Developing an Observation Instrument to Support Authentic Independent Reading Time during School in a Data-driven World

Lunetta M. Williams, lmwillia@unf.edu
University of North Florida, Florida, USA

Katrina W. Hall, khall@unf.edu
University of North Florida, Florida, USA

Wanda B. Hedrick, whedrick@unf.edu
University of North Florida, Florida, USA

Marcia Lamkin, mlamkin52@gmail.com
University of North Florida, Florida, USA

Jennifer Abendroth, j.abendroth@gmail.com
University of North Florida, Florida, USA

Abstract

The purpose of the present study was to develop an instrument to measure reading during in-school independent reading (ISIR). Procedures to establish validity and reliability of the instrument included videotaping and observing students during ISIR, gathering feedback from literacy experts, establishing interrater reliability, crosschecking instruments, and minimizing observer bias. A factor analysis revealed items loaded onto four factors: attention to text, outward indicators, physical interactions with text, and spontaneous sharing of content. Further analysis indicated that involvement during ISIR was highly correlated with items associated with attention to text and the readers' responses. This article provides an overview of the development and initial field-testing of the instrument with recommendations for future use and development.

Key words: independent reading, involvement, observation instrument, silent reading

In the real world and with some constraints, readers can choose what they read, where they read, and the amount of time they read. According to Morrow (2003), opportunities to choose to read at home and school are necessary components of the "cultivation and practice of literacy" (p. 857). Accordingly, in-school independent reading (ISIR) time has the potential to foster students' reading habits. Krashen (2004, 2005) argued that existing research supports daily independent reading time in school and that students who spent more time reading in school perform as well or better on standardized tests of reading comprehension than other students, an assertion supported by a number of researchers (e.g. Allington, 1977; Coddington, 2009; Gambrell, 1978; Hunt, 1970; Ivey & Broaddus, 2001; Ivey & Fisher, 2006). Allington (2012) stated that the act of reading fosters the ability to read accurately, fluently, and with comprehension.

Additionally, Hedrick and Cunningham (2002) revealed that wide reading, or reading large quantities and with some variety, accomplished partially through ISIR, might contribute to growth in listening comprehension. In a landmark study, Anderson, Wilson, and Fielding (1988) found that students who did not read during their free time eventually lost academic ground. Although researchers such as Gambrell (2007) support out-of-school reading, many students, particularly those from socioeconomically disadvantaged backgrounds, do not have access to books, time, or a place to read without interruption (Krashen, 2005; Neuman, 1999; Neuman & Celano, 2001). As such, ISIR is a necessary component in the classroom to ensure that all students, including those from less advantaged backgrounds, have an opportunity to read widely and without interruption.

If one goal of ISIR is for students to choose to read when they are not required to read, then we advocate for the practice to be implemented in an "authentic" manner; thus, we use the term "authentic independent reading." By authentic, we mean that while the reading might occur during a specified time in a school day, the teacher has created an environment in which students move beyond "obedience" and read not simply because it is on the schedule, but because they look forward to the time to read. Classroom teachers can foster authentic ISIR by providing time in school to read and discuss books, access to interesting books, and the opportunity for students to choose their books (Allington, 2012; Deci & Ryan, 1985; Gambrell, Palmer, Codling, & Mazzoni, 1996; Neuman, 1999; Wigfield, Eccles, & Rodrigues, 1998). In this environment, choice is key, as interest in a book can significantly influence students' answers to higher order comprehension questions (Spaulding, 1992). If students are not interested in the book, if they are distracted, or if the book is too difficult, they may engage in avoidance tactics, distract others, or master the art of "fake reading," thus defeating the purpose of the time (Kelley & Clausen-Grace, 2009, p. 314).

Students who read a lot and are involved in their reading will likely have increased reading achievement scores on standardized tests. Additionally, students who have opportunities to read independently may read more widely, which is critical to acquiring an extensive vocabulary (Nagy, Anderson, & Herman, 1987; Stanovich, 1986; West, Stanovich, & Mitchell, 1993). Furthermore, reading extensively increases world knowledge and a reliance on reading for informed decision-making (Stanovich, 2000).

In considering the above research findings, one might wonder if the relationship between wide reading and reading achievement was correlational and not causal (Carver & Liebert, 1995, p. 29). Members of the National Reading Panel (NRP) did in fact wonder this (National Institute of Child Health and Development, 2000). The Panel analyzed experimental/quasi-experimental investigations focused on methods of improving children's reading and writing abilities. Using 14 studies that met the design criteria, the Panel concluded that people need to question the results of the impact of wide reading on reading achievement. The Panel's final report stated:

> Literally hundreds of correlational studies find that the best readers read the most and that poor readers read the least. These correlational studies suggest that the more children read, the better their fluency, vocabulary, and comprehension. However, these findings are correlational in nature, and correlation does not imply causation. (p. 12)

While the NRP did not entirely negate the effect of independent reading on reading achievement, the Panel indicated the need for more experimental/quasi-experimental studies.

One challenge in justifying independent reading during the school day has been quantifying whether children are actually reading during that time. With a reliable, valid instrument that is non-intrusive, researchers can conduct experiments/quasi-experiments using ISIR as an independent variable to determine whether time involved in ISIR can affect student achievement. While we discovered some instruments that measure students' behaviors during ISIR in the review of literature, most studies did not report information about establishing validity, and several instruments included broad terms to indicate behaviors that could present practical problems for researchers (Topping, Samuels, & Paul, 2007). In the literature review, we will discuss these studies in more depth. The purpose of this study was to begin developing an observational instrument with evidence of validity and reliability to measure children's behaviors that suggest children are reading independently. This article provides an overview of the development and initial field-testing of the instrument as well as recommendations for future use and development.

Conceptual Framework

Three theoretical constructs framed this study. We began with Smith's "conditions of learning" in which involvement and active comprehension occur simultaneously, coinciding with an event such as ISIR (Smith, 1975; 2004). Smith (2004) noted that learning took place when the learner interacted productively with a demonstration or a cluster of acts. For example, conditions of learning were present when the learner encountered an unfamiliar word in a book and reread to figure out the meaning.

Additionally, we focused on reader response because of its resemblance to real-world reading behavior. Rosenblatt's (1978) transactional theory of reader response suggests that readers adopt an "aesthetic" or "efferent" stance during a reading event. "In the aesthetic stance attention is focused primarily on experiencing what is evoked, lived through, during the reading" (Rosenblatt, 1978, p. 184). A reader who adopts an efferent stance focuses on acquiring information or solutions. Readers may move along the continuum of aesthetic and efferent during the same event, and one stance is not mutually exclusive of the other. When readers

immerse themselves in the text during independent reading time, whatever the genre, they are engaged in an aesthetic experience that may include efferent elements. However, Rosenblatt noted that when readers were required to show their comprehension through worksheets, tests, or questions, they adopted a primarily "efferent" stance, in that they were reading for the information needed to answer external queries. The aesthetic or pleasure of the event lessened.

Lastly, extending this idea is Csikszentmihalyi's (1978) theory of "flow" in which students read, for example, because it is satisfying, it allows them to lose sense of ego and time, and it provides a sense of unity with surroundings. When a reader is in the state of flow, nothing matters but the book, which is the ultimate goal of independent reading time.

Review of the Literature

In this section, we included a review of the literature on instruments used to measure students' involvement in reading during or after reading instruction. Specifically related to our research question, we then reviewed instruments from studies focused on measuring students' involvement during in-school independent reading time. We used the word "involvement" to describe the observable, verbal and nonverbal behaviors that Csikszentmihalyi (1978) talks about in his flow theory, to theorize what readers may exhibit during reading.

Measuring Students' Involvement During or After Reading Instruction

Researchers have measured students' involvement during general reading instruction time at school by observing specific reading behaviors and attaining at least an 81% agreement rate (Lutz, Guthrie, & Davis, 2006; Powell, McIntyre, & Rightmyer, 2006; Waason, Beare, & Wassan, 1990; Wigfield, Guthrie, Perencevich, Taboada, Klauda, McRae, & Barbosa, 2008). In Wigfield et al. (2008), evidence for construct and criterion-related validity was included for the Reading Engagement Index (REI) instrument. Using a four-point response scale, teachers assessed students on the REI by responding to items such as, "This student is often reading independently," "This student works hard in reading," and "This student thinks deeply about the content of texts." Internal consistency according to Cronbach's alpha was .92.

Several researchers have also conducted studies to measure students' reading involvement after the time that they read (Cipielewski & Stanovich, 1992; Kamil, 2008; McBride-Chang, Manis, Seidenberg, Custodio, & Doi, 1993; Taylor, Frye, & Maruyama, 1990). Many of these studies used instruments such as daily reading logs to measure students' involvement when reading, but researchers did not record specific behaviors. Furthermore, researchers did not report validity measures in these studies, and only one study, Cipielewski & Stanovich's (1992) Author Recognition Test (ART) and Title Recognition Test (TRT), included reliability measures. Cipielewski & Stanovich listed popular and fictitious authors and titles on the ART and TRT on checklists, respectively, while students checked items that were familiar. These instruments seem to correlate well with measures of reading involvement as avid readers were more likely to read more widely and recognize more authors and book titles than those who read less often or widely or who were not involved in their reading. The Cronbach's alpha reliability of the ART which contains 40 items and the TRT with 38 items were .70 and .77, respectively.

Measuring In-School Independent Reading Involvement

Several studies have measured behaviors that indicated involvement during ISIR. Trudel (2007) examined whether a structured independent reading program with teacher guidance for students' text selections and time for student reflection was more effective for 16 third and fourth grade students than a traditional silent reading program (SSR) in which students and the teacher read self-selected books. Teachers recorded student behaviors such as a student being on task when reading, responding, or sharing with a peer. While the study did not report interrater agreement, findings revealed that the majority of students were on task more often during the structured independent reading program than during SSR. Kelley and Clausen-Grace (2006; 2008) studied third graders during independent reading time and developed a checklist to measure silent reading behaviors to identify students who were not reading or actively interacting with the text. Again, the researchers did not report interrater agreement, but they did find that when the teacher was able to identify the students' interaction level, she was able to foster a higher level of engagement, which they defined as characteristics of those readers who "choose to read because they are interested in a text and they enjoy reading" (Kelley & Clausen-Grace, 2009, p.313).

The following studies that measured involvement during ISIR also reported reliability measures. Carver and Liebert (1995) examined the effects of reading books at different levels of difficulty on reading ability among 43 students. Teachers assigned half the students to read books slightly above their level, while the other half read books below their level, though the researchers later realized that the books were all relatively easy. The researchers observed and rated students during independent reading time using codes that included "not reading," "selecting a book," "chatting or talking," or "sleeping." Interrater agreement among three raters was 98 percent. The researchers concluded that reading ability did not significantly improve from reading easy books and that both groups were usually reading during independent reading time. Using a small sample ($n = 3$), Bryan, Fawson, and Reutzel's (2003) study examined the value of literature discussions on students' reading behaviors identified from a previous study (Waason et al., 1990), including the amount of time it took for the students to begin a task, missing instructional materials, and distracting student-caused sounds. Implementing literature discussions increased behaviors that suggested participants read, and reduced off-task behaviors during the independent reading time, which indicated a Pearson correlation coefficient of .98 between different raters' judgments.

In summary, while these studies measured students' behaviors during independent reading time, researchers did not report information about establishing validity, and some instruments included general terms to note behaviors that might not have indicated that students were actually reading text. When researchers directly observed students, it was difficult to know whether the student was "fake reading" (Kelley & Clausen-Grace, 2009). Based on our review, there were no portable, valid instruments focused on ISIR employed in experimental/quasi-experimental studies that the NRP requested. Researchers needed a more specific instrument to measure observable behaviors that indicated a student was actually reading. This study provides researchers with the development of such an instrument.

Method

The following section describes the method of the study, which included developing the instrument and then field testing the instrument in a variety of schools. The question that guided our study was: During in-school independent reading (ISIR), among a sample of third graders, what verbal and non-verbal behaviors suggested their involvement in reading? The goal was to develop an instrument for which we could begin to establish content validity and reliability and support authentic ISIR.

Development of the Instrument

When initially developing the instrument, we systematically observed students during ISIR to note behaviors suggesting involvement. Beginning with a blank sheet of paper and pencil, we recorded behaviors and conducted procedures to gain evidence for content validity and interrater reliability.

Setting and participants. The initial setting was an inclusive, third grade, co-taught classroom in an elementary school in a large Northeast Florida school district. The school's student population was ethnically diverse, according to the district's webpage (7% Asian, 16% Black, 7% Hispanic, 5% Mixed, 65% White), with 28% of the student population eligible for free or reduced-price lunch. Prior to this study, the school had received an "A" for five consecutive years, a grade assigned by the state department of education and based largely on students' performance on statewide-standardized tests.

We selected the school for the initial development largely based on our established relationship with the administration and teachers and because the school's administration and teacher support and flexibility were imperative to the success of the observation instrument's development stage. The principal recommended a classroom with highly qualified teachers who sufficiently managed independent reading time and a majority of students who were able to read independently during this time. We selected third grade because students had received reading instruction for a few years and could make appropriate book choices, which allowed for more successful participation in independent reading (King, 1967).

Twenty-seven of 28 students agreed to participate and returned signed consent forms from their caregivers. Ten were boys; seventeen were girls. The majority of the students were White ($n = 19$), but others were identified as Black or as having Arabic, Asian, or Hispanic origins. All students were reading at a minimum of a first-grade reading level.

The district scheduled ISIR within a common three-part reading workshop format. First, the teachers worked with the whole class on a mini-lesson. Next, during the work session, the students read independently, read with a partner, or conferenced with the teachers in small groups. The final part of the workshop, the closing, involved the teachers working with the whole group to share and reflect (Calkins, 2000). During our observations, the students participated simultaneously in 20 minutes of ISIR. The teachers provided students with access to books that varied in reading difficulty, genre, and content. Teachers allowed students to read

books from home or from the classroom or school library. For 20 minutes, all students read silently while the teachers occasionally walked around and redirected students whom were off task. After 20 minutes, students recorded the title of the book and the number of pages read.

Procedures. The following section outlines the sequence for the development of the instrument. The sequence included videotaping and observing students during ISIR, developing a coding system, gaining evidence for content validity, and establishing interrater reliability.

Videotaping and observations of ISIR. Our group of observers included three university faculty members holding doctoral degrees and one undergraduate honors candidate majoring in education. We opted to use a combination of real-time observation and videotaping in order to allow for later analysis of the video as we established interrater reliability. Because the teachers hosted the local university's college of education undergraduate pre-service teachers, the students were accustomed to adult visitors in the classroom. To lessen the cameras' impact on students' behavior, we set up two cameras in opposite corners of the classroom a week before the taping began. The teachers provided a list of students and a classroom map. In teams of two, we videotaped and observed students during their 20 minute ISIR over the course of five weeks. During our observations, we found we needed to allow a three-minute transition period for students to begin reading, which was consistent with the research on transitions covered in more depth in the discussion section (Doyle & Carter, 1984). We compared notes and reviewed the recordings together to develop a list of behaviors that suggested involvement in ISIR.

Coding system. The coding system for the rating scale evolved from the observations completed over time. Initially we developed twelve items, which we then judged on a scale ranging from 0-4, rating a student as exhibiting behaviors "none of the time" (0), "somewhat" (1), "moderately" (2), "a lot of the time" (3), and "extreme" (4). The final modified scale ranged from one to four with behaviors being scored as occurring "none" (1), "some" (2), "a lot" (3), and "all the time" (4).

After observing and coding two sessions, we developed a draft instrument to guide the remaining sessions. The teachers observed a 20-minute video segment with us, taking notes and rating students on their reading involvement. The teachers provided a holistic score of two students and a list of behaviors that they believed suggested involvement or lack of involvement. We used the teachers' observations to confirm or disconfirm items as we refined the emerging instrument. After our discussion with the teachers, we added "turning the page" and "density of text on a page" as issues to consider. For example, a student reading a novel might turn fewer pages than a student reading a nonfiction text with many illustrations. At this point, we developed a draft of the instrument to observe students' involvement in reading (Figure 1). The discussion section will address the behaviors described in Figure 1.

Figure 1. Draft Observation Instrument

Instrument Directions: Using the 4-point scale circle the number that best describes the item descriptor. Give the student credit if there is a reasonable doubt by giving the next higher number; do not give partial credit.

		none 1	some 2	a lot 3	all of the time 4
1.	Book position conducive to reading	1	2	3	4
2.	Student body position appears to be conducive to being involved in book	1	2	3	4
3.	Student's eyes focus on the page	1	2	3	4
4.	When student's eyes focus on the page they are moving in appropriate response to external text features	1	2	3	4
5.	Student uses finger or some item to stay on line	1	2	3	4
6.	Student intentionally moves back or forward page(s) in book, but returns to reading place	1	2	3	4
7.	Student uses subvocalizations (lips are moving as if reading to self)	1	2	3	4
8.	Student appears to respond "affectively" to book	1	2	3	4
9.	Student is distracted by external distractions	1	2	3	4
10.	Student's own movements cause him or her to stop looking/focusing on page	1	2	3	4
11.	Student becomes "socially" involved in another student's book	1	2	3	4
12.	Student attempts "socially" to involve another student in his/her book	1	2	3	4
	OVERALL INVOLVEMENT	1	2	3	4

Content validity. After developing the list of items, we shared the instrument with 12 literacy experts including professors from various universities, a district reading coach, and classroom teachers with advanced degrees in literacy. Based on their feedback, we reworded and clarified some language that we discuss in the findings section.

Interrater reliability. We determined that the video cameras needed to zoom in more closely in order for observers to distinguish facial expressions and eye movements as we isolated behaviors related to reading involvement. We selected four students (two boys identified as White, and two girls, one identified as White with a disability, and one identified as Black) to videotape at a separate table. Video cameras, positioned at opposite ends of the table, captured two students' behaviors closely; thus, eye movements were evident when reviewing the tapes. We chose these students because we agreed with their teachers that they represented a range of students involved during ISIR. We did frequent crosschecking as we observed. During this period, we met and viewed 33 randomly selected segments of 30 seconds of video. Two observers coded each segment independently. We performed an interrater reliability analysis for the instrument (SPSS Version 18) using agreement percentages among items and the Cohen's Kappa statistic to determine consistency among the raters.

Correlation of observation items. After using the observations from the 33 segments to establish interrater reliability, we examined the correlation of items on the observation instrument using Spearman *rho* due to the number of observation instances.

Field Testing

To ensure that the instrument worked across student populations, minimize observer bias, and begin building evidence of reliability, we increased the number of observations. We trained four additional observers to begin observing students in three other classrooms across three schools using 30-second video segments of ISIR time. We also obtained acceptable levels of interrater reliability.

Setting and participants. We recruited three third-grade teachers from three elementary schools located in the same district as the initial school where the draft instrument was created. The schools had racially and ethnically diverse student populations. According to the district's website, School One's student population was 24% Black, 15% Hispanic or Mixed-race, and 61% White, with 45% of the student population receiving free or reduced-price lunch. The school had received a grade of "A" largely based on state standardized tests scores for the previous academic year. In School Two, 27% of students identified as Black, 25% Hispanic, and 38% of the students identified as White; the remaining students identified as being of Mixed-race or of other origin. Seventy-nine percent of the school's population received free or reduced-price lunch. The school had received a grade of "B" from the state. The third school was comprised of 62% Black students, 16% of students identified as being Hispanic, Mixed-race, or of other origin, and 22% White. Ninety-two percent of students received free or reduced-price lunch. School Three had received a "B" grade.

The principals identified teachers using the same criteria as the initial school, and the teachers agreed to participate. As with the initial school, the teachers who agreed to participate regularly hosted the local university's college of education pre-service teachers, and the students were used to adult visitors in the classroom. All students in the three classrooms were invited to participate in the study (*n* = 54). We selected 15 students from those who returned permission paperwork, five from each classroom whom teachers identified as reading on grade level. To obtain as diverse a sample as possible, we considered gender, race, and students identified as having special learning needs. Two White females, one identified as having a learning disability, four Black females, three females of Hispanic origin, three White males, and three Black males, one diagnosed as autistic, were selected. During data collection, one Hispanic female and one Black male without a disability were absent, reducing the number to 13 students. Because a large number of students in the classrooms returned signed permission slips to participate, we were able to focus unobtrusively on the targeted children. The students did not know whom we were observing until we interviewed them.

Students had access to books through the classroom and school libraries, and teachers allowed them to bring books from home. Each teacher reminded the students to read books that they had not read before, so responses would relate to the day's reading. For 20 minutes, all students read silently. At the end of the sessions, the students recorded the titles of their books and the number of pages read.

Procedures. The following section outlines the sequence of procedures for the field-testing. The sequence included establishing evidence of instrument reliability, instrument crosschecking, and minimizing observer bias.

Establishing evidence of instrument reliability. We completed 121 total observations. At this point, we concurred that our instrument worked effectively to observe ISIR involvement, and we conducted a confirmatory factor analysis. The factor analysis is a widely used statistical technique designed to explore a data forest, particularly in studies involving instrument development (Costello & Osborne, 2005). In order to carefully define the scale characteristics of the observation instrument, we conducted a factor analysis with varimax rotation with data from the 121 observations.

Instrument crosschecking and minimizing observer bias. We observed ISIR for nine 20-minute sessions in each classroom over the course of two weeks, using the initial instrument. No videotaping occurred because we had established interrater reliability and did not need further observation segments to analyze.

We interviewed students after ISIR and confirmed responses by using their books as references. During the interviews, we asked two reader response-type questions based on a form created by Taberski (2000). The questions included: 1) Tell me a little about what you read today; and 2) Are you enjoying this book? Why or why not? Realizing that students might have reread books, we asked if the book was new for them. We chose this verbal interaction because it closely resembled a real-world situation of discussing books, and research has suggested that children's talk can indicate their literary understanding (Sipe, 2000). Research supported the efficacy of a conversational approach and personal response in evaluating students' understanding of text (Daniels, 1994; Eeds & Wells, 1989). We elected to use verbal interviews, since, developmentally, third graders' ability to self-report can be unreliable due to their age (Assor & Connell, 1992; Lutz et al., 2006). While students' interpersonal or verbal skills varied, the teachers included book talks in their instruction, and the third graders were familiar with the practice of responding to books verbally. As these students were already required to log their reading, adding a writing component, even a brief self-report, changed the nature of the reading event and moved it farther from the real-world reading experience that our instrument attempted to measure.

We made notes based on student responses and, using a four-point scale similar to one used by Lutz et al. (2006), coded 13 students' responses to the reader-response questions. Students received a score of 4 if they were extremely talkative about their books and provided thorough responses. They received a score of 3 if they were somewhat talkative and provided fairly detailed responses. Students received a score of 2 if they provided little information about their books; and a score of 1 if they had no response or if it was clear that they had been uninvolved. We then conducted a Spearman correlation analysis in order to examine the relationships, if any, between scores on the observation instrument and reader response questions.
To minimize observer bias, the three faculty researchers acted as raters and interviewers. As raters, we observed students during ISIR. Each researcher, in the role of rater, selected a student on whom to focus for 10 minutes. After ISIR, the rater who had not observed the student took on

the role of interviewer, asked the student the reader response questions, and recorded the students' responses. Two researchers independently scored the responses.

Findings

In this section, we discuss findings that include content validity, interrater reliability, correlation of observation items, establishing evidence of instrument validity, instrument crosschecking, and minimizing observer bias.

Content Validity

The experts agreed the indicators described involved behaviors exhibited during independent reading and identified no new behaviors to add or consider, thus establishing content validity. A few experts provided suggestions to clarify language used in the instrument. One expert recommended we change the term "Student uses subvocalizations" to "Student vocalizes or moves lips as if reading to self" to more accurately describe these events. We revised the language of some items, creating our second draft of the instrument (Figure 2).

Figure 2. Draft Observation Instrument with Expert Feedback

Instrument Directions: Using the 4-point scale circle the number that best describes the item descriptor. Give the student credit if there is a reasonable doubt by giving the next higher number; do not give partial credit.

		none 1	some 2	a lot 3	all of the time 4
1.	Book position is conducive to reading	1	2	3	4
2.	Student body position appears to be conducive to being involved in book	1	2	3	4
3.	Student's eyes are focused on the page	1	2	3	4
4.	When student's eyes are focused on the page they are moving in appropriate response to external text features	1	2	3	4
5.	Student uses finger or some item to stay on line	1	2	3	4
6.	Student intentionally moves back or forward page(s) in book, but returns to reading place	1	2	3	4
7.	Student vocalizes or moves lips as if reading to self	1	2	3	4
8.	Student appears to respond "affectively" to book (sighs, laughs, cries, frowns, smiles, etc.)	1	2	3	4
9.	Student is distracted by external distractions	1	2	3	4
10.	Student's own movements *cause* him or her to *stop* looking/focusing on the page (scratching, tapping, turning too many pages accidently, looking around)	1	2	3	4
11.	Student becomes "socially" involved in another student's book	1	2	3	4
12.	Student attempts "socially" to involve another student in his/her book	1	2	3	4
	OVERALL INVOLVEMENT	1	2	3	4

Interrater Reliability

Interrater reliability on the formative 33 segments was 93.9% exact agreement, with a Cohen's Kappa of 0.875. We found no indications of systematic observer inconsistencies, observer fatigue, or deterioration in observers' reliability across time.

Correlation of Observation Items

The results of the Spearman *rho* correlation are displayed in Table 1.

Table 1

Correlations of Observation Ratings from the Observational Instrument

	Spearman's *Rho* Values with Significance Noted								
	Book position	Student body position	Eyes focused	Eyes moving	Student vocalizes	Student affective	Student distracted	Student movements	Overall
(1) Book position		.946	.795	.727			-.486	-.532	.724
(2) Student body position			.792	.722			-.507	-.509	.698
(3) Eyes focused				.882			-.675	-.637	.872
(4) Eyes moving							-.564	-.592	.861
(7) Student vocalizes						.548			
(8) Student affective									
(9) Student distracted								.553	
(10) Student movements									-.597

Note. For all stated *rho*, $p \leq .01$ (2-tailed). $N = 33$ for all analyses.

Nine of the thirteen items correlated significantly to at least one other item in the nine items. The following items correlated significantly to at least one of the other items: 1 "Book position conducive to reading"; 2 "Student body position appears to be conducive to being involved in book"; 3 "Student's eyes are focused on the page"; 4 "When student's eyes are focused on the page, they are moving in appropriate response to external text features"; 7 "Student vocalizes or moves lips as if reading to self"; 8 "Student appears to respond affectively to book"; 9 "Student is distracted by external distractions"; 10 "Student's other movements cause him or her to stop looking/focusing on page," and 13 "Overall involvement". The following items did not correlate significantly to any of the other 13 items: 5 "Student uses finger or some item to stay on line"; 6 "Student intentionally moves back and forward page(s) in book, but returns to reading place"; and 12 "Student attempts 'socially' to involve another student in his/her book". Item 11 "Student

becomes 'socially' involved in another student's book," could not be determined because there was no variance in the observations.

Establishing Evidence of Instrument Reliability

The following section explains the results of the underlying Pearson correlation matrix from the factor analysis. As seen in Table 1, nine of the thirteen items correlated significantly in the independently executed Spearman analysis with at least one other item with a Spearman *rho* of $r_s>.3$. Four items did not correlate significantly with any other item. Based on the computation of correlations, we proceeded with the factor analysis. The statistical output of the factor analysis included a Pearson correlation matrix with eight additional significant correlations, seen in Table 2.

Table 2

Additional Correlations of Observation Ratings from the Observational Instrument

	Pearson's *rho* values with significance noted							
	Student finger	Student pages	Student vocalizes	Student affective	Student distracted	Student movement	Student involves	Overall
(4) Eyes moving	.212*							
(5) Student finger		.251**	.265**					
(6) Student pages								
(7) Student vocalizes				.241**				
(8) Student affective					-.211*			
(9) Student distracted								
(10) Student movement								
(11) Student social						.186*	.270**	.187*

Note. * = $p \le .05$, ** = $p \le .01$. N = 121 for all analyses.

Item 4, "When student's eyes are focused on the page, they are moving in appropriate response to external text features," significantly correlated with item 5 "Student uses finger or some item to stay on line" ($r_s=.212\ p < .05$). Item 5 significantly correlated with item 6 "Student intentionally moves back and forward page(s) in book, but returns to reading place" ($r_s=.251\ p < .01$) and item 7 "Student vocalizes or moves lips as if reading to self" ($r_s=.265, p< .01$). Item 7 significantly correlated with item 8 "Student appears to respond 'affectively' to book (sighs, laughs, cries, frowns, smiles, etc.)" ($r_s=.241\ p < .01$). Item 8 also significantly correlated with item 9 "Student is distracted by external distractions" ($r_s=-.211\ p < .05$).

Additionally, item 11 "Student becomes 'socially' involved in another's book" correlated significantly with three items. It significantly correlated with item 10 "Student's own movements *cause* him or her to *stop* looking/focusing on page (scratching, tapping, turning too many pages accidentally; looking around)" (r_s=.186 $p < .05$). Item 11 significantly correlated with item 12 "Student attempts 'socially' to involve another student in his/her book" (r_s=.270 $p < .01$). Lastly, item 11 significantly correlated with item 13 "Overall involvement" (r_s=.187 $p < .05$).

In the factor analysis, the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.78, above the recommended value of .6, and Bartlett's test of sphericity was significant (χ^2 (78) = 709.98, $p < .01$). We interpreted all resulting factors with eigenvalues greater than 1.0. The initial eigenvalues showed that the first factor explained 35% of the variance; the second factor 12% of the variance; a third factor 9% of the variance; and the fourth factor 8%. These factors cumulatively explained 64% of the variance. Table 3 presents the rotated factor matrix for these four factors.

Table 3

Rotated Component Matrix

	Component			
	1	2	3	4
Overall	.904			
Item 3	.895			
Item 2	.845			
Item 4	.831			
Item 1	.808			
Item 9	-.699			
Item 10	-.465		.350	
Item 7		.777		
Item 8		.719		
Item 6			.833	
Item 5		.383	.601	
Item 12				.816
Item 11			.257	.757

Note. Extraction method = Principal component analysis. Rotation method = Varimax with Kaiser Normalization. Rotation converged in five iterations.

The communalities for each variable, that is, the proportion of variance that each item has in common with the other items, were all above .3, further confirming that each item shared some common variance with other items (see Table 4). The proportion of variance unique to each item is the respective item's total variance minus the communality.

Table 4

Factor Loadings and Communalities Based on a Principal Components Analysis with Orthogonal Varimax Rotation for 13 Items from the Observational Instrument

Items from Observational Instrument	Attention to text	Outward indicators	Physical interaction with text	Spontaneous sharing of text	Communality
Book position is conducive to reading.	0.81				0.69
Student body position appears to be conducive to being involved in book.	0.85				0.74
Student's eyes are focused on the page.	0.90				0.79
When student's eyes are focused on page, they are moving in appropriate response to external text features.	0.83				0.72
Student uses finger or some item to stay on line.		0.38	0.60		0.53
Student intentionally moves back or forward page(s) in book, but returns to reading place.			0.83		0.69
Student vocalizes or moves lips as if reading to self.		0.78			0.63
Student appears to respond "affectively" to book (sighs, laughs, cries, frowns, smiles, etc.).		0.72			0.53
Student is distracted by external distractions.	-0.70				0.50
Student's own movements *cause* him or her to *stop* looking/focusing on page (scratching, tapping, turning too many pages accidently, looking around).	-0.47		0.35		0.40
Student becomes "socially" involved in another student's book.				0.78	0.65
Student attempts "socially" to involve another student in his/her book.				0.82	0.70
OVERALL INVOLVEMENT	0.90				0.82

Note. Factor loadings < .2 are suppressed. $N = 121$ for all analyses.

We identified four factors: attention to text (Factor 1), outward indicators (Factor 2), physical interaction with text (Factor 3), and spontaneous sharing of text (Factor 4). All items excluding "Student's own movements cause him or her to stop looking/focusing on page" loaded to the respective factors at a level of 0.6 and higher, with ten items loading to a factor at 0.7 and higher. Hair, Anderson, Tatham, and Black (1998), consider 0.6 to be a very high factor. The following section discusses the items located in each factor.

Attention to text: Factor 1. Seven items in our instrument loaded onto a central factor "Attention to text," or actions typically associated with book reading. "Overall involvement" had a factor loading of 0.90 onto Factor 1. While "Overall involvement" is a holistic score and not considered an observable item, it does capture the observer's impressions and was used in the factor analysis along with the twelve observable items.

Two items loaded negatively onto Factor 1. "Student is distracted by external distractions" loaded at -0.70, and "Student's own movements cause him or her to stop looking/focusing on page" loaded at -0.47. We noted that this was the weakest item included in Factor 1, but it fell above the cut-off score of 0.4 for a central factor as suggested by Raubenheimer (2004). This item also loaded positively onto "Physical interaction with text" (Factor 3) at 0.35, but did not fall above the cut-off score of 0.4. We determined that, theoretically, this item was more indicative of attention to, or distraction from, the text.

Outward indicators: Factor 2. Two items loaded onto Factor 2, interpreted as "Outward indicators." These behaviors were facial responses that suggested involvement. "Student vocalizes or moves lips as if reading to self" loaded at 0.78. "Student appears to respond 'affectively' to a book" loaded at 0.72.

Physical interactions with text: Factor 3. Two items related to physical interactions with text that would not typically be necessary to read a book loaded onto Factor 3. "Student uses finger or some item to stay on line" loaded onto Factor 3 at 0.60, and "Student intentionally moves back or forward pages in book, but returns to reading place" loaded at 0.83.

Spontaneous sharing of content: Factor 4. Two items loaded onto Factor 4, interpreted as spontaneous sharing of content. These behaviors appeared to stem from a student's excitement about a book or passage and his or her desire to share with someone. "Student becomes 'socially' involved in another's book" loaded at 0.78, and "Student attempts 'socially' to involve another student in his/her book" loaded at 0.82. We determined that behaviors such as students' attempting to talk to peers during ISIR might indicate involvement because students were trying to discuss their books. We then proceeded with the reordered and final instrument (see Figure 3).

Instrument Cross-checking and Minimizing Observer Bias

Raters' scores on Question 1, "Tell me a little about what you read today", correlated at $r=.958$ and on Question 2, "Are you enjoying this book? Why or why not?", at $r=.967$. The raters' judgment of students' overall involvement based on their responses to the two questions correlated at $r=1.0$. We correlated the observation item ratings with the reader response scores (see Table 5).

Figure 3. Final Observation Instrument

Name _____ Date _____ Time _____ Rater Initials _____

Instrument Directions: Using the 4-point scale circle the number that best describes the item descriptor. Give the student credit if there is a reasonable doubt by giving the next higher number; do not give partial credit.

		1 none	2 some	3 a lot	4 all of the time
	ATTENTION TO TEXT				
1.	Book position is conducive to reading[1]				
2.	Student body position appears to be conducive to being involved in book[2]				
3.	Student's eyes are focused on the page				
4.	When student's eyes are focused on the page they are moving in appropriate response to external text features[3]				
5.	Student is distracted by external distractions[4]				
6.	Student's own movements *cause* him or her to *stop* looking/focusing on the page (scratching, tapping, turning too many pages accidently, looking around)[5]				
	OUTWARD INDICATORS				
7.	Student vocalizes or moves lips as if reading to self				
8.	Student appears to respond "affectively" to book (sighs, laughs, cries, frowns, smiles, etc.)				
	PHYSICAL INTERACTION WITH TEXT				
9.	Student uses finger or some item to stay on line				
10.	Student intentionally moves back or forward page(s) in book, but returns to reading place				
	SPONTANEOUS SHARING OF CONTENT				
11.	Student becomes "socially" involved in another student's book				
12.	Student attempts "socially" to involve another student in his/her book				
	OVERALL INVOLVEMENT				

1

---

[1] Book is open in a way that appears to be easy to read/see

[2] Distance between book and student is optimal for reading

[3] Student turns page to go forward in book (and it is appropriate)

When reading narrative text appropriately, student's eyes travel left to right/top to bottom and examine pictures on page

When student reads non-narrative text appropriately, eyes move around page, but are focused on illustrations/headings, etc.

Student turns pages forward or backward in book and examines or views them intently, but continues to look at book

[4] Easily distracted by everyday classroom noises and remains uninvolved Not involved in book and briefly returns to book when noticed by teacher

[5] Scratching, tapping, turning too many pages accidentally

Non-involved, but appears to be due to illness, etc.

Looks around

Looks forward in book to see how many pages to the end of the chapter or book (not reading pages, but counting)

Table 5

Statistically Significant Correlations among 13 Items on Observational Instrument and Reader Response Interview Question Scores by Two Independent Reviewers for 13 Instances

	RRQ1Rater1	RRQ1Rater2	RRQ2Rater1	RRQ2Rater2	OverallRater1	OverallRater2
Item 1	.682*	.670*	.711**	.727**	.785**	.785**
Item 2	.682*	.670*	.711**	.727**	.785**	.785**
Item 3	.635*	.583*	.607*	.578*	.678*	.678*
Item 4	.774**	.735*	.755**	.746**	.780**	.780**
Item 5						
Item 6						
Item 7						
Item 8						
Item 9	-.675*	-.629*	-.566*		-.669*	-.669*
Item 10	-.770**	-.742**	-.756**	-.758**	-.805**	-.805**
Item 11						
Item 12						
Overall	.774**	.735**	.755**	.746**	.780**	.780**

Note. ** = $p \leq 0.01$ (2-tailed), * = $p \leq 0.05$ (2-tailed).

We interpreted the correlations between observational instrument items and reader response scores according to factor. Instrument items that loaded to "Attention to text", Factor 1 (Items 1, 2, 3, 4, 9, 10, and Overall), were highly correlated with the reader response scores. The correlations were significant at $p<.001$. They ranged from $r=-.566$ to $r=-.805$ on the Overall reader response scores. Items in Factors 2 (Items 7 and 8), 3 (Items 5 and 6), and 4 (Items 11 and 12) did not correlate significantly with reader response scores. Students' overall scores on the reader response questions related to overall involvement on the observation instrument at $r=.780$.

Discussion

The study's findings suggest that involvement in independent reading can be observed and measured. More specifically, overall involvement loaded most powerfully onto items associated with "Attention to text" (Factor 1). Additionally, the readers' responses gathered in brief interviews correlated strongly with Factor 1. Based on the findings, behaviors that might be taken for granted, such as the reader's book and body position, contributed to the measurement of overall involvement. Many readers demonstrated some variation of these behaviors, similar to Lutz et al.'s (2006) finding that eye movement and posture toward the speaker indicated being on-task.

As expected, it was important to see students' eye movements to discern overlapping behaviors that indicated involvement in reading. Students almost always focused on a page at some point during the interval, but they might have been "fake reading" (Kelly & Clausen-Grace, 2006). In order to be reading, their eyes needed to be moving appropriately. This finding connects with research on reading eye movements, or saccades, which reveals that when reading English, the reader's eyes move from left to right and top to bottom, but do not move "relentlessly forward." Instead, they regress and pause, with poor readers backtracking more than good readers do (Rayner, 1998; Samuels, Rasinski, & Hiebert, 2011).

We found that students' eyes moved in an appropriate manner as determined by the reading material's text structure. In narrative books, this generally meant the student's eyes moved left to right and top to bottom, pausing to look at pictures. Narrative books with atypical formats such as Pilkey's (1997) *The Adventures of Captain Underpants* and Kinney's (2007) *Diary of a Wimpy Kid* series caused readers to flip forward or backward. Similarly, in non-narrative books, the students' eyes might have moved around the page, stopping to focus on illustrations, graphs, or headings. The student might purposefully turn to different pages in the book, such as the glossary. In narrative books with atypical formats or informational books, students' actions such as fanning the pages, turning the book upside down, or flipping to the glossary in the back might be interpreted as not being involved. In these instances, however, the movements were indications of involvement, which reinforced the idea that observers should be familiar with the nature and layout of a student's book in order to accurately label involvement levels.

As we developed the instrument, we noted that students' own movements, such as scratching, or external factors, such as a loud noise, caused distractions. The items developed were consistent with those from other studies (Bryan et al., 2003; Kelley & Clausen-Grace, 2009; Powell et al., 2006; Wasson et al., 1990).

Student transition times into ISIR were consistent with Doyle and Carter's (1984) findings on the nature of academic work, in that the difficulty or unfamiliarity of a task affected the speed with which students started. In this study, students were familiar with teacher expectations for ISIR, so waiting three minutes before using the instrument was sufficient to capture behaviors indicative of their overall involvement.

While the behaviors in "Attention to text" (Factor 1) were necessary to be involved with a book, the behaviors in "Outward indicators" (Factor 2) and "Physical interaction with text" (Factor 3) might have been present but not necessary to be involved. For example, many children, strong readers in particular, did not use their fingers to stay on line while reading, a behavior included in Factor 2. Although high rating on some items suggested involvement, low ratings on them did not necessarily signal low involvement.

Some students exhibited behaviors associated with "Outward indicators" (Factor 2) which were "student vocalizes or moves lips as if reading to self" and "student appears to respond 'affectively' to the book." We noticed that typically the less proficient readers exhibited vocalizing behaviors, perhaps because they were trying to stay focused or struggling with the words. Most instances of responding affectively seemed to be when children were reading books with humor, such as books from the *Captain Underpants* series. Some students also exhibited behaviors associated with "Physical interaction with text" (Factor 3) which included "using a finger to stay on line" and "intentionally moving back or forward in the book and returning to the reading place." Most students who exhibited using a finger to stay on line were less proficient readers. The type of book seemed to be important when students exhibited the behavior of flipping back and forth in the book. In this study, nonfiction books evoked this behavior more often. The instances observed in Factors 2 and 3 were low, but we retained the items on the instrument in the event that they might be behaviors evident in a larger population.

Behaviors that fell under "Spontaneous sharing of content" (Factor 4) were consistent with Guthrie's (2004) assertion that social interaction was often connected with involved reading; while reading can be a solitary activity, involved readers might discuss books with others. We initially thought students' attempts to talk during ISIR suggested lack of involvement, but later realized that in some instances the students were trying to share their books. Although we did not often observe spontaneous sharing of content, we attributed this to the classroom contexts; the teachers we observed did not include opportunities for discussion about books during or after ISIR.

The "Overall engagement" item rating is an acknowledgement of the concept that effective teachers have a "with-it-ness" and intuitively know when their students are engaged and conversely, when they are not (Kounin, 1977). The first six descriptors were necessary to determine overall engagement. For example, the reader's body position being conducive to reading was important, but was different for each reader. Also, it was important that the book was open and within the reader's view. For instance, if the student laid his or her head on the desk and could barely see the page, it was likely that the student was not engaged. Items five and six, which included if the student was distracted by external distractions and/or his or her own movements, and if so, how often, seemed most revealing. It was easiest to score students who never read the book (1/none) and students who never looked up from the book (4/all of the time). The challenge was differentiating between 2/some and a 3/a lot. In general, we rated a student who had up to four short distractions in the time-span as 3/a lot and those who had five or more distractions or a couple of very long distractions as 2/some.

The instrument took 10 minutes to administer, was unobtrusive, and allowed the researchers to confirm or disconfirm students' involvement during ISIR. Further, using data from the instrument could be a step toward developing a gold standard to ensure that students are reading during ISIR. The results might be used to diagnose individual students who are able to read independently at any grade level. If students display behaviors suggesting that they are not independently reading, researchers can suggest ways to foster "flow" with text. For example, after noting that several students exhibit outward behaviors, opportunities could be provided for students to share or talk about the books they are reading with classmates (Almasi, 1996). Students might need a quiet atmosphere or might listen to music with headphones while they read. If a student is not involved during ISIR, then teachers may provide instruction on book choices. When students are involved in ISIR, then teachers can work to not interrupt their "flow" with questions or tasks that can result in a less pleasurable experience (Csikszentmihalyi, 1978; Smith, 2004).

Limitations

As with all classroom research, there were limitations. The presence of video cameras and researchers might have distracted the students. To address this potential disturbance, we placed the cameras in the classroom several days before the study began. Although adults were often in the classroom because of the university's college of education field placements, we visited the classrooms frequently so that students became accustomed to our presence. The classroom teachers in this study regularly implemented ISIR, making the findings limited to similar classroom contexts.

Another limitation was that the study was small in scale. We expected the twelve items to load onto no more than two factors and speculated that a larger number of factors might be attributed to low instances of the behavior or context. For example, since teachers did not allow students to talk to each other during ISIR, there were very few instances of children sharing with each other. Even though some factors had only two items measuring reading involvement, we observed the items during this study, and a panel of experts considered them valid. We suspect that with a more diverse setting, the factors would appear more often, and by using this instrument, researchers would have a place to capture these behaviors. Although we included a diverse group of third grade students, more research is needed using a larger sample to provide further empirical support for the validity of the instrument and generalization of the findings. In this study, teachers instructed students to write about their reading experience after ISIR. Including an instrument item focused on exhibiting reading involvement through writing (such as underling, jotting down notes, etc.) was beyond the scope of this study. However, if researchers observed writing behaviors in the future, perhaps they could be noted under the factor, "Physical interaction with text."

Conclusion

Students who are involved during ISIR may enjoy their reading, read more, and read widely with teacher assistance, resulting in increased achievement. Cunningham & Stanovich (2003) argued, "It is well accepted that comprehension ability and reading volume are in a reciprocal relationship" (p. 670). While measuring reading ability has been paramount in recent years, independent reading time has unfortunately taken a backseat to activities that are more easily quantifiable. We began developing this instrument in an attempt to respond to the National Reading Panel's request for more experimental studies focused on the effect of independent reading on reading achievement, and to quantify behaviors during ISIR that researchers can measure and use in these studies.

In the future, replicating this study in different settings and with participants from various grade and reading levels would help to increase the instrument's generalizability. Perhaps more research on this instrument's use during ISIR in content area classes where nonfiction texts are predominantly used and when students are instructed to write as they read for recall or graphic organizer purposes would be beneficial. Our initial results suggest that researchers may use this instrument to quantify behaviors that indicate involved reading. While the instrument has evidence of validity and reliability, it is also portable and includes specific, observable behaviors of involved readers.

References

Allington, R. (1977). If they don't read much, how they ever gonna get good? *Journal of Reading, 21*(1), 57-61.

Allington, R. (2012). *What really matters for struggling readers: Designing research based programs* (3rd ed.). Columbus, OH: Allyn & Bacon.

Almasi, J. F. (1996). A new view of discussion. In L.B. Gambrell & J. F. Almasi (Eds.). *Lively discussions! Fostering engaged reading* (pp. 2-24). Newark, DE: International Reading Association.

Anderson, R., Wilson, P., & Fielding, L. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23*(3), 285-303. doi: 10.1598/RRQ.23.3.2

Assor, A., & Connell, J. P. (1992). The validity of students' self-reports as measures of performance-affecting self-appraisals. In D. H. Schunk & J.Meece (Eds.), *Student perceptions in the classroom* (pp. 25-46). Mahwah, NJ: Lawrence Erlbaum. doi: 10.1177/0163278702250093

Bryan, G., Fawson, P., & Reutzel, D. R. (2003). Sustained silent reading: Exploring the value of literature discussion with three non-engaged readers. *Reading Research and Instruction, 43*(1), 47-73. doi:10.1080/1938807030955840

Calkins, L. M. (2000). *The art of teaching reading.* New York: Allyn and Bacon.

Carver, R., & Leibert, R. (1995). The effect of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly, 30*(1), 26-48. doi: 10.2307/747743

Cipielewski, J., & Stanovich, K. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology, 54*(1), 74-89. doi: 10.1016/0022-0965(92)90018-2

Coddington, C. S. (2009). *The effects of constructs of motivation that affirm and undermine reading achievement inside and outside of school on middle school students' reading achievement.* (Doctoral dissertation). Retrieved from https://login.dax.lib.unf.edu/login?url=http://search.proquest.com.dax.lib.unf.edu/docview/304921537?accountid=14690.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation, 10*(7). Retrieved from http://pareonline.net/getvn.asp?v=10&n=7.

Csikszentmihalyi, M. (1978). Intrinsic rewards and emergent motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 205-216). Mahwah, NJ: Lawrence Erlbaum.

Cunningham, A. E., & Stanovich, K. E. (2003). Reading matters: How reading engagement influences cognition. In J. Flood, D. Lapp, J. Squire, & J. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 666-675). Mahwah, NJ: Lawrence Erlbaum.

Daniels, H. (1994). *Literature circles: Voice and choice in the student-centered classroom.* Portland, ME: Stenhouse.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Doyle, W., & Carter, K. (1984). Academic tasks in the classroom. *Curriculum Inquiry, 14*(2), 129-149.

Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English, 23*(1)*,* 4-29.

Gambrell, L. B. (1978). Getting started with sustained silent reading and keeping it going. *The Reading Teacher, 32*(2), 328-331.

Gambrell, L. B. (2007). Reading: Does practice make perfect? *Reading Today, 24*(6), 16.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher, 49*(7), 518–533.

Guthrie, J. T. (2004). Teaching for literacy engagement. *Journal of Literacy Research, 36*(1), 1-30. doi: 10.1207/s15548430jlr3601_2

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis with readings* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hedrick, W. B., & Cunningham, J. W. (2002). Investigating the wide effect of wide reading on listening comprehension of written language. *Reading Psychology, 23*(2), 107-126. doi: 10.1080/027027102760351025

Hunt, L. C. (1970). The effect of self-selection, interest, and motivation upon independent, instructional, and frustrational levels. *The Reading Teacher, 24*(2), 146-151, 158.

Ivey, G., & Broaddus, K. (2001). "Just plain reading": A survey of what makes children want to read in middle school classrooms. *Reading Research Quarterly, 36*(4), 350-377. doi:10.1598/RRQ.36.42

Ivey, G., & Fisher, D. (2006). *Creating literacy-rich schools for adolescents.* Alexandria, VA: Association of Supervision and Curriculum Development.

Kamil, M. (2008). How to get recreational reading to increase reading ability. In Y. Kim, V. J. Risko, D. L. Compton, D. K. Dickinson, M. K. Hundley, R. T. Jimenez, K. M. Leander, & D. Well Rowe (Eds.), *57th Yearbook of the National Reading Conference* (pp. 31-40). Oak Creek, WI: National Reading Conference.

Kelley, M., & Clausen-Grace, N. (2006). R[5]: The Sustained Silent Reading makeover that transformed readers. *The Reading Teacher, 60*(2), 148-156. doi: 10.1598/RT.60.2.5

Kelley, M., & Clausen-Grace, N. (2008). *R[5] in your classroom: A guide to differentiating reading and developing avid readers*. Newark, DE: International Reading Association.

Kelley, M., & Clausen-Grace, N. (2009). Facilitating engagement by differentiating independent reading. *The Reading Teacher, 63*(4), 313-318. doi: 10.1598/RT.63.4.6.

Kinney, J. (2007). *Diary of a wimpy kid.* New York: Harry Abrams, Inc.

King, E. (1967). Critical appraisal of research on children's reading interests, preferences, and habits. *Canadian Research and Educational Digest, 7,* 312.

Kounin, J. (1977). *Discipline and group management in classrooms*. NY: Holt, Rinehart and Winston.

Krashen, S. (2004). False claims about literacy development. *Educational Leadership, 61*(6), 18-21.

Krashen, S. (2005). Is in-school free reading good for children? Why the National Reading Panel Report is (still) wrong. *Phi Delta Kappan, 86*(6), 444-447.

Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *Journal of Educational Research, 100*(1), 3-20. doi: 10.3200/JOER.100.1.3-20

McBride-Chang, C., Manis, F., Seidenberg, M., Custodio, R., & Doi, L. (1993). Print exposure as a predictor of word reading and reading comprehension in disabled and nondisabled readers. *Journal of Educational Psychology, 85*(2), 230-238. doi: 10.1037/0022-663.85.2.230

Morrow, L. (2003). Motivating lifelong voluntary readers. In D. Lapp & J. Flood (Eds.), *The handbook of language arts* (pp. 857-867). Mahwah, N.J.: Lawrence Erlbaum.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237-270. doi: 10.2307/1162893

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nichd.nih.gov/publications/nrp/smallbook.cfm.

Neuman, S. (1999). Books make a difference: A study of access to literacy. *Reading Research Quarterly, 34*(3), 286-301. doi: 10.1598/RRQ.34.3.3

Neuman, S., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly, 36*(1), 8-26. doi:10.2307/747906

Pilkey, D. (1997). *The adventures of Captain Underpants*. New York: Scholastic.

Powell, R., McIntyre, E., & Rightmyer, E. (2006). Johnny won't read, and Susie won't either: Reading instruction and student resistance. *Journal of Early Childhood Literacy, 6*(1), 5-31. doi: 10.1177/1468798406062172

Raubenheimer, J. E. (2004). An item selection procedure to maximize scale reliability and validity. *South African Journal of Industrial Psychology, 30*(4), 59-64.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Rosenblatt, L. (1978). *The reader, the text, the poem: The transactional theory of the literary work.* Carbondale: Southern Illinois University Press. doi:10.1006/ceps.1999.1009

Samuels, S. J., Rasinski, T. V., & Hiebert, E. H. (2011). Eye movements and reading: What teachers need to know. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (4th ed., pp. 25-50). Newark, DE: International Reading Association.

Sipe, L. R. (2000). The construction of literary understanding by first and second graders in oral response to picture storybook read-alouds. *Reading Research Quarterly, 35*(2), 252-275. doi:10.1598/RRQ.35.2.4

Smith, F. (1975). *Comprehension and learning*. New York, NY: Holt, Rinehart and Winston.

Smith, F. (2004). *Understanding reading: A psycholinguistic analysis of reading and learning* (6th ed.). Mahwah, NJ: Lawrence Erlbaum.

Spaulding, C. L. (1992). The motivation to read and write. In J.W. Irwin & M.A. Doyle (Eds.), *Reading/writing connections: Learning from research* (pp. 177-201). Newark, DE: International Reading Association.

Stanovich, K. (1986). Matthew effects on reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407. doi: 10.1598/RRQ.21.4.1

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York, NY: Guilford.

Taberski, S. (2000). *On solid ground: Strategies for teaching reading K-3.* Portsmouth, NH: Heinemann.

Taylor, B. M., Frye, B. J., & Maruyama, G. M. (1990). Time spent reading and reading growth. *American Educational Research Journal, 27*(2), 351-362. doi: 10.23071163013

Topping, K. J., Samuels, J., & Paul, T. (2007). Does practice make perfect? Independent reading quantity, quality and student achievement. *Learning and Instruction, 17*(3), 253-264. doi: 10.1016/j.learninstruc.2007.02.002

Trudel, H. (2007). Making data-driven decisions: Silent reading. *The Reading Teacher, 61*(4), 308-315. doi: 10.1598/RT.61.4.3

Waason, B., Beare, P., & Wassan, J. (1990). Classroom behavior of good and poor readers. *Journal of Educational Research, 83*(3), 162-165.

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly, 28*(4), 34-50. doi: 10.2307/747815

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. *Review of Research in Education: Vol. 23.* (pp.73-118). Washington, DC: American Educational Research Association.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating the effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools, 45*(5), 432-445. doi: 10.1002/pits.20307