorough evaluation of a child's phonic knowledge. While the authors of CRI-CFC (Cooter et al., 2007) and BRLI (Bader, 2005) provide supplementary phonics assessments in their manuals, there are other more systematic and comprehensive assessments of this aspect of reading available. For this reason, these supplementary assessments are not recommended for evaluating this pillar of reading.

It should be noted that the miscue analysis and word list components (see the Vocabulary section) featured in most of the IRIs allow the evaluator to gain valuable insights on patterns related to students' word recognition abilities, including insights related to phonics. In addition, miscue analyses of passages read orally provide the advantage of allowing the tester to observe how a child actually applies phonics skills while reading familiar and unknown words in connected text. Because of this powerful function, the miscue analysis portion of an IRI should not be skipped or overlooked.

Fluency

With the exception of CRI-SW (Silvaroli & Wheelock, 2004), each of the IRIs includes some measure of fluency. All but Woods and Moe (2007) suggest, at a minimum, tracking the reading rate, and all but Applegate et al. (2008), who includes an oral reading rubric in the manual, provide norms or guidelines in their manuals for interpreting scores. In some of the IRIs, checklists are provided listing additional aspects of fluency to evaluate, such as pitch, stress, intonation, and use of punctuation, among other qualities observed, to check off as applicable. Woods and Moe also include a four-point fluency scoring guide. Given the relevance of fluent reading to reading comprehension (Allington, 1983), these measures provide valuable data for interpreting the results of an IRI assessment and are recommended.

Choosing an IRI

One of the purposes of this article is to cross-compare current IRIs with a goal of providing assistance in selecting one that best fits a teacher's needs. Although each IRI has its strengths and limitations, there are also unique characteristics to consider that may sway someone toward using one instrument or another.

For reading professionals who work with diverse populations and are looking for a diagnostic tool to assess the five critical components of reading instruction, the CRI-CFC, in Spanish and English (Cooter et al., 2007) for regular and special education students, as well as some sections of the BRLI (Bader, 2005), are attractive options. Most likely, those who work with middle and high school students will find the QRI-4 (Leslie & Caldwell, 2006) and ARI (Woods & Moe, 2007) passages and assessment options appealing. The CRI-2 (Applegate et al., 2008) would be a good fit for reading professionals concerned with thoughtful response and higher-level thinking.

In addition, the variety of passages and rubrics in BRI (Johns, 2005) and contrasting format options in CRI-SW (Silvaroli & Wheelock, 2004) would provide flexibility for those who work with diverse classrooms that are skills-based and have more of a literacy emphasis. For literature-based literacy programs, the IRI-BR (Burns & Roe, 2007) with its appendix of leveled literature selections is a valuable resource for matching students with appropriate book selections after students' reading levels are determined. In all cases, caution is advised for assessment components lacking technical rigor or for use of alternate forms without proven reliability.

Some of the IRIs had features worth noting because they made the complex manuals and various components easier to navigate and use. Some of these features include the fold-out tabs in CRI-SW (Silvaroli & Wheelock, 2004); indexes (Johns, 2005; Leslie & Caldwell, 2006), which most of the IRIs do not include; and inside-cover quick reference guides (Bader, 2005; Burns & Roe, 2007; Johns, 2005; Woods & Moe, 2007). Some handy resources located conveniently in appendixes include extra passages and rubrics, checklists, and scoring guides (Johns, 2005; Burns & Roe, 2007) and various summary forms (Cooter et al., 2007; Johns, 2005). As a feature of its newest edition, CRI-2 (Applegate et al., 2008) offers a variety of tools on its companion website, including access to an Automated Scoring Assistant software to help manage assessment data collected. It should be noted that the theoretical orientation of the evaluator and the technical features (e.g., validity and reliability) of the instruments are fundamental factors to consider in choosing an IRI.

For literacy-related professionals seeking ways to better address the instructional needs of children facing the greatest challenges in their journey to become successful readers, IRIs can serve as valuable diagnostic tools. Perhaps this summary of some key information will provide assistance to others in the selection of IRIs well suited to their particular educational settings and classroom contexts.

Nilsson, N.L. (2008). A Critical Analysis of Eight Informal Reading Inventories. The Reading Teacher, 61(7), pp. 526ï¿½536.