Reading at Students’ Frustrational, Instructional, and Independent Levels: Effects on Comprehension and Time On-Task

Megan Ann Treptow
University of Minnesota

July, 2006

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Reading at Students’ Frustrational, Instructional, and Independent Levels:
Effects on Comprehension and Time On-Task

Megan Ann Treptow

University of Minnesota

Abstract

A single-subject design was used to examine the effects of reading activities at three levels of challenge: the frustrational, instructional, and independent levels, on reading comprehension and task engagement. The three levels were operationalized according to Gickling & Havertape’s (1981) ratios of known to unknown words in text, with 93-97% known words constituting the instructional level, greater than 97% known the independent level, and less than 93% known the frustrational level. Third grade students were selected based on screenings for both off-task behavior and reading difficulties. It was hypothesized that students would continue to exhibit high levels of off-task behavior at the frustrational level, but demonstrate increased time on-task when given materials at their instructional levels. Comprehension was hypothesized to be very low at the frustrational level and higher at the instructional level. At the independent level, comprehension was expected to remain high while the students’ time on-task was expected to drop to lower levels as the students became less engaged in tasks. Moderate to large effect sizes supported these hypotheses for two out of three students selected. Overall results were consistent with Gickling & Armstrong’s (1978) original demonstration, although smaller in magnitude of effects.
Educators are challenged to address the needs of students with a wide range of learning rates, skill levels, and background experiences within the regular education classroom. The *No Child Left Behind Act* (2001) increased this pressure by setting grade level standards at which each child must perform, stressing that all children must learn to read by the third grade. At the same time, the numbers of students labeled as learning disabled (LD) have more than doubled in the last 20 years, with over 3 million children currently receiving special education services as LD and numbers continuously increasing (National Center for Educational Statistics, 2004). Of these students, an estimated 80% are labeled as LD primarily because of reading difficulties (President’s Commission on Excellence in Special Education, 2002). Similarly, the numbers of students receiving special education in the category of other health impaired (OHI) have more than tripled in the last 10 years, with much of this increase owed to the increase in students labeled with attention deficit/hyperactivity disorder (ADHD; President’s Commission on Excellence in Special Education). Indeed, the most frequent concerns for which students are referred for special education evaluations include reading problems, low task completion rates, and classroom conduct problems (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002).

The use of ineffective instructional practices could lead some students to fall behind in the early grades. These low-achieving and struggling students have been referred to as “curriculum casualties” (Gickling & Thompson, 1985, p. 208) who fail to keep pace with the demands of grade level instruction. Children who struggle in reading in the early grades will often continue to fall further and further behind throughout their school years (Stanovich, 1986), and a high correlation exists between reading difficulties and classroom behavior and attention

The recent *Individuals with Disabilities Education Improvement Act* (H.R. 1350, 2004) includes provisions for the prevention and remediation of learning problems through a response-to-intervention (RTI) approach, in which scientific, research-based interventions are employed, and students’ learning in response to these interventions is monitored. The hope of many is that the RTI approach will reduce the numbers of students requiring special education services (Vaughn & Fuchs, 2003; President’s Commission on Excellence in Special Education, 2002). However, such an approach requires a strong research base of effective instructional practices and educational interventions. Shapiro (2004) outlined a process for problem-solving student difficulties that emphasized the appropriateness of the student’s instructional placement within the curriculum. This study endeavors to test a method of matching instruction in reading to individual students’ abilities, so as to provide optimal learning and task engagement for struggling students in the hopes of preventing increased classroom difficulties.

**Academic Learning Time**

Perhaps the most frequent struggle faced by classroom teachers is that of maintaining students’ engagement in classroom learning tasks. Researchers have estimated engagement rates among students to range from as low as 50% to 90%, depending on classroom and student factors (Hollowood, Salisbury, Rainforth, & Palombaro, 1995). Often referred to as academic learning time (ALT), the amount of time students spend actively, successfully, and productively engaged in learning has been consistently demonstrated to be a strong determinant of academic achievement (Denham & Lieberman, 1980; Fisher & Berliner, 1985; Gettinger & Seibert, 2002). Greenwood, Terry, Marquis, and Walker (1994) demonstrated the important role of student
engagement not only as a construct affecting school outcomes, but also as a construct affected by instruction. Thus, ALT is an important predictor of academic achievement that can be altered along with instructional variables.

Several factors have been identified as contributing to ALT, including allocated time, instructional time, engagement rate, and the rate of academic success and productivity (Gettinger & Siebert, 2002). According to Gettinger and Siebert, allocated time refers to the amount of time teachers allocate for instruction, instructional time is the proportion of that time that is actually used for instruction, and the engagement rate is the proportion of instructional time during which students are engaged in learning. Moreover, a distinction is made between two forms of engagement: procedural engagement and substantive engagement. Procedural engagement, often termed time on-task, includes directly observable behaviors (e.g., looking at the instructional stimuli or completing written work), while substantive engagement relies on the quality of the instructional tasks and the level of students’ investments in them (Gettinger & Siebert, 2002). The relevance of the tasks to the student’s learning and the student’s ability to comprehend the tasks are thus critical factors (Gickling & Thompson, 1985).

For learning tasks to be meaningful, they must be well-matched to the student’s interests, ability levels, background knowledge, and instructional needs. When this is true, the fourth variable contributing to ALT, the rate of academic success and productivity, is likely to be higher than when it is not (Gettinger & Siebert, 2002). The rate of academic responding, in addition to an appropriate instructional match, was identified by Christenson and Ysseldyke (1989) as an important instructional variable with strong links to achievement. Research found that students gained the most from their learning time when materials were at an appropriately matched difficulty level and they experienced a high rate of successful responding (Gettinger &
Seibert, 2002). This was evidenced by Wyne and Stuck (1979), who found that increasing task-oriented behavior and academic responding rates resulted in significant gains in reading achievement among struggling students over a period of 24 weeks, in comparison to matched students not receiving the intervention.

**Instructional Level**

Several theories about the appropriate instructional difficulty level of curricular materials have been proposed. Vygotsky theorized that children become more absorbed in learning and learn more effectively when focused on tasks that are within their own “zone of proximal development”, or the set of skills between those the child can do independently and those they can do with some assistance (1935/1978). According to Vygotsky (1934/1986),

…the only good kind of instruction is that which marches ahead of development and leads it; it must be aimed not so much at the ripe as at the ripening functions. It remains necessary to determine the lowest threshold at which instruction in, say, arithmetic may begin, since a certain minimal ripeness of functions is required. But we must consider the upper threshold as well; instruction must be oriented toward the future, not the past. (p. 188)

Others have further discussed the level of instruction according to the degree of instructional match or mismatch between the skills of the student and the learning demands of the instructional task. Betts (1946) initially defined the instructional level as a task that is sufficiently familiar yet still provides some degree of challenge to bring about optimal learning for the student. When the student experiences too little challenge or too great a challenge, the task is said to be at an independent level or a frustrational level, respectively. These levels of task challenge have been operationalized according to the percentage of known versus unknown
items in the learning material (Gickling & Havertape, 1981). Based on Betts’ (1946) observations that students’ comprehension of reading material began to falter when the number of unknown words exceeded 5%, and that when the percentage of known words was 98% or better students had no difficulty reading, Gickling and Havertape (1981) operationally defined the instructional level for reading material as containing 93-97% known words, with a 3-7% margin of challenge. Independent level was defined as containing 98-100% known words, and frustrational level as containing fewer than 93% known words.

Gickling and Armstrong (1978) tested these definitions of instructional difficulty levels, obtaining remarkable results on the on-task behavior, task completion, and task comprehension of first and second grade students identified as struggling readers. The researchers manipulated the ratio of known to challenging elements in reading assignments to create frustrational, instructional, and independent levels. As expected, students demonstrated the highest comprehension at the independent level, consistently close to 100%. They demonstrated consistently low levels of comprehension, not exceeding 20 percent, during frustrational reading material conditions, and during instructional reading material conditions, comprehension ranged from 87-94%. The most striking results were on students’ time on-task, with rates much higher during instructional level tasks than during either frustrational or independent level tasks. When curriculum materials were maintained at the ratios of instructional level, the students maintained rates of on-task behavior ranging from 84-94%, compared to 40-52% during baseline, around 45% when materials were at frustrational level, and around 53% when materials were at the independent level. Task completion rates were also affected, with rates during both the instructional and independent level conditions staying high, at around 96-98%, compared to around 45% for the frustrational condition and a range of 52-75% during baseline.
Classroom Behavior Problems

Students with high-incidence disability labels, including learning disabilities, ADHD, and emotional-behavioral disorders, experience elevated rates of inattentiveness and off-task classroom behavior compared to non-labeled peers (Hintze & Matthews, 2004). A recent survey of teachers indicated that approximately 16% of elementary school children display frequent off-task behavior along with inattention and/or poor concentration, ranking these as among the most common difficulties experienced in the classroom (Wolraich, Hannah, Baumgaertel, & Feurer, 1998). Psychoeducational and clinical data have suggested that learning difficulties and attention problems are interrelated, and often coexist (Mayes et al., 2000). Mayes and colleagues found that a learning disability was present in 70% of clinically referred children with ADHD, and 82% of children with learning disabilities also had ADHD.

Reading difficulties, in particular, may be closely linked to behavioral difficulties in children. Poor readers were found to spend more time off-task in the classroom and volunteer answers less often than skilled readers (Wasson, Beare, & Wasson, 1990). Reading problems have been associated with higher rates of inattentiveness and disruptive behavior, poorer school attendance, and increased risks of juvenile offending among adolescents (Maughan et al., 1996). Reading ability has been found to predict teacher-reported behavior problems in school-aged children above and beyond intelligence test scores, even after controlling for family variables and pre-school problem behavior (Stanton et al., 1990).

Meta-analytic research found that academic intervention strategies worked as well as contingency management approaches and better than cognitive-behavioral modification strategies in reducing ADHD-related problem behaviors (DuPaul & Eckert, 1997). Thus, it may be just as effective to target rate of academic responding, a behavior that is incompatible with
disruptive activity, as it is to focus on reducing ADHD behaviors directly when the goal is behavior change. A previous study found that students with severe behavior disorders exhibited less disruptive behavior and spent more time on-task when completing math assignments that they could complete with 90% or greater accuracy than when completing math assignments that they could complete with only 75% accuracy or less (DePaepe, Shores, Jack, & Denny, 1996). Other research found that maintaining curricular materials at students’ instructional levels resulted in higher levels of on-task behavior among learning disabled children than a token economy (Gickling & Havertape, 1981). Among students with ADHD, presenting curricular materials at instructional level (controlled instruction) without medication resulted in rates of on-task behavior commensurate with those exhibited while on medication, and task completion and task comprehension rates were higher when using controlled instruction alone than when using medication alone (Thompson, Gickling, & Havertape, 1983).

Functional analysis and functional assessment are frequently used methods of examining the maintaining features of problem behaviors. One possible function of students’ higher rates of off-task behavior during tasks at the frustrational level is to receive negative reinforcement in the form of escape from aversive task demands. McComas, Hoch, Paone, and El-Roy (2000) discussed the role of antecedent variables in establishing operations that set the stage for the reinforcing effect to occur. For example, the reinforcing value of escape may be greater during tasks that cannot be completed with a high rate of accuracy or success than during tasks that can be completed successfully. Thus, negatively reinforced off-task behavior may be more likely to occur during the former condition than during the latter. Similarly, when tasks are too easy and students become bored, the reinforcing value of off-task behavior may increase, making it more likely to occur. In either case, a change in instructional antecedents (e.g., the difficulty level of
material) could potentially diminish the reinforcing effects of escape to a point at which escape-maintained undesirable behavior is not displayed. Using establishing operations analyses, McComas et al. (2000) demonstrated the effectiveness of changing antecedent instructional variables on reducing the occurrence of escape behavior among students with developmental disabilities. Thus, the evidence suggests that, for many students, the rate of task success serves as an antecedent to appropriate classroom conduct, rather than the reverse relationship (Gickling & Thompson, 1985).

At the same time that negative reinforcement may maintain off-task behavior for many students when tasks are either too difficult to be completed successfully or too easy to be of interest, reinforcement may also maintain on-task behavior and task completion when assignments are within the appropriate range of challenge. Skinner and colleagues (Cates & Skinner, 2000; Skinner, Fletcher, Wildmon, & Belfiore, 1996; McCurdy, Skinner, Grantham, Watson & Hindman, 2001) demonstrated that interspersing known, or easier, items within math assignments increased student preference for the assignment. McCurdy et al. (2001) posited that, when students are working on assignments that require them to complete many discrete tasks, the completion of each task within the assignment may serve as a reinforcing event, thus reinforcing on-task behavior. Meta-analytic research of studies that involved interspersing easier, or known, items within mathematics assignments supported the discrete task completion hypothesis (Skinner, 2002). Therefore, increasing the rate of successful task completion within assignments could also increase the rate of reinforcement and thus the probability of students choosing to engage in assigned work.
Reading Comprehension and Academic Achievement

Behavioral difficulties have been consistently linked to academic difficulties, and academic difficulties can result from a mismatch between student skill and the instructional material (Daly, Martens, Kilmer, & Massie, 1996; Daly, Witt, Martens, & Dool, 1997). Stanovich (1986) described Matthew effects in reading, or “rich-get-richer” effects, in which the very children who are reading well read more, develop larger vocabularies and broader knowledge bases, and hence read even better. This reciprocal relationship was also proposed to effect poor readers, who read slowly and with less enjoyment, and therefore read less and have subsequent reduced growth in reading ability (Stanovich, 1986). It was suggested that the Matthew effect results when skilled readers read more and thereby acquire an expanded knowledge base, which facilitates the induction of new word meanings from text, thus expanding students’ vocabularies and further increasing their reading fluency. However, skilled readers appeared to learn new words from context with a greater efficiency than less skilled readers even when differences in knowledge base were controlled (Stanovich, 1986). Therefore, it was proposed that reading fluency influences vocabulary acquisition more directly, having a concurrent influence on reading comprehension, and therefore affecting the overall learning acquired from text. Thus, instructional adaptations based on matching task difficulty with student skills led to improved student learning (Daly & Martens, 1994; Daly et al., 1996).

The ability to derive the meanings of unknown words from unfamiliar passages has demonstrated a correlation of .65 with reading comprehension (Sternberg, 1985). Reading fluency has also been consistently linked to comprehension (Fuchs, Fuchs, Hosp & Jenkins, 2001; McCormick & Samuels, 1979; Rasinski, 1990), with a minimum reading fluency rate of approximately 50 words per minute found to be necessary for adequate comprehension of
reading material among third- and fourth-grade students (Burns, Tucker, Hauser, Thelen, Holmes, & White, 2002). It follows that, if the difficulty level (percentage of known words) of reading material influences a student’s reading fluency rate (words read accurately per minute), then it may in turn influence both the student’s ability to learn meanings of unknown words from context and their overall comprehension of the reading material.

Gickling and Armstrong (1978) demonstrated the effect of instructional difficulty of reading material on students’ comprehension by manipulating the ratio of known to challenging elements within reading assignments. Other researchers have found improved progress in students’ reading fluency and comprehension by presenting material at the instructional level through pre-teaching unknown words from the text to achieve the ratio of 93-97% known words (Burns, 2002; Burns, 2004a; Burns, Dean, & Foley, 2004; Shapiro, 1992). In one study that facilitated an instructional level in reading over 15 weeks by pre-teaching unknown words to third grade students identified as LD, the students in the pre-teaching condition made significantly greater progress within the reading curriculum than students serving as controls (Burns, 2004a). Moreover, the correlation between the number of passages read at instructional level and students’ progress within the curriculum was .80 and progress rates of 65.5% of students in the treatment condition exceeded previously established trajectories, compared to 10% of students in the control group (Burns, 2004a). Using the same strategy with third and fourth grade students identified as LD in one session yielded effect sizes of .38 and 1.76 for reading fluency and comprehension, respectively (Burns, Dean, & Foley, 2004). Thus, elevating reading materials to instructional level by pre-teaching unknown words was consistently successful with regular education students, students with various disabilities (mental retardation,
learning disabilities, and behavior disorders), and in various subject areas (Browder & Shear, 1996; Burns, 2002; Gickling & Havertape, 1981; McCurdy et al., 2001).

**Curriculum-Based Assessment**

Curriculum-based assessment (CBA; Gickling & Havertape, 1981) is a method of assessing student skill level in order to design appropriately challenging instruction to move the student forward relative to his or her present abilities within the curriculum. The goal of CBA is to design instructional practices that connect the prior knowledge and skills of the student to the task, thereby ensuring an optimal instructional match (Gravois & Gickling, 2002). This is especially relevant when the typical curricular materials may present too great a challenge for the student’s present skills (Gravois & Gickling, 2002). CBA involves examining the immediate learning task and determining the ratio of known to challenging material based on skills the student can perform accurately and fluently (Gickling & Havertape, 1981). Next, materials can be designed to be at the student’s appropriate instructional level and instruction can focus on specific skills that need to be acquired (Gravois & Gickling, 2002). In reading, this means determining the percentage of words the student reads correctly within the reading material, along with the student’s reading fluency, comprehension, and background knowledge, and using this information to match the reading material to the student’s abilities (Burns, 2004b). Emphasis is placed on ensuring that the student is reading materials at their instructional level, or materials in which 93-97% of the words are known, and a 3-7% margin of challenge is maintained.

Assessing the instructional level for reading with CBA has been shown to be sufficiently reliable for instructional decision-making (Burns, Tucker, Frame, Foley, & Hauser, 2000). Additionally, data collected with CBA have demonstrated adequate validity through content
(Burns, 2004b) and criterion-related (Burns & Mosack, in press) analyses. As previously discussed, modifying instruction with CBA data led to improved student outcomes (Burns, 2002; Burns, 2004a; Shapiro, 1992; Shapiro & Ager, 1992), suggesting instructional validity for CBA (Burns, 2004b) as defined by Messick (1995).

Some confusion has existed in the literature and practice regarding CBA’s similarity to other approaches to assessment based in curriculum materials (Burns, MacQuarrie, & Campbell, 1999). At around the same time that CBA emerged as a model for collecting assessment information relevant for curricular placement and instructional delivery, the term curriculum-based measurement (CBM) emerged as a method for monitoring students’ progress and evaluating the effectiveness of instruction (Deno, 1985). Since then, the two terms have often been used interchangeably, despite important conceptual differences between the two approaches (Burns et al., 1999). While both CBA and CBM use materials either directly from or closely matching the curriculum materials used in everyday instruction, the forms and purposes of the data differ (Burns et al., 1999). The data collected in CBA focus on the accuracy of student responses and on the ratio of known to challenging items within the material, with the goal of designing effective instruction for the student. In contrast, the data collected in CBM focus on fluency and rate of correct responding, with the goal of monitoring students’ growth in reading (Burns, 2004b).

As students become more skilled in academic tasks, their fluency on those tasks will increase (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good & Jefferson, 1998; McCormick & Samuels, 1979). Therefore, data provided by CBM serve to monitor students’ progress in academic skills and evaluate the effectiveness of the instruction provided (Burns et al., 1999). Both relatively quick to administer, CBA and CBM can be used together in a system of
intervention decision-making and progress monitoring, with CBA used to design the instruction and CBM used to monitor its ongoing effectiveness (Burns, 2002; Shapiro, 2004).

Some have suggested that CBM can be used for instructional placement decisions by using fluency rates of 40-60 words per minute to represent the instructional level in first and second grades and 70-100 words per minute in third through sixth grades (Fuchs & Deno, 1982). In other words, reading material that children in second grade could read at a rate of 50 words per minute would represent an instructional level. Although modifications based on fluency rates have increased student learning (Daly & Martens, 1994; Daly et al., 1996; VanDerHeyden & Burns, in press) and fluency measures consistently demonstrate adequate psychometric properties (Marston, 1989), the reliability and validity of fluency-based instructional decisions remain unknown. In other words, a second-grade student’s fluency rate of 50 words per minute may be stable data, but whether or not that represents an instructional level is as of yet unknown.

Since its introduction, CBA has received widespread support in the empirical literature as a reliable and valid method for the assessment of academic skills problems (Burns, 2004b; Shapiro, 2004; Shapiro, Angello, & Eckert, 2004). There seems to be a high degree of treatment utility for CBA, as evidence suggests that curriculum manipulations based on CBA data have resulted in academic and behavior improvements (Burns, 2002; Gickling & Rosenfield, 1995; Shapiro, 1992; Shapiro, 2004; Shapiro & Ager, 1992). In addition to its usefulness for making instructional or program modifications, CBA has demonstrated particular usefulness in designing and monitoring the effectiveness of interventions in an RTI model of special education eligibility determination (Burns, Dean, & Klar, 2004). However, despite high ratings by teachers and school psychologists regarding CBA’s usefulness and acceptability as an assessment approach, CBA has yet to be as widely implemented and tested as it has been discussed (Shapiro et al.,
2004; Shapiro, 2004). In a recent national survey of schools psychologists, 46% indicated that they do not use CBA in their practice (Shapiro, et al., 2004). Despite its current inclusion in over 90% of school psychology graduate training programs, the most common reason cited for not using CBA was insufficient training in the technique (Shapiro et al., 2004). The second most common reason, cited by 17% of nonusers, was the view of CBA as too time consuming (Shapiro et al., 2004), despite that the administration of CBA takes less time than the administration of most norm-referenced tests.

More research is needed on the implementation of CBA in classroom settings for adapting the instruction of struggling students, to examine both potential barriers and potential benefits to its use. Additionally, few studies have tested the ratios of known to unknown items in curricular material suggested by Gickling and Havertape (1981) to constitute the frustrational, instructional, and independent levels. This study will further test those ratios by replicating Gickling and Armstrong’s (1978) original demonstration on the effects of these instructional ratios on students’ learning and engagement, using CBA to design instructional materials.

**Purpose of the Present Study**

The purpose of the present study was to examine the effects of reading activities at three levels of challenge: the instructional, independent, and frustrational levels, on students’ reading comprehension and task engagement. Students were selected who exhibited high levels of off-task classroom behavior and whose reading skills placed them at risk for school failure. As a replication of Gickling & Armstrong (1978), the suggested ratios of known to unknown words in text were implemented, using CBA to determine each student’s frustrational, instructional, and independent levels within a reading series. It was hypothesized that those students who exhibited the highest levels of off-task classroom behavior may have frequently experienced
frustration due to the levels of challenge posed by classroom tasks. If given materials at frustrational level, they would thus continue to exhibit high rates of off-task behavior as in typical classroom tasks, but when given materials at their instructional levels, these students were expected to engage in the material and exhibit much higher rates of time on-task. Students’ comprehension of reading materials was also expected to improve, from low at the frustrational level to quite high at the instructional level. At the independent level, while comprehension was expected to remain high, the students’ time on-task was predicted to drop to lower rates as the students became less engaged in the tasks. Thus, the research hypotheses that focused the study were a) students typically exhibiting high rates of off-task classroom behavior and reading difficulties will exhibit the highest rates of on-task behavior while participating in reading material containing 93% to 97% known words, as compared to passages with more than 97% known and less than 93% known words, and b) these students will exhibit the highest reading comprehension when reading passages containing 93% to 97% known words or more than 97% known words, as compared to passages with less than 93% known words. Implications for the students’ learning and for instructional practices will be discussed.

Method

Selection of Participants

Three students from an inclusive third grade classroom containing a total of 23 students were selected for participation based on screenings for both off-task behavior during reading instruction and reading difficulties. Momentary time sampling procedures were used with 10-second intervals to record the time on-task of each student in the classroom on two different days during independent work in reading. Each observation was approximately 10 minutes in length, for a total of two 10-minute observations of each student. The results of these observations were
then combined with curriculum-based measurement (CBM) data collected by the school to select those students who exhibited both higher levels of off-task behavior during reading instruction and lower CBM scores than the averages for the class.

Six students met these criteria and were then screened further using the DIBELS (Dynamic Indicators of Basic Early Literacy Skills; Good & Kaminski, 2002) measure of oral reading fluency and one additional 10-minute observation. The student with the lowest observed percentage of time on-task was excluded from selection as a participant due to concerns about the type and severity of his disability and its impact on his behaviors. Two other students exhibiting low levels of time on-task during observations declined to participate in the study. Three remaining students whose overall levels of time on-task and reading fluency were the lowest were selected for participation. These students included one African American girl and two Caucasian American boys. One of the boys was diagnosed with a learning disability in reading and received special education services, and the other two students were not diagnosed with any disability for which they received special education services.

**Materials**

Reading passages were selected from the reading series, *Read Naturally* (RN; Read Naturally, 2003). RN is a reading program that includes a collection of short reading passages at sequenced levels of reading difficulty, designed to allow for instruction in and monitoring of students’ progress in developing reading fluency and comprehension. The program includes 24 passages at each of 13 reading levels, ranging from grade 1 to grade 8, and an additional set of passages in the phonics series that begin at the .8 (pre-primer) level. The passages ranged in length from 50 words at the .8 level to 200 words at level 5. Each passage was followed by five comprehension questions, four multiple choice and one open-ended.
Independent Variable

Reading material difficulty served as the independent variable and had three levels: frustrational reading level, instructional reading level, and independent reading level. Frustrational reading level was defined as 80-92% known words, instructional level as 93-97% known words, and independent level as 98-100% known words (Gickling & Armstrong, 1978). Curriculum-based assessments (CBA; Gickling & Havertape, 1981) were conducted individually with each of the three selected students prior to the beginning the manipulations. Each student was asked to read aloud for 1 minute from two randomly selected passages at his or her grade level and two at levels above and/or below his or her grade level. The examiner recorded the words as known if they were read rapidly and correctly, and unknown if they were read incorrectly (e.g., went for want), skipped, or if the student hesitated for more than three seconds on the word (Gickling & Havertape, 1981). The percentage of known words for each passage was calculated to determine which level, frustrational, instructional, or independent, the passages represented for the student. This process was then repeated by using easier and/or more difficult passages until passages at the student’s frustrational, instructional, and independent levels were identified.

Design and Procedures

A single-subject, multi-element design was used in which the three treatment phases were presented three times each in randomized order for each student. Observational data from the selection phase, collected during typical classroom seatwork tasks in reading, were used as baseline data on time on-task for each student. The three treatment phases took place over a total of nine sessions in whole-class setting. Students were asked to read silently two randomly selected passages from the RN series and complete the subsequent comprehension questions in
writing. Passages that were already read by a student were not repeated, but other passages at the appropriately matched levels were randomly selected from within the RN series.

**Dependent Variables**

**Time on-task.** Each student’s time on-task was observed and recorded using time-sampling procedures with 10-second intervals for the duration of the time taken to complete the task. On-task behavior was defined as actively attending to the assigned instructional material (Shapiro, 2004). Examples include looking at the reading material, writing, listening to instructions from the teacher, and raising a hand for assistance from the teacher (Shapiro, 2004). Off-task behavior was defined as the student not “having his head and/or eyes oriented toward assigned material” (Skinner, Rhymer, & McDaniel, 2000, pg. 23). Examples of off-task behavior include talking about anything other than the assigned reading, leaving the seat for non-relevant reasons, aimless movement of the reading passages (e.g., flipping back and forth), gazing away from the reading passages, reading something other than the assigned passages, and focusing attention to the activities of others (Shapiro, 2004).

**Reading Comprehension.** Comprehension of each passage was scored according to the number of comprehension questions answered correctly out of 5. Answers were judged as correct or incorrect based on the scoring key included with the *Read Naturally* series (Read Naturally, 2003). The total comprehension score for each session was recorded as the total out of 10 for the two passages administered.

**Manipulation checks**

To check that the stories administered were indeed at each students’ frustrational, instructional, and independent levels, each student was asked to read aloud for one minute to the
examiner one randomly selected passage at each level following completion of the task. The examiner recorded the percentage of known words obtained.

**Observation Integrity**

**Training of observers.** All observations were conducted by graduate students in a school psychology training program, trained in observation techniques. All but one of the observers were blind to the task difficulty conditions received by students at each session.

**Interobserver/Interrater Agreements.** Agreements with an independent observer were collected for 1/3 of the observations of the selected students at each phase, and 25% of the screening observations prior to the selection of students. Inter-observer agreements were calculated by computing the ratio of agreements to the total number of recording intervals (Steege, Brown-Chidsey, & Mace, 2002).

Interobserver agreements (IOA) of on-task behavior during the selection phase ranged from 90% to 100% agreement, and averaged 94.3%. IOA during the three treatment phases ranged from 96% to 100%, with a mean of 99%. When disagreements occurred in the total observed percentage of time on-task, the average of the two observers was used.

Agreements of 100% were obtained with an independent rater for the percentage of known words read by students during manipulation check procedures. Interrater agreements of 100% were obtained for students’ scores on comprehension questions.

**Threats to Internal Validity**

Like any research methodology, single-subject designs have their limitations in the inferences that can be drawn from them. However, single-subject designs have increasingly been viewed as a useful and valid means of examining the effects of treatments on individuals’ behaviors (Stoiber & Kratochwill, 2000). Kratochwill (1992) provided an overview of the
proliferation of writings on single-subject research designs. Many authors who have written extensively on research designs agree that, when certain conditions are met, including the repetition of treatment phase alterations, repeated measurements within phases, and controlling other factors, consistent changes in individuals’ behaviors between phases can reasonably be attributed to the change in treatment (Kazdin, 1982; Kratochwill, 1985; Kratochwill, Mott, & Dodson, 1984; Shaughnessy, Zechmeister, & Zechmeister, 2003). While the focus in single-subject designs is on determining if behavior change in a given individual is brought about by a treatment change, and not on finding changes in groups of individuals, similarly consistent treatment effects seen across several individuals can strengthen our confidence in making causal inferences regarding the effects of that treatment (Kratochwill, 1985).

A major requirement of single-subject designs discussed by Kazdin (1982) is that phases be alternated so that performance improves at some points and reverts toward baseline levels at other points. The design of the present study addressed this requirement by repeating the treatment phases of frustrational, instructional, and independent reading levels, under the hypothesis that students’ measured behaviors would revert back to previous levels as the phases were repeated. While initial data on students’ time on-task were collected prior to the treatment phases, the reading level of instructional materials during these observations was uncontrolled. For this reason, the controlled frustrational reading level phases served as a type of baseline condition. An additional design requirement discussed by Kazdin (1982) is that performance be observed and measured both continuously over the periods while the treatments are in effect and on several occasions. This requirement was met in the present design by observing students’ time on-task for the duration of each treatment condition, assessing comprehension with 5 questions approximately every 100-200 words, within the range recommended by Fuchs, Fuchs,
and Maxwell (1988), and by measuring students’ behavior on 3 different days at each phase level. Furthermore, the present design attempted to employ objective data collection procedures rather than subjective descriptions of behaviors or analyses of responses, an additional criteria discussed by Kazdin (1982) and Kratochwill (1985).

**Results**

**Classroom Screening**

The mean percentage of time on-task for all students in the classroom during reading instruction observations was 79.52%, with a standard deviation of 16.13% and a range from 30% of intervals on-task to 97% of intervals on-task. The majority of students (13 out of 23) were on-task 80% of the time or more during observations.

Results for the three students selected for participation are presented below. All effect sizes were calculated according to the no assumptions approach (Busk & Serlin, 1992).

**Jeremy**

Jeremy’s mean percentage of time on-task during baseline observations was 54.96%, with a standard deviation of 19.73% of intervals across the three observations. Screening with the DIBELS measure of oral reading fluency placed him in the at risk range according to the third grade end of the year benchmark, with a score of 75 words per minute (Good & Kaminski, 2002).

Jeremy’s percentages of time on-task and task comprehension for each session are displayed in figure 1. His mean percentage of time on-task during frustrational level tasks was 89.59% (SD = 7.23%). His mean percentage of time on-task was highest during instructional level tasks at 94.96% (SD = 2.98%). During independent level tasks, his mean percentage of time on-task was slightly lower than for frustrational level tasks at 87.68% (SD = 7.98%). These
data produced effect sizes of .74 for time on-task at the instructional level compared to the frustrational level and .26 for the independent level compared to the frustrational level. The effect size for time on-task at the instructional level compared to the independent level was .91. At the instructional level, 33% of the data points were non-overlapping with those at the frustrational level. Between the instructional level and the independent level, 33% were non-overlapping.

Jeremy’s mean percentage of task comprehension was 40% (SD = 20%) during frustrational level tasks, 50% during instructional level tasks (SD = 10%), and 86.67% during independent level tasks (SD = 5.77%). These data produced effect sizes of 1.0 for comprehension at the instructional level compared to the frustrational level and 2.33 for comprehension at the independent level compared to the frustrational level. Thirty-three percent (33%) of the data points at the instructional level were non-overlapping with those at the frustrational level, and 100% of the data points at the independent level were non-overlapping with data points at either the frustrational or instructional levels.

Results of manipulation checks supported the determination that the passages used to produce the frustrational, instructional, and independent levels for Jeremy indeed reflected the appropriate ratios of known to unknown words in text. The percentages of known words obtained were within the expected ratios for 100% of the passages selected.

Jessica

Jessica’s mean percentage of time on-task during baseline observations was 78.14%, with a standard deviation of 10.67% of intervals across the three observations. Screening with the DIBELS measure of oral reading fluency placed her in the at risk range according to the third
grade end of the year benchmark, with a score of 69.64 words per minute (Good & Kaminski, 2002).

Jessica’s percentages of time on-task and task comprehension for each session are displayed in figure 2. Her mean percentage of time on-task during frustrational level tasks was 93.09% (SD = .31%). Her mean percentage of time on-task during instructional level tasks was 96.31% (SD = 3.22%). During independent level tasks, her mean percentage of time on-task was 92.58% (SD = 7.43%). These data produced effect sizes of 10.39 for time on-task at the instructional level compared to the frustrational level, and 1.65 for time on-task at the independent level compared to the frustrational level. The effect size for time on-task at the instructional level compared to the independent level was .50. One hundred percent (100%) of the data points at the instructional level were non-overlapping with those at the frustrational level. Although the mean percentage of time on-task at the frustrational level was higher than at the independent level, 67% of the data points at the independent level were non-overlapping in the positive direction with the data points at the frustrational level. Between the instructional and the independent levels, 33% of the data points were non-overlapping.

Jessica’s mean percentage of task comprehension at the frustrational level was 46.67% (SD = 11.55%). At the instructional level, her mean percentage of task comprehension was 70% (SD = 17.32%). And at the independent level, her mean percentage of task comprehension was 76.67% (SD = 11.55%). These data produced effect sizes of 2.02 for comprehension at the instructional level compared to the frustrational level and 2.60 for comprehension at the independent level compared to the frustrational level. At the instructional level, 66% of the data points were non-overlapping with those at the frustrational level. At the independent level,
100% of the data points were non-overlapping with those at the frustrational level, and 33% were non-overlapping with those at the instructional level.

Results of manipulation checks again supported the determination that the passages used to produce the frustrational, instructional, and independent levels for Jessica indeed reflected the appropriate ratios of known to unknown words in text. The percentages of known words obtained were within the expected ratios for 100% of the passages selected.

**Benjamin**

Benjamin’s mean percentage of time on-task during baseline observations was 68.31%, with a standard deviation of 26.51% of intervals across the three observations. Screening with the DIBELS measure of oral reading fluency placed him in the at risk range according to the third grade end of the year benchmark, with a score of 17 words per minute on third grade level reading probes. When given first grade level reading probes, at his instructional level according to curriculum-based assessment results, he read 33 words per minute, placing him in the range designated as “some risk” even according to DIBELS criteria for first graders (Good & Kaminski, 2002).

Benjamin’s percentages of time on-task and task comprehension for each session are displayed in figure 3. His mean percentage of time on-task during frustrational level tasks was 81.22% (SD = 16.93%). His mean percentage of time on-task during instructional level tasks was 92.02% (SD = 3.0%). During independent level tasks, his mean percentage of time on-task was 96.52% (SD = 4.10%). These data produced effect sizes of .64 for time on-task at the instructional level compared to the frustrational level and .90 at the independent level compared to the frustrational level. However, none of the data points at the instructional level were non-overlapping with those at the frustrational level. At the independent level, 67% of the data
points were non-overlapping with those at the frustrational level and 67% were non-overlapping with those at the instructional level.

Benjamin’s mean percentage of task comprehension was 20% (SD = 10%) for frustrational level tasks, 53.33% for instructional level tasks (SD = 15.28%), and 53.33% for independent level tasks (SD = 15.28%). These data produced effect sizes of 3.33 for comprehension at the instructional level compared to the frustrational level and 3.33 for comprehension at the independent level compared to the frustrational level. One hundred percent (100%) of the data points at both the instructional and independent levels were non-overlapping with those at the frustrational level. However, none of the data points between the instructional and independent levels were non-overlapping.

Results of manipulation checks indicated that Benjamin struggled with reading even those passages determined to be at his instructional and independent levels. Although initial CBA data indicated that level 1 stories within the RN series constituted the instructional level for Benjamin, during manipulation checks he read the level 1 passage with only 92.6% known words. Stories at the .8 level, the lowest level within the RN series, were used to reflect the independent level for Benjamin, although manipulation checks indicated that these may have more accurately reflected his instructional level, as he read the .8 level story with 94% known words. Thus, results for Benjamin may not accurately reflect his functioning at the instructional and independent levels.

Trends Across Students

The mean percentages of time on task for each of the three students by phase are plotted together in figure 4. These data show that both Jeremy and Jessica were most on-task during instructional level tasks, while Benjamin was most on-task during tasks intended to represent his
Mean percentages of time on-task for each condition averaged across all three students are displayed in figure 5. At baseline, the mean percentage of time on-task across all students was 67.13% (SD = 20.07%). During frustrational level tasks, the mean percentage of time on-task across all three students was 87.97% (SD = 10.61%). The mean percentage of time on-task across all three students was highest during instructional level tasks at 94.43%, and had the smallest variability with a standard deviation of 3.27%. The mean percentage of time on-task during independent level tasks was 92.26% (SD = 5.15%).

The overall mean effect size for time on-task at the instructional level compared to the frustrational level was 3.92, for time on-task at the independent level compared to the frustrational level was .94, and for time on-task at the instructional level compared to the independent level was 1.14. The mean percentage of data points for time on-task at the instructional level not overlapping with those at the frustrational level was 44.33%. The mean percentage of non-overlapping data points for time on-task between the instructional and the independent levels was 55%. The mean percentage of data points for time on-task at the independent level not overlapping with those at the frustrational level in the positive direction was 44%.

Figure 6 displays the mean percentages of task comprehension across all three students. The mean percentage of task comprehension for frustrational level tasks was 36% (SD = 17%), for instructional level tasks was 61% (SD = 14.53%), and for independent level tasks was 72% (SD= 18%). The mean effect size for comprehension at the instructional level compared to the frustrational level was 2.12, and for comprehension at the independent level compared to the frustrational level was 2.75. The mean percentage of non-overlapping data points for comprehension between the frustrational level and the instructional level was 67%. The mean
percentage of non-overlapping data points for comprehension between the frustrational level and the independent level was 100%.

**Discussion**

As hypothesized, results for Jeremy and Jessica indicated that they spent the largest percentage of time on-task while completing tasks at the instructional level (containing 93-97% known words), as compared to the frustrational level (less than 93% known words) or the independent level (97-100% known words). This effect did not appear to be as large in magnitude as that obtained by Gickling and Armstrong (1978), in which 100% of the data points at instructional level were non-overlapping with those at the frustrational and independent levels when session results were averaged across students. However, by presenting the results of their study averaged across students by session, the amount of variability within and between students was not evident, and no mention of effect sizes was made. Figure 7 shows the current results averaged across students by session, as the results of Gickling and Armstrong (1978) were presented. When presented this way, the current results appear to be more similar to those presented by Gickling and Armstrong (1978). However, the current results are not discussed averaged across students by session, as the order of presentation of conditions was randomized for each student, and thus all students did not receive the same condition at the same time. This was done in an attempt to eliminate the possibility of a contagion effect, in which students’ off-task behavior is increased by the off-task behavior of others in the classroom, which may have also contributed to the large effects presented by Gickling and Armstrong (1978). Nevertheless, the current results lend support to the ratios of known to unknown items in text proposed by Gickling and Havertape (1981) to constitute the instructional, independent, and frustrational levels.
It was expected that the frustrational level would represent a continuation of the baseline in students’ percentages of time on-task. However, all three students displayed higher percentages of time on-task during the frustrational level condition than during baseline, when the usual classroom curricular materials were presented. This could be explained by the novelty of the Read Naturally stories to the students, which might have maintained their attention even at frustrational level more than the normal classroom routine. Additionally, the Read Naturally stories were presented by university students, which may have seemed interesting to them and motivated them to do well. The way in which the classroom teacher introduced the activity also seemed to give the students the impression that it was a test, which may have additionally improved students’ behavior above what would be typical. A higher degree of variability was also present in students’ time on-task during baseline conditions, suggesting that a high degree of variability in task demands may also have been present during the regular classroom curriculum conditions, especially compared to the controlled task demands at the frustrational, instructional, and independent levels.

The results for Benjamin indicated that he spent the most time on-task while completing tasks at the independent level. However, manipulation check data indicated that the stories used to represent his instructional level may have actually still been frustrational for him, and those stories at the independent level may have more accurately represented his instructional level. This likely resulted from Benjamin’s struggles with even the lowest leveled stories in the Read Naturally series. Taking this into consideration, the results obtained for Benjamin were consistent with the hypothesis that students would be most on-task while completing tasks containing 93-97% known words. As tasks became more difficult for Benjamin, he became more off-task. It is unknown how Benjamin would have performed had an activity that was truly
at his independent level been presented, although it is likely that his results would have mirrored those of the other students, with time on-task lower at the independent level than at the instructional level.

Effects on comprehension for each student were consistent with the expectation that comprehension would be the highest during the instructional and independent levels. Comprehension for each student improved noticeably from the frustrational level to the instructional level. Comprehension for Benjamin did not improve additionally from the instructional level to the independent level, as the stories presented at the independent level were likely not much easier for him than those at the instructional level. However, comprehension for both Jeremy and Jessica continued to improve from the instructional to the independent level as stories became easier for them to read.

Results averaged across students by condition supported the hypotheses for both time on-task (figure 5) and comprehension (figure 6). The overall mean percentage of time on-task was highest at the instructional level, slightly lower at the independent level, and lowest at the frustrational level and during baseline. Overall means for comprehension, as expected, were highest at the independent level and lowest at the frustrational level. These results, when aggregated across students, were more consistent with predictions than the results for the individual students. Data presented by Gickling and Armstrong (1978) were only presented in their aggregated form across students, with no mention of the variability either within or between students. Effects such as these can appear larger and more consistent when presented as group means than when presented individually by student, which allows individual differences and variability to become apparent. This provides support for the value of single subject designs in examining the effectiveness of interventions for implementation with individual students.
An unanticipated but interesting effect seen in the data is the lower variability in students’ time on-task at the instructional level compared to the other levels, both within and between students. The standard deviation from the mean for students’ aggregate percentage of time on-task was lowest at the instructional level, and for two out of three of the individual students, the standard deviation was also the smallest at the instructional level. Thus, it may be that students’ on-task behavior is more consistent at the instructional level than at the independent or frustrational levels. The researchers noted that sometimes the students seemed to take the frustrational level passages as a special challenge, while other times they did not seem to even try to read them. And at the independent level, students sometimes appeared motivated to complete the stories and sometimes did not. However, passages at the instructional level seemed to hold students’ attention more consistently overall.

The results of this investigation have implications for providing effective instruction in inclusive classrooms, and for potentially reducing the numbers of students labeled with behavioral or learning disabilities. As teachers face classrooms of students with widely ranging skill levels, differentiating instruction to the individual needs of students is crucial for ensuring that each child is given the opportunity to progress from their current level, and that no child is left behind academically. Curriculum based assessment (CBA), while not widely implemented, has been supported in the research as a useful means of matching instruction to students’ individual skill levels to allow for optimal learning to take place (Burns, 2002; Shapiro, 2004; Shapiro, et al., 2004). The feasibility of matching instruction to individual students’ skill levels within inclusive general education classrooms has been criticized (Shapiro et al., 2004). However, this study was carried out in a classroom of 23 students, without additional planning time required by the teacher. Curriculum-based assessments took only a few minutes to conduct.
with each student, and materials at students’ individually matched levels within the reading series were provided during individual seatwork based on these data. The results of this study thus provide support for both the effectiveness and feasibility of using CBA to differentiate instruction to varying skill levels in inclusive classrooms.

A wealth of research has demonstrated that classroom behavior problems are closely linked to academic difficulties (Maughan et al., 1996; Mayes et al., 2000; McComas et al., 2000; Stanton et al., 1990; Wasson et al., 1990). This investigation further supports the conclusion that both classroom behavior difficulties and task comprehension can be improved by matching instructional materials to individual student’s skills. A variety of high incidence disabilities may begin with academic skills difficulties and off-task classroom behavior that worsen over time as students fall further and further behind the level of instruction occurring in their classrooms (Mayes et al., 2000; Stanovich, 1986). Providing individually matched instructional materials through the use of CBA may hold promise for preventing these classroom difficulties early on and thus reducing the numbers of students who are referred for special education services under high incidence disability labels. The provision of such prevention and early intervention services to reduce the labeling of students with disabilities has been identified as an important goal for special education by the President’s Commission on Excellence in Special Education (2002). If providing learning materials at students’ instructional levels can increase time on-task and comprehension rates, then students can likely learn at faster and more efficient rates. This may eliminate the need to move more slowly with struggling students, as is often done in remedial classes and will only increase the skill deficits between struggling students and their peers. If struggling students are able to progress at a more regular pace through the curriculum, they may eventually catch up to the level of their grade-level peers.
Limitations of the Current Research

A significant limitation of the current research is the minimal number of data points obtained for students at each phase due to time constraints on the collection of data. With only three data points at each phase, variability within the data makes it difficult to discern with confidence the true levels and trends in students’ performance. Thus, more consistent results might be obtained by continuing the phases over more sessions. Additional limitations include the lack of a true independent level obtained for one of the students, and that results were only obtained for a total of three individual students. While it is unknown if these trends would hold up across larger numbers of students, the more students with which this effect is obtained, the more confident we can be in our conclusions.

Future research might attempt to examine the effects of instructional difficulty levels on group means for time on-task and comprehension, using larger numbers of students. This could provide a useful means of examining various ratios of known to unknown items and their effects for various groups of students, including students with various disability labels, and of various ages and ability levels. Student engagement in learning at various difficulty levels might be gauged not only with time on-task, but also with rates of academic responding during whole class instruction. The instructional ratios should also be further examined for their applicability in various content areas, including mathematics and spelling, and for various types of tasks.

Maintaining students’ engagement in academic tasks is an ongoing struggle for educators. For optimal learning to take place, not only procedural engagement, but substantive engagement, is key (Gettinger & Siebert, 2002). Thus, students must not only work on the assigned tasks, but also comprehend the learning material. For optimal engagement, individualizing instruction should involve not only matching the difficulty level of material to
students’ skill levels, but also matching content to students’ interests and making learning experiences relevant to students’ experiences outside of the classroom (Gettinger & Siebert, 2002; Gickling & Thompson, 1985). When all of these components are in place, students are likely to actively engage in learning, self-motivate to achieve, and optimize their learning experience (Stipek, 1996). Thus, all of these factors are important, and researchers and practitioners should work to develop ways to achieve an instructional match for students not only by skill level, but by students’ interests and relevance to their life experiences as well.
References


Available: http://dibels.uoregon.edu/.


Individuals with Disabilities Education Improvement Act of 2004, H.R. 1350.


President’s Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families.* Available at:


**Figure Captions**

*Figure 1.* Percentages of time on task and task comprehension for each session, by condition, for Jeremy.

*Figure 2.* Percentages of time on task and comprehension for each session, by condition, for Jessica.

*Figure 3.* Percentages of time on task and comprehension for each session, by condition, for Benjamin.

*Figure 4.* Mean percentages of time on task at each condition, for each student.

*Figure 5.* Mean percentages of time on task across all three students, by condition.

*Figure 6.* Mean percentages of task comprehension across all three students, by condition.

*Figure 7.* Mean percentages of time on task and comprehension for each session, averaged across students.
Percent time on task

Baseline
Frustrational
Instructional
Independent