2008-11-25

The Promise of Academic Learning Time in a Dose-Response Model of Early Reading Achievement

Benjamin Heuston
Brigham Young University - Provo

THE PROMISE OF ACADEMIC LEARNING TIME IN A DOSE-RESPONSE MODEL OF EARLY READING ACHIEVEMENT

Edward Benjamin Hull Heuston

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirement of the degree of

Master of Psychology

Department of Psychology

Brigham Young University

December 2008

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Edward Benjamin Hull Heuston

The thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Harold Miller, Chair
Date	Timothy B. Smith
Date	Mikle South

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Edward Benjamin Hull Heuston in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date | Harold Miller, Chair

Accepted for the Department

Date | Harold Miller
Graduate Coordinator

Accepted for the College

Date | Susan Rugh
Associate Dean, College of Family, Home and Social Sciences

ABSTRACT

THE PROMISE OF ACADEMIC LEARNING TIME IN A DOSE-RESPONSE MODEL OF EARLY READING ACHIEVEMENT

Edward Benjamin Hull Heuston

Department of Psychology

Master of Science

Reading has long been acknowledged to be a critical skill that is best acquired early in life. According to the most recent National Assessment of Educational Progress (NAEP) reports, American public school children continue to struggle to master the basics of reading. Although federal funding in real dollars has increased consistently over time, reading scores have not followed suit, suggesting that fiscal resources have not been applied successfully to the variables that are directly related to reading acquisition and achievement. The current state of affairs therefore suggests the need for identifying a fiscally-targetable, instructionally-relevant variable with a direct, causal relationship to early-reading achievement.

One way to determine whether such a relationship exists between two variables is by means of dose-response methodology. Although this methodology has not been broadly implemented in educational research, it is attractive because it allows for the formal characterization and comparison of cause-effect relationships, and may also inform practice in readily implementable ways.

Researchers have noted that time spent learning (TSL), and in particular academic learning time (ALT), is a promising candidate for a dose-response relationship with student achievement in early reading. Although ALT holds promise, there have traditionally been significant difficulties in operationalizing and quantifying it. The growing prevalence of academic software in the American public classroom holds promise for overcoming these challenges and provides an opportunity to test the hypothesis that there is a dose-response relationship between ALT and student achievement in early reading.

ACKNOWLEDGEMENTS

To my wife for her incisive edits and unwavering support; to my children for their patience and love; to my graduate committee for their advice and encouragement; to my advisor for endless L&Gs and for helping me to elevate; to my graduate coordinator for imposing order; to my father for giving me the time; to my mother for setting the bar; and to my sister who always knew.

Table of Contents

Page

List of Figures

List of Tables

"Should someone find it desirable to do so, a degree of scientific rationality can be brought to bear on the educational system" – D. C. Berliner

The Importance of Reading

It is axiomatic that reading is a fundamental and valuable skill (Adams, 1990; Snow, Burns, & Griffen, 1998; Stanovich, 2000). Grover Whitehurst (2003), director of the U.S. Department of Education's Institute of Education Sciences, noted that "reading . . . is absolutely fundamental . . . [T]he inability to be fluent is to consign children to failure in school and to consign adults to the lowest strata of job and life opportunities". It is clear from these comments that reading skills have an impact far beyond the classroom. This point was made even more strongly by Pressley (1998), who asserted that "one part of any strategy to prevent disadvantaged children from being upwardly mobile would be to deny them effective literary instruction" (p. 37). Thus, reading is not just a valuable academic skill, it is a defining life skill and therefore demands our best efforts to understand how best to promote it among those most in need.

Reading vs. Speaking

"Some people there are who, being grown, forget the horrible task of learning to read. It is perhaps the greatest single effort that the human undertakes, and he must do it as a child." – John Steinbeck

Most students learn to speak without explicit instruction, but only a small fraction can master the complexities of literacy on their own. Using Geary's (2000) terminology, human speech is a *biologically primary* capability: it needs no formal instruction and seems to be "hard-wired" for the majority of people; it is best thought about as an evolved trait that helps define a species as opposed to a skill that is developed or acquired during a lifetime. The American educational system relies heavily on the fact that the vast majority of children will be able to communicate verbally with their teachers on the first day of school and although refinements might be in order, the basic ability to speak is generally taken as a given.

On the other hard, reading and writing are *biologically secondary* capabilities: they require explicit, systematic instruction before they can be mastered and are better conceptualized as skills that can be learned or acquired only through extended and extensive effort. However, they are not defining traits enjoyed by all. Reading is not an assumed skill and for good reason – it requires years of concentrated instruction and practice in order to be fully developed.

Reading Acquisition and the Matthew Effect

"For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." – Matthew 25:29

One notable aspect of the acquisition of reading skills is that it is subject to the Matthew Effect (Walberg & Tsai, 1983; Stanovich, 1986), so termed in reference to a passage from the New Testament wherein the rich get richer and the poor get poorer. In an educational context it refers to the fact that "[s]tudents who are behind at the beginning of schooling . . . usually learn at a slower rate; those who start ahead gain at a faster rate, which results in . . . the academically rich getting richer" (Walberg, 2003, p. 12).

This phenomenon has been well-documented in the study of reading. In his 2000 review of reading research, Stanovich summarizes the effect thus:

In short, many things that facilitate further growth in reading comprehension ability . . . are developed by reading itself. The increased reading experiences . . . have important positive feedback effects. Such feedback effects appear to be potent sources of individual differences in academic achievement (p. 163).

Bast and Reitsma (1998) found that “good readers tended to read more frequently than poor readers. These leisure time reading activities were related to differences in the size of the vocabulary at the end of second grade. In turn, vocabulary affected subsequent comprehension in reading” (p. 1387). Cunningham, & Stanovich (1997) found in their longitudinal study that “[f]irst-grade reading ability was a strong predictor of all the 11th-grade outcomes and remained so even when measures of cognitive ability were partialed out” (p. 934). Thus, to promote the acquisition of reading skills by students in American public schools, it is important to focus on helping students acquire reading skills at the beginning of their schooling careers so they can benefit from the positive-feedback effects that are connected with early-reading skill acquisition.

The Simple View of Reading

Having established the broad importance of the skill of early reading, it remains to more closely define what exactly composes it. Labeling “the skill of reading” as a singular skill is an oversimplification. Reading is composed of a variety of skills that work together seamlessly and transparently in the fluent reader; its composite nature is generally only observable when the text is beyond the reader’s mastery level (Adams, 1990). To better understand the various components of reading, it may be helpful to introduce a theoretical model.

Gough & Juel (1991) decompose the skill of reading into two distinct abilities, namely *word recognition* – the ability to "grasp what word each letter string represents" – and *comprehension* – the ability to "decide what those words collectively mean" (p. 47). The authors situate these variables in relation to one another according to the following equation: R = D x C, where R refers to reading, D to word recognition, and C to comprehension. This so-called *Simple View of Reading* has been validated subsequently by research (Catts, Adlof, & Weismer, 2006; Pressley, Billan, Perry, Reffitt, & Reynolds, 2007; Stanovich, 2000) and provides a useful starting place for investigating which skills should be promoted in order to improve reading acquisition in America.

Comprehension

For children entering the American public school system, the first year of formal education is generally kindergarten. Gough, Hoover, & Peterson (1996) note that average kindergartners are not constrained in their reading by their comprehension abilities but rather by their word-recognition abilities: "The typical text that confronts the child at this age is very simple, with a difficulty level well beneath the mean. If the text were read to the child, it would be understood by almost every normal child" (p. 5). Returning to the equation of R = D x C, if comprehension is effectively perfect (i.e., equal to 1.0) for these children, then the equation simplifies to R = D. Thus, at least across the early elementary grades, it is to be expected that individual differences will be correlated primarily with word-recognition skills as opposed to comprehension skills.

Word-Recognition Skills

Gough and Juel (1991) note the relative importance of word recognition skills for beginning readers:

> The first grade child already knows, in their spoken or phonological form, most of the words that he will encounter in print for the next 3 years. What he doesn't know is their printed form. If he had a means of converting the novel printed form into phonological form, then he could recognize it. (p. 51)

To develop these means, children must first deepen their knowledge of the spoken language and from there build bridges of understanding to the realm of print. This complex journey has been lucidly chronicled by the National Reading Panel [NRP] (2000), whose comprehensive report undergirds the following overview of the process of acquiring early-reading skills.

Phonemes vs. Graphemes

The basic building blocks of spoken language are known as *phonemes*, which are combined to form larger units of speech such as syllables and words. "Phonemes are different from graphemes which are units of written language and represent phonemes in the spellings of words" (NRP, 2000, p. 2-10). In a language like English, the letters that comprise a single grapheme (and therefore map to an individual phoneme) can and often do differ. Thus, words can be spelled with the same number of letters and phonemes, such as /c/ /a/ /t/ and C-A-T, or they might differ significantly, as in /h/ /I/ and H-IGH.

Phonological and Phonemic Awareness

In beginning the transition from speaker to reader, the child first must begin to parse the sound stream of language. This appears to be a biologically secondary (i.e., non-intuitive) skill that is generally only developed through explicit instruction and effortful practice. “Being able to distinguish the separate phonemes in pronunciations of words so that they can be linked to graphemes is difficult. This is because spoken language is seamless and there are no breaks in speech signaling where one phoneme ends and the next one begins” (NRP, 2000, p. 2-11).

The two skills that children generally need to acquire in order to begin to read are *phonological awareness*, which “refers to a general appreciation of the sounds of speech as distinct from their meaning,” and *phonemic awareness*, which further refines phonological awareness by referring specifically to “an understanding that words can be divided into a sequence of phonemes” (Snow et al., 1998, p. 51). Thus phonological awareness is a broader category than phonemic awareness and includes the larger units of speech, such as syllables and rhymes; phonemic awareness refers specifically to “the ability to focus on and manipulate phonemes in spoken words” (NRP, 2000, p. 2-1). According to Snow et al. (1998), “[c]hildren with phonemic awareness are able to discern that camp and soap end with the same sound, that blood and brown begin with the same sound, or, more advanced still, that removing the /m/ from smell leaves sell.” (Snow et al., p. 53).

The Alphabetic Principle and Decoding

Once a child has achieved a mastery of phonemic awareness, she is positioned to understand the *alphabetic principle*, namely, that the graphemes she sees on the page map onto the phonemes that she hears. This fundamental understanding is required before further progress can be made. "A beginning reader must at some point discover the alphabetic principle . . . [t]his principle may be induced; it may be acquired through direct instruction . . . but it must be acquired if a child is to progress successfully in reading" (Stanovich, 2000, p. 162). The alphabetic principle provides a Rosetta stone of sorts to the beginning reader, forming the basis for synchronizing and translating between written and spoken language.

Once in possession of this core insight, a burgeoning reader can then develop the ability to take the written (or graphemic) form of a word and translate it into its spoken (or phonological) form, a skill known as *decoding*. A word that has been successfully decoded has also been *identified*– word identification refers to pronunciation, not to comprehension (Snow et al., 1998). This delineation is obvious in the common occurrence of *word calling* in students who are on the cusp of fluent decoding. Word calling is typified by an ability to correctly decode a word but an inability to comprehend the word. This lack of comprehension is not due to the word's absence from the student's oral vocabulary but rather to the lack of decoding fluency: "It is quite possible for accurate decoding to be so slow and capacity-demanding that it strains available cognitive resources and causes comprehension breakdowns" (Stanovich, 1986, p. 373).

Summary

The Simple View of Reading bifurcates the skill of reading into the broad categories of word recognition and comprehension, both of which are necessary for a child to be fluent. Children beginning to learn how to read often have oral comprehension that far exceeds their ability to recognize words. Thus approaches to early-reading skill acquisition by and large focus on word recognition and its components. Summing up the relevant research, Stanovich (2000) notes: "The causal model . . . [of] phonological awareness facilitating decoding skill, which in turn determines word recognition ability, which in conjunction with listening comprehension determines reading comprehension . . . has largely stood the test of time" (p. 61).

Reading Skill Development Timeline

The cumulative nature of reading skills makes them particularly sensitive to initial effects, and they are therefore most easily acquired early in life (Fuchs et al., 2001; Gallacher, 2008; Snow et al., 1998). "It is . . . important that the prerequisite phonological awareness and skill at spelling-to-sound mapping be in place early in the child's development because their absence can initiate a causal chain of escalating negative side effects" (Stanovich, 2000, p. 162). As for what "early" means, Stanovich notes that "extremely large differences in reading practice begin to emerge as early as the middle of the first-grade year," and: "Thus, soon after experiencing greater difficulty in breaking the spelling-to-sound code, poorer [first-grade] readers begin to be exposed to less text than other peers" (p. 162). Snow et al. (1998) also found that "For most children, an awareness of the phonological structure of speech generally develops gradually over the preschool years" (p. 51), but "the acquisition of 'real' reading typically begins at about age 5 to 7, after the child has entered Kindergarten" (p. 68).

It is apparent, then, that in American public schools phonological awareness is more of a school-readiness skill, but instruction in phonemic awareness, with its much more fine-grained ability to parse speech, should begin in earnest when children first enter formal schooling in kindergarten or 1st grade. The decoding of novel words begins in the 1st grade and continues through 3rd grade, after which time it should be rapid and automatic. (For detailed developmental and educational milestones associated with reading from birth through the end of 3rd grade, see Appendix A).

The State of Early Reading in American Public Schools

Having identified the building blocks and timing of early-reading skills, the natural next step is to evaluate how American public school students are currently performing on these skills. One way to assess the present state of early-reading achievement in the United States is to look at children's scores on early-reading assessments. Although there is a wealth of reading tests in use throughout the country, the only reading test consistently administered at the national level is the National Assessment of Educational Progress (NAEP), commonly known as "The Nation's Report Card" [National Center for Education Statistics (NCES), 2008a].

National Assessment of Educational Progress

NAEP is a broad term encompassing a variety of tests and measures that that are administered on a periodic basis to students in both public and private schools across the United States. The tests provide information on "subject-matter achievement, instructional experiences, and school environment" and are administered at carefully chosen times that "represent critical junctures in academic achievement" (NCES, 2008a).

In line with its billing as a national test, NAEP is given to randomized strata of students and schools across the United States. Employing a matrix-sampling design, NAEP is structured so that each participating student sees only a portion of the possible test items. Based on this approach, NAEP is able to provide comparative and trending data on the performance of states, regions, and groups, but it is unable to provide information at individual school or student levels (NCES, 2008a).

NAEP Reading

According to the National Assessment Governing Board (NAGB; 2004), the NAEP reading assessment "measures the outcomes of instruction as reflected in the behaviors of readers." There are two different varieties of NAEP reading tests: the main NAEP and the long-term trend NAEP. The main NAEP is the only one that provides data on achievement levels (NCES, 2008b); therefore the long-term trend NAEP will not be considered further.

The earliest point at which the main NAEP measures reading is the 4th grade, when, according to federal reading standards, children should already be functioning as independent readers. Viewed through the lens provided by the Simple View of Reading, NAEP assumes that by the 4th grade children should have mastered word recognition to the point that reading ability is constrained primarily by comprehension.

This expectation is reflected in the philosophy of the 4th-grade NAEP reading test, which measures reading in two different contexts: reading for literary experience and reading for information (see Appendix B for a representative sample of NAEP items). In both cases the actual decoding of the words themselves is assumed. According to the

NAGB (2004), the intent of the NAEP reading test is therefore not to focus on the words themselves, but rather on the text and its message:

> All NAEP questions emphasize critical thinking and reasoning . . . constructed-response questions require students to integrate information from the text with their background knowledge, reorganize ideas, and critically consider the text. In an assessment of reading, it is important to have items that can directly and accurately reflect how readers use multiple strategies to build understanding. (chap. 2)

This emphasis on strategies and critical thinking skills indicates that the 4th-grade NAEP reading test is focused on reading comprehension and assumes that the skills of phonological and phonemic awareness, decoding, and word recognition have already been mastered.

NAEP and Early-Reading Achievement

The timing of the main NAEP reading assessment is unfortunate with respect to evaluating the current state of early-reading achievement, as it cannot provide synchronous data on the current state of early reading. Instead reading scores on the 4th-grade NAEP simply mark an endpoint to the collective processes that occurred earlier.

The 4th-grade main NAEP is similarly unable to identify any of the early-reading skills that research has identified as critical to future reading success because they are not explicitly measured. At best it may indicate that 4th graders in America perform poorly in reading comprehension, but whether the deficit lies in comprehension per se or is due to an underlying deficit in phonological awareness or decoding cannot be disambiguated with certainty.

NAEP and Early-Reading Trends

Acknowledging these limitations, NAEP nevertheless provides a useful broad perspective on what children's reading comprehension skills currently are and also how they have changed over time. According to the 2005 NAEP results, more than one-third of American 4th-graders score in the lowest category (Below Basic), indicating they cannot read and understand a simple paragraph (see Table 1).

This result is alarming not only because it indicates that many children are not learning to read successfully, but when combined with the Matthew Effect, it indicates that these underperforming children will likely never read at an acceptable level. In addition, it appears that these results are not an anomaly – reading scores and achievement-level distributions have changed little over the past 13 years at any of the grade levels NAEP measures.

Summary

The most recent national achievement tests in reading demonstrate that roughly one-third of 4th-grade students in American public schools are unable to demonstrate the lowest level of reading proficiency, while only one-third are meeting or exceeding the federal reading standards. Furthermore, these results have stayed relatively stable across an extended period of time. Although these results do not speak directly to early-reading skills, they are assumed to be a direct outgrowth from an earlier failure to acquire the prerequisite early-reading skills.

Percentile and achievement level	Grade 4 1992[1]	Grade 4 1998[1]	Grade 4 1998	Grade 4 2002	Grade 4 2005	Grade 8 1992[1]	Grade 8 1998[1]	Grade 8 1998	Grade 8 2002	Grade 8 2005	Grade 12 1992[1]	Grade 12 1998[1]	Grade 12 1998	Grade 12 2002	Grade 12 2005
Average score															
Total	**217**	**217**	**215**	**219**	**219**	**260**	**264**	**263**	**264**	**262**	**292**	**291**	**290**	**287**	**286**
Percentile[2]															
10th	170	167	163	170	171	213	217	216	220	216	249	242	240	237	235
25th	194	193	191	196	196	237	242	241	244	240	271	268	267	263	262
50th	219	220	217	221	221	262	267	266	267	265	294	293	293	289	288
75th	242	244	242	244	244	285	288	288	288	286	315	317	317	312	313
90th	261	263	262	263	263	305	305	306	305	305	333	337	336	332	333
Percentage at achievement level															
Achievement level															
Below *Basic*	38	38	40	36	36	31	26	27	25	27	20	23	24	26	27
At or above *Basic*	62	62	60	64	64	69	74	73	75	73	80	77	76	74	73
At or above *Proficient*	29	31	29	31	31	29	33	32	33	31	40	40	40	36	35
At *Advanced*	6	7	7	7	8	3	3	3	3	3	4	6	6	5	5

[1] Testing accommodations (e.g., extended time, small group testing) for children with disabilities and limited-English-proficient students were not permitted.
[2] A score location at or below which a specified percentage of the population falls. In 1992, for example, 10 percent of 4th-graders scored at or below 170, while 90 percent of 4th-graders scored above 170.
NOTE: Beginning in 2002, the NAEP national sample for grades 4 and 8 was obtained by aggregating the samples from each state and the District of Columbia, rather than by obtaining an independently selected national sample. As a consequence, the size of the national sample for grades 4 and 8 increased, and smaller differences between years or between types of students were found to be statistically significant than would have been detected in previous assessments. See *supplemental note 4* for more information on NAEP.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1992–2005 Reading Assessments, NAEP Data Explorer.

Table 1.
NAEP Reading Performance of Students in Grades 4, 8, and 12
Note: From *The Condition of Education 2007 (NCES 2007-064)*, p. 134, by U.S. Department of Education, National Center for Education Statistics., 2007,Washington, DC: U.S. Government Printing Office.

National tests thus provide a compelling case that action needs to be taken to improve the state of early reading in American public schools. Prior to suggesting a course of action, however, it is important to review what national efforts have already been undertaken to try and improve early-reading achievement.

Federal Efforts to Improve Early-Reading Achievement

Beginning with the landmark 1965 Elementary and Secondary Education Act (ESEA), the federal government has aggressively sought to improve the equity of the educational landscape in America for all children. In attempting to do so it has relied on two major strategies, namely funding and accountability. The clear trend over time has been to increase federal funding dramatically in exchange for concomitant accountability in the area of student performance. While earlier legislation was focused on general academic improvement, the most recent legislation has targeted early-reading achievement directly.

Elementary Secondary Education Act

"From our very beginnings as a nation, we have felt a fierce commitment to the ideal of education for everyone." – Lyndon B. Johnson

The ESEA was passed in 1965 as a centerpiece of President Lyndon B. Johnson's "War on Poverty." According to Guthrie & Springer (2004), "President Johnson deeply believed that household poverty prevented many American children from participating fully in the nation's riches and that a principal instrument for overcoming this deficit was to enable poor children to engage successfully in the education system" (p. 31). The primary reason for this lack of engagement was hypothesized to lie in the scarcity of resources allocated to poorer schools. A new program, Title I, was authorized to address the problem. Carter (1984) observed that "[o]ne of [Title I's] major justifications was the desire to improve the educational opportunities of the poor and educationally disadvantaged. Funds were made available to all states and in turn to local school districts to support additional instruction at schools in economically poorer areas" (p. 6). The fiscal impact of the ESEA was immediate and dramatic: "In just 1 year federal spending on education doubled from $1 billion to $2 billion and it grew to nearly $3 billion by the end of the decade" (Viteritti, 2004, p. 69).

Such a large spending package did not come without stipulation: "In view of the large sums involved and the concern of some members of Congress that the local school districts might not use the money as intended, the Act required that the Title I program be evaluated" (Carter, 1984, p. 6). However, the very nature of the funding bill, was problematic from an evaluation standpoint as clear methods and outcomes were not articulated (Carter, 1984). This resulted in the conclusion that "…Title I was better defined as a funding program than as an educational treatment" (p. 11). Given this fundamental flaw, it is not surprising that researchers subsequently found that Title I as originally enacted was not successful: "By the early 1980s, studies were indicating that the billions of dollars spent on compensatory education for poor children were bearing no tangible results in the classroom" (Viteritti, 2004, p. 69).

A Nation at Risk

> The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people – *A Nation at Risk*

In 1983, the National Commission on Excellence in Education (NCEE) issued the galvanizing report *A Nation at Risk: The Imperative for Educational Reform* (NAR). This call-to-arms document raised the specter of America falling behind the rest of the world economically and educationally. At its core, NAR argued that "downwardly spiraling pupil performance had rendered the U.S. education system dysfunctional, thereby threatening the nation's technological, military, and economic preeminence"; in this precarious position, "only by elevating education achievement could the United States avoid subordinating itself to its educational superiors and economic competitors" (Guthrie & Springer, 2004, p. 8).

The framers of NAR left no doubt as to its incisiveness and proposed the following axiom as a prelude to the document:

> All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost. This promise means that all children by virtue of their own efforts, competently guided, can hope to attain the mature and informed judgement needed to secure gainful employment, and to manage their own lives, thereby serving not only their own interests but also the progress of society itself.

By defining the target audience as "all," NAR takes the original focus of the ESEA on the disadvantaged and dramatically expands it. This new "entitlement" demanded that quality educational instruction be made available to all children regardless of their background.

Like the ESEA, NAR increased federal spending, but in exchange it required a much stronger level of oversight in connection with these dollars. According to Viteritti (2004), "NAR articulated a demand for educational excellence and an understanding that for reform to be meaningful it must result in changes that have tangible academic results" (p. 65). In particular, NAR's recommendations centered around five goals: strengthening the content of what is taught; increasing both standards and expectations; increasing students' time on task; improving the preparation of teachers; and improving educational leadership and fiscal support of policy. While none of these goals was singled out as a panacea, the inference was that, by achieving all of them, the crisis of low student achievement would be resolved.

NAR catalyzed change in the American public school system, moving policy from its sole focus on funding to one that included performance. In their summation of its impact 20 years later, Guthrie & Springer (2004) note:

> From today's vantage point, the most positive result of NAR seems to have been that it triggered a move away from measuring the quality of schools by the resources they receive and onto a plane where school performance is judged on outcomes students' [*sic*] achieve. (p. 9)

This shift of focus from inputs (funding) to outputs (achievement) fundamentally changed the conversation about American school reform. Instead of assuming that more money was in and of itself going to solve the problem, NAR defined the metric of educational effectiveness to be student achievement, thereby allowing for efficiency and efficacy to enter the national dialogue. NAR was, in effect, the genesis of the modern accountability movement.

In attempting to raise student achievement, NAR urged the targeting of educational resources in a manner that maximized their efficacy. This required that programs and initiatives be ranked in comparison to one another. In order for comparisons to be made, standards needed to be developed, implemented, and coordinated. While NAR in of itself did not explicitly articulate or achieve these goals, it set American public policy firmly down this path. For Wong, Guthrie, & Harris (2004), "[p]erhaps the greatest legacy of NAR is the subsequent implementation of its recommendations about raising standards. High school graduation requirements, more stringent content requirements, and other policies aimed at raising standards have become commonplace since the report's release" (p. 3).

Another policy stance that finds its roots in NAR is the increased focus on lower-performing students. This was clearly not the stated intent of NAR, which sought to provide all children with excellent educational opportunities, but NAR's emphasis on test scores as a measure of the nation's strength inexorably led to a more intensive examination of the performance of students whose test scores were typically the lowest—socially and economically disadvantaged youth. So, although it was not the NCEE's primary intent, it would be fair to credit the report with spurring a trend that also led to demands for improving education for children at the bottom of the achievement distribution. (Guthrie & Springer, 2004, p. 6)

Thus although NAR's stated purpose appears in principle to be much more egalitarian, in practice it closely aligned with the preceding ESEA legislation and its focus on underprivileged children.

No Child Left Behind

"Taken together, these reforms express my deep belief in our public schools and their mission to build the mind and character of every child, from every background, in every part of America." - George W. Bush

The No Child Left Behind Act of 2001 (NCLB), a reauthorization of the ESEA, built upon and deepened the theme of accountability, outlining a broad program of standards and testing. For Wong & Nicotera (2004), "[t]he primary focus of NCLB is to improve the academic achievement of all students by enhancing state systems of accountability, requiring clearly defined statewide standards for academic proficiency, mandating teacher and paraprofessional quality standards, and enacting annual testing in third grade through eighth grade with results disaggregated by subgroup" (p. 101). The mandate to help "all" students preserves the same broad scope established by NAR, but the language employed clearly focuses on those children who are in danger of being "left behind" (i.e., those who are chronically lower-performing).

In addressing accountability, NCLB emphasizes increased funding and autonomy for local educational agencies in exchange for continuous improvement in student achievement scores (Department of Education, 2004). These improvements, known by the term *adequate yearly progress* (AYP), need to be demonstrated on an annual basis using state-selected assessments in grades 3-8. In order to ensure that this progress is meaningful and not simply due to the idiosyncrasies of the test used, under NCLB all states are required to participate in the National Assessment of Educational Progress (NAEP) testing program.

NCLB requires AYP to be shown not just for an "average" student but for important subgroups as well. By requiring disaggregation by subgroup, NCLB legislation allows for much more targeted accountability to be brought to bear on the education system. Schools not able to demonstrate adequate yearly progress (AYP) in student achievement are subject to a loss of funding and potentially of local control.

In addition to this broad focus on standards and accountability, NCLB also brought a strong focus to reading in general and to early reading specifically through the enactment of Reading First and Early Reading First legislation. According to the Department of Education's Reading First website (DOE, 2008b), the Reading First program:

> focuses on putting proven methods of early reading instruction in classrooms. Through Reading First, states and districts receive support to apply scientifically based reading research—and the proven instructional and assessment tools consistent with this research—to ensure that all children learn to read well by the end of third grade.

Reading First accomplishes these goals by providing grants to state education agencies that then competitively award them to local education agencies. From 2002-2008, funding for Reading First totaled $6.4 billion (DOE, 2008b).

Reading First's goals are supported in part by those of Early Reading First (ERF), a program that has a mission "to ensure that all children enter kindergarten with the necessary language, cognitive, and early reading skills for continued success in school" (DOE, 2008a). This goal is "based on the understanding that literacy is a learned skill, not a biological awakening, the initiative promotes coherent, skill-based instruction in the years before kindergarten" (DOE, 2008a). In particular, the skills that are targeted are: "Oral language (vocabulary, expressive language, listening comprehension); Phonological awareness (rhyming, blending, segmenting); Print awareness; and Alphabetic knowledge" (DOE, 2008a). Early Reading First accomplishes these goals by awarding competitive grants to "local education agencies and public or private organizations that serve children from low-income families"; from 2002-2008, funding for Early Reading First exceeded $680 million (DOE, 2008a).

Summary

From 1965 until 2008 there has been a significant increase in federal education funding. By the time that NCLB was enacted in 2001, total ESEA funding was $300 billion (Guthrie & Springer, 2004). According to the NCES (2007), "total expenditures per student rose 27 percent in constant dollars between 1989-90 and 2003-04, from $7,692 to $9,762" (p. 75; see Figure 1).

In parallel with this larger funding came a heightened desire to see meaningful improvements in student achievement. Spending legislation was initially very broad and accountability was ill-defined and consequently informal. Over time both areas of focus have increased markedly in specificity, culminating in NCLB legislation that is both rigorous and narrowly targeted: programs seeking funding through Reading First must demonstrate that they are "scientifically based" and the instructional and assessment tools must be "proven." The validity of these monikers is then itself brought under scrutiny by requiring schools that receive federal funding to participate in defined testing programs (DOE, 2008b).

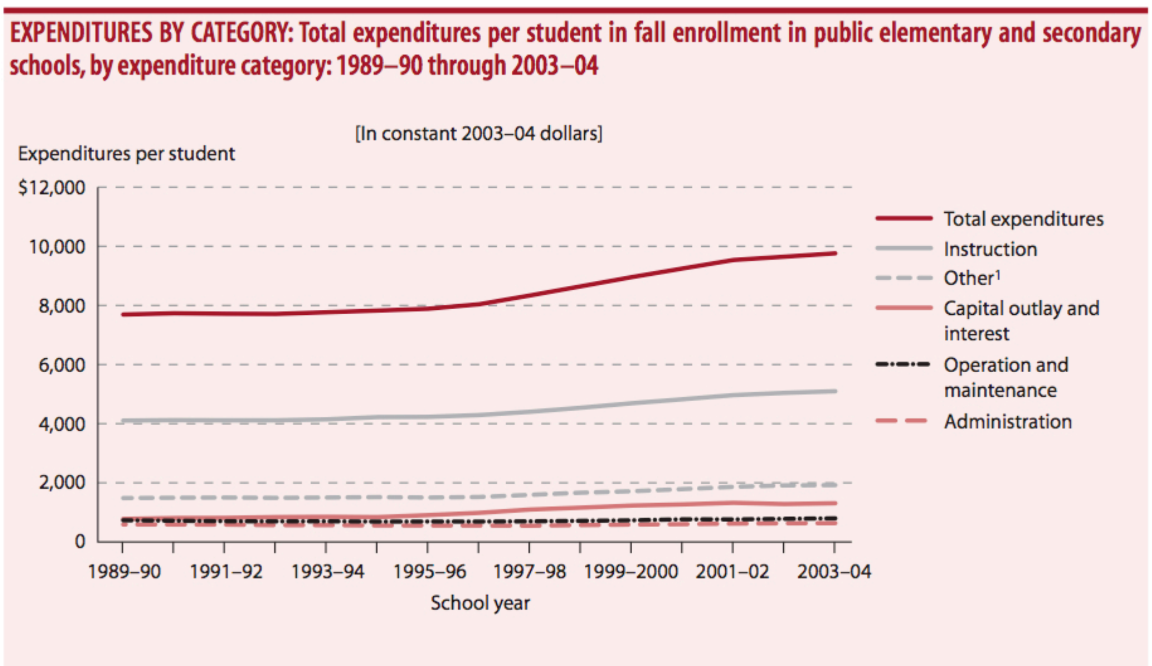


Figure 1.
Total Student Expenditures
Note: From *The Condition of Education 2007 (NCES 2007-064)*, p. 75, by U.S. Department of Education, National Center for Education Statistics., 2007,Washington, DC: U.S. Government Printing Office.

Although the ESEA did not place any special emphasis on early-reading achievement, subsequent legislation has progressively focused on attempting to improve early-reading achievement by increasing both funding and accountability. These efforts have leveraged the recommendations of early reading experts as to what skills need to be promoted and when, and future spending on early reading programs is contingent on their being research-based so that such spending will actually result in an increase in early-reading achievement. These carefully orchestrated initiatives give reason to believe that current federal funds are potentially being spent in the most efficient way possible to accomplish the policy goal of improving early-reading achievement.

The Need for a New Direction

Insanity: doing the same thing over and over again and expecting different results.
– Albert Einstein

Beginning with the ESEA in 1965, American federal policy has focused on increasing educational funding with the initial intent of helping the disadvantaged and, over time, all children to an excellent education. This funding has consistently become more targeted in nature, culminating in the recent Reading First and Early Reading First legislation that specifically targets those programs and approaches that have been scientifically demonstrated to raise reading achievement scores. Unfortunately, early-reading achievement has not meaningfully improved in step, suggesting that sheer funding alone is not the root of the problem, a position long held by economists of education (Hanushek, 1997, 2003; Hoxby, 2003; Minter-Hoxby, 1996). Accountability and expectations of student achievement have also risen in connection with funding, but at the present point early-reading achievement has not (NCEE, 2008). Although an argument could always be made that funding and accountability have not been pursued vigorously enough, in line with the maxim that "research should precede policy" (Blackman, 1996), prudence suggests that it is time to examine research-based alternatives for raising early-reading achievement (Cohen, Raudenbush, & Ball, 2003).

In turning to the research literature, it is clear that early-reading skill acquisition is a complex, multiyear process that occurs across a variety of contexts and is therefore potentially affected by a wide array of factors. Factors that are demonstrably associated with student achievement are staggeringly diverse, ranging from home environment to school environment to instructional methods to individual student characteristics (Walberg, 2003). Acknowledging that psychology counts predicting and controlling behavior among its goals (Stanovich, 2003), what is needed is a scientifically-based research method that can help determine which variables are most closely aligned with early-reading skill acquisition.

Dose-Response Methodology

One such method is *dose-response methodology*. Although it has not been broadly implemented in educational research, dose-response methodology holds great promise because it allows for the exploration and characterization of causal relationships between specific inputs and their associated outputs, specifically, between environmental variables and an individual's behavior (Tallarida & Jacob, 1979). In order to understand dose-response methodology, it is important to first understand what is meant by a dose.

The Concept of Dose

Alle Ding' sind Gift und nichts ohn' Gift; allein die Dosis macht, daβ ein Ding kein Gift ist
– Paracelsus

Paracelsus is commonly known as the father of toxicology (Pachter, 1951) due to his insight (loosely translated from the quotation above) that "it is the dose that makes the poison". This insight explodes the common approach of strictly classifying a substance as either "salubrious," "benign," or "poisonous." It has long been known that mercury is poisonous to humans. What Paracelsus demonstrated was its medicinal properties that were only apparent at very low dosages (Grell, 1998). Thus what was earlier thought of as a one-dimensional picture (either a compound was harmful or not) was shown to be multidimensional.

When applied to education, this insight continues to have relevance. It is a truism that most variables (e.g., funding and accountability) are not black or white, but rather are best represented as spectral. While there are exceptions (e.g., gender), generally it is important to include in any analysis not just whether a given variable of interest is present but also to what degree, or at what dosage level, it is present.

The Promise of Dose-Response Methodology

In connection with education reform and early reading, it is important to note again that the levels of funding and accountability have increased over time. Because there has been no concomitant change in reading achievement, it is reasonable to assume that reading achievement is dose-invariant with respect to these two variables, at least across the dosages that have been applied over the past 40 years. That is not to deny that, if more dollars and oversight could be provided, achievement might change. But if the dosage levels cannot be reasonably (i.e., cost-effectively and practically) changed, then for all intents and purposes the variables in question could be dismissed as having no effect on the measured outcome.

The Dose-Response Approach

The logic that undergirds the dose-response methodology is this: By altering the dosage level of a drug and measuring the impact of that change on a given variable, a researcher can begin to characterize a relationship between the drug and the outcome variable. When undertaken in a formal experimental way, dose-response methodology can help determine whether a variable has a direct, causal relationship with another variable. Often such exploration begins at very modest dosage levels, perhaps even at a level at which it is assumed that there will be no change in observable behavior. This establishes baseline behavior and provides context for the behavioral readings at higher dosages.

Dosage level is then systematically varied, and the resultant changes in behavior are recorded for each level. The level is increased until the point where further increases do not result in corresponding changes in observable behavior. At that point the experimental portion of the investigation is complete. Other factors (cost, practicality of administering additional dose levels, etc.) might also establish both the beginning and ending dosage levels used in the investigation.

To the extent that the only experimental difference between the various dosages is the amount of the doses themselves, resultant changes in behavior across these dosages can be validly ascribed to this difference. To the extent that behavioral changes can be reliably induced by changes in the dosage level, a dose-response relationship is said to have been established.

Dose-response relationships can themselves be compared against one another in order to establish the relative efficacy of different drugs vis-a-vis the same behavior, as is the case in behavioral pharmacology (Poling & Byrne, 2000). Indeed, the Russian pharmacologist Zavadaskii arguably established modern behavioral pharmacology by doing just that. While working in Pavlov's laboratory, he studied "the effects of alcohol, morphine, cocaine, and caffeine on the conditioned salivary reflex" in dogs (Laties, 1979). Such comparisons are greatly facilitated by the use of dose-response curves.

Dose-Response Curves

> I dare affirm a man shall more profit in one week by figures and charts well and perfectly made than he shall by the only reading or hearing the rules of that science by the space of half a year at the least
> – Thomas Elyot

In examining data generated by dose-response methodology, it is helpful to construct a *dose-response curve*. Dose-response curves allow visual inspection of the data, which is itself a powerful technique for exploring and organizing quantitative data (Tufte, 1983; Singer & Willet, 2003).

In constructing a dose-response curve, the independent variable (dose) is plotted on the x-axis and the dependent variable (response) on the y-axis. Dose-response curves are generally monotonically increasing in a sigmoidal shape (see Figure 2) and have a number of important characteristics. The first is the *threshold dose*, which indicates the minimal dosage level that elicits a change in the behavioral response. Levels that are less than the threshold dose (i.e., those that lie to the left on the curve) are deemed to have no impact on the behavior in question and therefore are ineffective. A second characteristic is the *point of maximal response*, also known as the shoulder of the curve. Dosage levels that are greater than this amount (i.e., those that lie to the right on the curve) do not have any additional effect on the dependent variable. The section of the curve that lies between the threshold and the maximal dose levels indicates the drug's relative *potency*, with a narrower section indicating greater potency (Tallarida & Jacob, 1979).

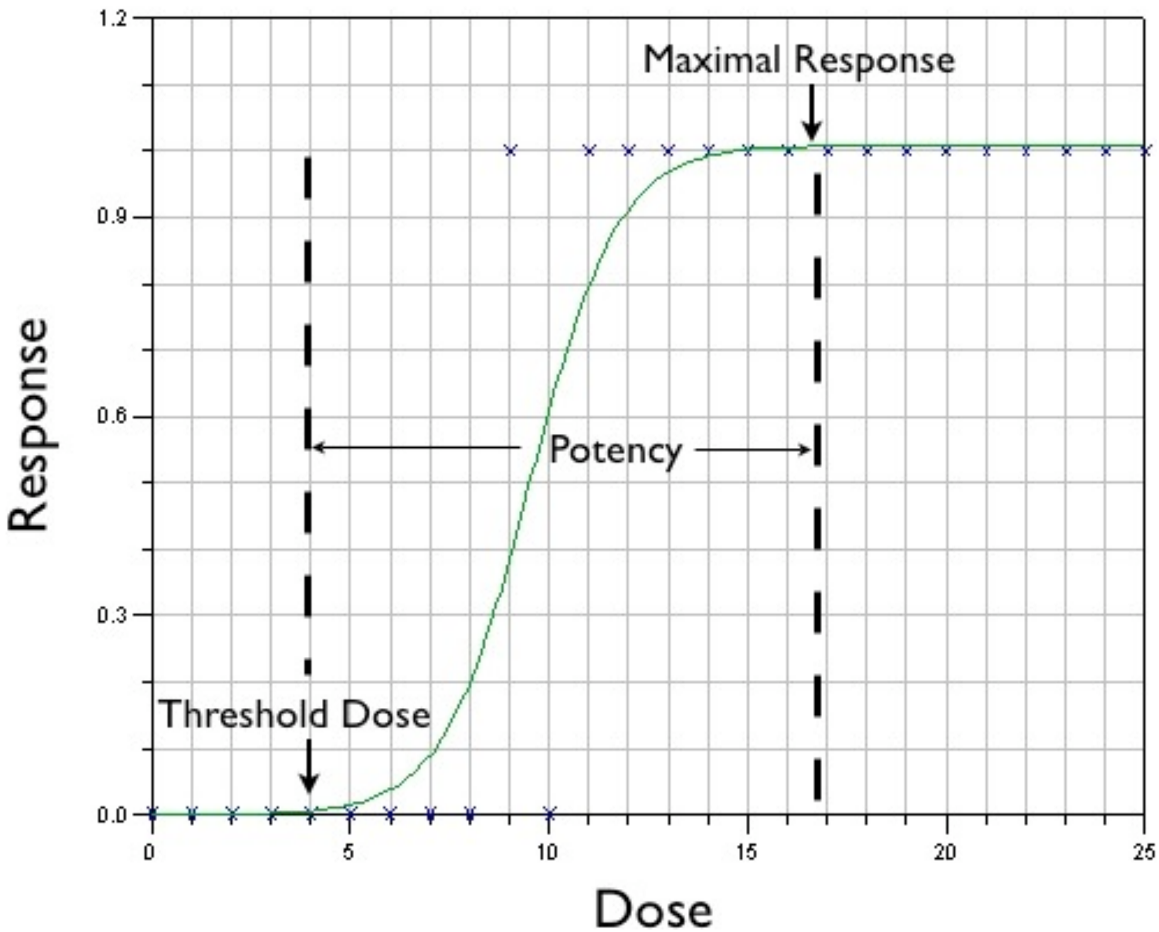


Figure 2.
A Typical Dose-Response Curve
Note. From DTREG predictive modeling software (http://www.dtreg.com/logistic.htm) by Phillip H. Sherrod. © 2008 Phillip H. Sherrod. Reprinted and adapted with permission.

Dose-response curves can themselves be compared by plotting multiple curves on the same set of axes (see Figure 3). Such analyses are a powerful tool for contrasting the relative merits of various drugs (each dose-response curve represents a different drug used with the same individual or population), various conditions (where each dose-response curve represents a particular drug's effect under a different condition), or various individuals or populations (where each dose-response curve represents the effect of a particular drug for a different individual or population; see Figure 3).

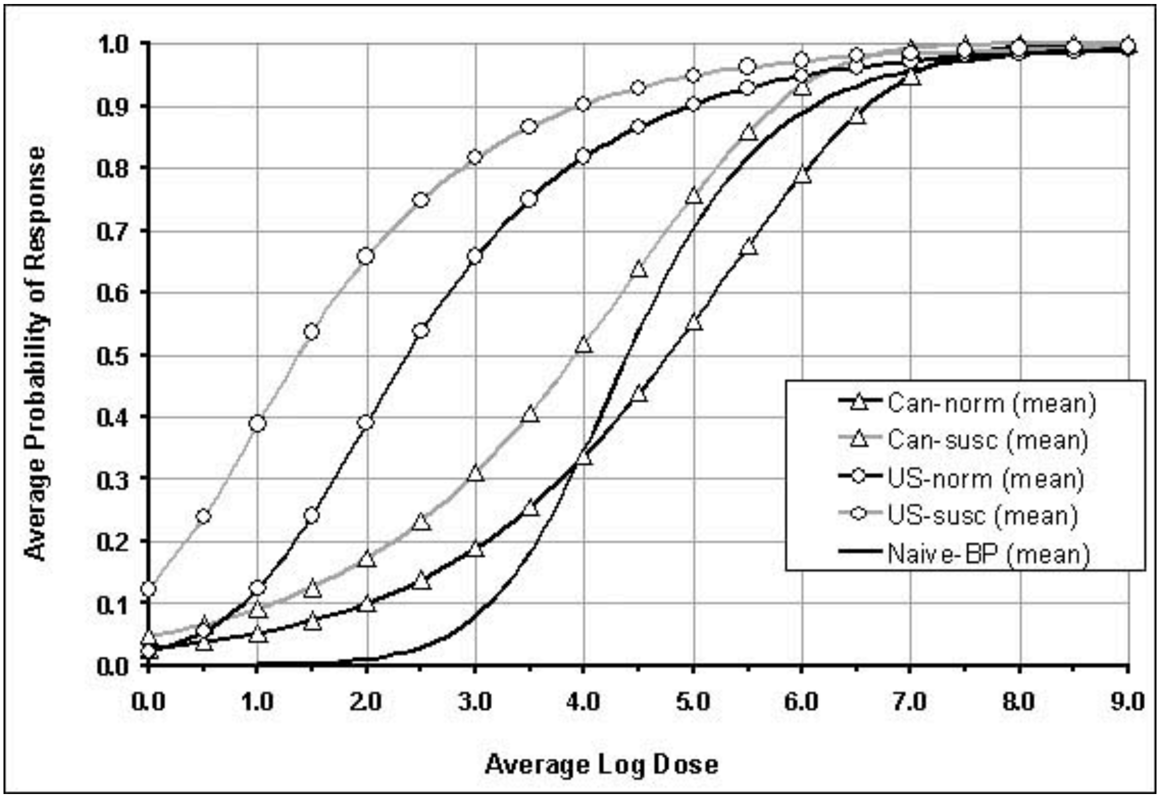


Figure 3.
Comparison of Multiple Dose-Response Curves.
Note. From Food and Agriculture Organization of the United Nations and World Health Organization, Risk Assessments of Salmonella in Eggs and Broiler Chickens - 2: Web address: http://www.fao.org/docrep/005/y4392e/y4392e06.jpg; 2002. ©2002, FAO. Reprinted with permission.

Examining the State of Early Reading Using Dose-Response Methodology

Although its roots lie in pharmacology, dose-response methodology is not confined to situations wherein the dosage levels are strictly applied to drugs. Dose-response methodology may be applied whenever a change in a manipulable variable's magnitude results in a measurable outcome. This appears to be the case in American education: over the past 40 years there have been government-mandated manipulations of two distinct systemic variables (i.e., funding and accountability), and there have been repeated measures of early-reading achievement over that same time period (i.e., 4th-grade main NAEP scores).

As already noted, early-reading achievement has not changed substantially despite the dramatic increases in both accountability and school funding. Interpreting this finding in terms of dose-response methodology, it is evident that, despite significant changes in dosage levels, there has not been a corresponding difference in output. This suggests that funding and accountability are either unrelated to early-reading achievement or that the dosage levels across which they have been measured reside either to the left of the threshold (in which case more of either or both would be required to affect achievement) or to the right of the shoulder (indicating that further increases in these variables would not have an impact).

What appears to be missing is a variable that has a direct, causal relationship to early-reading achievement; in other words, a variable that would exhibit a dose-response relationship with early-reading achievement. If such a variable could be isolated, and if it were fiscally-targetable, then it might provide hope for improving the current state of early reading in U.S. public schools.

The Promise of Time as an Input Variable

In beginning the search for a variable that is directly associated with early reading achievement, it is helpful to start at as plainly as possible. What variable might be logically related to early reading? Or, at an even more basic level, what variable would be related to learning in general?

It is a truism that if you want to get better at something you should spend more time doing it – after all, “practice makes perfect”. Indeed, this has been shown to be the case for a wide variety of skills (Levitt & Dubner, 2005). It therefore seems reasonable to think that, if you want to get better at reading, you should spend more time reading. Indeed, out of this simple idea came an entirely new approach to education reform that argued not just that time spent learning was important, but that it, in and of itself, is the key determinant of whether learning occurs.

Carroll’s Model of School Reform

The value of a school lies in its ability to provide high-quality instruction to each of its students (Corcoran & Goetz, 1995). In his model of school reform, Carroll (1963) noted that this instruction itself could be parsed into individual *learning tasks*. “The learner’s task of going from ignorance of some specified fact or concept to knowledge or understanding of it, or of proceeding from incapability of performing some specified act to capability of performing it, is a learning task” (p. 723). Thus, early-reading acquisition and its constituent skills (e.g., phonemic awareness, decoding) can be conceptualized as a series of learning tasks that each student needs to master.

What does all of this have to do with time? Carroll’s core insight was that the mastery of these learning tasks was related to the amount of time a student spent mastering them: the more time a student spends mastering a particular learning task, the greater his mastery. Carroll noted, however, that each student would potentially need a different amount of time to learn a given learning task to criterion. Mastery therefore becomes a function of the time spent vs. the time needed. This relation can be expressed mathematically as follows:

$$\text{Degree of Learning} = f\left[\frac{\text{Time Spent Learning}}{\text{Time Needed to Learn}}\right]$$

Hence, the degree to which a student masters a given learning task (Degree of Learning) is assumed to be a function of the amount of time the student is engaged in learning (Time Spent Learning) relative to how much time the student needs to master that particular learning task (Time Needed to Learn). While a number of factors may be thought to be missing from this straightforward equation (e.g., the quality of the instruction, the aptitude of the learner, etc.), Carroll asserted that any additional factor ultimately will alter either Time Spent Learning or Time Needed to Learn. Or, in his own words: “Briefly, our model says that the learner will succeed in learning a given task to the extent that he spends the amount of time that he needs to learn the task” (p. 725).

The complex process of learning thus reduces to this deceptively simple function, which, as Berliner (1990) points out, has the ingenious quality of being based on the metric of time. Carroll noted that this was purposeful: “It will be seen that as many as possible of the basic concepts in the model are defined so that they can be measured in terms of time in order to capitalize on the advantages of a scale with a meaningful zero point and equal units of measurement” (p. 723). An additional advantage to such an approach is that it greatly simplifies the number of concepts that need to be addressed: in the end it is only the time that matters, or in Carroll’s words: “[o]ne of the bolder hypotheses implicit in the model is that the degree of learning, other things being equal, is a simple function of the amount of time during which the pupil engages actively in learning" (p. 732).

Time Spent Learning

A distinct strength of Carroll's equation is that it is testable (Berliner, 1990; Gettinger, 1984, 1985; Millman, Bieger, Klag, & Pine, 1983). Researchers have subsequently found that *time spent learning* (TSL), defined as the cumulative time students spend in learning-related activities (e.g., seatwork and homework), is directly related to educational achievement (Abadzi, 2004; Brown & Saks, 1986; Gettinger, 1984; Parkerson, Lomax, Schiller, & Walberg, 1984; Scheerens & Bosker, 1997). The link between this measure, also known as *opportunity to learn* (Berliner, 1990), and learner outcomes is strong enough that insufficient opportunity to learn became one of the cornerstones for a successful class action suit by parents against the state of California (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003). As summed up by Walberg (2003): "The positive effect of time is perhaps most consistent of all causes of learning" (p. 7). It is therefore not surprising that researchers have repeatedly called for increasing TSL in an effort to boost student achievement (Abadzi, 2004; Berliner, 1990; Cohen, Raudenbusch, & Ball, 2003; Paik, Wang, & Walberg, 2002; Rowan, Camburn, & Correnti, 2004).

From a theoretical standpoint, TSL is necessarily correlated with measures of learning. For example, TSL has been repeatedly demonstrated to be associated with eventual learning outcomes. Therefore, it appears to be a promising dosage variable for early-reading achievement (Adams, 1990; Gest & Gest, 2005).

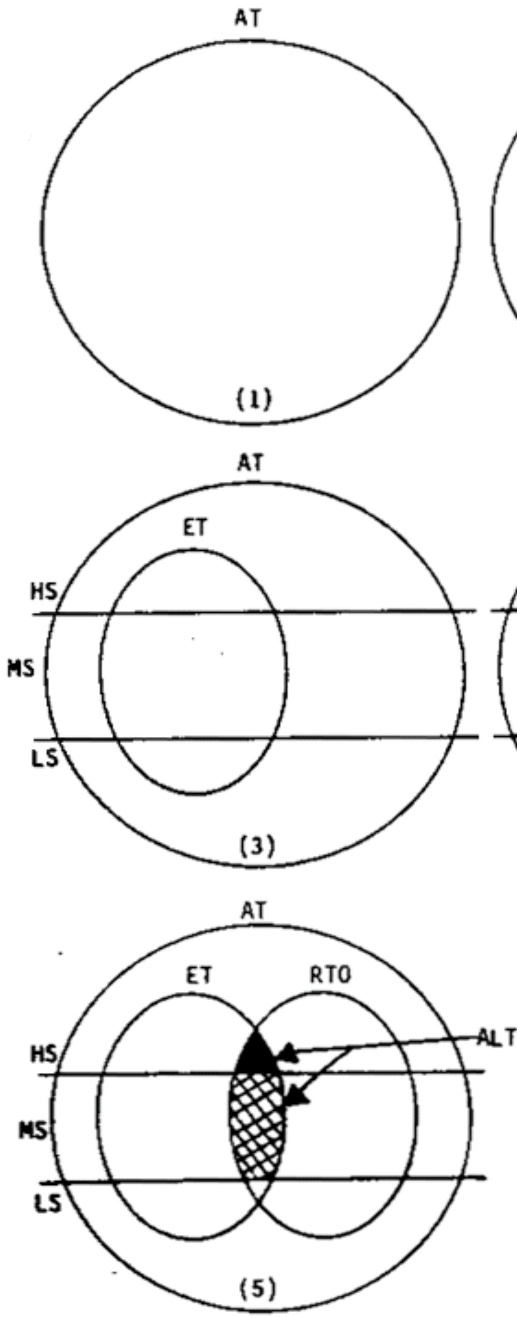
Limitations of Time Spent Learning

Although state and federal legislation mandates the amount of time students spend at public schools, not all of the time spent there is directly connected to learning (Carroll, 1963; Baker, Fabrega, Galindo, & Mishook, 2004). Teachers and students engage in a variety of activities throughout the school day, some of them menial (from an academic standpoint), others quite meaningful. The flawed assumption that the effect of one hour of time at school is equivalent to the effect of any other hour there raises the question of how time spent instructing and TSL should be measured (Gettinger, 1985). As Berliner (1990) noted, "instructional time should be considered a superordinate concept, and in this way is no different from the concept of 'mammal,' which encompasses organisms as disparate as elephants, mice, platypuses, bats and us, homo sapiens" (p. 4). It seems reasonable to assume that TSL could similarly be thought about as a superordinate concept. The question thus becomes one of identifying the appropriate subordinate elements in the school day that have a direct relationship with early reading achievement.

Academic Learning Time

In his summary of research on instructional time, Berliner (1990) identified a half-dozen potential definitions of instructional time. Of particular relevance to student achievement is his definition of *academic learning time* (ALT), a concept that initially came from the Beginning Teacher Evaluation Study (Denham & Lieberman, 1980). Berliner defines ALT as "that part of allocated time in a subject-matter area . . . in which a student is engaged successfully in the activities or with the materials to which he or she is exposed, and in which those activities and materials are related to educational outcomes that are valued" (p. 5). A visual demonstration of ALT is provided in Figure 4, where it can be seen that ALT is a the subset of the overall allocated time wherein a student is engaged in tasks that are related to outcomes; for young children, there is the additional constraint that this engagement is highly successful (depicted by the small black triangle).

Under this definition, ALT should therefore be directly applicable to student achievement. Instead of just being a measure of time on task (or TSL), ALT becomes the equivalent of "time on [the] right tasks" (Berliner, 1990, p. 18). Moreover, harking back to Carroll's original theory of the learning process, time on task is all that matters. Consequently Berliner can claim that "[u]nless ALT is affected in some way, there will be no changes in student achievement at all" (p. 22). Based on this assertion, ALT might ultimately be the *only* salient input variable. In order to utilize ALT as a potential dosage variable, however, it first needs to be operationalized.



<u>Notes</u>

AT = Allocated time
ET = Engaged time, always a subset of AT
HS = High success experience
MS = Medium success experience
LS = Low success experience
RTO = Time spent on materials or in activities related to outcomes
ALT = Academic Learning Time: That part of (1) allocated time in which students (2) are engaged, (3) are succeeding at what they are doing, and (4) what they are doing is related to outcomes that are desirable. (5) ALT for young children (5th grade and below) is the small dark section; ALT for older students is both the dark and the cross-hatched sections.

Figure 4.
A Visual Representation of ALT
Note: From Berliner (1990), p. 19. Copyright 1990 by David C. Berliner. Reprinted with permission.

Operationalizing ALT

"I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of *science*." – Lord Kelvin

ALT differs from conventional time on task or TSL in a number of important ways. According to Berliner's original definition, in order for a student's time to be classified as ALT four distinct criteria be met:

(1) The time must be instructional in nature.

(2) The student must be engaged.

(3) The instruction must be appropriate for the student.

(4) The instructional content must be aligned directly with desired outcomes (i.e., student-achievement measures).

It is not sufficient for just one of these criteria to be met – if any of the four criteria goes unmet, then the time in question cannot be labeled as ALT.

These criteria are consistent with the claim of a dose-response relationship between ALT and student achievement: as ALT increases, student achievement should improve. Based on these criteria, ALT holds real promise as a dosage variable.

The Challenges of Quantifying ALT

Having established the desirability of ALT as a dosage variable, the next question is how to reliably and validly measure it. Any quantification of ALT must address the four criteria mentioned above, each of which presents its own difficulties.

Instructional Time

The first criterion is that ALT needs to be instructional. In order to establish how much time is spent on instructional activities in a classroom, researchers have, for cost and logistical reasons, relied either on teacher self-reports or on direct observation of the teachers and students in the classroom (Bromme & Hömberg, 1990; Rowan, Camburn, & Correnti, 2004). In both of these studies, teachers' estimates of instructional time in the classroom were found to be reasonably accurate using 3rd-party observation, indicating that teachers have a good grasp of instruction at the whole-class level.

However, this accuracy breaks down when teachers are asked to recall instructional time for individual students. After reviewing the research in this area and combining it with their own empirical findings, Bromme & Hömberg (1990) theorized that teachers' poor performance (e.g., teachers noticing only 3% of individual students' successes and failures) is not due to teachers' failure to accurately evaluate students' understanding, but rather that "teachers had observed the class as a unit. They perceived the learning progress of the 'collective student,' that is, an abstract subject composed of the various students in the lesson dialogue" (pp. 183-84). Teachers appear to be focused on the overall instructional flow in the classroom and notice individual students only as they specifically inform and direct that flow.

Noting that teachers are not a valid source of instructional time at the student level poses a significant challenge for ALT research. Carroll's (1963) equation was not envisioned for a "collective student" that only varied between classes. Rather it was built upon the supposition that the Time Needed to Learn would vary between individual students within a given classroom. Thus any attempt to generalize ALT across multiple students fundamentally undermines the underpinnings of the construct itself. Not being able to leverage the teacher – the person who is most familiar with the instruction and with individual students – places severe constraints on ALT research in traditional classrooms and casts doubt on whether it can be accomplished in a reliable, valid, and cost-effective way.

Engaged Time

The second criterion is that the student needs to be engaged. In Carroll's (1963) words:

> "spending time" means actually spending time on the act of learning. "Time" is therefore not "elapsed time" but the time during which the person is oriented to the learning task and actively engaged in learning. In common parlance, it is the time during which he is "paying attention" and "trying to learn." (p. 725)

On the surface this criterion seems straightforward, but, upon reflection, less so. How can a researcher objectively measure a student's level of engagement? Should a researcher suppose that a student is "actively engaged" if the student is watching the teacher or if the student appears to be writing something? Drawing? Has her chin in her hand and appears to be thinking?

A recent major review of the research on student engagement confirms the relevance of such questions, noting that the very definition of "engagement" is problematic and recommending that it be divided into three separate constructs (Fredericks, Blumenfeld, & Paris, 2004). Developing an operational definition of engagement in a traditional classroom therefore poses a significant challenge that has yet to be solved by the research community (Ball & Rowan, 2004). Until such challenges are met, validly quantifying the amount of ALT in a traditional classroom appears to be an unattainable goal.

Instructional Difficulty

The third criterion is that the instruction must fall within an acceptable band of difficulty for the student, a concept akin to Vygotsky's (1978) *zone of proximal development*. In essence, this requires the researcher to know the range of capabilities of each student and to determine whether the instruction falls within this individualized range. Time spent on topics of an appropriate difficulty level would therefore be admissible for inclusion towards an overall calculation of ALT, while time spent on topics that are too easy or too difficult would not qualify and therefore would be excluded from ALT calculations.

This raises the practical question of how a researcher would be able to ascertain whether a given learning objective is at an appropriate instructional difficulty level. In practice, this is generally assayed by requiring a student to demonstrate acceptable mastery of the learning task as defined by an appropriate assessment. This demonstration on the part of the student back to the teacher is a form of *feedback* (Waldrop, 2001), which allows the teacher to discover whether a student has learned what was intended and is ready to move on or whether additional time and practice might be called for.

Unfortunately, teachers do not have the capacity to garner and respond to feedback at the individual level. Therefore they spend the bulk of their time teaching to the "collective student" (Bromme & Hömberg, 1990; Conant, 1973). This is due in part to the overwhelming amount of time and energy individual feedback requires. For example, acceptable mastery has been defined for younger students as an 80% success rate (Gettinger & Seibert, 2002). Under this definition of mastery, multiple responses would be required to determine a success rate, with confidence in the measurement increasing as the number of learning trials increases (Scheuren, 2005). Assuming a class of 30 students and 5 responses per student (the minimum number required to measure an 80% success rate), a teacher would need to solicit, receive, and evaluate 150 responses for each learning task, which is clearly an impractical expectation.

Assessment Alignment

The fourth and final criterion specifies the need for direct alignment between instructional content and outcome assessment. In practice, a researcher would need to regularly ensure that any topic taught by the teacher eventually will be assessed in a diagnostic way. Time spent on topics that are not ultimately assessed would not be counted as part of an individual student's ALT.

While such an arrangement does not on the surface appear to be problematic, it has disturbing implications for ALT on a broader scale. ALT research would necessarily vary, and potentially vary quite radically, depending on the standards for instruction and the outcome measures that were used. In order to establish ALT research in a given curricular area, it would be prudent to establish a uniform context for instruction and assessment. Such a context would require a common lexicon for instructional content, instructional methods, and outcome measures. To date, however, there is no common lexicon for teachers, much less for educational researchers (Anderson, 2002). Although

NAEP is a national assessment, it is not particularly useful for individual schools or districts as "NAEP does not provide scores for individual students or schools" (NAEP, 2008). Thus, American public schools currently have no single, valid metric by which to gauge individual ALT. In addition, although assessments can be standardized across various settings, the validity of a given assessment in terms of ALT (i.e., its alignment with what occurs in individual classrooms and at the individual-student level within those classrooms) is not uniform. This is due to the ethic of atomized teaching that is strongly exhibited in American public schools (Elmore, 2002). In sum, at present there is no uniform way to reliably quantify ALT from one instructional setting to the next.

Limitations of ALT Research

Clearly these challenges are significant and bring into question the practicality of using ALT as a metric in research settings. It is therefore not surprising that, despite its promise as a research variable, there have been relatively few studies of ALT to this point. The following recent examples illustrate the very real problems associated with attempting to measure ALT in traditional classroom settings.

Instructional Logs

In their survey of research on measuring instruction, Rowan, Camburn, and Correnti (2004) define *enacted curriculum* as "the amount of instructional time devoted to teaching various strands and/or topics in the school curriculum" (p. 76). Enacted curriculum is therefore analogous to allocated time in Berliner's model.

While Rowan et al. note the importance of measuring the enacted curriculum, they bemoan the fact that "the procedures used [in educational research] to measure the enacted curriculum remain much as they were 2 decades ago" (p. 76), namely, annual teacher questionnaires combined with small numbers of qualitative classroom observations, both of which are methodologically problematic. These approaches also raise theoretical questions in that they focus on the teacher or classroom as the unit of measurement as opposed to the student. This is particularly alarming because research has demonstrated that students can receive substantially different instructional opportunities within the same classroom (Rowan, Correnti, & Miller, 2002).

In order to address these shortcomings, Rowan et al. (2004) propose using instructional logs or time diaries. These detailed forms (see Appendix C for the ones used by the researchers) are filled out just after instruction is given, thereby ensuring that the information is as accurate as possible. Forms are filled out regularly to ensure that they are reliable estimates, a distinct advantage over 3rd-party observations in the classroom, which are often, for cost reasons, few and far between. Additionally, logs allow for estimates of the enacted curriculum to be gathered at the student level instead of just at the classroom level, although, to lessen the burden on the teacher, the logs are filled out only for a single, randomly-selected student on any given day.

While there is much to praise about the notion of instructional logs, they do not provide a satisfactory measure of ALT. Of the four components that need to be included, logs as described and implemented by Rowan et al. mostly address the first and fourth components, namely, instructional time and curricular alignment. Although enacted curriculum is defined in terms of time and would therefore seem to be compatible with the more granular notion of ALT, the instructional log as constructed (see Appendix C) only asks about time for the whole day and does not attempt to segment it among various learning tasks or objectives. This indicates that the research question being asked is more about coverage of curriculum than ALT.

As enacted, instructional logs do not speak to the portion of time during the instruction that the student was engaged, nor to whether the instruction was appropriate for the individual student (i.e., whether the student was successful). Due to the amount of observational time and effort that a teacher would need to spend in order to meet these additional criteria, it is difficult to see how instructional logs in and of themselves could measure ALT validly. Indeed, one of the primary issues with having a teacher measure ALT in a traditional classroom is that it effectively requires a teacher to spend more time observing and recording than teaching.

Increasing ALT

In their chapter on ALT, Gettinger and Seibert (2002) focus on helping teachers organize and promote ALT in their classrooms. The authors' description of ALT is more in alignment with Berliner's original definition than Rowan et al.'s (2004) enacted curriculum and consequently has distinct advantages. Moreover, the authors prioritize and rank the relative importance of the various components of ALT:

> Although the amount of time teachers allocate (allocated time) and use for instruction (instructional time), as well as the proportion of time during which students are engaged (engagement rate), are all positively correlated with learning, it is the proportion of engaged time that is productive, active, and successful that relates most strongly to achievement. (p. 3)

Indeed, Gettinger and Seibert refine and extend Berliner's model by splitting engagement into procedural engagement and substantive engagement and by defining "instructionally appropriate" for young children as an 80% success rate.

Unfortunately, while they address ways of increasing ALT in the classroom, none of the techniques the authors describe refers to quantifying the amount of ALT or measuring it in a way other than in relation to the amount of instruction that is already occurring in a particular classroom. Thus, while this information is helpful to a teacher for maximizing ALT in his or her own classroom, it ultimately falls short of helping to quantify ALT in a reliable and valid way that could allow for comparisons across multiple environments.

Summary

Assuming a traditional group-based instructional setting, quantifying ALT in a valid, reliable, and cost-effective manner appears to be an unrealistic goal. The nexus of the problem seems to reside in the nature of the instruction: when a teacher is instructing, it is not possible to have that teacher simultaneously observe and record what is happening for each child in the classroom without seriously disrupting the instructional process itself. The traditional alternative, which is to place one or more 3rd-party observers in the classroom, is not cost-effective nor is it scalable.

ALT research is further hampered by a lack of assessment standards. Allowing assessments to vary from one instructional setting to the next opens the door to the situation in which the exact same learning experience for a specific student turns out to have a substantially different value in terms of ALT. For ALT research to be useful on a larger scale would require more consistency in assessment so that valid comparisons can be made between instructional settings.

The Promise of a Computer-Based Approach to Quantifying ALT

Computer-based instruction (CBI) has long provided a promising avenue for exploring a relationship between ALT and achievement. As noted by Atkinson and Hansen (1966) over 40 years ago, instructional systems "make it possible to obtain rigorous behavioral measures", with the result that "subject-matter learning can be studied under conditions of greater control and with more precision in response-recording than has ever been possible even in the psychologist's laboratory" (p. 8). The relevance of this investigative advance with regards to ALT is best understood by returning to the four criteria for quantifying ALT and examining them in the context of CBI systems.

Instructional Time

Computers can vary the instructional time they provide with precision. Instructional programs can be set in advance to run for specified periods of time or for a variable amount of time based on a variety of student-specific variables (Atkinson, 1974). Such control makes CBI an ideal context for dose-response methodology.

Computers are also capable of handling the amount of information necessary for analysis of achievement at the individual level, as opposed to just the group or classroom level. Computers can record each student's responses in real time while simultaneously presenting instructional material – in essence, they can multitask without the crippling penalties that people incur when trying to do the same (Dzubak, 2008). In addition, the instructional content delivered by the computer is identical every time, so there is no question as to each student's opportunity to learn. Based on these strengths, it is reasonable to assume that computer-based ALT can be easily manipulated, making it ideal for dose-response methodology.

While the time using the CBI system should not be interpreted as comprising the total ALT a student receives (i.e., the instruction delivered by the teacher and by other means would not be included in this quantification), the power of this approach is that, at the very least, this particular component of overall ALT can be accurately measured, as has been demonstrated with deaf children learning early skills in mathematics (Suppes, Fletcher, & Zanotti, 1975; 1976). Such an approach dramatically improves on the traditional classroom environment where instructional time cannot be effectively measured at the student level.

Engaged Time

Contemporary CBI at its best is interactive and engaging, meaning that it both requires and inspires the student's frequent attention (Dickey, 2005; Spence & Usher, 2007). Although time spent using the software could not be directly equated with engaged time, such a measure may be a more reasonable approximation than today's "time-in-class" approach (Zhang, Almeroth, & Bulger, 2005). With software that is highly interactive and engaging, it seems reasonable that TSL would asymptotically approach engaged time and therefore qualify for inclusion in the calculation of ALT.

Instructional Difficulty

Computers can instruct in an adaptive manner, meaning the computer selects which instructional element to present to a student based on the student's past performance as well as on the various difficulty ratings of the learning elements themselves (Macken, Suppes, & Zanotti, 1980), much as item response theory (IRT) does for computer-adaptive testing in the area of assessment (Embretson & Reise, 2000). While a teacher in a traditional classroom often provides the same instruction to the entire class (Bromme & Hömberg, 1990), a computer has the capacity to individualize instruction for each student and can do so without disrupting the learning of other students. This approach promises to provide what each student needs in the moment – a goal long considered the holy grail of instruction (Bloom, 1984; Fletcher, 2003).

Harking back to Carroll's (1963) original model, adaptive instruction also helps students receive the amount of instructional time they need in order to be successful. In her initial study on time spent learning and time needed to learn, Gettinger (1984) found that "[t]ime needed is an equally important factor in determining the value of time spent in learning and its relationship to achievement" (p. 626). Expanding on that initial study, Gettinger (1985) subsequently found that 4th- and 5th-grade students often needed more instructional trials to achieve mastery in reading and spelling than they believed they needed, spending on average only 68% of the time needed. When these students were given one to two fewer learning trials than they needed, achievement and retention decreased 11% and 16% respectively. Gettinger's conclusion was that: "The disparity between potential and actual achievement observed on this experimental task may be magnified over days, months, or years of schooling if teaching time allocations and amount of academic engaged time do not approximate time needed" (p. 10). By ensuring that students do not move on before they are ready, computer-based adaptive instruction should consequently result in increased student achievement.

Assessment Alignment

The fourth criterion is addressed in multiple ways by shifting not just the instruction but also the assessment to a computer-based approach. Computer-based assessment eliminates traditional issues of scalability by allowing for rapid, accurate, and inexpensive duplication. At the same time it eliminates disparities in presentation: there is only one assessor. While this does not necessarily eliminate bias, it does make it a constant.

Regarding the alignment of instructional content to outcome measures, a computer-based approach does not guarantee alignment, but given judicious design during the construction of both the instructional and assessment components, alignment can be built into the system. In many ways, this is how computer-adaptive instruction is accomplished – students are regularly assessed to ensure that the learning tasks they spend time working on correspond to the proper difficulty level. When all of the learning tasks are directly related to the eventual outcome measure and are properly situated in relation to one another, achievement on an outcome measure can be inferred from a student's position and trajectory in the overall curricular spectrum (Suppes & Zanotti, 1996).

Summary

CBI systems appear to provide a number of solutions for the challenges that confront researchers in a more traditional classroom setting. These solutions require that the instruction be delivered as well as be quantified by the computer. The precision and control afforded by CBI approaches make them ideal for use with dose-response methodologies, especially when the assessment itself is similarly computer-based.

Limitations of a Computer-Based Approach to Quantifying ALT

While researchers initially viewed CBI as an opportunity to completely replace the traditional classroom environment (McDonald, Yanchar, & Osguthorpe, 2005; Atkinson & Hansen, 1966), it quickly became apparent that "some aspects of instruction could be done very effectively using a computer, but that there were other tasks for which the computer did not have any advantages and possibly had some disadvantages over classroom teaching" (Atkinson, 1974, p. 169). Recognizing this limitation, CBI was subsequently relabeled "computer-assisted instruction" (CAI), or instruction "that supplements classroom teaching and concentrates on those tasks in which individualization is critically important" (p. 169). In this same spirit, it seems prudent to reflect on the potential limitations of a computer-based approach to quantifying ALT in early reading.

Qualitative Differences in Instruction

It is possible that there are meaningful qualitative differences between how computers and teachers instruct children to read. In other words, ALT provided by a teacher (whether it is difficult to measure or not) might not be equivalent to ALT provided by a computer. The potential solution outlined above does not solve the problem of measuring ALT as delivered by a teacher in a traditional classroom. Indeed, much of the solution hinges on changing the instructional paradigm so that instruction is provided by the computer and not by a teacher.

These potential instructional differences also extend to the social setting of the instruction. Students in a traditional classroom learn primarily in groups, but computers in their current instantiation generally instruct each child individually. Thus, the data generated by CBI might be more directly applicable to tutoring environments as opposed to traditional classroom settings.

For some researchers these might be untenable trade-offs as they limit investigation to instructional scenarios where it is feasible for the instruction in question to be delivered by a computer. In addition, the questions that a CBI approach answers are not necessarily directly relevant to a traditional classroom setting as they employ an approach (namely, individualized instruction) that generally is not feasible to implement in a traditional American classroom.

Pace of Change and Versioning

Computers deliver the same instructional program to every child, every time, but these programs, along with the computers themselves, are in a state of constant flux. What this often results in is a significant gap between what is currently available and what researchers have evaluated. According to Fletcher (2003), "[b]y the time an evaluation study is performed, documented, and reported in a form accessible to developers and potential users, the application originally under consideration is likely to be 5 or more years old" (p. 95). While such rapid obsolescence does not invalidate experimental findings, it does present challenges to decision makers who need to act on the opportunities at hand as well as to researchers who often do their work iteratively.

These challenges are compounded by the nature and magnitude of technological change. It would be unthinkable to ask teachers to work twice as quickly, but that is the speed at which computers are regularly changing and there are reasons to believe that this pace of change will continue to accelerate (Kurzweil, 2005). Similarly, asking teachers to develop entirely new capabilities is not an option, but such drastic changes are routinely expected of technology, as can be seen by the dramatic changes in cell phones, which have gone from being primarily a voice-communication device to a multimedia device that browses the web, takes pictures, plays music and videos, and allows for texting and e-mail. Thus, although it has made sense for decades to talk about what teachers can accomplish in the classroom, it does not make sense to engage in the same conversations about software or computers because they are not similarly constrained.

What this rapid pace of change means for research in ALT and student achievement in early reading is not entirely clear. Tentatively, it appears that research done in this area could be limited in terms of external validity due to a lack of strict replicability. This lack ostensibly would stem from technological obsolescence (i.e., the software might not run on newer hardware) as well as from a lack of relevance (i.e., there is now a newer version and to continue testing the older version seems beside the point).

One possible way to ameliorate these difficulties would be to focus on a *class* of applications as opposed to a *specific* application. Thus, although "useful information on a specific application is frequently unavailable until the state of the art passes it by", "the principles underlying the design of the application and their success may well be of continuing interest to designers and potential users" (Fletcher, 2003, p. 95). While such an approach limits the specificity of the investigation, it might eventually prove a necessary strategy for dealing with the realities of a world where the current pace of change will change it more in the next 100 years than it changed over the past 20,000 (Kurzweil, 2005).

Summary

The limitations of CBI for investigating the relationship between ALT and early-reading achievement do not appear to seriously undermine the research undertaking itself, although they do place significant constraints on the generalizability of the findings. These limitations are not peculiar to the field of education; rather they are due to the idiosyncrasies of technology itself which, due to its unprecedented pace of change, rapidly undermines the usefulness of research information. One potential way to minimize such limitations is to trade off specificity of information about a particular application or system for information about a broader class of applications or systems that are built on similar principles.

Research on CBI

Having established that ALT *could* be measured using CBI systems, the question remains as to how CBI methods have performed in other research – after all, if CBI has not been shown to be effective, then measuring the amount of time spent doing it might not be valuable.

The past 40 years have demonstrated that CBI approaches are effective in a wide variety of educational settings and across a wide variety of ages:

> Overall, a rule of "thirds" emerges from CBI assessments. Findings suggest that use of interactive instructional technologies reduces the cost of instruction by

> about one-third, and it either reduces time of instruction by about one-third or it increases the amount of skills and knowledge acquired by about one-third. (Fletcher, 2005, p. 20)

While these numbers are impressive, recent research suggests that more modern iterations of CBI might well be significantly more effective than earlier prototypes, opening the door for even larger discrepancies between traditional and CBI models (Fletcher, 2003).

In terms of early-reading instruction, research has long suggested the cost-effectiveness of using CBI to augment more traditional classroom instruction.

> [T]he yearly cost is roughly $97.00 per student. If this is multiplied by three, we have a figure of $291.00, a cost that places students at grade level by the end of the third grade who would normally be over a year behind. There is no doubt that such a cost is acceptable (Atkinson, 1974, p. 177)

While this figure is undoubtedly higher in 2008 dollars, the cost-effectiveness of CBI is clearly compelling, and the investigation of a link between computer-delivered ALT and student achievement in early reading appears to be worthwhile.

The Characteristics of an Ideal Solution

In order to investigate a link between computer-delivered ALT and student achievement in early reading, one or more appropriate CBI systems would need to be identified. While an ideal CBI system might not exist at present, the literature reviewed provides a number of characteristics that such a system would exhibit. Once again, Berliner's (1990) four-faceted definition of ALT provides an operational touchstone.

Instructional Time

The first hallmark of an ideal CBI system would be that its instructional content would be aligned with what reading researchers have empirically found to be correlated with early-reading skill acquisition. That is, it would cover all of the skills of phonological awareness, phonemic awareness, decoding, etc. (Gough & Juel, 1991; Snow et al., 1998; Stanovich, 2000). Said in a different way, an ideal CBI system would satisfy the pre-conditions for funding through Reading First or Early Reading First.

In addition, an ideal CBI system would accurately and precisely quantify the amount of time that each student spends on each learning task. While time on the system might equal instructional time, that is a desirable but not a necessary characteristic – as long as a researcher can partial out the amount of time spent specifically on instructional activities, the system could be used to quantify ALT.

Engaged Time

The nature of the instruction provided through an ideal CBI system needs to be engaging and interactive. Acknowledging that the very nature of the construct of "engagement" is problematic at this point, suffice it to say that students would display the outward characteristics generally associated with meaningful engagement: eyes would remain on the screen and hands on the mouse and/or keyboard as appropriate. It is to be noted that this characteristic as operationalized might not be one that can be internally confirmed by the system itself and would therefore require an outside observer.

Instructional Difficulty

An ideal CBI solution would individualize instruction, that is, it would solicit and correctly interpret feedback for each student for each learning task and then would leverage that feedback in order to determine which learning task would be most appropriate for that student to work on next. In this way, an ideal CBI solution would always stay within a student's difficulty level. Students of differing ability and background would therefore be expected to receive differentiated instruction based upon their own unique learning history (Atkinson & Hansen, 1966).

Assessment Alignment

An ideal CBI system for early reading would include a computer-based early-reading achievement test that would align directly with the system's early-reading tasks. In order to leverage dose-response methodology this assessment would need to be able to be given on multiple occasions to the same student without decreasing its validity.

Summary

A dose-response investigation of the link between computer-delivered ALT and early-reading achievement does not require the use of an ideal CBI system; after all, educational researchers have long noted the importance of TSL, which is an imperfect measure. However, the more a given CBI system exhibits these idealized features, the more tightly its use should be aligned with early-reading achievement.

Conclusions

In terms of the original goal of improving early-reading achievement in American public schools, the literature reviewed here leads to the following conclusions:

(1) Early reading is a critical academic skill that has profound implications for a child's quality of life both inside and outside the classroom.

(2) Policy approaches over the past 40 years have focused on increasing funding and accountability, but such efforts have not resulted in raising early-reading achievement scores, suggesting that other variables need to explored.

(3) Dose-response methodology provides a means for establishing causal relations between input and output variables.

(4) Time, and more specifically ALT, appears both from a theoretical and from a research standpoint to be a variable that is causally related to early-reading achievement.

(5) Although measuring ALT has historically been problematic, a CBI approach appears to overcome many of these hurdles.

(6) CBI approaches are both effective and cost-effective for early-reading skill acquisition.

(7) A CBI approach combined with dose-response methodology could allow the exploration of a relationship between ALT and early-reading achievement.

In this way, CBI solutions may ultimately link input (dollars) to student learning (ALT) to eventual learning outcomes (student achievement).

Appendix A – Reading Milestones

Note: From Snow et al., 1998, pp. 61, 80-83. © 1998 by the National Academy of Sciences. Courtesy of the National Academies Press, Washington, D.C. Reprinted with permission.

TABLE 2-1 Developmental Accomplishments of Literacy Acquisition

Birth to Three-Year-Old Accomplishments

- Recognizes specific books by cover.
- Pretends to read books.
- Understands that books are handled in particular ways.
- Enters into a book-sharing routine with primary caregivers.
- Vocalization play in crib gives way to enjoyment of rhyming language, nonsense word play, etc.
- Labels objects in books.
- Comments on characters in books.
- Looks at picture in book and realizes it is a symbol for real object.
- Listens to stories.
- Requests/commands adult to read or write.
- May begin attending to specific print such as letters in names.
- Uses increasingly purposive scribbling.
- Occasionally seems to distinguish between drawing and writing.
- Produces some letter-like forms and scribbles with some features of English writing.

Three- and Four-Year-Old Accomplishments

- Knows that alphabet letters are a special category of visual graphics that can be individually named.
- Recognizes local environmental print.
- Knows that it is the print that is read in stories.
- Understands that different text forms are used for different functions of print (e.g., list for groceries).
- Pays attention to separable and repeating sounds in language (e.g., Peter, Peter, Pumpkin Eater, Peter Eater).
- Uses new vocabulary and grammatical constructions in own speech.
- Understands and follows oral directions.
- Is sensitive to some sequences of events in stories.
- Shows an interest in books and reading.
- When being read a story, connects information and events to life experiences.
- Questions and comments demonstrate understanding of literal meaning of story being told.
- Displays reading and writing attempts, calling attention to self: "Look at my story."
- Can identify 10 alphabet letters, especially those from own name.
- "Writes" (scribbles) message as part of playful activity.
- May begin to attend to beginning or rhyming sound in salient words.

TABLE 2-2 Accomplishments in Reading

Kindergarten Accomplishments

- Knows the parts of a book and their functions.
- Begins to track print when listening to a familiar text being read or when rereading own writing.
- "Reads" familiar texts emergently, i.e., not necessarily verbatim from the print alone.
- Recognizes and can name all uppercase and lowercase letters.
- Understands that the sequence of letters in a written word represents the sequence of sounds (phonemes) in a spoken word (alphabetic principle).
- Learns many, though not all, one-to-one letter sound correspondences.
- Recognizes some words by sight, including a few very common ones (a, the, I, my, you, is, are).
- Uses new vocabulary and grammatical constructions in own speech.
- Makes appropriate switches from oral to written language situations.
- Notices when simple sentences fail to make sense.
- Connects information and events in texts to life and life to text experiences.
- Retells, reenacts, or dramatizes stories or parts of stories.
- Listens attentively to books teacher reads to class.
- Can name some book titles and authors.
- Demonstrates familiarity with a number of types or genres of text (e.g., storybooks, expository texts, poems, newspapers, and everyday print such as signs, notices, labels).
- Correctly answers questions about stories read aloud.
- Makes predictions based on illustrations or portions of stories.
- Demonstrates understanding that spoken words consist of a sequences of phonemes.
- Given spoken sets like "dan, dan, den" can identify the first two as being the same and the third as different.
- Given spoken sets like "dak, pat, zen" can identify the first two as sharing a same sound.
- Given spoken segments can merge them into a meaningful target word.
- Given a spoken word can produce another word that rhymes with it.
- Independently writes many uppercase and lowercase letters.
- Uses phonemic awareness and letter knowledge to spell independently (invented or creative spelling).
- Writes (unconventionally) to express own meaning.
- Builds a repertoire of some conventionally spelled words.
- Shows awareness of distinction between "kid writing" and conventional orthography.
- Writes own name (first and last) and the first names of some friends or classmates.
- Can write most letters and some words when they are dictated.

TABLE 2-2 Continued

First-Grade Accomplishments

- Makes a transition from emergent to "real" reading.
- Reads aloud with accuracy and comprehension any text that is appropriately designed for the first half of grade 1.
- Accurately decodes orthographically regular, one-syllable words and nonsense words (e.g., sit, zot), using print-sound mappings to sound out unknown words.
- Uses letter-sound correspondence knowledge to sound out unknown words when reading text.
- Recognizes common, irregularly spelled words by sight (have, said, where, two).
- Has a reading vocabulary of 300 to 500 words, sight words and easily sounded out words.
- Monitors own reading and self-corrects when an incorrectly identified word does not fit with cues provided by the letters in the word or the context surrounding the word.
- Reads and comprehends both fiction and nonfiction that is appropriately designed for grade level.
- Shows evidence of expanding language repertory, including increasing appropriate use of standard more formal language registers.
- Creates own written texts for others to read.
- Notices when difficulties are encountered in understanding text.
- Reads and understands simple written instructions.
- Predicts and justifies what will happen next in stories.
- Discusses prior knowledge of topics in expository texts.
- Discusses how, why, and what-if questions in sharing nonfiction texts.
- Describes new information gained from texts in own words.
- Distinguishes whether simple sentences are incomplete or fail to make sense; notices when simple texts fail to make sense.
- Can answer simple written comprehension questions based on material read.
- Can count the number of syllables in a word.
- Can blend or segment the phonemes of most one-syllable words.
- Spells correctly three- and four-letter short vowel words.
- Composes fairly readable first drafts using appropriate parts of the writing process (some attention to planning, drafting, rereading for meaning, and some self-correction).
- Uses invented spelling/phonics-based knowledge to spell independently, when necessary.
- Shows spelling consciousness or sensitivity to conventional spelling.
- Uses basic punctuation and capitalization.
- Produces a variety of types of compositions (e.g., stories, descriptions, journal entries), showing appropriate relationships between printed text, illustrations, and other graphics.
- Engages in a variety of literary activities voluntarily (e.g., choosing books and stories to read, writing a note to a friend).

TABLE 2-2 Continued

Second-Grade Accomplishments

- Reads and comprehends both fiction and nonfiction that is appropriately designed for grade level.
- Accurately decodes orthographically regular multisyllable words and nonsense words (e.g., capital, Kalamazoo).
- Uses knowledge of print-sound mappings to sound out unknown words.
- Accurately reads many irregularly spelled words and such spelling patterns as diphthongs, special vowel spellings, and common word endings.
- Reads and comprehends both fiction and nonfiction that is appropriately designed for grade level.
- Shows evidence of expanding language repertory, including increasing use of more formal language registers.
- Reads voluntarily for interest and own purposes.
- Rereads sentences when meaning is not clear.
- Interprets information from diagrams, charts, and graphs.
- Recalls facts and details of texts.
- Reads nonfiction materials for answers to specific questions or for specific purposes.
- Takes part in creative responses to texts such as dramatizations, oral presentations, fantasy play, etc.
- Discusses similarities in characters and events across stories.
- Connects and compares information across nonfiction selections.
- Poses possible answers to how, why, and what-if questions.
- Correctly spells previously studied words and spelling patterns in own writing.
- Represents the complete sound of a word when spelling independently.
- Shows sensitivity to using formal language patterns in place of oral language patterns at appropriate spots in own writing (e.g., decontextualizing sentences, conventions for quoted speech, literary language forms, proper verb forms).
- Makes reasonable judgments about what to include in written products.
- Productively discusses ways to clarify and refine writing of own and others.
- With assistance, adds use of conferencing, revision, and editing processes to clarify and refine own writing to the steps of the expected parts of the writing process.
- Given organizational help, writes informative well-structured reports.
- Attends to spelling, mechanics, and presentation for final products.
- Produces a variety of types of compositions (e.g., stories, reports, correspondence).

TABLE 2-2 Continued

Third-Grade Accomplishments

- Reads aloud with fluency and comprehension any text that is appropriately designed for grade level.
- Uses letter-sound correspondence knowledge and structural analysis to decode words.
- Reads and comprehends both fiction and nonfiction that is appropriately designed for grade level.
- Reads longer fictional selections and chapter books independently.
- Takes part in creative responses to texts such as dramatizations, oral presentations, fantasy play, etc.
- Can point to or clearly identify specific words or wordings that are causing comprehension difficulties.
- Summarizes major points from fiction and nonfiction texts.
- In interpreting fiction, discusses underlying theme or message.
- Asks how, why, and what-if questions in interpreting nonfiction texts.
- In interpreting nonfiction, distinguishes cause and effect, fact and opinion, main idea and supporting details.
- Uses information and reasoning to examine bases of hypotheses and opinions.
- Infers word meanings from taught roots, prefixes, and suffixes.
- Correctly spells previously studied words and spelling patterns in own writing.
- Begins to incorporate literacy words and language patterns in own writing (e.g., elaborates descriptions, uses figurative wording).
- With some guidance, uses all aspects of the writing process in producing own compositions and reports.
- Combines information from multiple sources in writing reports.
- With assistance, suggests and implements editing and revision to clarify and refine own writing.
- Presents and discusses own writing with other students and responds helpfully to other students' compositions.
- Independently reviews work for spelling, mechanics, and presentation.
- Produces a variety of written works (e.g., literature responses, reports, "published" books, semantic maps) in a variety of formats, including multimedia forms.

Appendix B – Sample NAEP Questions and Items

Note: From (NAGB, 2008).

Exhibit 7. Sample NAEP Items, by Element of Literary Text and Aspect of Reading

Element of Literary Text	Aspect of Reading			
	Forming a General Understanding	**Developing Interpretation**	**Making Reader/Text Connections**	**Examining Content and Structure**
Theme	What is the moral in the story? Use evidence from the story in your response.	How does the setting help to illustrate the theme of the story?	Do you think the lesson in this story is true today? Why or why not?	Explain what makes this story a fable.
Major characters	What was the major character's opinion of _____?	What causes the main character to do _____? Use evidence from the story in your response.	How do you think the character's actions might be different today? Support your response with evidence from the story.	How does the author's description of _____ help explain the character's actions?
Major events	Write a short summary of the major events in the story.	What happens after _____?	How do you think the story would have ended if _____ had not happened?	How do the first events help you predict the ending?
Problem	How does _____ make the problem worse? Use evidence from the text to support your response	How did _____ help solve the problem?	How does the problem in the story compare with another story you have read? Include evidence from the text and another story.	Why does the author explain the problem in the first part of the story? Explain with evidence from the story.
Vocabulary	Which words describe what the story is mostly about? Use evidence from the text to support your response.	Which words let you know that time has gone by? Explain with evidence from the story.	Explain the double meaning of _____. Tell which meaning better explains the major ideas in the passage.	Why does the author use the words _____ to describe how _____ feels?

Exhibit 8. Sample NAEP Items, by Element of Informational Text and Aspect of Reading

Element of Informational Text	Aspect of Reading			
	Forming a General Understanding	Developing Interpretation	Making Reader/Text Connections	Examining Content and Structure
Central purpose	What might be the author's message in this article?	How does the author support the message?	Do you agree with the author's message? Give evidence from the text.	Based on what you read, what might be the reason the author wrote this?
Major ideas	Give a summary of the major ideas.	How does the big idea in the first section relate to the big idea in the last section?	Who might need or want this information? Use details from the text in your answer.	What did the author do to present information clearly?
Supporting ideas	Identify ideas that most closely relate to the topic. Give evidence from the text to support your choice.	How does the author show you that the main idea is important?	Which details about the _____ help you to have a clear image of the topic? Explain why you chose them.	What information did the author have to know before writing the article?
Adjunct aids	The chart in this article is mostly used to _____?	How does the information in the chart support the information in the article?	Why did the author include the picture with the chart? Explain using what you know and information from the text.	What is the significance of the map to the article? Explain.
Vocabulary	Which words describe what the passage is mostly about? Use the evidence from the text to support your choice.	Which words do you think mean the same as the title? Tell why you think so.	Explain the double meaning of _____. Tell which meaning better explains the major ideas in the passage.	Why did the author give a definition of _____ in paragraph 2?

Appendix C – Language Arts Log

Note: From Rowan et al. (2004), pp. 95-98. © 2000 The Regents of the University of Michigan. Reprinted with permission.

LANGUAGE ARTS LOG

Study of Instructional Improvement

Carefully place your student label here

For office use only

1. How much total time did the target student spend on language arts today? Please include all language arts instruction the target student received including routine times such as morning board work, even if the instruction took place in another room or by another teacher.
(Print the number of minutes using all three boxes. For example, write 015 if you taught for 15 minutes.)

☐☐☐ If 0 minutes, skip to Question 3.

2. Of the language arts time recorded in Question 1, how much time were you either the teacher or an observer of the teaching?
(Print the number of minutes using all three boxes. For example, write 015 if you taught for 15 minutes.)

☐☐☐ If more than 0 minutes, skip to Question 4.

3. Please mark the reason(s) why you recorded 0 minutes in Question 1 or 2, and then stop here.
(For any of the following items you choose, place an "X" in the corresponding box. Mark all that apply.)

- ☐ Target student was absent
- ☐ I was absent
- ☐ School was not in session (e.g., vacation period)
- ☐ There was a field trip, assembly, visitor, or other special event
- ☐ Target student participated in standardized testing/test preparation
- ☐ Target student received "pull out" instruction
- ☐ Other ____________________

4. To what extent were the following topics a focus of your work with the target student in reading/language arts today? *(Place an "X" in one of the boxes for each item.)*

	A major focus	A minor focus	Touched on briefly	Not taught today	Complete section(s) if this topic was a major or minor focus
a. Comprehension..........	☐	☐	☐	☐	A
b. Writing..........................	☐	☐	☐	☐	B
c. Word analysis.............	☐	☐	☐	☐	C
d. Concepts of print........	☐	☐	☐	☐	None
e. Reading fluency..........	☐	☐	☐	☐	None
f. Vocabulary..................	☐	☐	☐	☐	None
g. Grammar.....................	☐	☐	☐	☐	None
h. Spelling........................	☐	☐	☐	☐	None
i. Research strategies....	☐	☐	☐	☐	None

If you marked major focus or minor focus for Questions 4a, 4b, or 4c, please turn the page and answer the questions for the section(s) indicated in the color boxes above. All others STOP HERE.

A - Comprehension

A0. Was the work in comprehension in... *(Mark all that apply.)*

- ☐ Listening comprehension (A0a)
- ☐ Reading comprehension (A0b)

A1. What areas of comprehension did the target student work on today?

(For each area you choose below, place an "X" in a box to indicate whether it was a focus of instruction or was touched on briefly.)

	A focus of instruction	Touched on briefly
Activating prior knowledge or making personal connections to text (A1a)	☐	☐
Making predictions, previewing, or surveying (A1b)	☐	☐
Vocabulary-comprehension relationships (A1c)	☐	☐
Students generating their own questions (A1d)	☐	☐
Reading for pleasure or information (A1e)	☐	☐
Self-monitoring for meaning (A1f)	☐	☐
Using visualization or imagery (A1g)	☐	☐
Using charts, graphs, figures, tables, or other visual aids in text (A1h)	☐	☐
Using concept maps, story maps, or text structure frames (A1i)	☐	☐
Answering questions that have answers directly stated in the text (A1j)	☐	☐
Answering questions that require inferences (A1k)	☐	☐
Explaining how to find answers or information (A1l)	☐	☐
Sequencing information or events (A1m)	☐	☐
Identifying story structure (A1n)	☐	☐
Practicing other skills such as identifying similes or understanding referents (A1o)	☐	☐
Comparing and/or contrasting information or texts (A1p)	☐	☐
Summarizing important details (A1q)	☐	☐
Analyzing and evaluating text (A1r)	☐	☐
Examining literary techniques or author's style (A1s)	☐	☐
Written literature extension project (A1t)	☐	☐
Non-written literature extension project (e.g., puppet show, play, shadow box, book talk) (A1u)	☐	☐

A2. Did the materials used by the target student in work on comprehension include any of the following? *(Mark all that apply.)*

- ☐ Informational text (A2a)
- Narrative text
  - ☐ with controlled vocabulary (sight words and/or words easily sounded out) (A2b)
  - ☐ with patterned or predictable language (A2c)
  - Literature-based or thematic text
    - ☐ short selection (A2d)
    - ☐ chapter book (A2e)

A3. In which of the following ways did the target student demonstrate comprehension? *(Mark all that apply.)*

- ☐ Answered brief oral questions (A3a)
- ☐ Discussed text with peers (A3b)
- ☐ Did a think-aloud or explained how they applied a skill or strategy (A3c)
- ☐ Generated questions about text (A3d)
- ☐ Answered multiple-choice questions (A3e)
- ☐ Completed sentences filling in the blanks (A3f)
- ☐ Worked on concept maps, story maps, or text structure frames (A3g)
- ☐ Wrote brief answers to questions (A3h)
- ☐ Wrote extensive answers to questions (A3i)
- ☐ Worked on a literature extension project (A3j)

36757

Office Use Only

A - Comprehension (cont'd)

A4. **Did your instruction in comprehension include any of the following?** *(Mark all that apply.)*

- ☐ I demonstrated or explained a skill (e.g., how to determine the main idea, how to make an inference) (A4a)
- ☐ I demonstrated or explained how to use a reading strategy (e.g., previewing, generating questions about text) (A4b)
- ☐ I explained why or when to use a reading strategy (A4c)
- ☐ I helped students practice a skill or strategy (A4d)
- ☐ I administered a comprehension test (A4e)

Proceed to Section B and/or C, ONLY IF you marked "major focus" or "minor focus" for Questions 4b or 4c.

B - Writing

B1. **What areas of writing did the target student work on today?**
(For each area you choose, below, place an "X" in a box to indicate whether it was a focus of instruction or was touched on briefly.)

	A focus of instruction	Touched on briefly
Generating ideas for writing (B1a)	☐	☐
Organizing ideas for writing (B1b)	☐	☐
Literary techniques or author's style (B1c)	☐	☐
Writing forms or genres (e.g. letter, drama, editorial, Haiku) (B1d)	☐	☐
Writing practice (B1e)	☐	☐
Revision of writing - elaboration (B1f)	☐	☐
Revision of writing - refining or reorganizing (B1g)	☐	☐
Editing capitals, punctuation, or spelling (B1h)	☐	☐
Editing word use, grammar, or syntax (B1i)	☐	☐
Sharing writing with others (e.g., author's chair, share-pair, performances) (B1j)	☐	☐

B2. **Did the target student's writing consist of . . .** *(Mark all that apply.)*

- ☐ Letter strings or words (with or without illustration) (B2a)
- ☐ Separate sentence(s) (with or without illustration) (B2b)
- ☐ Separate paragraph(s) (B2c)
- ☐ Connected paragraphs (B2d)

B3. **Did your instruction in writing include any of the following?** *(Mark all that apply.)*

- ☐ I demonstrated or did a think-aloud using my own writing (B3a)
- I explained how to write, organize ideas, revise or edit...
  - ☐ using student writing (B3b)
  - ☐ using a published author's writing (B3c)
- ☐ I took dictation from the student (B3d)
- ☐ I led the student and his/her peers in a group composition (B3e)
- ☐ I commented on <u>what</u> the student wrote (not how) (B3f)
- ☐ I described what the student did well in his/her writing (B3g)
- ☐ I commented on <u>how</u> the student could improve his/her writing (B3h)
- ☐ I provided a writing or proofreading guide (B3i)

C - Word Analysis

C1. What areas of word analysis did the target student work on today?
(For each area you choose below, place an "X" in a box to indicate whether it was a focus of instruction or was touched on briefly.)

	A focus of instruction	Touched on briefly
Letter-sound relationships (C1a)	☐	☐
Sound segmenting:		
Counting the number of sounds in words (C1b)	☐	☐
Sound spelling/invented spelling/developmental spelling (C1c)	☐	☐
Segmenting a part of the word (for example, 'many' without 'm' is 'any,' or 'upstairs' without 'stairs' is 'up') (C1d)	☐	☐
Other segmenting tasks (C1e)	☐	☐
Sound blending:		
Blending initial sound with a rhyming word (onset-rime) (C1f)	☐	☐
Blending individual phonemes (sounds) into real words (C1g)	☐	☐
Blending phonemes (sounds) into nonsense words (C1h)	☐	
Blending syllables (C1i)	☐	☐
Other blending tasks (C1j)	☐	☐
Word recognition, sight words (C1k)	☐	☐
Structural analysis, examining word families, prefixes, suffixes, contractions, etc. (C1l)	☐	☐
Use of context, picture, and/or sentence meaning and structure to read words (C1m)	☐	☐
Use of phonics-based or letter-sound relationships to read words in sentences or stories (C1n)	☐	☐

C2. Did the materials used by the target student in work on word analysis contain any of the following? *(Mark all that apply.)*

- ☐ Sounds only (C2a)
- ☐ Pictures or objects to identify letters, words (C2b)
- ☐ Isolated words and letters (C2c)
- ☐ Individual sentences (C2d)
- Connected text (for example, stories, articles, poems, etc.)
  - ☐ with controlled vocabulary (sight words and/or words easily sounded out) (C2e)
  - ☐ with patterned or predictable language (C2f)
  - ☐ that is literature-based or thematic (C2g)

C3. What did you do when a student got stuck or made errors in word analysis? *(Mark all that apply.)*

- ☐ I corrected the student's errors or modeled the correct answer (C3a)
- ☐ I told the student to try again (C3b)
- ☐ I prompted the student to use the context (other words in sentence, pictures, what they already know) to read the word (C3c)
- ☐ I gave oral cues - sounding out parts of the word for them (C3d)
- ☐ I ignored the error and waited for the student to self-correct (C3e)

C4. Did your instruction in word analysis include any of the following? *(Mark all that apply.)*

- ☐ I listened to the target student read (C4a)
- ☐ I took running records or conducted a miscue analysis (C4b)
- ☐ I administered a word analysis test (C4c)

Office Use Only

References

Abadzi, H. (2007). *Absenteeism and beyond: Instructional time loss and consequences*. World Bank Policy Research Working Paper No. 4376. Retrieved October 23, 2008 from http://ssrn.com/abstract=1021370

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory into Practice*, *41*(4), 255-260.

Atkinson, R. C. (1974). Teaching children to read using a computer. *American Psychologist*, *29*(3), 169-178.

Atkinson, R. C. & Hansen, D. N. (1966). Computer-assisted instruction in initial reading: The Stanford project. *Reading Research Quarterly*, *2*(1), 5-25.

Baker, D., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-national evidence. *Prospects*, *34*(3), 311-334.

Ball, D. L. & Rowan, B. (2004). Introduction: Measuring instruction. *The Elementary School Journal*, *105*(1), 3-10.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew Effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, *34*(6), 1373-99.

Berliner, D. C. (1990). What's all the fuss about instructional time? In M. Ben-Peretz & R. Bromme(Eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions* (pp. 3-35). New York: Teachers College Press.

Blackman, P. H. (1996). Epidemiology of firearms and homicide: The need for basic science. In P. K. Lattimore & C. A. Nahabedian (Eds.), *Nature of homicide: Trends and changes – Proceedings of the 1996 meeting of the Homicide Research Working Group* (pp. 190-204). Santa Monica, CA: National Institute of Justice.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4-16.

Bromme, R. & Hömberg, E. (1990). Mathematics teachers' perception of time in class. In M. Ben-Peretz & R. Bromme (Eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions* (pp. 161-188). New York: Teachers College Press.

Brown, B. W., & Saks, D. H. (1986). Measuring the effects of instructional time on student learning: evidence from the beginning teacher evaluation study. *American Journal of Education*, *94*(4), 480-500.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723-733.

Carter, L. F. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher*, *13*(7), 4-13.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the Simple View of Reading. *Journal of Speech, Language, and Hearing Research*, *49*(2), 278-293.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119.

Conant, E. H. (1973). Teacher and paraprofessional work productivity: A public school cost effectiveness study. Lexington, Mass: Lexington Books.

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance schools. *Educational Researcher*, *24*(9), 27-31.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934-45.

Denham, C., & Lieberman, A. (1980). *Time to learn*. Washington, D.C.: U.S. Dept. of Education, National Institute of Education, Program on Teaching and Learning.

Department of Education (DOE). (2004). *NCLB Executive Summary*. Retrieved on October 28, 2008 from http://www.ed.gov/nclb/overview/intro/execsumm.html

Department of Education (DOE). (2008a). *Early Reading First.* Retrieved on October 28, 2008 from http://www.ed.gov/programs/earlyreading/index.html

Department of Education (DOE). (2008b). *Reading First.* Retrieved on October 24, 2008 from http://www.ed.gov/programs/readingfirst/index.html.

Dickey. M. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Design*, *53*(2), 67–83.

Dzubak, C. M. (2008). Multitasking: The good, the bad, and the unknown. *Synergy*. 1 (2), 1-12. Retrieved November 10, 2008 from http://www.myatp.org/Synergy_1/Syn_6.pdf.

Elmore, R. (2002). The limits of "change." *Harvard Education Letter*. Jan/Feb 2002. Retrieved November 10, 2008 from http://www.edletter.org/past/issues/2002-jf/limitsofchange.shtml.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates.

Fletcher, J. D. (2003). Evidence for learning from technology-assisted instruction. In H. F. O'Neil, Jr. and R. Perez (Eds.), *Technology applications in education: A learning view* (pp. 79–99). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, J. D. (2005). *The Advanced Distributed Learning (ADL) vision and getting from here to there*. Retrieved on November 10, 2008 from http://handle.dtic.mil/100.2/ADA452053.

Fredericks, J. A., Blumenfeld, P. C. & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109.

Fuchs, D., Fuchs, L. S., Thompson, A., Otaiba, S. A., Yen, L., Yang, N. J., et al. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*, 251-267.

Gallacher, M. S. (2008). Predicting sixth grade performance on criterion-referenced reading tests with third grade test scores. Thesis (M.A.)—Brigham Young University, 2008.

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, *11*(36).

Geary, D. C. (2000). Principles of evolutionary educational psychology. *Learning and Individual Differences*, *12*(4), 317-345.

Gest, S. D., & Gest, J. M. (2005). Reading tutoring for students at academic and behavioral risk: Effects on time-on-task in the classroom. *Education & Treatment of Children*, *28*(1), 25-47.

Gettinger, M. (1984). Achievement as a function of time spent in learning and time needed for learning. *American Educational Research Journal*, *21*(3), 617-28.

Gettinger, M. (1985). Time allocated and spent relative to time needed for learning as determinants of achievement. *Journal of Educational Psychology*, *77*(1), 3-11.

Gettinger, M. & Seibert, J.K. (2002). Best practices in increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV: Volume I* (4th ed., pp. 773-787). Bethesda, MD: National Association of School Psychologists.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a Simple View of Reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ: L. Erlbaum Associates.

Gough, P. B. & Juel, C. (1991). The first stages of word recognition. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 47-56). Hillsdale, NJ: L. Erlbaum Associates.

Grell, O. P. (1998). Paracelsus: The man and his reputation, his ideas and their transformation. Leiden: Brill.

Guthrie, J. W., & Springer, M. G. (2004). *A Nation at Risk* revisited: Did “wrong” reasoning result in “right” results? At what cost? *Peabody Journal of Education, 79*(1), 7-35.

Hanushek, E. A. (1997). Outcomes, incentives, and beliefs: Reflections on analysis of the economics of schools. *Educational Evaluation and Policy Analysis*, *19*(4), 301-08.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal – London-*, *113*(485), F64-F98.

Hoxby, C. M. (2003). The economics of school choice. A National Bureau of Economic Research conference report. Chicago: University of Chicago Press.

Kurzweil, R. (2005). The singularity is near: When humans transcend biology. New York: Viking.

Laties VG. (1979). I. V. Zavadskii and the beginnings of behavioral pharmacology: An historical note and translation. *Journal of the Experimental Analysis of Behavior*, *32*(3), 463-72.

Levitt, S. D., & Dubner, S. J. (2005). Freakonomics: A rogue economist explores the hidden side of everything. New York: William Morrow.

Macken, E., Suppes, P., & Zanotti, M. (1980). Considerations in evaluating individualized instruction. *Journal of Research and Development in Education*, *14*(1), 79-83.

McDonald, J. K., Yanchar, S. C., & Osguthorpe, R. T. (2005). Learning from programmed instruction: Examining implications for modern instructional technology. *Educational Technology Research and Development*, *53*(2), 84-98.

Millman, J., Bieger, G. R., Klag, P. A., & Pine, C. K. (1983). Relation between perseverance and rate of learning: A test of Carroll's model of school learning. *American Educational Research Journal*, *20*(3), 425-434.

Minter Hoxby, C. (1996). "Are efficiency and equity in school finance substitutes or complements?". *Journal of Economic Perspectives*, *10*(4), 51-72.

National Assessment Governing Board (NAGB). (2004). *Reading framework for the 2005 National Assessment of Educational Progress* (chap. 2). Retrieved October 16, 2008 from http://www.nagb.org/pubs/r_framework_05/ch2.html

National Center for Education Statistics (NCES). (2007). *The condition of education 2007*, NCES 2007-064. Retrieved October 23, 2008 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007064

National Center for Education Statistics (NCES). (2008a). *NAEP overview.* Retrieved October 23, 2008 from http://www.nces.ed.gov/nationsreportcard/about/

National Center for Education Statistics (NCES). (2008b). *What are the differences between long-term trend NAEP and main NAEP?* Retrieved October 27, 2008 from http://www.nces.ed.gov/nationsreportcard/about/ltt_main_diff.asp

National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Commission on Excellence in Education (NCEE). (2008). *Reading First impact study: Interim report* (NCEE 2008-4016). Retrieved October 28, 2008 from http://ies.ed.gov/ncee/pdf/20084016.pdf

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction – Reports of the Subgroups. Rockville, MD: National Institutes of Health.

Pachter, H. M. (1951). Parscelsus: magic into science; being the true history of the most renowned, widely traveled, very learned and pious gentleman, scholar, and most highly experienced and illustrious physicus, the Honorable Philippus Theophrastus Aureolus Bombastus ab Hohenheim, Eremita, called Paracelsus. New York: Henry Schuman.

Paik, S., Wang, D., & Walberg, H. J. (2002). Timely learning improvements. *Educational Horizons*, *80*(2), 69–71.

Parkerson, J. A., Lomax, R. G., Schiller, D. P., & Walberg, H. J. (1984). Exploring causal models of educational achievement. *Journal of Educational Psychology*, *76*(4), 638-46.

Poling, A. D., & Byrne, T. (2000). *Introduction to behavioral pharmacology*. Reno, NV: Context Press.

Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. New York: Guilford Press.

Pressley, M., Billan, A. K., Perry, K. H., Reffitt, K. E., & Reynolds, J. M. (Eds.) (2007). *Shaping literacy achievement: Research we have, research we need*. New York: Guilford Press.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal*, *105*, 75-102.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the "Prospects" study of elementary schools. *Teachers College Record*, *104*(8), 1525-67.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, OX: Pergamon.

Scheuren, F. (2005). *What is a survey?* Retrieved November 10, 2008 from http://www.amstat.org/sections/srms/pamphlet.pdf

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford: Oxford University Press.

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

Spence, D., & Usher, E. (2007). Engagement with mathematics courseware in traditional and online remedial learning environments: Relationship to self-efficacy and achievement. *Journal of Educational Computing Research*, *37*(3), 267-288.

Stanovich, K. E. (1986). Matthew Effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360-407.

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

Stanovich, K. E. (2003). *How to think straight about psychology*. Boston: Allyn and Bacon.

Suppes, P., Fletcher, J. D., & Zanotti, M. (1975). Performance models of American Indian students on computer-assisted instruction in elementary mathematics. *Instructional Science*, *4*, 303-313.

Suppes, P., Fletcher, J. D., & Zanotti, M. (1976). Models of individual trajectories in computer-assisted instruction for deaf students. *Journal of Educational Psychology*, *68*(2), 117-27.

Suppes, P., & Zanotti, M. (1996). Mastery learning in elementary mathematics: Theory and data. In P. Suppes & M. Zanotti (Eds.), *Foundations of probability with applications* (pp. 149-188). New York: Cambridge University Press.

Tallarida, R. J., & Jacob, L. S. (1979). *The dose-response relation in pharmacology*. New York: Springer-Verlag.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Viteritti, J. P. (2004). From excellence to equity: Observations on politics, history, and policy. *Peabody Journal of Education*, *79*(1), 64-86.

Vygotsky, L. S. (1978). Mind and society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Walberg, H. J., & Tsai, S.-L. (1983). Matthew Effects in education. *American Educational Research Journal*, *20*(3), 359-73.

Walberg, H. J. (2003). *Improving educational productivity*. Publication Series No. 1. Philadelphia: Laboratory for School Success. Retrieved November 5, 2008 from http://www.temple.edu/lss/pdf/publications/pubs2003-1.pdf

Waldrop, M. M. (2001). The dream machine: J. C. R. Licklider and the revolution that made computing personal. New York: Viking.

Whitehurst, G. J. (2003). September 10, 2003 interview with David Boulton. Retrieved October 16, 2008 from http://www.childrenofthecode.org/interviews/whitehurst.htm

Wong, K. K., Guthrie, J. W., & Harris, D. N. (2004). A nation at risk: A 20-year reappraisal. *Peabody Journal of Education*, *79*(1), 1-6.

Wong, K. K. & Nicotera, A. C. (2004). Educational quality and policy redesign: Reconsidering the *NAR* and federal Title I policy. *Peabody Journal of Education*, *79*(1), 87-104.

Zhang, H., Almeroth, K. C., & Bulger, M. (2005). An activity monitoring system to support classroom research. In P. Kommers & G. Richards (Eds.), *Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2005* (pp. 1444-1449). Chesapeake, VA: AACE.