2010-03-12

Effects of Computer-Based, Early-Reading Academic Learning Time on Early-Reading Achievement: A Dose-Response Approach

Benjamin Heuston
Brigham Young University - Provo

Effects of Computer-Based Early-Reading Academic Learning Time

on Early-Reading Achievement: A Dose-Response Approach

Edward Benjamin Hull Heuston

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Harold Miller
Mikle South
Ross Flom
Joseph Olsen
Tim Smith

Department of Psychology

Brigham Young University

April, 2010

ABSTRACT

Effects of Computer-Based Early-Reading Academic Learning Time

on Early-Reading Achievement: A Dose-Response Approach

Edward Benjamin Hull Heuston

Department of Psychology

Doctor of Philosophy

Academic learning time (ALT) has long had the theoretical underpinnings sufficient to claim a causal relationship with academic achievement, but to this point empirical evidence has been lacking. This dearth of evidence has existed primarily due to difficulties associated with operationalizing ALT in traditional educational settings.

Recent advancements in computer-based instruction provide an unprecedented opportunity to model ALT and to test the underlying theory. A widely-used computer-based early-reading curriculum was operationalized using Berliner's model of ALT (Berliner, 1991). This curriculum was then mapped to a computer-based assessment to determine an appropriate method of quantifying early-reading ALT. Software limitations required that ALT be quantified as a summative measure.

Data were collected from 1,347 prekindergarteners and were analyzed using a dose-response approach that associated usage of the curriculum with a generalized variable of early-reading achievement. Gains across four early-reading skills were demonstrated via linear regression to be predicted by minutes of usage (Adj. $R^2 = .078$). A sample optimized to test the hypothesis showed a stronger correlation (Adj. $R^2 = .096$). Time spent using the Free Play version of the curriculum did not uniquely predict additional variance. Similarly, gains on reading skills that were not taught explicitly by the curriculum were not predicted by overall usage. These three results were interpreted as supporting the ALT learning model.

Post-hoc analyses were performed on curriculum-usage compliance and on within-curriculum progress, both of which were statistically significant when added to the basic dose-response model. Multiple exploratory best-fit models were constructed. The strongest accounted for just under 20% of the overall variance (Adj. $R^2 = .186$).

Effect sizes were in the medium-to-large range for the entire sample (D = 0.71) with significant improvement for the optimized sample (D = 1.26). Children in the optimized sample who used the program over 20% more than recommended had even stronger gains (D = 1.67).

The ability to remotely and accurately quantify interaction with a computer-based curriculum and assessment in the home defines a new vista in ALT research.

Keywords: [Academic Learning Time, Dose Response, Early Reading, Compliance, Computer-Based Instruction, Early Childhood]

ACKNOWLEDGEMENTS

To my chair and mentor Dr. Miller, for insisting on the highest caliber of thinking and for repeated efforts to educate me on the importance of clear and precise prose. To Dr. Olsen for his flexibility, patience, and wisdom with regards to all things statistical. To my parents for the opportunity, the vision (3 Ne. 6:12), and the example of what it means to be dedicated. To my wife and children for study breaks and for letting Dad work. D&C 59:7, 21.

Table of Contents

List of Figures

List of Tables

Introduction: The Importance of Time in Skill Acquisition

Carroll (1963) postulated that academic skills are mastered in proportion to the amount of time spent learning them. According to his model, "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). The brilliance of this model is that while the amount of time required for any given student to master a given learning task might vary based on learning history, aptitude, quality of instruction, etc., in the end it all comes down to time. Ensuing educational research has repeatedly verified the importance of including learning time as a variable, leading Walberg to note that "[t]he positive effect of time is perhaps most consistent of all causes of learning" (2003, p. 7).

Berliner and Academic Learning Time

Subsequent researchers have built upon and refined Carroll's basic model of time and learning, most notably Berliner, whose 1990 review of instructional time detailed the construct of *academic learning time* (ALT). Berliner's definition of ALT is four-fold:

(1) The time in question must be instructional in nature.

(2) The learner must be engaged across the time period.

(3) The instructional difficulty must be appropriate for the learner.

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

Thus, ALT refers specifically to that time when relevant, appropriate, and assessed instruction is provided to an actively engaged student. If any of these four requirements is not met, then the instructional time does not count as ALT for that particular student.

By being defined more narrowly and explicitly than Carroll's original construct of "time spent learning," ALT becomes more than just time on task: it becomes "time on [the] right tasks" (p. 18). Every second of ALT is, by definition, helping a student move closer to mastery of specific, quantifiable skills and content. According to Berliner, ALT is therefore not just useful, it is absolutely essential to the learning process: "[u]nless ALT is affected in some way, there will be no changes in student achievement at all" (p. 22).

The Challenge of Quantifying ALT

Although ALT is an attractive variable from a theoretical standpoint, it is very difficult to measure in traditional academic settings. This difficulty stems in part from the fact that ALT needs to be quantified at the individual and not at the group level.

To understand what this would look like from a practical standpoint, consider a typical elementary classroom with 30 students whose teacher is instructing them in reading. Although the teacher has significant knowledge about the students in the classroom, researchers have found that teachers are highly inaccurate when it comes to evaluating individual student progress (Bromme & Hömberg, 1990). In order to quantify ALT, another observer, namely a researcher, is required.

In order to accurately quantify ALT, this researcher would need to have 30 clocks running simultaneously, one for each student in the class. The clocks would all switch off if the teacher embarked on a non-reading topic (not instructional) or if the reading skill being taught was not going to be included on the upcoming test (not assessed), and would switch back on whenever the topic is reading-related and is included on the test.

At the same time, each individual student's clock would switch off whenever s/he is disengaged or when the level of the instruction is either too easy or too difficult for that particular student, and would switch back on when that student is meaningfully engaged and the material is of an appropriate difficulty level.

These last two criteria are especially difficult not only because they need to be measured at the student level, but simply because they need to be measured at all. The connection between observable behavior (e.g., watching the teacher) and engagement is not always clear-cut, raising the question of how it should be measured (Ball & Rowan, 2004; Fredericks, Blumenfeld, & Paris, 2004). As for establishing the appropriateness of a given learning task for a student, Gettinger & Seibert (2002) recommend that young students should be able to perform a given learning task at an 80% level of mastery. Such an approach would require a minimum of 5 questions to be given to each student, resulting in a staggering 150 questions needing to be asked, answered, and evaluated for the whole classroom – for each learning task.

Limits of the Manual Model

These challenges are evident in a recent attempt to measure instructional time in the classroom (Rowan, Camburn, & Correnti, 2004). After reviewing the shortcomings of traditional methods for measuring instructional time in the classroom, Rowan, et al. proposed a new method that involves a tool known as *instructional logs*. An instructional log is a time diary that teachers fill out on a daily basis and that records the topics taught during the course of the day. These daily class-level observations are then supplemented on a daily basis by detailing the instructional topics presented to a single, randomly-chosen student.

Clearly such an approach falls well short of quantifying ALT. While it quantifies instructional time at the class level, it does not speak to engagement or instructional difficulty at all, either at the class or at the student level. Given the constraints of a traditional classroom and traditional research methods, it is difficult to see how student-level ALT could be effectively quantified.

The Promise of Computer-Based Instruction

Current computer-based instruction provides a promising avenue for addressing these challenges (Heuston, 2008). Instructionally, computers provide curriculum in a programmatic manner, allowing for a precise measure of instructional time and for the ability to map curricular elements to an eventual assessment (Atkinson, 1974; Suppes & Zanotti, 1996). This ability to quantify instructional time at the student level does not impact the presentation of the curriculum – unlike a traditional teacher, a computer can truly multitask. Computers generally cannot directly measure engagement, but engagement can be inferred from patterns of responses and progress through the curriculum. Moreover, computer-based instruction can be highly engaging (Dickey, 2005). Computer-based instruction can be delivered individually and at its best is delivered adaptively, thereby addressing the concern about instructional appropriateness (Macken, Suppes, & Zanotti, 1980).

Operationalizing and Testing Berliner's ALT

Given these advantages, computer-based instruction should allow for a more precise quantification of ALT than previously has been possible. The aim of this dissertation was to demonstrate this potential by:

- operationalizing a computer-based curriculum and a computer-based assessment so they could be used to investigate Berliner's notion of ALT; and
- conducting a pilot study of ALT using these tools with the intent of testing the hypothesis that ALT is the causal variable in learning.

The academic subject that I focused on was early reading, and the computer-based curriculum and computer-based assessment I selected by which to operationalize ALT were *Rusty and Rosy Learn With Me*^{TM} (RRLWM) and *Waterford Assessments of Core Skills*^{TM} (WACS), respectively.

Disclaimer

The selection of these tools for operationalization was not happenstance. I am currently the President and Chief Operating Officer of the non-profit Waterford Research Institute, which actively develops and maintains these programs, and where I have worked full-time in various capacities for over a dozen years. As such, I am not a disinterested third-party investigator. The benefits of my position from the standpoint of this study were that I was able to gain unfettered access to the program and that I had an intimate knowledge of not just the operations of but also of the design and intent of the program. The obvious challenge that my position presented was that it was impossible to remain entirely objective about the program or the data it generated. My dissertation committee was well aware of my position and the potential conflicts inherent in my circumstances, and satisfied themselves as to both the rigor of the undertaking as well as the defensibility of the analysis that accompanies this dissertation. To be clear, I was not directly involved in any of the data collection or administration of the study itself.

Part One: Operationalization of RRLWM Reading and WACS

This section details the operationalization of a computer-based early-reading curriculum and a computer-based early-reading assessment for the purposes of measuring early-reading ALT. While the actual calculation of ALT is mathematically straightforward, there are numerous assumptions that undergird the calculation. The intent of this section is to describe these assumptions to the reader to help provide context for the ensuing study. Thus, the operationalizations both enabled and informed the subsequent pilot study.

A Note About Program Versions

Computer-based technologies that are actively being developed change at a pace that creates difficulties for academic researchers (Fletcher, 2003). RRLWM and WACS are not exempt from these challenges—the versions used during the pilot study (from April, 2009 through December, 2009) were 1.3.0.17 for the RRLWM client and 1.3.0.16 for the RRLWM application server, and 2.0.0.53 for the WACS media and 2.0.1.41 for the WACS application server. At the time of publication of this dissertation in April, 2010, just a short 12 months after the beginning of data collection, these versions will have already been superseded by more current versions and will no longer be available. The long-term relevance of this study is therefore to be found more in the principles it suggests as opposed to the finer details of the study itself.

Overview of RRLWM

This overview is divided into two segments. The first describes the components and overall structure of RRLWM's reading program. The second segment places the program into a broader educational context by reviewing previously published research on the effects of the reading curriculum on young learners.

The overview focuses solely on those portions of RRLWM that are relevant to the operationalization and subsequent pilot study. Interested readers are referred to the official product website (http://www.waterford.org/products/rusty-and-rosy) for further detail.

The Structure of RRLWM

RRLWM is composed of two distinct curricular portions – the Early Reading Program (ERP) and Early Math & Science. The operationalization and pilot study dealt only with the ERP portion.

ERP covers early-reading skills through the 2nd grade. The curriculum is comprehensive in nature and is split into three levels, each of which covers a full school year of curriculum. Recommended usage of ERP is 15 minutes per day for pre-kindergarteners and kindergarteners, and 30 minutes per day for 1st- and 2nd-graders. These recommendations combine developmental realities (i.e., young children do not have hour-long attention spans) and the recognition that there is a finite amount of value expected from the curriculum (i.e., infinite usage is not expected to result in infinite gains).

When using RRLWM, there are two distinct ways for children to access the reading content. The first is to allow the computer to determine the individual child's instructional path through the curriculum. Using this method, reading activities are presented in a structured manner, and children move through the curriculum from easier to harder objectives. This is the default mode for the curriculum and is optimized for learning efficiency.

The second way to access the reading content of RRLWM is to allow the child to repeatedly choose what s/he would like to work on, a mode known as "free play." The content available in this mode is based on the progress a child has made on the structured side. As children progress through the structured content, they "unlock" corresponding activities on the free-play side. Not all of the structured activities are available in free play – only those assumed to be more entertaining in nature (e.g., songs, games, etc.) are accessible. By enabling the child to repeatedly choose among the most appealing activities, free play is optimized for engagement.

Educational Context and Rationale

This study provided the first major review of the educational efficacy of RRLWM. There are a number of reasons for this. The first is that RRLWM is relatively new. A second and related reason is that there are comparatively few users of RRLWM. A third reason is that RRLWM is targeted at children learning in the home, an environment that tends to not generate as much research interest.

Although RRLWM has not been studied extensively, the reading portion of RRLWM is the same curriculum that is contained in the *Waterford Early Reading Program* (WERP), which has been actively developed, marketed, and evaluated since its inception over 15 years ago. For this reason, it may be instructive to review where WERP has been used and what researchers have reported about its effectiveness in promoting students' gains in early-reading achievement.

WERP is widely used in elementary classrooms in the United States: "[a]ccording to company statistics, at the close of the 2002 fiscal year, the Waterford Early Reading Program was in 5% of the elementary schools nationwide. During 2002, 2,700 schools, 12,750 classrooms, and approximately 326,000 children nationwide were using the Waterford Early Reading Program" (Tracey & Young, 2007, p. 451). These numbers have increased markedly in the subsequent eight years, suggesting that WERP will continue to play an important role in the education of young readers in America for many years to come.

With respect to the concept of ALT, time spent using WERP in elementary classrooms has been correlated to student gains on early-reading achievement tests. For example, Powers and Price-Johnson (2006) found a positive correlation in kindergarteners between early-reading achievement and time spent using WERP, although no formal modeling of the relationship was attempted, and Hecht and Close (2002) similarly found that “[t]he amount of time that children used the WERP [Level 1] was correlated with all posttest measures of emergent literacy skills” (p. 111).

Operationalizing RRLWM

Having established a rationale for studying the use of RRLWM’s reading curriculum, it remains to formally map RRLWM to Berliner’s four-part definition of ALT. This mapping will take place in two steps. First RRLWM’s content and capabilities will be logically mapped onto the first three of the four requirements. Second, these mappings will then be expressed as a procedure for calculating early-reading ALT.

Mapping RRLWM to Berliner’s ALT

As previously indicated, there are four criteria for counting time as ALT – the time must be instructional, the learner must be actively engaged in the learning process, the content must be appropriate for the learner, and the content must be aligned with the assessment. By itself, RRLWM satisfies the first three requirements for measuring ALT.

RRLWM and instructional time. The first requirement is that the time spent using the curriculum needs to qualify as early-reading instruction. The simplest way to measure this is to determine whether the curriculum aligns with what reading researchers and policy makers have established as guidelines for successful early-reading curricula.

RRLWM's reading curriculum aligns directly with both state and federal early-reading standards, including the National Reading Panel's (2000) five core components of early-reading instruction, namely, phonemic awareness, phonics, vocabulary, fluency, and comprehension. Indeed, WERP has repeatedly qualified for Reading First funds, a competitive federal grant program that focuses on putting scientifically-proven methods of early-reading instruction into classrooms (DOE, 2008). RRLWM's use of the same curriculum suggests that it similarly satisfies Berliner's first criterion for ALT.

RRLWM and engaged time. One of the greatest challenges in measuring ALT in traditional classrooms is quantifying child engagement, as engagement is itself an unobservable event. This problem is not solved by simply moving to a computer-delivered curriculum, but it is ameliorated when programs require regular, substantive student-computer interactions. Such observable interactions (e.g., selecting an answer using a mouse) can be viewed as indicative of meaningful engagement.

RRLWM requires such engagement in order for a child to progress through the curriculum. Children take part through the use of a mouse (for most activities), a keyboard (for keyboarding activities), and a microphone (for activities that involve recording). In addition, RRLWM requires engagement sufficient to result in mastery of the material in order for progress to occur.

RRLWM and instructional difficulty. RRLWM has an adaptive curriculum that adjusts to fit the demonstrated level of mastery for each child. The curriculum adjusts itself in two distinct ways: placement and sequencing. *Placement* generally occurs when a child first uses the program and involves a brief battery of activities that identify the child's general level of reading ability and then assigns the child to one of a number of pre-defined starting points within the curriculum. After this initial placement, individualization occurs through *sequencing*, a dynamic that determines, based on each child's unique learning history, what the next learning task for the child should be. In determining whether a child has mastered a given learning task, RRLWM uses a success rate of 80%, the same rate that Gettinger & Seibert (2002) cited as appropriate for young children.

Calculating RRLWM-specific ALT

Having demonstrated that RRLWM logically satisfies the first three of Berliner's ALT requirements, the next step is to detail how each of the three requirements is involved in the overall calculation of early-reading ALT.

Instructional time. The amount of instructional time provided to each child is assumed to be the session time, or the time the child spends using the reading curriculum. This time is recorded in the database by the management system. Thus, the overall instructional time is simply the concatenation of the individual session times for each child. Time spent using WACS is not included in calculation of ALT because WACS does not give instructional feedback and therefore is not considered to be an instructional program per se.

It is important to distinguish between time spent using the structured (or sequenced) portion of the program and time spent using the unstructured (or free-play) portion. The structured portion refers to those activities wherein the child is being actively sequenced by the program. Time spent using the free-play activities is categorically different because it is determined by the child's interests and not necessarily by the child's ability level. In addition, as mentioned earlier, free-play activities are solely composed of previously-seen material and do not provide formal instruction. In addition, RRLWM does not record which free-play activities a child uses. Consequently, free-play time could include time spent on math or science content and therefore might not be fully reflective of reading-content usage. Due to these issues, free-play time was not included in the overall calculation of early-reading ALT.

This relatively simplistic definition of instructional time (i.e., any time spent using the structured reading curriculum) is sufficient only if the task at hand is viewed as learning more about the skill of “early reading” as broadly defined. Such an approach is likely not entirely appropriate for RRLWM, which is composed of hundreds of diverse elements that span not only the range of early-reading skills, but also include activities related to school readiness (e.g., colors and shapes) or that involve other skills (e.g., typing).

A solution to this problem would be to record time not just at the session level, but also at the level of the individual activity. These activities could then be mapped to early-reading skills of specific interest to the operationalization of ALT, and the time spent on activities that are not aligned with those skills would not be included. The version of RRLWM used, however, reports time uniformly for sessions but not for activities, so the calculation of instructional time cannot occur at the activity level and needs to remain at the session level. If a future version of RRLWM erases this limitation, then a more granular measurement of ALT will be possible.

Engaged time. RRLWM is designed specifically to engage young children and employs a variety of tools to accomplish this goal. First, engagement is fostered by graphics and music that appeal to young children. Second, the curriculum is individualized to fit each child’s unique learning history within the curriculum. Third, the program requires regular, substantive interaction in order for the child to progress through the curriculum.

Taken collectively, these strategies appear to be effective in fostering engagement in formal learning environments. Generally students who use WERP demonstrate outward signs of engagement, including visually focusing on the computer; using the mouse, keyboard, and microphone to interact with the program; and moving their lips as they silently read. In addition, young students often sing along to familiar songs (e.g., the ABC song) while using the program.

RRLWM also has limited ability to infer engagement and takes steps to reengage children (e.g., by repeating instructions when responses are delayed). In the event that such steps are not successful, the program pauses and requires parent intervention before the child can resume. Time spent in a paused state is not included in the session time calculation, helping to ensure that the time recorded for each child is reflective of the time that child spent actively using the program.

Instructional difficulty. As noted earlier, children using the program are automatically placed and sequenced based on their individual responses. For this reason, the time spent using the sequenced portion of the program was assumed to equate to the time in which the child is receiving instruction of appropriate difficulty.

This assumption can be called into question if the program is not used sufficiently often enough or if the initial placement for a child is too low. In the first case, a lack of consistent use of the program is generally seen as a threat to the overall operation of the software. The primary reason for this is that RRLWM generally is not the sole source of instruction for a child. Therefore, to cite an extreme case, assume a kindergartner were to take RRLWM's placement test and find herself at the very beginning of the curriculum, but then did not use the program for 2 years. If she were to return to the program as a 2nd grader, the previous placement would most likely severely underestimate the child's current ability level (due to early-reading ALT experienced outside of RRLWM), and the child would then be forced to sequentially work her way through the curriculum to a now-appropriate level.

The second issue arises primarily due to the philosophy of the sequencing feature, which errs on the side of ensuring that a child is not advanced too quickly (which could result in failure or discouragement). This is accomplished by requiring a relatively high level of mastery (80%) before the child can progress to more difficult tasks. If the child somehow reaches activities that are too difficult, the sequencer has three options: (a) to repeat the instruction in question, generally with different content; (b) to return to prerequisite activities and ensure that they are properly mastered; or (c) to simply move the child on to other material and note that the child is having a problem that the computer cannot remedy.

The sequencer does not contain an automatic mechanism to determine whether a child initially is placed too conservatively in the curriculum. The simple solution to this problem is to have the student retake the placement activity to ensure that she is placed correctly, but this is a manual process and requires parental involvement.

Summary. The use of RRLWM to measure early-reading ALT is relatively straightforward – the amount of early-reading ALT is simply the amount of time spent using the structured reading curriculum itself. That being said, this measure is fraught with assumptions, one of which (namely, that the unstructured portion of the curriculum should not count as ALT) was explicitly tested as part of the pilot study. The greatest single limitation of this approach may be its inability to track time at the activity level, resulting in the inability to directly link time spent on a specific skill (e.g., Blending) with the amount of time spent on instructional material related directly to that skill. Thus, this operationalization of ALT speaks only to the skill of "early-reading instruction" as broadly defined.

Overview of WACS

RRLWM does not contain a comprehensive early-reading assessment and therefore cannot on its own fulfill the remaining requirement for measuring ALT. To that end, WACS, a computer-based early-reading assessment, was also used. As with RRLWM, this overview is not intended to be comprehensive. Interested readers are referred to the official product website (http://www.waterford.org/products/wacs) for further details.

WACS Structure

WACS consists of 11 distinct subtests, each of which measures a different reading skill. Each skill has its own unique items that are ranked relative to one another in terms of difficulty. The lowest possible score on any of the subtests is 1,001, and the highest possible score is 7,000, although no subtest spans this entire range.

Although the subtests are distinct, WACS as a whole has been shown to have a single underlying factor, suggesting that each skill is a partial manifestation of a unitary reading-skill continuum (Shamir, Johnson & Brown, 2009). Thus, scores from some or all of the completed subtests can be averaged to gain a more general picture of early-reading skills. This is particularly important for the current study because, as mentioned in the RRLWM overview, RRLWM usage time is measured only at the session level as opposed to at the activity level, and therefore usage can be measured for groups of skills but not for individual skills themselves.

In addition to these subtests, WACS also contains an instructional component that helps teach children how to use a mouse. The intent of this instruction is to help ensure that a child’s lack of familiarity with a mouse does not unduly bias the test scores.

WACS Sequencing

Children do not generally receive all of the subtests of WACS during a given administration. Rather they see a series of subtests deemed appropriate for their grade level but that are adjustable based on their performance. Thus, a Fall pre-kindergartener begins the test with Letter Sound, but a Fall 2nd-grader begins with Nonwords. The subtests presented thereafter can be based upon the performance in the previous subtest(s), but the paths of children of the same grade level are dictated by the same sequencing logic (see Appendix A for the prekindergarten sequencing logic).

Within a given subtest, test items are chosen randomly from a pool of items that is determined by the child's estimated ability level as well as by the items that have already been presented to the child. WACS has a large-enough item pool to ensure that no item will be seen more than once during the test. Additionally, WACS keeps track of the items already presented to a child during earlier testing sessions and uses this information to minimize the number of times a given item is presented across testing sessions as well. The intent of this design is to ensure that children's scores are based as much as possible on the underlying skill as opposed to any individual test item.

If a child misses the first four items on a subtest, s/he is assigned a floor score (see the "4 & Out" column in Appendix B), and the subtest ends. The intent of this design is to maximize the efficiency of the overall test as well as to avoid measurement inaccuracies that could be caused by having a student continue to work on items that are too difficult.

Interpretation of WACS

Scores on each individual subtest are scaled continuously from a lower threshold, with higher scores indicating greater skill levels. As mentioned earlier, scores that fall below this threshold occur only when a child misses the first four questions, and are therefore a potential indicator that the child is below the measurement threshold for the given subtest.

In longitudinal settings, where children take the same test across multiple points in time (Singer and Willet, 2003), a higher score on a subtest is indicative of relative growth in that skill area, while a lower score indicates a relative decline. In formal educational settings, scores are generally expected to increase over time, although declines are not unusual, especially when test administrations occur relatively close together.

In interpreting scores for an individual subtest, the following rules of thumb apply. Beginning in prekindergarten at 1,001, every 1,000 points on WACS is the equivalent of a grade level. Thus a student performing between 2,001 and 3,000 is demonstrating kindergarten-level mastery, and a student scoring between 4,001 and 5,000 is demonstrating 2nd-grade mastery. These 1,000-point grade levels are further subdivided into 3 intervals of 333.33 points each, with the first interval (1 – 334.33) indicating Fall norms, the next interval (334.34 – 667.66) Winter norms, and the final interval (667.67 – 1,000.0) Spring norms. Thus, a student scoring 3,107 on Letter Recognition is demonstrating 1st-grade Fall mastery, but a score of 3,568 on Real Words demonstrates 1st-grade Winter mastery.

There is always a danger when a continuous scale is interpreted in non-continuous ways. Score intervals should not be taken as absolutes but rather as general interpretive guides. A fraction of a point can move a student from one interval to another, even though the absolute difference between the two scores is trivial. In estimating the performance of a given student over time, calculations should therefore not utilize intervals or grade levels, which are discontinuous, but should use the continuous score continuum instead.

WACS is a robust indicator of a child's early-reading skill levels, but it is not appropriate as the sole measure, nor would it be appropriate for use as a high-stakes measure. This caution is especially warranted given the format of the test, which generally gives a child only three alternatives to choose from, resulting in a 33% chance of a false positive. Although guessing was not found to be a significant enough factor to warrant departing from a Rasch (one-parameter) model for the test as a whole (Shamir, Johnson, & Brown, 2009; see Embretson & Reise, 2000 for a more comprehensive overview of item-response theory), guessing can inflate scores at the individual level, suggesting that they should be interpreted with care.

WACS and Early-Reading ALT

With the first three requirements for operationalizing ALT satisfied by RRLWM, the fourth remains: to ensure that the time spent with the curriculum is spent on skills and content that are actually assessed. In this section the operationalization of WACS is detailed with respect to RRLWM.

Mapping RRLWM onto WACS

WACS was designed to assess the same early-reading skills that underpin the RRLWM curriculum. Pairing RRLWM and WACS is therefore desirable from the standpoint that much of the time spent using the curriculum should be relevant for the assessment and therefore would be suitable for inclusion in the overall calculation of ALT. In order to more fully explore the nature of this alignment, the curriculum objectives should be explicitly mapped onto each of the individual assessment tasks. The mapping of RRLWM's reading curriculum to WACS is detailed in Appendix C and Appendix D.

This mapping shows that there is indeed a sizable overlap between the curricular objectives and the skills that are assessed. This overlap is demonstrated by the fact that there are no skills that WACS assesses that are not taught by the reading curriculum (Appendix C). The overlap, however is not complete, as is shown by the lengthy list of objectives that ERP teaches but that WACS does not assess (Appendix D).

The lack of a complete overlap between the curriculum and the assessment suggests that the dose-response relation between RRLWM and WACS might be less efficient than it would be were they completely aligned (Anderson, 2002). If this is indeed the case, then it suggests that the pilot study might have underestimated the true strength of the dose-response relationship.

On the other hand, too close a fit between an assessment and a curriculum raises the specter that the test might not be an accurate metric of what a child actually knows. Although such concerns are minimal for skills such as letter recognition, where the content is comprehensive, they are heightened for certain skills, such as vocabulary, where the testing methodology necessarily involves sampling.

If for no other reason than the fact that they were created by the same company, there remains the question of whether RRLWM "teaches to the test" – in short, whether the test is to some extent invalid for the purposes of measuring the efficacy of the program because the two are so closely matched.

In order to address these concerns, as well as to provide for the wider interpretability of WACS scores, WACS was cross-validated during the 2008-2009 school year with a number of standardized paper-and-pencil early-reading tests, including Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Iowa Tests of Basic Skills (ITBS), and Stanford Achievement Test, Tenth Edition (SAT-10; see Shamir, Johnson, & Brown, 2009). For the cross-validation, students were given both WACS and one of these pencil-and-paper tests, and the overall scores on each test were then compared both in the fall and in the spring. Scores were found to correlate significantly ($p < .001$; r from .44 to .76) (see Appendix E for additional detail). The conclusion is that WACS is a reasonable proxy for other standardized early-reading assessments and is therefore not inappropriate for use with RRLWM.

The Use of WACS in the Early-Reading ALT Calculation

The selection of an assessment is crucial to the calculation of ALT, as only the time spent with those curricular elements that are ultimately assessed should be eligible for inclusion in the final calculation. In selecting an assessment, it is therefore important to ensure that it assesses the skills that the curriculum purportedly fosters.

Having mapped RRLWM to WACS, it is obvious that although all of the skills tested by WACS are taught to some extent by the reading curriculum, not all of the skills included in the curriculum are assessed. Unfortunately, as RRLWM does not log the time spent in each individual activity, it is not possible to exclude the time spent on these activities from the overall ALT calculation. Thus, for the pilot study, time spent using any of the activities (as opposed to just those that were tested) was included in the overall calculation. The danger in doing so was that it might dilute the strength of the relationship between usage and gain because the time spent in learning activities that were not assessed (and therefore should not be in the overall ALT calculation) would count as if it were being spent in activities that the children were tested on.

Summary

By combining the reading portion of RRLWM with WACS, early-reading ALT can be measured reasonably. Though this approach fails to include early-reading ALT that occurs in the traditional classroom, the home, and other formal and informal venues, it nevertheless comprises a starting point for investigating Berliner's model of learning as it relates to early-reading ALT. The result of operationalizing Berliner's definition of ALT in this way is that a calculation of ALT for a given curriculum is always assessment-dependent. Thus this operationalization is specific to the current version of WACS as the assessment instrument for the current version of RRLWM.

Part Two: The Pilot ALT Study

This section describes the implementation of the newly-operationalized measure of early-reading ALT in a pilot study. The study used dose-response methodology to examine the relation between computer-based early-reading ALT and early-reading achievement. Dose-response methodology has been used extensively in human pharmacological research (see Poling & Byrne, 2000, for an overview) and involves varying the dosage of a quantifiable variable over time and measuring the resultant changes in order to quantify the relationship between input (dose) and output (response).

Hypotheses and Limitations

The central hypothesis ("Hypothesis 1") of the pilot study was that increases in computer-based early-reading ALT (dose) will result in corresponding increases in early-reading achievement (response). A second hypothesis ("Hypothesis 2") was that the use of the sequenced portion of the reading software will be more predictive of early-reading skill achievement gains than the unsequenced free-play portion. The final hypothesis ("Hypothesis 3") was that early-reading ALT will be less related to other computer-based reading-related achievement gains.

A major limitation in the attempt to model these relations was that achievement was measured at only two points in time. This necessitated a linear model (Willett, 1989), which might be inconsistent with learning-growth models (Walberg, 1981). Another limitation was the inability to model reductions of dose level. Thus, although in a pharmacological context a drug can be introduced and then removed (for instance, through metabolic processing), there is no obvious analogy when it comes to acquiring an academic skill.

Another limitation that is endemic to ALT research is the tautological nature of the construct itself. Any variable that facilitates learning – a better presentation, prior knowledge, exceptional aptitude – lowers the amount of ALT required for that specific learning task for that specific student at that particular point in time. The converse, of course, is also true – poor instructional presentation, lack of necessary familiarity with the topic at hand, or low aptitude all raise the required ALT. The all-encompassing nature of the construct makes ALT highly context-specific and therefore protean.

These challenges are similar in their nature to technology's versioning problem that was referenced in the introduction. One strategy suggested by Fletcher (2003) was to focus on the principles that research yields over time as opposed to getting bogged down in the specifics of a particular instantiation of technology. This study employed a similar tactic – rather than focusing solely on individuals, the study combined performance (as measured at the individual level) into an overall average effect. Thus, while ALT was approximated for each individual child, the net effect of increasing ALT was addressed on a much larger scale. The advantage of this approach is that it helps to mute idiosyncratic individual differences. The disadvantage is that it assumes that the resulting measure of ALT commonly applies to each child – that a "better presentation" has the same meaning for each child (i.e., that the variable transcends individual contexts).

Another limitation was the lack of a randomized control group. In some ways, this limitation is muted when dealing with ALT because ALT is presumed to vary by child by task (due to aptitude level, prior learning history, etc.), rendering the construction of an appropriate control group problematic at best. Nevertheless, the absence of a control group made it more difficult to eliminate alternative explanations for achievement growth based on developmental maturation or other confounds.

Finally, it is important to reemphasize that this study did not attempt to quantify all instances of early-reading ALT. Children involved in the study received early-reading instruction from a variety of formal and informal sources with differing levels of regularity. In seeking to establish whether there is a relationship between computer-based early-reading ALT and early-reading achievement scores, the pilot study explicitly set aside the task of formally modeling or controlling for other sources of early-reading ALT and their impact on early-reading achievement.

Method

The data for this study were generated in connection with the Utah Preparing Students Today for a Rewarding Tomorrow or UPSTART program. The UPSTART program was created by Utah House Bill 200 (http://le.utah.gov/~2008/bills/hbillint/HB0200.htm), which established "a pilot project known as UPSTART which uses a home-based educational technology program to develop school readiness skills of preschool children" (lines 14-16). The stated intent of the UPSTART program is to:

> (a) evaluate the effectiveness of giving preschool children access, at home, to interactive individualized instruction delivered by computers and the Internet to prepare them academically for success in school; and
>
> (b) test the feasibility of scaling a home-based curriculum in reading, math, and science delivered by computers and the Internet to all preschool children in Utah. (lines 66-70)

The use of these data in this dissertation was in no way intended to be an evaluation of the UPSTART program, which has its own goals, not to mention its own independent evaluators (lines 21-22). Rather, these data were used because they had the attributes necessary for investigating the link between early-reading ALT and early-reading achievement.

Recruitment

The UPSTART program was located entirely within the state of Utah and targeted 4- and 5-year-old children who had not yet attended kindergarten (lines 54-56 of the bill). Extensive marketing efforts were undertaken to recruit and enroll children. Methods included flyers (see Appendix F), public service announcements, newspaper advertisements, Google AdWords, and direct contact with superintendents from all Utah school districts. Marketing materials were translated into Spanish, and enrollment was similarly available in English and Spanish.

Special efforts were made to reach lower-income and minority populations. For instance, flyers were sent to 193 food pantries, 8 United Way locations, 43 used clothing shops, 110 library locations, 1,007 health-clinic locations, 45 State Department of Health locations, 290 commercial child-care locations, 337 Boys & Girls Club locations, 56 Native American tribe locations, 927 home child-care locations, 17 university child-care locations, and 89 children's places locations (e.g., children's clothing and bookstores, toy stores, etc.). In addition, to help ensure that lower-income children could participate, the UPSTART program included funding for up to one-third of the participants in the program (lines 139-140) for the purchase and installation of computers and internet access for "families that cannot afford the equipment and service" (lines 82-83).

Selection

Due to limited funding, not every child who expressed interest could be admitted to the program. Admission was contingent both on geography (defined both by school district and by the rural / urban designation) and on income (with 200% of the Federal Poverty guidelines as the cutoff-point for low income). When equally-qualified children competed, admission occurred on a first-come, first-served basis. Enrollment in the initial phase of the program was opened in April, 2009 and remained open until all available slots were filled. Children were successfully enrolled from every Utah school district. The demographic summary of the children who enrolled are shown in Table 1.

Table 1
Summary of Demographic Information for UPSTART Participants

	InPreK	Gender (M/F)	LowIncome	English	NonWhite
UPSTART	47%	54% / 46%	61%	92%	20%

Program Description

Parents qualified and their children were enrolled over the phone in a conversation that utilized a common script that outlined the details of the program and that asked for specific demographic information (see Appendix G for details of the demographic coding). Qualifying participants in the program were provided with the necessary hardware, software, and internet access. Initial parent training and program information was available on the official website (http://www.utahupstart.org/index.html). In addition, 35 regional “town hall” training sessions were held during June-August, 2009 (see Appendix H for an overview of the training).

Assessment

Parents were instructed to ask their children to take WACS prior to beginning the RRLWM Reading Level 1 curriculum. Parents were further instructed not to help their children during the test. In cases where extremely high scores were received, parents were contacted and, if help had been provided during the test, those parents were asked to ask their children to retake the assessment alone, as it was assumed that parental help had biased the test results. In those cases, the original score was replaced with the second score.

Prior to exiting UPSTART for kindergarten (for 5-year-olds) or after approximately six months in the program (for 4-year-olds), parents were again instructed to ask their children to take WACS. Although parents were again instructed to not help their children during the test, there was no way to effectively control for such an occurrence. The cutoff date for inclusion in the dataset was December 22, 2009.

There was no programmatic way to enforce the administration of WACS, and consequently not all children participated. In cases where WACS was not administered within the expected timeframe, parents repeatedly were asked both via e-mail and phone to administer WACS, but children were not dropped from the program for not completing WACS.

Program Implementation

Parents were instructed to ask their children to use the sequenced portion of the Reading portion of RRLWM for 15 minutes per day, 5 days per week. Use of the Math & Science curriculum and of the free-play mode was optional. Usage was measured by the reading program software, and these data, along with WACS scores, were collected using the internet.

With rare exceptions, children began the curriculum without using the placement tool and therefore started the curriculum at the beginning of Level 1. The placement tool was used to override this initial placement only when children scored exceptionally high on the initial WACS test and when parents confirmed the validity of the score.

Substantial efforts were made to encourage and motivate both the participating parents and children during the course of the UPSTART program. Parents received a customized e-mail every Friday indicating whether usage of the reading software during that week had been sufficient or not (see Appendix I for two examples). Families where children began to show a pattern of lack of use were contacted to determine their level of commitment. Those who were unable to improve to acceptable levels were asked to return the materials so that other children could benefit from the program.

Each month children received certificates in the mail that heralded their progress through the curriculum or encouraged them to spend more time in the coming month (see Appendix J for an example of each). Children who used the structured reading curriculum for less than 60 minutes received the “We missed you!” letter, while medals were awarded to those who had usage of 61-385 minutes (Bronze), 386-642 minutes (Silver), or over 642 minutes (Gold). In other words, assuming 31 calendar days in a month, children were sent a reminder letter for less than 1.94 minutes of daily use or received a Bronze medal for 1.94 – 12.4 minutes, a Silver medal for 12.4 – 20.7 minutes, or a Gold medal for more than 20.7 minutes of daily use.

Survey

Dan Jones & Associates (www.djasurvey.com) was contracted by the Waterford Institute to conduct a survey of UPSTART participants. The survey was conducted during December, 2009 and involved 321 current and recently-exited participants. The margin of error of the study was +/- 4.65%. Complete scripts and results for selected survey items are included in Appendix K.

Hypotheses

Hypothesis #1. Computer-based early-reading ALT should result in early-reading gains. UPSTART was intended to academically prepare prekindergarteners for entry into kindergarten. Thus, children generally started the program at the beginning of the Level 1 reading curriculum. There are four principal early-reading skills that children are taught and held accountable for in Level 1: Letter Recognition, Letter Sounds, Initial Sound, and Blending. In order to summarize the impact of the program on these early-reading skills, a new variable (Level1Gain) was created by averaging the gains across the four skills. Based on Carroll’s and Berliner’s theories, I therefore hypothesized that the time spent using Level 1 would be correlated with gains on these four skills and on the summative variable Level1Gain.

Hypothesis #2. I hypothesized that time spent using the unstructured portion of the curriculum (designated FreePlayUsage) would not be as strongly predictive of gains. Such a finding would be expected because FreePlayUsage was a measure of time spent in activities that were (a) not instructional, and (b) had already been seen (and potentially mastered). As such, it was expected that these activities would not be as educationally efficient as the structured activities, and therefore the relationship between FreePlayUsage and achievement would not be as robust (see Johnson, Perry & Shamir, in press, for related results).

Hypothesis #3. In connection with Berliner's assertion that learning does not occur without a change in ALT, I hypothesized that children in UPSTART would not show comparable gains in skills they were not directly taught or held accountable for. Thus, there should be a lower correlation between usage of Level 1 and gains in Vocabulary, Reading Comprehension, Listening Comprehension, Nonwords, Sight Words and Real Words.

Paring

The evaluation of each hypothesis began by using as much of the available data as possible. Where appropriate, the data were modified in an attempt to eliminate biases that may have obscured underlying patterns in the data. An example of such modification would be when children below a certain threshold score were excluded. Whenever this occurred, the reasoning for the modification was provided, and the new size of the sample (N) was reported.

After the modification occurred, the remaining data were labeled "pure" with respect to that particular factor. For example, if all children who had worked on the later levels of the reading program were excluded from the analysis, the remaining sample was termed

"Level1Pure". In cases where tables or figures used a particular subsample of children, the designation of the subsample is included in parentheses at the end of the title.

The "pure" designation was used because the analyses were frequently multi-step, and it otherwise was easy to lose track of the steps taken to produce a given sample of children. In addition, not only were the individual analyses themselves multi-step, but some analyses required samples of children from different analyses to be crossed, as, for example, in the usage-to-gain (i.e., dose-response) analysis.

This approach of modifying the sample size through the use of specific rules was planned from the outset of the study, but the actual rules and groupings that made sense for the resultant dataset could not be entirely anticipated. As such, paring is necessarily a post-hoc undertaking.

Supplementary Investigations

In addition to the core hypotheses, a number of supplemental investigations were undertaken during the data analysis phase. It is important to note that these were post-hoc investigations and as such need to be viewed as exploratory and not confirmatory.

Compliance. Another way to try to expose the hypothesized usage-to-gain relation was to measure the fidelity of the RRLWM implementation itself. This notion of compliance aimed at capturing how closely the curriculum was implemented with respect to the intent of the intervention organizer—specifically, the use of the structured reading curriculum for 15 minutes per day, 5 days per week. Due to the rolling nature of enrollment in UPSTART, there was no consistently recommended number of days in the program. Therefore, the analysis of compliance focused on minutes per day, which was calculated by dividing the total usage by the total number of calendar days spent in the program.

To explore the impact of compliance on the dose-response relation, each child's usage was sorted into one of three categories – below recommended, recommended (i.e., within a specified deviation from the recommended level), or exceeding recommended. In order to construct these three categories, the specified deviation from the recommended level was allowed to vary between 5% and 95% in increments of 5%, resulting in 19 distinct compliance levels. A compliance level of 70% indicated that usage below 70% was undercompliant (30% under recommended), usage between 70% and 130% compliant, and usage over 130% overcompliant.

Each of the 19 levels of deviation was then entered as a second step in the usage-to-gain regression analysis in order to discover the level that was the most statistically significant. Children were then coded using this level of deviation as either undercompliant (-1), compliant (0), or overcompliant (1) with respect to their usage of the program, and this variable was included in subsequent analyses.

Progress monitoring. As children used the reading curriculum, the software assessed their mastery of the various learning objectives. Level 1 of the structured reading curriculum contains 380 such objectives. As previously noted, the mastery of an objective (as demonstrated by a child's score on an activity assessment) allows a child to progress through the curriculum, while a lower score results in remedial sequencing. The number of objectives that are mastered is therefore an indication of overall progress.

In this way the number of discrete learning objectives that a child ultimately masters could function as a moderating variable in the usage-to-gain relationship. Among possible reasons why a moderating relationship might not exist is that not all of the learning objectives take the same amount of time to complete or have the same amount of educational import. In order to explore the effect of this variable on program effectiveness, the number of objectives mastered was included as a predictor variable for gains in early-reading skills (Level1Gain).

Model of best fit. The final supplemental investigation took the hypotheses and the exploratory analyses and attempted to trade them off against one another in an effort to discover a best-fitting model. This was a data-driven as opposed to a theory-driven analysis.

The methodology for this investigation was relatively straightforward. Variables of interest were added stepwise to an overall regression to determine their relative abilities to predict overall variance. Where multiple variables were possible, the order in which the variables were entered was similarly varied to allow for all possible combinations to be explored. The determination of the overall best-fitting model was based on a combination of ability to predict the dependent variable and overall parsimony.

Study Design

Kazdin (2003) provides a useful framework for evaluating a research design in terms of four types of validity: internal, external, construct, and statistical conclusion (see Table 2). These measures are not binary in nature, making it more useful to talk about the strengths and weaknesses of the proposed design with regard to each type of validity as opposed to whether the design includes a particular type.

Table 2

Types of Experimental Validity and the Questions They Address

Type of Validity	*Questions Addressed*
Internal Validity	To what extent can the intervention, rather than extraneous influences, be considered to account for the results, changes, or group differences?
External Validity	To what extent can the results be generalized or extended to people, settings, times, measures, and characteristics other than those in this particular experimental arrangement?
Construct Validity	Given that the intervention was responsible for change, what specific aspect of the intervention or arrangement was the causal agent, that is, what is the conceptual basis (construct) underlying the effect?
Statistical Conclusion Validity	To what extent is a relation shown, demonstrated, or evident, and how well can the investigation detect effects if they exist?

Source: Kazdin, RESEARCH DESIGN IN CLINICAL PSYCHOLOGY, Table 2.1 p.23, © 2003. Reprinted by permission of Pearson Education, Inc.

The formal implementation of computer-based early-reading ALT as a dose variable for early-reading student achievement represents a new vista in educational research. For such cases Kazdin (2003) notes the importance of prioritizing internal validity over external validity (p. 51).

Internal validity. In Carroll's and Berliner's theories, the only independent variable of interest is instructional time. In the present study children were only allowed to participate in the program until they started kindergarten. For that reason, I assumed that during the study only small amounts of formal reading instruction were likely to be provided by sources outside of the curriculum itself. To control for any schooling effects that might nonetheless arise, enrollment in preschool was included as part of the demographic analysis.

External validity. A significant threat to the external validity of the pilot study was the demographics of the population involved. Although the demographics were representative of the population of Utah, they were not representative of the nation as a whole. In particular, sample sizes were not sufficient for the various subcategories of ethnicity to be evaluated. The ability to predict the effect of this program on populations with stronger representations of these subcategories is therefore blunted.

That being said, the program itself is eminently replicable. Indeed, replicability is one of the hallmarks of technological solutions (Heuston, 1996). Thus, although some of the particulars (e.g., demographics) may be difficult to replicate, the core instructional and assessment portions of the study are not.

Construct validity. As noted earlier, one potentially confounding issue was the amount of reading instruction that a child received outside of the program. The most obvious potential source of formal reading instruction was participation in a preschool. This concern was somewhat mitigated by the fact that the majority of the time spent using RRLWM was during the summer months, but it was a threat to construct validity nonetheless, and participation in preschool (designated as the variable InPreK) was therefore one of the demographic variables tracked in the study and investigated in the analysis.

Another likely source of reading instruction was the child's parents. This variable was not controlled for.

Statistical conclusion validity. A standard approach to statistical validation is Cohen's (1992) method of statistical power analysis:

> Statistical power analysis exploits the relationships among the four variables involved in statistical inference: sample size (N), significance criterion (α), population effect size

> (ES), and statistical power. For any statistical model, these relationships are such that each is a function of the other three. For example, in power reviews, for any given statistical test, we can determine power for given α, N, and ES. (p. 156).

For the purposes of this study, the standard tolerances of $\alpha = .05$ and power = .80 were used. Based on Cohen's Power Table (p. 158), group sizes of 393, 64, and 26 were required to detect an ES that is small, medium, or large, respectively.

Magnitude questions

In looking at dose-response, it is important to ask not just whether there are gains, but what the magnitude of the gains is. The most straightforward way to analyze gains is in terms of the measure itself, that is, if children gained an average of 34 points on WACS, what does that signal about their early-reading ability?

Although such an analysis may be worthwhile, it falls short of placing the results of the study in a larger context. Briefly, it is difficult to say how such gains would compare to gains achieved using other interventions that also measured growth but that used a different metric than WACS.

One way to address this concern is to calculate an effect size (see Cohen, 1992), which uses the variability of the data itself as a measuring stick, thereby allowing results from very different contexts to be compared (Walberg, 2003). A standard effect-size measure is Cohen's D, which represents the difference in means between two groups divided by the standard deviation.

This measure can be calculated in a variety of ways, and I have included three of them. The first was Glass's delta, which uses the standard deviation of the pre-test as the denominator.

The second used a pooled standard deviation. Both of these methods are regularly used to calculate effect sizes for independent groups, or, in other words, for groups whose scores are not expected to be correlated (e.g., between a treatment and a control group). In the pilot study, however, the group that took the pre-test and the group that took the post-test were composed of the same individuals, and therefore their scores were expected to be correlated. This correlation between pre- and post-test scores can result in either of the independent methods underestimating the true effect size.

To address this concern, an online repeated-measures effect-size calculator (Cohensdrepeatedmeasures.xls), provided by James Neill's tutorial on effect sizes (http://wilderdom.com/courses/surveyresearch/tutorials/5/), was used to calculate the third and final effect size. While I report effect sizes obtained using all three methods, the repeated-measures approach is considered the most accurate, given the design of the pilot study.

Although the calculation of the effect sizes is statistically correct, it is difficult to know what, if any, gains might have been made without the introduction of the curriculum. From a theoretical standpoint, unless ALT is applied, gains should not occur. Therefore, if the assumption is correct that the early-reading ALT for these children outside of UPSTART is minimal, then the lack of a control group should not be a major concern within the model. Unfortunately, without a control group or a valid norm group, this assumption could not be explicitly tested.

Results

The results are presented in four sections. The first reports results related to RRLWM usage. The second details changes in early-reading achievement scores. The third provides results related to each of the three hypotheses. The fourth reports the supplementary analyses of compliance, progress monitoring, and an overall best-fitting model.

Usage Results

Program usage was calculated between the first and second administrations of WACS. Of the 1,073 children who took the pre-test and 849 who took the post-test, 785 children took both tests. The two administrations were separated by an average of 134.4 (S.D. = 30.86) calendar days. As previously noted, due to the rolling nature of the admissions, there was no specific recommended number of calendar days spent with the curriculum. One child took WACS but did not use the reading curriculum, leaving 784 children for potential inclusion in the usage analysis.

These 784 children were roughly half of the 1,347 children who originally enrolled in the study. Such large changes in sample size can often be a source of bias in analyzing the results. In order to better understand the demographic changes brought about by this change in size, a CROSSTABS analysis was run to compare the groups. The results appear in Table 3.

Table 3

CROSSTABS Analysis of Pre-Post Sample vs. Original Sample Demographics

Demographic Variable	Pearson Chi-Square	df	Asymptotic Significance (2-sided)	Directional Effect
NonWhite	42.49	1	.000***	Fewer NonWhite
InPreK	2.64	1	.104	None
LowIncome	8.27	1	.000***	Fewer Low Income
Gender	1.66	1	.197	None
English	61.97	1	.000***	Fewer Non-English

*p <= .05; **p<=.01; ***p<=.001

The results indicate that there were no statistically significant differences between the two samples in terms of level of participation in preschool or gender, but there were proportionally fewer non-white children, fewer low-income children, and fewer children who do not speak English at home in the pre/post sample as compared to those excluded from the sample.

As described earlier, curriculum usage could be defined in terms of unstructured free play and structured reading. The free-play activities were used on average 588.1 (S.D. = 892.8) minutes across the 134.4 days, or 4.38 minutes per day. The huge standard deviation indicated an extremely wide range of data values and the possibility that outliers might be present. SPSS's EXAMINE command identified 44 extreme cases (identified as lying outside the one-and-a-half interquartile range). After removal of these cases, free-play activities were used by FreePlayUsagePure children an average of 445.3 (S.D. = 465.5) minutes or 3.31 minutes per day. The continued large standard deviation is a function of the large number of children who used free play for minimal amounts of time – of the 740 children in the sample, 50 did not even use it at all (see Figure 1).

Figure 1.
Histogram of FreePlayUsage (FreePlayUsagePure)

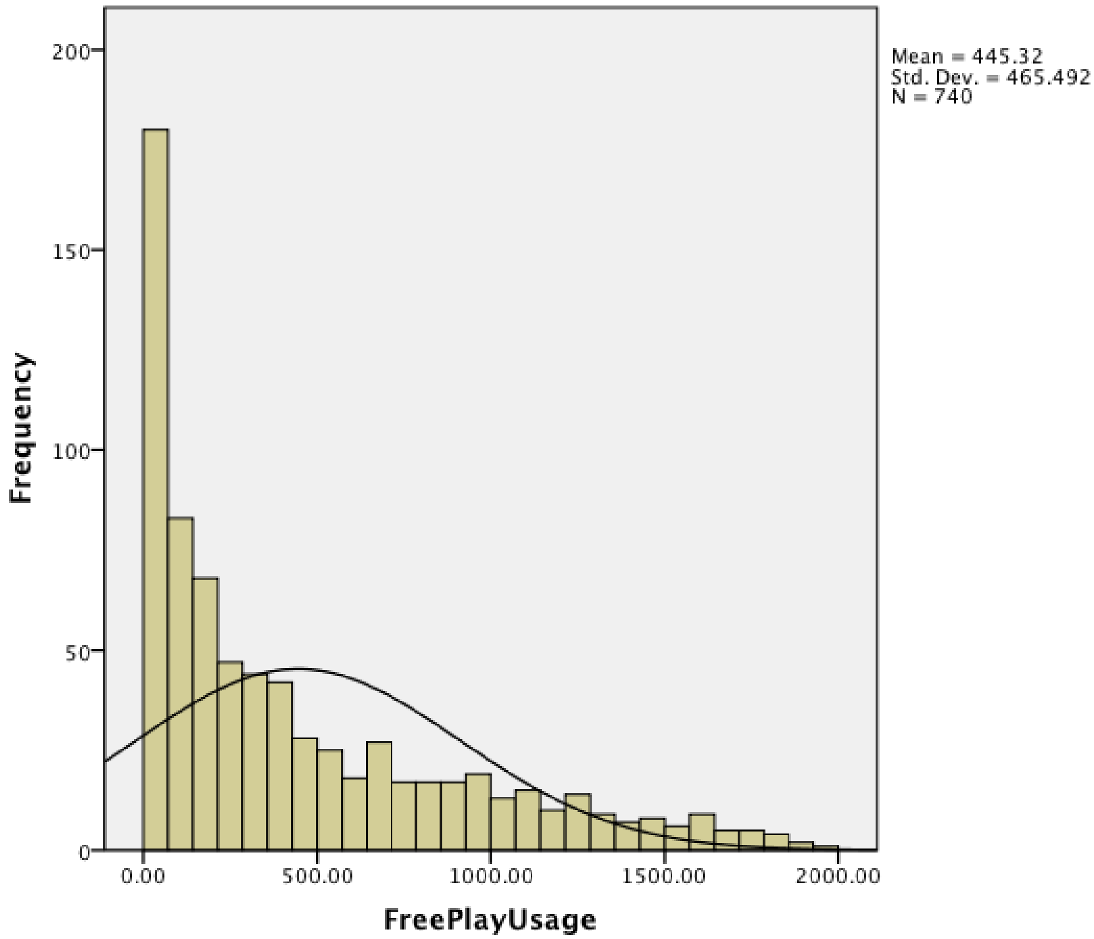


Figure 1. FreePlayUsage is measured in minutes. Frequency is measured in number of children.

The non-normality of the distribution is demonstrated by the wide discrepancy between the shape of the histogram and the superimposed normal curve.

The second definition of usage was time spent using the structured reading curriculum. Overall usage of all three levels of the curriculum was represented by the variable ReadingUsage which had a mean of 1,950 (S.D. = 955.4), or 14.51 minutes per calendar day. The recommended use of the sequenced reading program was 15 minutes per day, 5 days per week or 10.71 minutes per calendar day. Combining the ReadingUsage and ProgramDays variables resulted in a proxy value of the recommended usage for the average child of 10.71 minutes * 134.4 days or 1,440 minutes.

The standard deviation of ReadingUsage, while not as extreme as that for the free-play portion of the curriculum, was nonetheless quite high—EXAMINE identified 23 extreme cases. After removal of these cases, the resulting sample (N=761) used the reading curriculum an average of 1,855 (S.D. = 769.6) minutes or 13.80 minutes per day, roughly 30% more than the recommended proxy value (see Figure 2).

Figure 2.
Histogram of ReadingUsage (ReadingUsagePure)

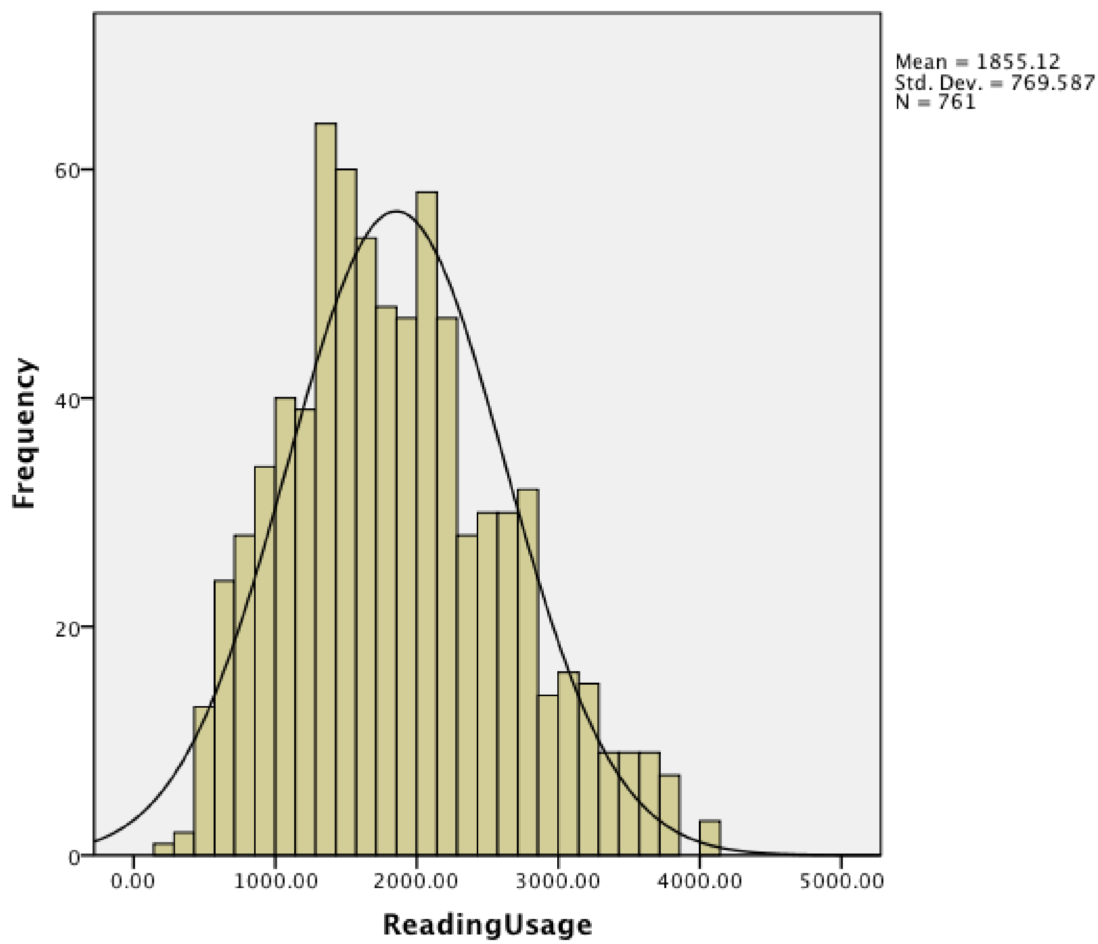


Figure 2. ReadingUsage is measured in minutes. Frequency is measured in number of children.

Unlike the FreePlayUsagePure results, the ReadingUsagePure results display a close approximation to a normal distribution.

A more targeted way to define the reading curriculum is simply as “early-reading instruction” and therefore to only look at usage of Level 1 (a variable labeled Level1Usage). Using the ReadingUsagePure data, the average child used Level 1 for 1,768 minutes (S.D. = 735.6) or 13.16 minutes per calendar day—almost 25% more than recommended. The results appear in Figure 3.

Figure 3.
Histogram of Level1Usage (ReadingUsagePure)

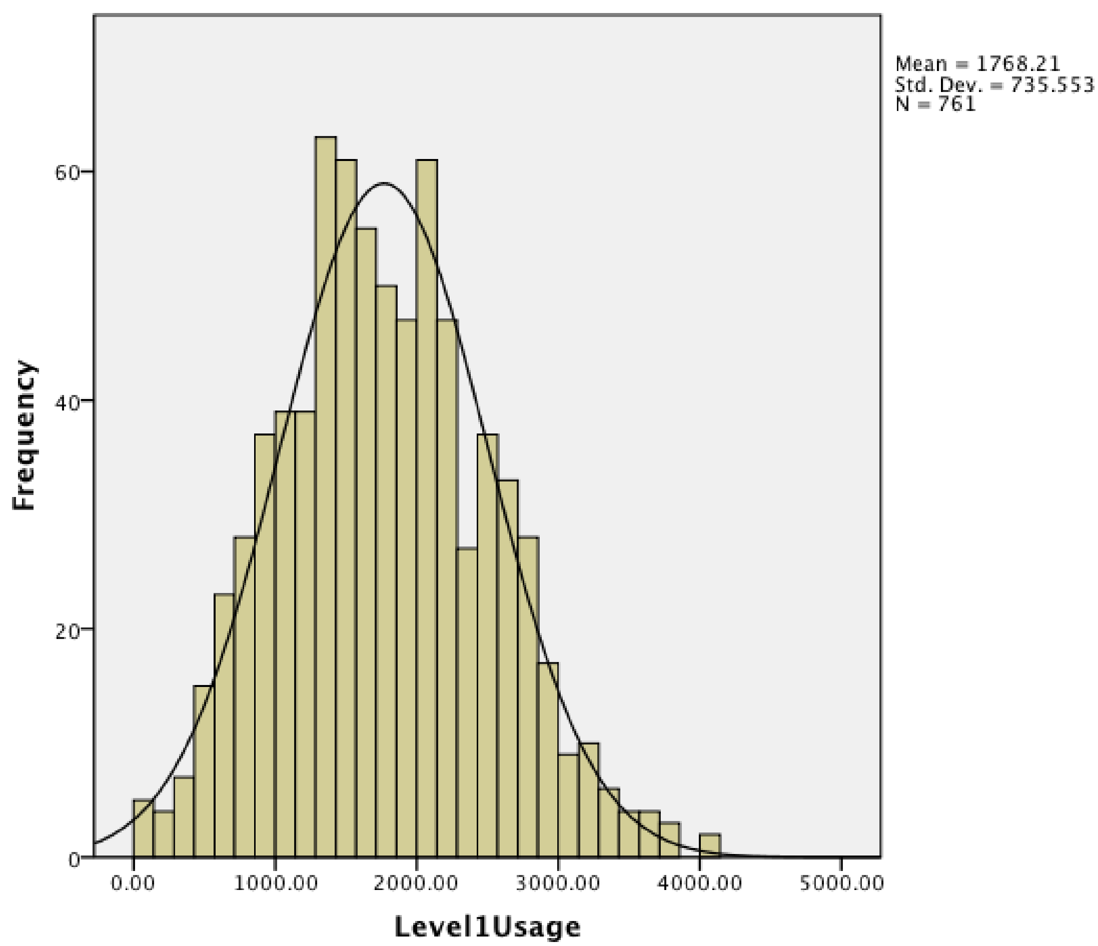


Figure 3. Level1Usage is measured in minutes. Frequency is measured in number of children.

To gain a clearer sense of “average usage,” extreme usage of either the free-play or the reading portion of the curriculum disqualified children from the remaining analysis in this section. This crossing of FreePlayUsagePure with ReadingUsagePure resulted in a “UsagePure” sample (N=725), whose usage summary appears in Table 4.

Table 4
Usage Statistics (UsagePure)

	FreePlay	Overall Reading	Level 1 Reading	Program Recommended	Program Days
Mean	435	1,822	1,736	1,426	133
S.D.	456	754	721		31
Avg./Day	*3.23*	*13.55*	*12.92*	*10.71*	

UsagePure children used the reading curriculum almost 28% more than recommended, and used Level 1 almost 22% more than recommended. In addition, approximately 20% of a child’s overall usage (FreePlayUsage + ReadingUsage) was spent in free-play activities. Of the time that UsagePure children spent in the structured portion of the reading curriculum, 1,736/1,822 or 95% of the time was spent using Level 1.

The number of children categorized as UsagePure (N=725) was roughly half of the original enrollees (N=1,347). A CROSSTABS analysis (see Table 5) was run with UsagePure coded as a dummy variable (0 = not UsagePure; 1 = UsagePure).

Table 5
CROSSTABS Analysis of UsagePure Grouping Effects on Demographic Variables

Demographic Variable	Pearson Chi-Square	df	Asymptotic Significance (2-sided)	Directional Effect
NonWhite	45.82	1	.000***	Fewer NonWhite
InPreK	.31	1	.577	None
LowIncome	13.82	1	.000***	Fewer Low Income
Gender	1.60	1	.206	None
English	64.39	1	.000***	Fewer Non-English

*p <= .05; **p<=.01; ***p<=.001

The results indicate that there were no statistically significant differences between the two samples in terms of level of participation in preschool or gender, but there were proportionally fewer non-white children, fewer low-income children, and fewer children who do not speak English at home in the UsagePure sample as compared to the non-UsagePure sample.

Gains Results

Of the 785 children who took WACS both as a pre- and as a post-test, 4 children took the kindergarten version of WACS and did not take the necessary subtests, leaving 781 children for inclusion in the gains analysis.

The children's pre- and post-test scores on the four Level 1 subtests as well as their combined average Level 1 score appear in Figures 4 - 8. To aid in interpretation, the scores are reported in intervals of 333.33 points. The ranges shown on the histograms represent the possible intervals on both the pre- and the post-tests, not just the intervals in which children actually scored (i.e., for each subtest, there were no intervals lower or higher than those presented in the histograms).

Children's pre- and post-test scores for the four Level 1 subtests and the combined Level 1 score along with the results of paired T-tests appear in Tables 6 - 10.

Letter recognition results.

Figure 4.
Letter Recognition Pre- and Post-Test Results

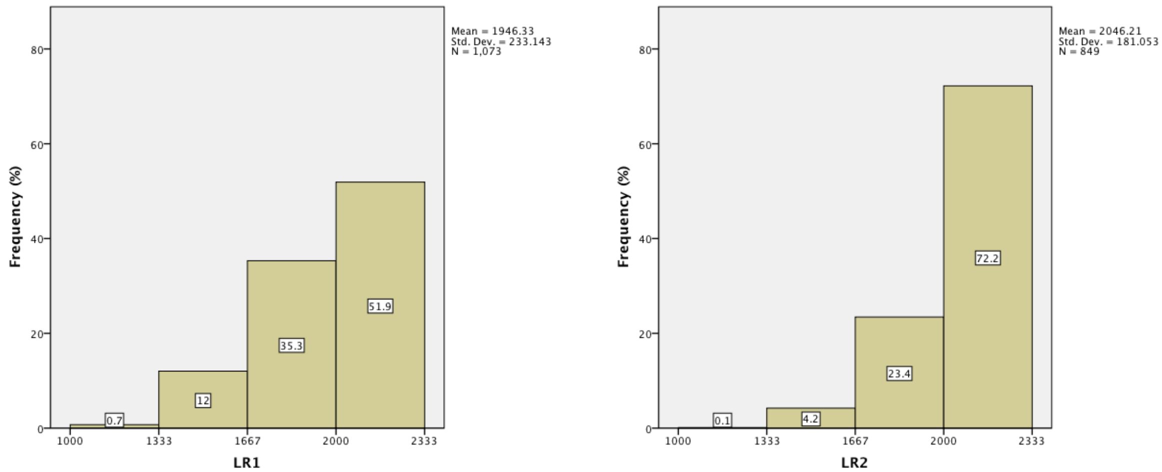


Figure 4. LR1 = WACS Letter Recognition pre-test score. LR2 = WACS Letter Recognition post-test score. Frequency is measured as percentage of children.

Table 6
Letter Recognition Pre- and Post-Test T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LR2	2,052.80	781	172.59	6.17
	LR1	1,957.05	781	223.40	7.99

Paired Samples Test

	Paired Differences							
				95% CI of the Difference				
	Mean	S.D.	Std. Error Mean	Lower	Upper	T	df	Sig. (2-tailed)
LR2 - LR1	95.745	230.597	8.251	79.548	111.943	11.604	780	.000

Children's scores grew significantly ($p < .000$) in Letter Recognition across their time in the reading program, averaging a gain of just under 96 points. However, the results appear to indicate the presence of a potential ceiling effect, with over half of the children scoring in the highest interval on the pretest.

Letter sound results.

Figure 5.
Letter Sound Pre- and Post-Test Results

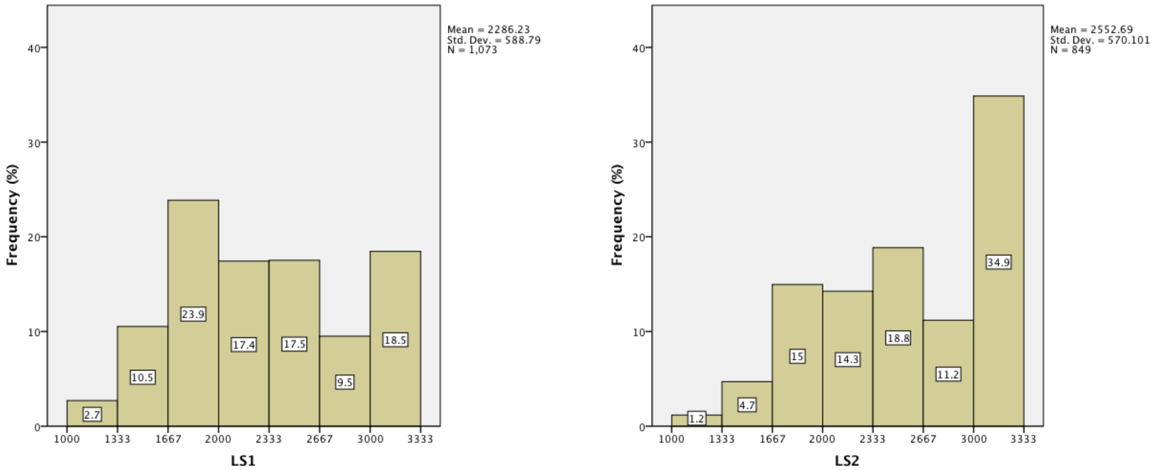


Figure 5. LS1 = WACS Letter Sound pre-test score. LS2 = WACS Letter Sound post-test score. Frequency is measured as percentage of children.

Table 7
Letter Sound Pre- and Post-Test T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LS2	2,570.53	781	560.93	20.07
	LS1	2,296.22	781	580.41	20.77

Paired Samples Test

	Paired Differences							
				95% CI of the Difference				
	Mean	S.D.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
LS2 - LS1	274.306	656.592	23.495	228.186	320.426	11.675	780	.000

Children grew significantly ($p < .000$) in their mastery of Letter Sounds, averaging a gain of 274 points.

Initial sound results.

Figure 6.
Initial Sound Pre- and Post-Test Results

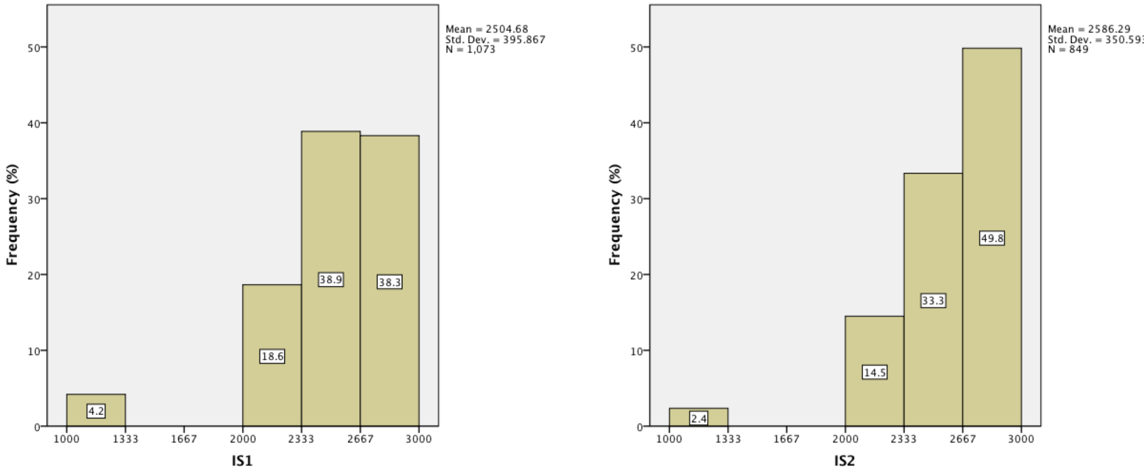


Figure 6. IS1 = WACS Initial Sound pre-test score. IS2 = WACS Initial Sound post-test score. Frequency is measured as percentage of children.

Table 8
Initial Sound Pre- and Post-Test T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	IS2	2,596.04	781	342.95	12.27
	IS1	2,499.05	781	408.85	14.63

Paired Samples Test

	Paired Differences							
				95% CI of the Difference				
	Mean	S.D.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
IS2 – IS1	96.987	478.628	17.127	63.367	130.607	5.663	780	.000

Children grew significantly ($p < .000$) in their ability to discriminate the Initial Sound in words, with an average gain of roughly 97 points. The histograms suggest a potential ceiling effect, with roughly 40% of children scoring in the highest interval on the pre-test.

Blending results.

Figure 7.
Blending Pre- and Post-Test Results

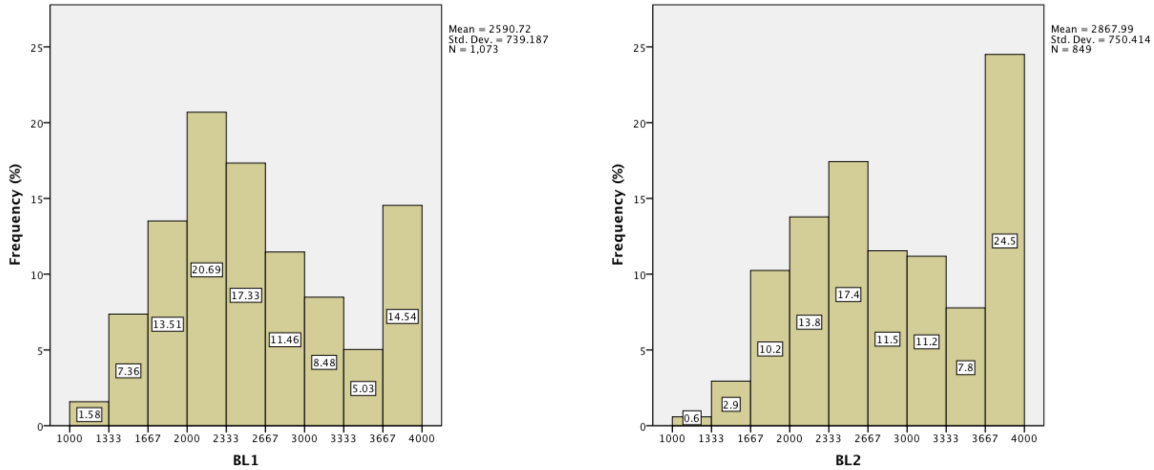


Figure 7. BL1 = WACS Blending pre-test score. BL2 = WACS Blending post-test score. Frequency is measured as percentage of children.

Table 9
Blending Pre- and Post-Test T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BL2	2,897.90	781	743.71	26.61
	BL1	2,598.96	781	738.96	26.44

Paired Samples Test

	Paired Differences							
				95% CI of the Difference				
	Mean	S.D.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
BL2 - BL1	298.936	936.740	33.519	233.137	364.734	8.918	780	.000

Children gained significantly in Blending ($p < .000$), gaining just under 300 points on average.

Overall level 1 results.

Figure 8.
Level 1 Combined Skill Pre- and Post-Test Histograms and T-Test

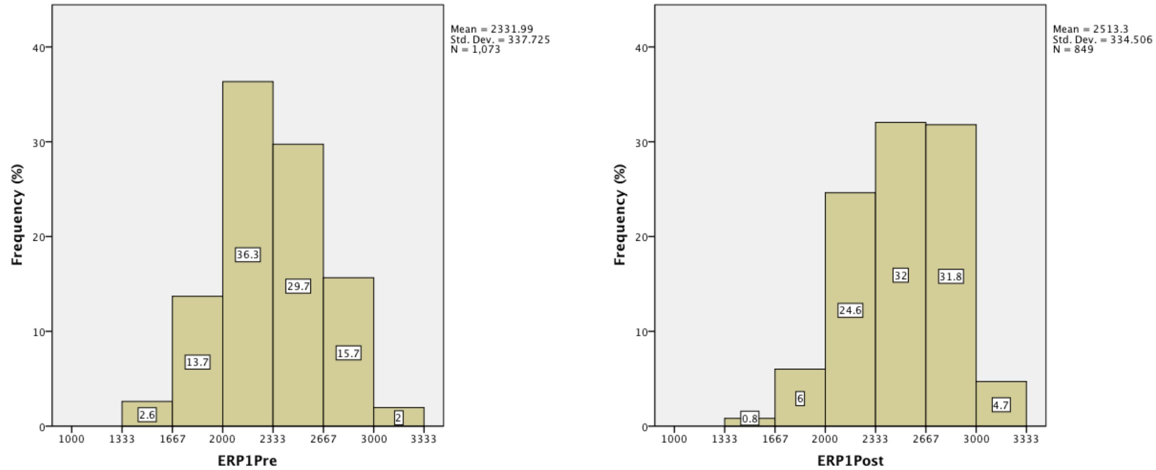


Figure 8. ERP1Pre = WACS Level 1 combined pre-test score. ERP1Post = WACS Level 1 combined post-test score. Frequency is measured as percentage of children.

Table 10
Level 1 Combined Pre- and Post-Test T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Level1Post	2,529.32	781	329.29	11.78
	Level1Pre	2,337.82	781	335.30	12.00

Paired Samples Test

	Paired Differences							
				95% CI of the Difference				
	Mean	S.D.	Std. Error Mean	Lower	Upper	t	Df	Sig. (2-tailed)
Level1Post – Level1Pre	191.494	355.674	12.727	166.510	216.477	15.046	780	.000

The minimum average score possible across the four Level 1 subtests was 1,001, and the maximum average score possible was 3,167. The average score across all four Level 1 skills grew significantly ($p < .000$), averaging 191 points. Table 11 details the correlations between the individual early-reading skills and the combined early-reading measure.

Table 11
Correlations of Early-Reading Individual and Summative Skills

	LR1	LS1	IS1	BL1	Level1Pre
N	*1073*	*1073*	*1073*	*1073*	*1073*
LR1	1	.369***	.254***	.195***	.515***
LS1	.369***	1	.258***	.277***	.727***
IS1	.254***	.258***	1	.246***	.584***
BL1	.195***	.277***	.246***	1	.774***
Level1Pre	.515***	.727***	.584***	.774***	1

*p <= .05; **p<=.01; ***p<=.001 (2-tailed)

Each of the four subtests is highly ($p < .000$) correlated to each of the other subtests, as well as to the early-reading summative measure, suggesting that Level1Gain is a reasonable summative measure for early-reading skills.

WACSPure. Figures 4 and 6 pointed to the possibility of a ceiling effect, that is, many of the students who took the pre-test received a score in the highest possible interval. One way to avoid the potential growth measure bias that such ceilings can produce would be to ensure that only students who had the opportunity for substantial growth in each skill was included in the gain analysis.

To construct such a sample, children were included only if they scored below the highest interval on the pre-test in each of the four skills taught by Level 1. Thus, each child would have had the opportunity to gain at least 333.33 points during the course of the program. The sample of 231 children in this category was labeled “WACSPure.” A MEANS report comparing the WACSPure and non-WACSPure samples as well as the original sample appears in Table 12.

Table 12
Individual and Summative Early-Reading Skill Gains by WACSPure

Group		LRGain	LSGain	ISGain	BLGain	Level1Gain
Non-WACSPure	Mean	31.82	238.71	53.17	225.40	137.27
(*N = 550*)	S.D.	204.55	641.89	448.41	963.54	339.36
WACSPure	Mean	247.95	359.06	201.32	474.03	320.59
(*N = 231*)	S.D.	217.76	684.33	530.49	846.18	361.05
All Children	Mean	95.75	274.31	96.99	298.94	191.49
(*N = 781*)	S.D.	230.60	656.59	478.63	936.74	355.67

The removal of children who initially scored in the highest interval consistently resulted in higher growth across each of the four skills as well as on the composite early-reading score (Level1Gain). Overall, the gain scores increased from an average of 191 points to 321 points.

A CROSSTABS analysis of the WACSPure sample's demographics appears in Table 13.

Table 13
CROSSTABS Analysis of WACSPure Sampling Effects on Demographic Variables

Demographic Variable	Pearson Chi-Square	df	Asymptotic Significance (2-sided)	Directional Effect
NonWhite	434.76	1	.000***	Fewer NonWhite
InPreK	8.43	1	.004**	Fewer in Preschool
LowIncome	2.28	1	.131	None
Gender	.01	1	.919	None
English	9.81	1	.002**	Fewer Non-English

*p <= .05; **p<=.01; ***p<=.001

Proportionally, the WACSPure sample included significantly fewer non-white, preschool, and non-English-speaking children than the Non-WACSPure sample. There were no statistically significant differences in gender or lower income children.

Pure. The question remains of whether there are any final selection criteria for a "pure" test of the dose-response hypothesis. After controlling for extreme use of the program (UsagePure), and for potential test issues (WACSPure), children who spent any time on the placement measure were excluded from the final sample, resulting in a Pure sample size of 208. The demographics of the Pure sample versus the entire sample are summarized in Table 14 below.

Table 14
Pure Sample Demographics

Group		NonWhite	InPreK	LowIncome	Gender	English	AgePre
Pure	Mean	.11	.38	.65	.54	.98	4.46
(*N = 208*)	S.D.	.31	.49	.48	.50	.14	.34
All	Mean	.20	.47	.61	.54	.92	4.59
(*N = 1343**)	S.D.	.40	.50	.49	.50	.27	.39

* AgePre's *N* was 1074.

The two-month difference in ages at the pre-test between the samples (4.59 vs. 4.46) was highly significant ($p < .000$; $df = 1280$; $t = 4.70$). The impact of this grouping on the other demographic variables was analyzed using CROSSTABS (see Table 15).

Table 15
CROSSTABS Analysis of Pure Sampling on Demographic Variables

Demographic Variable	Pearson Chi-Square	df	Asymptotic Significance (2-sided)	Directional Effect
NonWhite	12.540	1	.000***	Fewer NonWhite
InPreK	7.634	1	.006**	Fewer in Preschool
LowIncome	1.277	1	.258	None
Gender	.032	1	.857	None
English	12.785	1	.000***	Fewer Non-English

*p <= .05; **p<=.01; ***p<=.001

Proportionally, the Pure sample was composed of significantly fewer non-white, preschool attending, non-English-speaking children than the non-Pure sample was. There were no statistically significant differences in low-income or gender.

Effect Sizes.

Table 16
Gain Scores and Pre-Post Correlations by Pure Samplings

Grouping		Mean	N	S.D.	Pre- Post Correlation
All Children	Level1Pre	2,331.99	1,073	337.73	
	Level1Post	2,513.30	849	334.51	
	Level1Gain	191.49	781	355.67	.427
UsagePure	Level1Pre	2,340.48	725	336.47	
	Level1Post	2,519.26	721	330.16	
	Level1Gain	179.89	721	354.64	.435
WACSPure	Level1Pre	2,044.75	231	230.80	
	Level1Post	2,365.34	231	342.35	
	Level1Gain	320.59	231	361.05	.254
Pure	Level1Pre	2,043.65	208	234.53	
	Level1Post	2,356.75	208	343.73	
	Level1Gain	313.10	208	362.17	.260

Table 17
Cohen's D Effect Size Calculations

Effect Size Method	All Children	UsagePure	WACSPure	Pure
Glass's Delta	0.57	0.53	1.39	1.34
Pooled Variances	0.57	0.54	1.10	1.07
Repeated Measures	0.71	0.71	1.30	1.26

Hypothesis Results

This section focuses on the results that addressed the three main hypotheses.

Hypothesis #1. The first hypothesis was that there would be a dose-response relationship between usage and gain. In this case, usage was interpreted as Level 1 usage and gain was interpreted as gains on the combined measure of early-reading skills. The results of this regression are reported in Table 18.

Table 18
Linear Regression of Level1Usage to Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000

[a]. Predictors: (Constant), Level1Usage

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-37.985	30.720		-1.236	.217
	Level1Usage	.126	.015	.281	8.160	.000

[a]. Dependent Variable: Level1Gain

Level1Usage was a statistically significant ($p < .000$) predictor of Level1Gain, with a correlation ($R = .281$) that accounted for 7.8% of the overall variance. The equation of the regression line was Level1Gain = .126*Level1Usage – 37.985. Thus, for every 1,000 minutes of usage of Level 1, a child would gain 126 points on WACS, but children who did not use the program would lose approximately 38 points.

The Level1Usage – Level1Gain model accounted for only 8% of the overall variance, suggesting that much of the story still remains untold. The regression was then repeated with all six demographic variables entered as a second step (see Table 19).

Table 19

Linear Regression of Level1Usage to Level1Gain with Demographic Variables

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.909	.079	66.330	1	776	.000
2	.298[b]	.089	.081	341.307	.010	1.457	6	770	.190

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, NonWhite, Gender, InPreK, Bday, LowIncome, English

The inclusion of demographic variables did not significantly ($p = .190$) improve the model, suggesting that the more parsimonious model would exclude demographic variables from the analysis.

In order to highlight the dose-response relationship as much as possible, the original regression was re-run using only the Pure sample. The results appear in Table 20.

Table 20

Linear Regression of Level1Usage With Level1Gain (Pure)

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	df2	Sig. F Change
1	.312[a]	.097	.093	344.916	.097	22.233	1	206	.000

[a] Predictors: (Constant), Level1Usage

Coefficients[a]

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	T	Sig.
1	(Constant)	2.028	70.174		.029	.977
	Level1Usage	.167	.035	.312	4.715	.000

[a] Dependent Variable: Level1Gain

The model summary indicates that usage of Level 1 (Level1Usage) was a statistically significant ($p < .000$) predictor of overall gain on early-reading skills (Level1Gain). In addition, Level1Usage and Level1Gain were closely correlated ($R = .312$), with Level1Usage accounting for nearly 10% of the overall variance in the data. According to the coefficients table, the actual equation is Level1Gain = 2.028 + .167*Level1Usage. In other words, for every 1,000 minutes of usage of Level 1, children should be expected to gain an average of 167 points on WACS across the four early-reading skills taught in that Level. In addition, children who had no Level 1 usage would effectively have no gain whatsoever (gaining only 2 points on WACS).

Despite the significant strength of this relationship, the vast majority of variance in the model (over 90%) remains unaccounted for, suggesting that variables other than Level1Usage might do a better job of accounting for the variance. The regression was then performed with the same six demographic variables added as a second step. The results are shown in Table 21.

Table 21
Linear Regression of Level1Usage and 6 Demographic Variables With Level1Gain (Pure)

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.312[a]	.097	.093	344.916	.097	22.233	1	206	.000
2	.365[b]	.133	.103	343.038	.036	1.377	6	200	.225

[a] Predictors: (Constant), Level1Usage

[b] Predictors: (Constant), Level1Usage, English, InPreK, Gender, Bday, LowIncome, NonWhite

The addition of these demographic variables did not significantly improve the model's fit (p = .225), again indicating that the demographic variables do not predict differences in early-reading gains once usage has been taken into account.

In order to more fully explore the impact of the Pure sampling on the dose-response relationship, sample inclusion ("Pure") was dummy coded (0 = Not Pure; 1 = Pure) and the inclusion variable was added as a second step in the regression. The results of this regression are shown in Table 22.

Table 22

Linear Regression of Level1Usage and Pure With Level1Gain

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	Df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.342[b]	.117	.115	334.637	.038	33.644	1	777	.000

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, Pure

The sampling variable of Pure was statistically significant ($p < .000$). A general linear model (GLM) was then constructed with the design specified to include Level1Usage, Pure, and the interaction between Level1Usage and Pure (Level1Usage*Pure). The parameter estimates from this model are shown in Table 23.

Table 23

*GLM Parameter Estimates for Level1Usage, Pure, and Level1Usage*Pure*

Dependent Variable:Level1Gain

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval: Lower Bound	Upper Bound
Intercept	2.028	68.039	.030	.976	-131.534	135.591
Level1Usage	.167	.034	4.863	.000	.099	.234
[Pure=0]	-57.064	75.862	-.752	.452	-205.983	91.855
[Pure=1]	0[a]	.	.	.	.	.
[Pure=0] * Level1Usage	-.054	.038	-1.414	.158	-.129	.021
[Pure=1] * Level1Usage	0[a]	.	.	.	.	.

[a]. This parameter is set to zero because it is redundant.

The results indicate that the interaction between Pure and Level1Usage was not statistically significant ($p = .158$). This suggests that the impact of usage on gain is not affected by Pure grouping.

Hypothesis #2. This section focuses on the comparing the impact that different types of curriculum usage had on Level1Gain. There were three different ways that usage was categorized: FreePlayUsage was comprised solely of the time that a child spent using the unstructured portion of RRLWM; ReadingUsage was comprised of time spent using any of the levels of the structured portion of RRLWM; and Level1Usage was comprised of time spent using only Level 1 of the structured reading program.

Correlations for each of the three usage types (Free Play, Reading, and Level 1) with WACS subtests are shown for the four Level 1 skills in Table 24.

Table 24
Correlations of Usage Types and Level 1 Skill Gains

Usage Type	*N*	LR *780*	LS *780*	IS *780*	BL *780*	Level1 *780*
FreePlay		-.004	.130***	.036	.124***	.153***
Reading		.079*	.175***	.083*	.235***	.277***
Level 1		.127***	.183***	.097**	.217***	.281***

*p <= .05; **p<=.01; ***p<=.001

Although scores for all three categories of program usage were significantly correlated with gains on Level 1 skills., it is apparent that usage of Level 1 had the strongest overall correlations to the various Level 1 skill gains. To confirm this, a regression analysis was run with FreePlayUsage entered as the first predictor, followed by ReadingUsage, and then finally by Level1Usage. The results appear in Table 25.

Table 25

Linear Regression of Different Usage Types with Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.289[b]	.083	.081	340.955	.005	3.879	1	777	.049
3	.290[c]	.084	.080	341.066	.001	.492	1	776	.483

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, ReadingUsage

[c]. Predictors: (Constant), Level1Usage, ReadingUsage, FreePlayUsage

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-37.985	30.720		-1.236	.217
	Level1Usage	.126	.015	.281	8.160	.000
2	(Constant)	-40.381	30.687		-1.316	.189
	Level1Usage	.074	.031	.164	2.393	.017
	ReadingUsage	.050	.026	.135	1.970	.049
3	(Constant)	-37.122	31.047		-1.196	.232
	Level1Usage	.073	.031	.161	2.347	.019
	ReadingUsage	.046	.026	.125	1.783	.075
	FreePlayUsage	.011	.015	.027	.702	.483

[a]. Dependent Variable: ERP1Gain

As expected, Level1Usage was the strongest predictor, with ReadingUsage just on the edge of statistical significance. FreePlayUsage was not statistically significant in the overall dose-response model, indicating that it did not account uniquely for variance in the model after the other types of usage were taken into account.

As with Hypothesis #1, it was expected that a more accurate test could be conducted by using just the Pure sample. The results for this regression appear in Table 26.

Table 26
Linear Regression of Different Usage Types with Level1Gain (Pure)

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	df2	Sig. F Change
1	.312[a]	.097	.093	344.916	.097	22.233	1	206	.000
2	.312[b]	.097	.089	345.753	.000	.004	1	205	.952
3	.313[c]	.098	.085	346.485	.001	.135	1	204	.714

[a]. Predictors: (Constant), Level1Usage
[b]. Predictors: (Constant), Level1Usage, ReadingUsage
[c]. Predictors: (Constant), Level1Usage, ReadingUsage, FreePlayUsage

Inclusion of only the Pure children underscored the preeminence of Level1Usage as a predictor of Level1Gain, as the addition of either ReadingUsage or FreePlayUsage did not significantly improve the model.

Hypothesis #3. This section reports analyses related to determining the relation between usage of Level 1 and gains on other reading skills. As mentioned in the introduction, WACS is comprised of eleven different subtests, 10 of which are available to prekindergartners. Of these, six – Vocabulary, Reading Comprehension, Listening Comprehension, Sight Words, Real Words, and Nonwords – are not taught or assessed explicitly in Level 1. As with Level1Gain, a summative variable (NonLevel1Gain) was created to express the average achievement gain across these six individual skills.

In order to understand the relation between Level1Usage and gains in these reading skills, correlations were calculated. The results are reported in Table 27.

Table 27
Correlations of Level1Usage to Reading Skill Gains

	VO	RC	LC	SW	RW	NW	NonLevel1
N	*784*	*658*	*282*	*756*	*756*	*661*	*784*
	-.021	-.019	.074	.018	.025	.080*	.044

*p <= .05; **p<=.01; ***p<=.001

The correlations indicated that Level1Usage was only significantly correlated to gains in Nonwords. To more clearly characterize the relationship, the same correlations were repeated with the Pure sample (see Table 28).

Table 28
Correlations of Level1Usage to Reading Skill Gains (Pure)

	VO	RC	LC	SW	RW	NW	NonLevel1
N	*208*	*154*	*98*	*190*	*190*	*155*	*208*
	-.032	.036	.060	-.066	-.016	.104	-.006

*p <= .05; **p<=.01; ***p<=.001

When only Pure children are included there were no statistically significant correlations between usage of Level 1 of the structured reading program and gains in any of the Non-Level 1 reading skills.

Supplementary Investigation Results

This section has three major components. The first details investigations of the notion of program compliance and its effects on the dose-response relationship. The second component investigated the importance of curricular progress in the overall dose-response relationship. The third attempts to construct an overall model of best fit for the dose-response relationship by combining earlier findings with the exploratory results from both compliance and progress monitoring.

Compliance. Compliance relates actual usage of the curriculum to recommended usage. Its impact, therefore, was a blend of overall usage and days in the program. This impact was further modified by a tolerance level that was allowed to vary from 5% to 95%. The goal of this analysis was to first empirically discover which level of tolerance was maximally impactful in the dose-response relationship and then to determine whether compliance, as defined by this tolerance level, ultimately moderated the dose-response relation.

Figure 9 is a scatterplot of compliance tolerance levels and their corresponding levels of statistical significance when entered as a second predictor variable in a linear regression of Level1Usage with Level1Gain for the Pure sample.

Figure 9.
Scatterplot of Level1Usage Compliance With Quadratic Fit

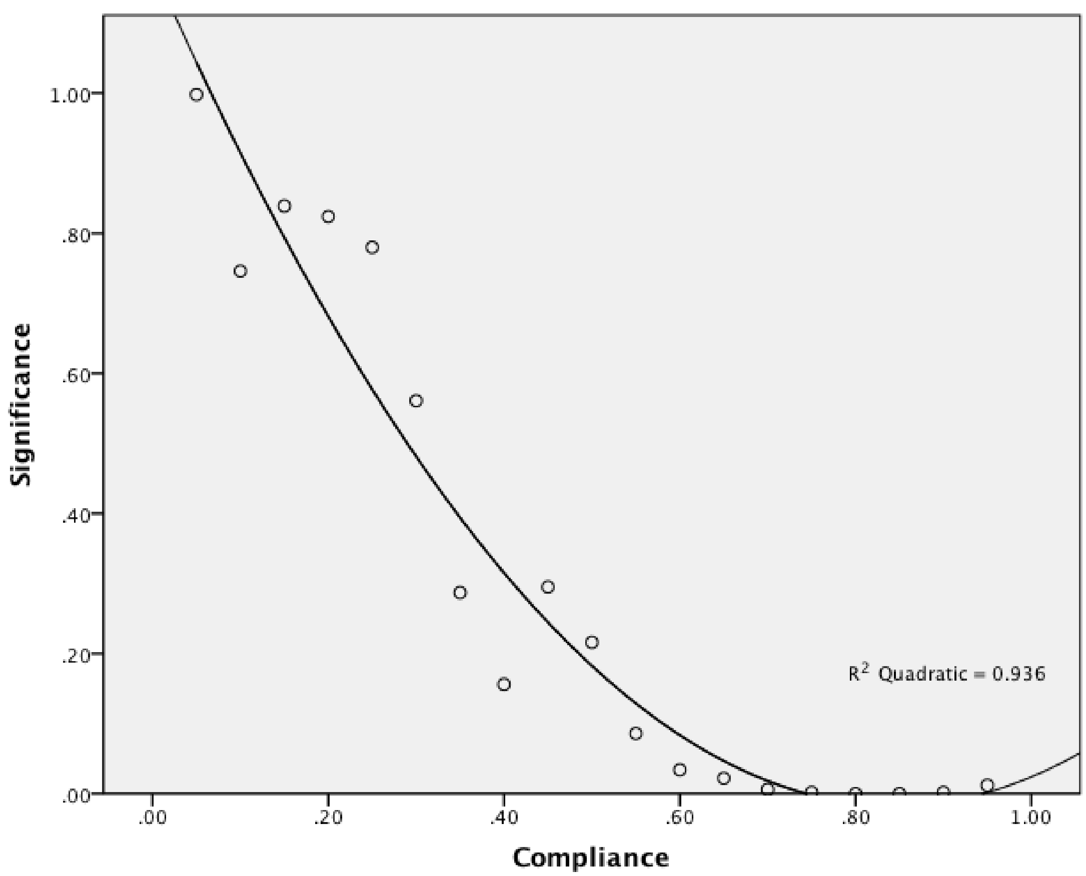


Figure 9. Compliance is measured as the percentage of recommended usage. Significance is a p-value.

The relationship between compliance tolerance levels and significance was curvilinear. A quadratic equation was fitted to the data ($R^2 = 0.936$). Figure 9 indicates that statistical significance was strongest at the 80% level of compliance. For brevity's sake, any reference to Compliance in the remainder of this section should be understood as "80% Compliance".

Figure 10 illustrates the relation between Level 1 usage, days in the program, and the three Compliance levels (-1.00 = less than 80% of recommended; .00 = 80 – 120% of recommended; 1.00 = greater than 120% of recommended).

Figure 10.
Scatterplot of ProgramDays to Level1Usage by 80% Compliance

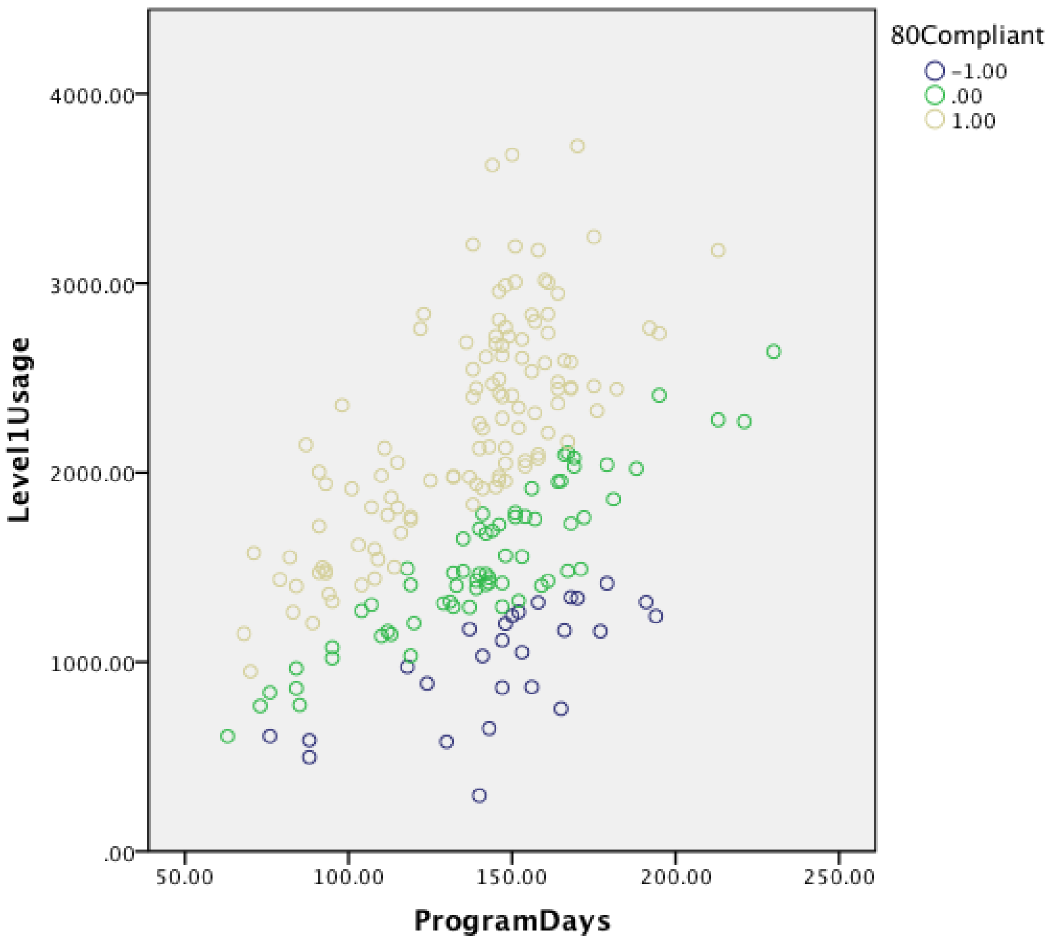


Figure 10. ProgramDays is measured in calendar days. Level1Usage is measured in minutes. Compliance is -1 (under 80% of recommended), 0 (80-120% of recommended), or 1 (over 120% of recommended).

Having established the level of tolerance that maximized the impact of compliance, all children were coded for Compliance. Compliance was entered as the second variable in a linear regression with Level1Usage as the first predictor variable. The results are shown in Table 29.

Table 29
Linear Regression of Level1Usage and Compliance with Level1Gain

					Change Statistics				
Model	R	R^2	Adj. R^2	Std. Error of the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.294[b]	.086	.084	340.410	.008	6.379	1	777	.012

[a] Predictors: (Constant), Level1Usage

[b] Predictors: (Constant), Level1Usage, 80Compliant

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-37.985	30.720		-1.236	.217
	Level1Usage	.126	.015	.281	8.160	.000
2	(Constant)	12.373	36.534		.339	.735
	Level1Usage	.087	.022	.193	3.951	.000
	80Compliant	57.406	22.729	.123	2.526	.012

[a] Dependent Variable: Level1Gain

Compliance was a statistically significant addition to the overall model ($p < .000$). The addition of Compliance reduced the Unstandardized Beta Coefficient (UBC) for Level1Usage from .126 to .087 and its t-value from 8.160 to 3.951, suggesting that some of the variance that was explained by Level1Usage is now explained by Compliance. This interpretation is further bolstered by the fact that the adjusted R^2 value only moved from .078 to .084. That these variables overlap is not surprising—after all, Compliance was constructed in part from Level 1 usage.

In order to establish whether the concept of Compliance was truly different from whether children used the curriculum sufficiently, a continuous variable was constructed by first establishing the recommended amount of time a child should use the program (ProgramDays * 10.71) and subtracting this recommended amount from the amount of time that the child actually used it (Level1Usage). This new variable (L1Recommended) was then entered after Level1Usage into a linear regression (see Table 30).

Table 30
Linear Regression of Level1Usage and L1Recommended With Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.281[b]	.079	.077	341.774	.000	.140	1	777	.708

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, L1Recommended

Addition of the continuous compliance variable did not significantly improve the earlier model ($p = .708$).

A similar test was conducted wherein the days in the UPSTART program was added into the dose-response model as a second step. These results are found in Table 31.

Table 31
Linear Regression of Level1Usage and ProgramDays With Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.281[b]	.079	.077	341.774	.000	.140	1	777	.708

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, ProgramDays

The addition of ProgramDays did not significantly improve the overall model ($p = .708$).

Having established that Compliance is a better predictor of Level1Gain than either L1Recommended or ProgramDays, the investigation of Compliance continued by repeating the earlier regression analysis with just the Pure sample, which produced the results shown in Table 32.

Table 32
Linear Regression of Level1Usage and Compliance With Level1Gain (Pure)

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.312[a]	.097	.093	344.916	.097	22.233	1	206	.000
2	.395[b]	.156	.148	334.314	.059	14.273	1	205	.000

[a] Predictors: (Constant), Level1Usage

[b] Predictors: (Constant), Level1Usage, 80Compliant

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.028	70.174		.029	.977
	Level1Usage	.167	.035	.312	4.715	.000
2	(Constant)	152.356	78.802		1.933	.055
	Level1Usage	.047	.047	.088	1.003	.317
	80Compliant	169.559	44.882	.330	3.778	.000

[a] Dependent Variable: Level1Gain

The use of the Pure sample accentuated the earlier trends, resulting in Level1Usage no longer being a significant factor (p = .317).

To explore the impact of the Pure sampling variable on the Compliance – Level1Gain relationship, a GLM was constructed with the design of Compliance, Pure, and Compliance*Pure. The parameter estimates for this GLM are found in Table 33.

Table 33
*GLM Parameter Estimates for Level1Usage, Pure, and Level1Usage*Pure*

Dependent Variable:Level1Gain

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Intercept	226.508	27.287	8.301	.000	172.943	280.073
80Compliant	200.134	33.025	6.060	.000	135.305	264.963
[Pure=0]	-111.597	31.328	-3.562	.000	-173.094	-50.100
[Pure=1]	0[a]	.	.	.	.	.
[Pure=0] * 80Compliant	-107.192	37.562	-2.854	.004	-180.928	-33.456
[Pure=1] * 80Compliant	0[a]	.	.	.	.	.

[a]. This parameter is set to zero because it is redundant.

The parameter estimates for the Compliance*Pure interaction were statistically significant ($p = .004$), suggesting that there are differential effects of Compliance on Level1Gain based on Pure grouping. The estimates for these parameters are reported in Table 34.

Table 34
*Parameter Estimates for Compliance*Pure*

Dependent Variable:Level1Gain

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	Upper Bound
[Pure=0] * 80Compliant	92.942	17.896	5.193	.000	57.811	128.072
[Pure=1] * 80Compliant	200.134	33.025	6.060	.000	135.305	264.963

Thus, the impact of Compliance on the combined Level 1 WACS measure was 93 points for children in the non-Pure sample, but 200 points for children in the Pure sample. This suggests that there is a more nuanced relationship between the Pure and Compliance variables, and that analyses involving them need to be approached with caution.

In further exploring Compliance, a linear regression was conducted with it as the sole predictor of Level1Gain for the Pure sample. The results of the regression are shown in Table 35.

Table 35
Linear Regression of Compliance With Level1Gain (Pure)

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	df2	Sig. F Change
1	.390[a]	.152	.148	334.319	.152	36.932	1	206	.000

[a] Predictors: (Constant), 80Compliant

Coefficients[a]

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	T	Sig.
1	(Constant)	226.508	27.210		8.324	.000
	80Compliant	200.134	32.932	.390	6.077	.000

[a] Dependent Variable: Level1Gain

The results of the regression indicate that the best linear fit for the data is the equation Level1Gain = 200.134*80Compliant + 226.508 and that Compliance explains just under 15% of the total variance in the model.

Next, a linear regression was run with Compliance entered first, followed by the six potential demographic predictors that were used earlier (see Appendix G). The results appear in Table 36.

Table 36
Linear Regression of Compliance With Level1Gain and Demographics (Pure)

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics: R^2 Change	F Change	df1	df2	Sig. F Change
1	.390[a]	.152	.148	334.319	.152	36.932	1	206	.000
2	.410[b]	.168	.139	335.998	.016	.658	6	200	.684

[a] Predictors: (Constant), 80Compliant
[b] Predictors: (Constant), 80Compliant, InPreK, English, Gender, Bday, LowIncome, NonWhite

The results indicate that the addition of these variables did not significantly improve the overall model (p = .684).

Table 37 shows how usage of the curriculum varied by level of Compliance and Pure.

Table 37
Level1Usage by Compliance by Pure

80Compliant	Pure	Mean	N	Std. Deviation
Under	Non-Pure	813.83	112	340.67
	Pure	997.52	26	310.69
	Total	*848.44*	*138*	*341.82*
Compliant	Non-Pure	1,470.11	150	379.21
	Pure	1,530.99	66	411.98
	Total	*1,488.71*	*216*	*389.57*
Over	Non-Pure	2,307.13	314	698.12
	Pure	2,251.63	116	570.89
	Total	*2,292.16*	*430*	*666.00*
Total	Non-Pure	1,798.79	576	826.58
	Pure	1,866.20	208	678.14
	Total	*1,816.67*	*784*	*790.06*

Taking the overall average of 134.4 days in the program, the difference between Undercompliance and Compliance was 640.2 minutes or 4.76 minutes per day, and the difference between Compliance and Overcompliance was 803.4 minutes or 5.98 minutes per day. These represent roughly 45% and 55% increases, respectively, over the recommended daily use of 10.71 minutes.

The differences between the gains for the three levels of Compliance were similarly striking. Children who were Undercompliant experienced almost no gains (11.81 points on WACS), but those who were Compliant and Overcompliant gained 163.1 and 263.5 points, respectively (see Table 38).

Table 38
Level1Gain by Compliance by Pure

80Compliant	Pure	Mean	N	Std. Deviation
Under	Non-Pure	5.38	110	320.68
	Pure	39.06	26	304.00
	Total	11.81	136	316.73
Compliant	Non-Pure	139.41	149	344.94
	Pure	216.52	66	346.61
	Total	163.08	215	346.48
Over	Non-Pure	202.02	313	336.12
	Pure	429.48	116	334.84
	Total	263.53	429	350.31
Total	Non-Pure	147.90	572	343.13
	Pure	313.10	208	362.17
	Total	191.95	780	355.67

These differences were heightened for the Pure sample, where children grew 39.1, 216.5, and 429.5 points respectively. These gains were plotted by group and lines were fitted for the Pure sample (red solid line) and for all children (blue broken line) (see Figure 11).

Figure 11.
Level1Gains by Compliance for Pure and All Children With Best-Fitting Lines

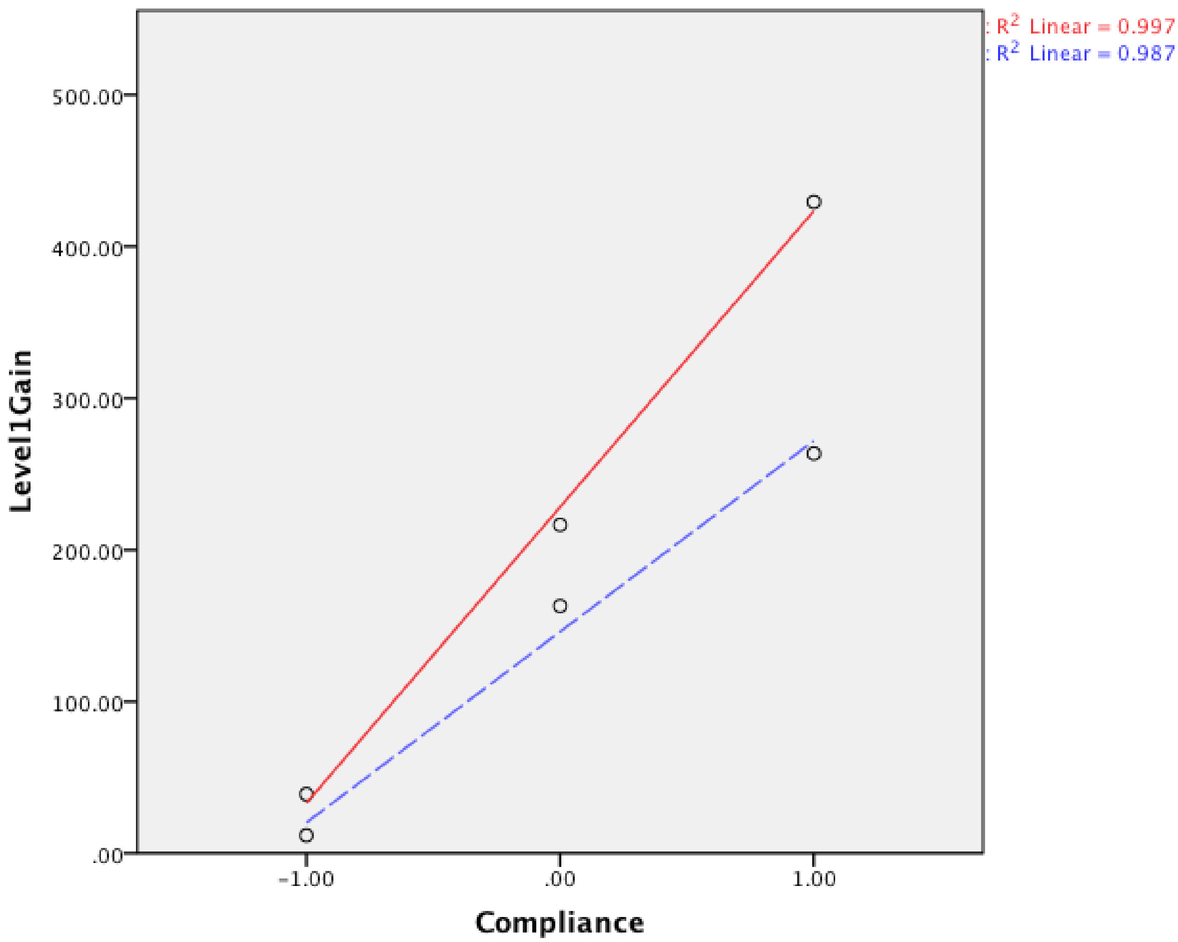


Figure 11. Compliance is measured as -1 (less than 80% of recommended), 0 (80-120% of recommended), or 1 (over 120% of recommended). Level1Gain is measured in WACS score units.

The differences across the three Compliance levels followed a similar pattern when converted into effect sizes (see Tables 39 and 40).

Table 39
Gain Scores and Pre-Post Correlations by Compliance by Pure

Group	Compliance		Mean	N	S.D.	Correlation
	Under	Level1Pre	2,046.52	26	201.89	
		Level1Post	2,085.58	26	281.33	
		Level1Gain	39.06	26	304.00	0.242
Pure	Compliant	Level1Pre	2,051.13	66	244.17	
		Level1Post	2,267.64	66	374.95	
		Level1Gain	216.52	66	346.61	0.437
	Over	Level1Pre	2,038.75	116	237.47	
		Level1Post	2,468.23	116	286.84	
		Level1Gain	429.48	116	334.84	0.195
	Under	Level1Pre	2,447.81	138	340.68	
		Level1Post	2,457.66	136	360.47	
		Level1Gain	11.81	136	316.73	0.594
All	Compliant	Level1Pre	2,297.81	216	330.39	
		Level1Post	2,459.27	215	341.80	
		Level1Gain	163.08	215	346.48	0.469
	Over	Level1Pre	2,323.07	430	328.01	
		Level1Post	2,586.52	429	301.12	
		Level1Gain	263.53	429	350.31	0.383

Table 40
Cohen's D Effect Size Calculations by Compliance by Pure

Group	Method	Under	Compliant	Over
	Glass's Delta	0.19	0.89	1.81
Pure	Pooled Variances	0.16	0.69	1.64
	Repeated Measures	0.19	0.93	1.67
	Glass's Delta	0.03	0.49	0.80
All	Pooled Variances	0.03	0.49	0.84
	Repeated Measures	0.04	0.66	1.07

Progress Monitoring. Progress through the reading program was gauged by the number of unique Level 1 learning objectives mastered (represented by the variable L1Obj). Figure 12 shows the frequency distribution for this variable.

Figure 12.
Histogram of L1Obj With Normal Curve

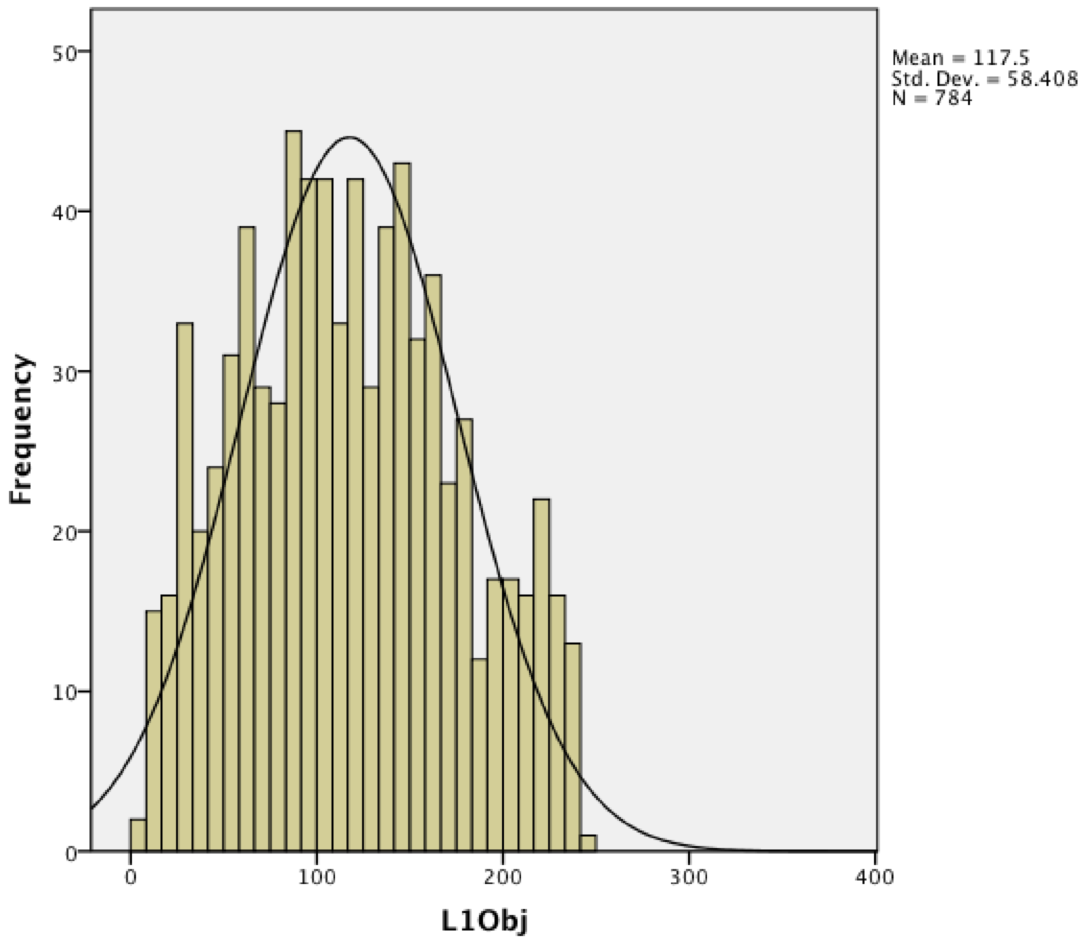


Figure 12. Level1Obj is measured in numbers of activities. Frequency is measured in number of children.

The distribution was approximately normal, and an EXAMINE analysis found no outliers. The lowest number of objectives mastered was 6, while the greatest number of objectives mastered was 249 out of the 380 possible.

Table 41 details the number of objectives mastered for various samples of children by Compliance level. Level1Usage and Level1Gain are included as variables, along with two new efficiency variables (UseEff and GainEff), which were calculated by dividing Level1Usage and Level1Gain by L1Obj, respectively.

Table 41
Means of L1Obj, Level1Usage, Level1Gain, UseEff, and GainEff For Pure and All Children by Compliance

Group	Compliance		L1Obj	L1Usage	Level1Gain	UseEff	GainEff
	Under	Mean	41.19	997.52	39.06	33.07	-1.71
		S.D.	26.34	310.69	304.00	19.99	14.86
		N	*26*	*26*	*26*	*26*	*26*
	Compliant	Mean	84.38	1,530.99	216.52	20.32	2.71
		S.D.	32.48	411.98	346.61	8.51	6.61
Pure		*N*	*66*	*66*	*66*	*66*	*66*
	Over	Mean	134.91	2,251.63	429.48	17.32	3.30
		S.D.	36.17	570.89	334.84	4.38	2.85
		N	*116*	*116*	*116*	*116*	*116*
	Total	Mean	107.16	1,866.20	313.10	20.25	2.49
		S.D.	47.80	678.14	362.17	10.35	6.89
		N	*208*	*208*	*208*	*208*	*208*
	Under	Mean	46.98	848.44	11.81	22.32	-1.89
		S.D.	26.41	341.82	316.73	14.01	13.82
		N	*138*	*138*	*136*	*138*	*136*
	Compliant	Mean	95.08	1,488.71	163.08	17.80	1.63
		S.D.	41.68	389.57	346.48	7.23	5.49
All		*N*	*216*	*216*	*215*	*216*	*215*
	Over	Mean	151.39	2,292.16	263.53	15.99	1.76
		S.D.	46.14	666.00	350.31	4.99	3.10
		N	*430*	*430*	*429*	*430*	*429*
	Total	Mean	117.50	1,816.67	191.95	17.60	1.09
		S.D.	58.41	790.06	355.67	8.23	6.96
		N	*784*	*784*	*780*	*784*	*780*

For both samples (All and Pure), the number of minutes required to master an objective (UseEff) decreased as Compliance increased, suggesting that the relative rate of learning increased. This effect was mirrored for achievement as well – as Compliance increased, the number of points a child gained on WACS per objective (GainEff) increased as well.

To explore mastery of objectives as a predictor of Level1Gain, a regression was run with L1Obj as a predictor. The results are displayed in Table 42.

Table 42
Linear Regression of L1Obj With Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.304[a]	.093	.091	339.020	.093	79.408	1	778	.000

[a] Predictors: (Constant), L1Obj

L1Obj was a significant predictor of Level1Gain ($p < .000$) and accounted for just over 9% of the overall variance in the model.

In order to understand how this new variable impacted the overall usage-to-gain relationship, a linear regression was run with Level1Usage entered first, followed by L1Obj for the Pure sample. The results are shown in Table 43.

Table 43
Linear Regression of Level1Usage and L1Obj With Level1Gain

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.313[b]	.098	.096	338.215	.019	16.581	1	777	.000

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, L1Obj

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-37.985	30.720		-1.236	.217
	Level1Usage	.126	.015	.281	8.160	.000
2	(Constant)	-57.347	30.786		-1.863	.063
	Level1Usage	.052	.024	.115	2.170	.030
	L1Obj	1.318	.324	.216	4.072	.000

[a]. Dependent Variable: ERP1Gain

The addition of L1Obj improved the model significantly ($p < .000$) and lowered the t-value of Level1Usage from 8 to 2, indicating a substantial overlap between these two variables. The regression was repeated with the Pure sample (see Table 44).

Table 44
Linear Regression of Level1Usage and L1Obj With Level1Gain (Pure)

					Change Statistics				
Model	R	R^2	Adj. R^2	Std. Error of the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.312[a]	.097	.093	344.916	.097	22.233	1	206	.000
2	.440[b]	.193	.186	326.858	.096	24.391	1	205	.000

[a] Predictors: (Constant), Level1Usage

[b] Predictors: (Constant), Level1Usage, L1Obj

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.028	70.174		.029	.977
	Level1Usage	.167	.035	.312	4.715	.000
2	(Constant)	-5.202	66.516		-.078	.938
	Level1Usage	-.054	.056	-.102	-.972	.332
	L1Obj	3.917	.793	.517	4.939	.000

[a] Dependent Variable: Level1Gain

The inclusion of L1Obj resulted in a significantly better model fit ($p < .000$) and in Level1Usage becoming nonsignificant ($p = .332$).

To investigate the impact of Pure sampling on the relationship between L1Obj and Level1Gain, a GLM was constructed with the design of L1Obj, Pure, and L1Obj*Pure. The parameter estimates from this GLM are reported in Table 45.

Table 45
*GLM Parameter Estimates for L1Obj, Pure, and L1Obj*Pure*

Dependent Variable:Level1Gain

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	-40.475	55.708	-.727	.468	-149.831	68.881
L1Obj	3.300	.475	6.947	.000	2.367	4.232
[Pure=0]	-21.863	63.430	-.345	.730	-146.378	102.652
[Pure=1]	0[a]	.	.	.	.	.
[Pure=0] * L1Obj	-1.567	.525	-2.987	.003	-2.597	-.537
[Pure=1] * L1Obj	0[a]	.	.	.	.	.

[a]. This parameter is set to zero because it is redundant.

The parameter estimates indicate that there was a significant ($p = .003$) interaction between the effects of L1Obj and Pure, suggesting that there was something about the grouping of Pure that had a differential impact on how L1Obj related to Level1Gain. The estimated marginal means for these values are shown in Table 46.

Table 46
*GLM Parameter Estimates for L1Obj, Pure, and L1Obj*Pure*

Dependent Variable:Level1Gain

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
[Pure=0] * L1Obj	1.732	.223	7.763	.000	1.294	2.170
[Pure=1] * L1Obj	3.300	.475	6.947	.000	2.367	4.232

Thus, Pure children scored 3.3 points higher on the combined Level1Gain WACS measure for each unique Level 1 objective mastered, but non-Pure children gained only 1.7 points. This interaction suggests that, for L1Obj, as for Compliance, the separation of children into Pure and Non-Pure groups might have had an unintended consequence for L1Obj and therefore should be undertaken with caution.

Model of Best Fit. Looking across results from the analyses presented thus far, the three candidates for constructing a best-fitting model are Level1Usage, Compliance, and L1Obj. None of the demographic variables improved any of the earlier models and were consequently excluded from this analysis. Due to the statistically significant interactions between Pure*Compliant and Pure*L1Obj, two separate models were explored, the first with all children and the second with the Pure sample.

When all of the children were included in the analysis, multiple variables remained statistically significant: when Level1Usage and Compliance were both entered into a linear regression model, both remained significant; the same was the case for Level1Usage and L1Obj. To establish the best model fit, the three candidate variables were entered stepwise into a linear regression as predictors of Level1Gain. As there has been no regression run to this point with Compliance and L1Obj, the ordering of the variables for this new regression was Compliance, L1Obj, and then Level1Usage. The results of the regression are shown in Table 47.

Table 47

Linear Regression of Compliance, L1Obj, and Level1Usage With Level1Gain

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics				
					R^2 Change	F Change	df1	df2	Sig. F Change
1	.261[a]	.068	.067	343.592	.068	56.743	1	778	.000
2	.312[b]	.097	.095	338.339	.029	25.341	1	777	.000
3	.316[c]	.100	.097	338.072	.003	2.232	1	776	.136

[a]. Predictors: (Constant), 80Compliant

[b]. Predictors: (Constant), 80Compliant, L1Obj

[c]. Predictors: (Constant), 80Compliant, L1Obj, Level1Usage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	146.369	13.710		10.676	.000
	80Compliant	121.345	16.109	.261	7.533	.000
2	(Constant)	4.534	31.243		.145	.885
	80Compliant	44.694	21.988	.096	2.033	.042
	L1Obj	1.451	.288	.238	5.034	.000
3	(Constant)	-28.268	38.165		-.741	.459
	80Compliant	30.749	23.871	.066	1.288	.198
	L1Obj	1.175	.342	.193	3.433	.001
	Level1Usage	.039	.026	.086	1.494	.136

[a]. Dependent Variable: Level1Gain

The results of the regression indicate that when L1Obj and Compliance are included, that both remain statistically significant, but when all three variables are included that L1Obj is the only variable that remains significant ($p = .001$). The third model, however, is not significantly better than the second model ($p = .136$).

In order to determine the relative strengths of the predictors, the regression was repeated, but with a different ordering of the three predictor variables. The results are shown in Table 48.

Table 48

Linear Regression of Level1Usage, Compliance, and L1Obj With Level1Gain

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.294[b]	.086	.084	340.410	.008	6.379	1	777	.012
3	.316[c]	.100	.097	338.072	.014	11.787	1	776	.001

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, 80Compliant

[c]. Predictors: (Constant), Level1Usage, 80Compliant, L1Obj

This ordering of the variables resulted in each subsequent model being significantly better than the last, suggesting that the model of Level1Usage and Compliance is inferior to that of L1Obj and Compliance. The results for the final combination of the 3 predictor variables is shown in Table 49.

Table 49

Linear Regression of Level1Usage, Compliance, and L1Obj With Level1Gain

Model	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics R^2 Change	F Change	df1	df2	Sig. F Change
1	.281[a]	.079	.078	341.585	.079	66.579	1	778	.000
2	.313[b]	.098	.096	338.215	.019	16.581	1	777	.000
3	.316[c]	.100	.097	338.072	.002	1.659	1	776	.198

[a]. Predictors: (Constant), Level1Usage

[b]. Predictors: (Constant), Level1Usage, L1Obj

[c]. Predictors: (Constant), Level1Usage, L1Obj, 80Compliant

These results suggest that the model comprised of L1Obj and Level1Usage was not significantly enhanced ($p = .198$) by the addition of Compliance.

Overall, these three regressions point to a two-variable model, composed of L1Obj and either Level1Usage (Adj. $R^2 = .096$) or Compliance (Adj. $R^2 = .095$) as the best-fitting model for all children. The fact that a model of Level1Usage and Compliance together is not as strong a model as either one with L1Obj is not surprising, given that Compliance is a derivative of Level1Usage and is therefore at least partly redundant.

With regards to just the Pure sample, Level1Usage was nonsignificant when either Compliance or L1Obj was added to the regression model. In order to establish whether these two effects were redundant, a linear regression was run using Pure children with Compliance entered first, followed by L1Obj. The results appear in Table 50.

Table 50
Linear Regression of Compliance and L1Obj With Level1Gain (Pure)

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.390[a]	.152	.148	334.319	.152	36.932	1	206	.000
2	.451[b]	.203	.195	324.873	.051	13.152	1	205	.000

[a] Predictors: (Constant), 80Compliant

[b] Predictors: (Constant), 80Compliant, L1Obj

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	226.508	27.210		8.324	.000
	80Compliant	200.134	32.932	.390	6.077	.000
2	(Constant)	17.223	63.477		.271	.786
	80Compliant	84.231	45.227	.164	1.862	.064
	L1Obj	2.421	.668	.320	3.627	.000

[a] Dependent Variable: Level1Gain

L1Obj significantly (p = .000) enhanced the model. Together with Compliance, L1Obj accounted for 19.5% of the overall variance of Level1Gain.

The inclusion of L1Obj markedly impacted Compliance in two ways: dropping its UBC from 200 to 84 and reducing its t-value from 6.08 to 1.86, resulting in Compliance just missing statistical significance (p = .064). The UBC for L1Obj indicates that, for each unique learning objective mastered, the WACS score for a child increased by roughly 2.4 points.

As Compliance was no longer statistically significant after the addition of L1Obj to the linear model, a regression of L1Obj with Level1Gain was conducted as the most parsimonious model. The results are shown in Table 51.

Table 51
Linear Regression of L1Obj With Level1Gain (Pure)

				Std. Error of	Change Statistics				
Model	R	R^2	Adj. R^2	the Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	.436[a]	.190	.186	326.814	.190	48.216	1	206	.000

[a] Predictors: (Constant), L1Obj

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-40.475	55.735		-.726	.469
	L1Obj	3.300	.475	.436	6.944	.000

[a] Dependent Variable: Level1Gain

The linear equation that best describes Level1Gain for children in the Pure sample was therefore:

$$\text{Level1Gain} = 3.3*\text{L1Obj} - 40.475$$

In other words, children gained 3.3 points on WACS for each unique Level 1 objective mastered. This linear model accounted for 18.6% of the overall variance in the data.

Discussion

Hypotheses Findings

Hypothesis #1. The first hypothesis asserted that Level1Usage would predict Level1Gain. Linear regression confirmed that Level1Usage was a statistically significant predictor ($p < .000$; Adj. $R^2 = .078$) of early-reading achievement gains. This finding was enhanced for children in the Pure sample (Adj. $R^2 = .093$). Demographic variables did not interfere with this model's ability to predict gain. There was no significant interaction between Pure sampling and the dose-response relationship.

Hypothesis #2. The second hypothesis claimed that FreePlayUsage would be an inferior predictor of Level1Gain because it is an inferior approximation of ALT. This was confirmed when FreePlayUsage's ability to predict early-reading achievement gains ($p < .000$; Adj. $R^2 = .022$; see Table 15) was supplanted when Level1Usage was entered first into the regression.

Hypothesis #3. The third hypothesis predicted that time spent using Level 1 would not be as strong a predictor of reading skills that were not explicitly part of its curriculum. This was statistically confirmed – Level1Usage was not a significant predictor of NonLevel1Gain ($p = .232$).

Hypotheses Limitations

All of the hypotheses were contingent upon an accurate quantification of early-reading ALT. Unfortunately, there are reasons to believe that the measure used to represent this quantification suffered from a number of shortcomings. The first and potentially largest issue is that the program was unable to measure time at the activity level, forcing the analyses to be performed at the session level. This resulted in the inability to measure how much time each child spent on specific learning tasks.

A second potential concern is the relatively high pre-test scores. Taken at face value, these scores indicate that many of the children did not need to start at the beginning of Level 1, thereby raising the specter of having a substantial portion of the measured time exempted from characterization as ALT. A related concern is the inability to control for potential parental influences on the test results.

A third concern is the amount of unexplained variance. Even the best model left over 80% of the variance in the model unaccounted for. If ALT is truly causal, it seems that it should account for much larger portion of the overall variance. The relatively modest level of variance accounted for might signal that ALT was not adequately operationalized.

A fourth concern is that almost half of the children did not participate meaningfully in the analysis. Although the original recruitment was robust, the high attrition rate could have resulted in selection bias, which could potentially provide an alternative explanation for the results.

A fifth concern is the lack of a peer group. The lack of a randomized control group meant that the model could not control for effects such as maturation. Although age-specific norms may often be helpful in such cases, the fact that WACS was normed on children in preschool settings raises an important question as to whether its norms are appropriate for children in a home.

In addition, there was no obvious solution for how to control for individual aptitude and the effect that it could have on the dose-response relationship. For older children, inclusion of IQ could function as a proxy of sorts, but IQ has been shown to be problematic for preschoolers: “Research suggests that within a span of a single year, obtained [IQ] scores may vary by as much as 1 standard deviation in 50% of the normal preschool population and as much as 2 standard deviations in 10%" (Hutchens, Hamilton, Town, Gaddis, & Presley, 1991, p. 14).

Supplementary Investigation Findings

As mentioned earlier, the supplementary investigations were undertaken to exploit the available data in order to elaborate on earlier analyses, to explore alternative hypotheses, and to identify possible alternative explanations of the results.

Compliance Findings. Compliance was not uniformly influential. In particular, permissive interpretations of compliance (i.e., those allowing a large deviation from recommended) did not add significantly to fit of the dose-response model (see Figure 9).

The most effective tolerance level for compliance was empirically estimated to be 80%. With the recommended usage set at 15 minutes per day, 5 days per week, this tolerance level translates into using the program 0-3 days per week (Undercompliant), 4-6 days per week (Compliant), or 7 or more days per week (Overcompliant). Measured in weekly minutes, less than 60 minutes was Undercompliant, 60 – 90 minutes was Compliant, and over 90 was Overcompliant. Although "overcompliance" is a term that normally might be construed as negative, in this case it was associated with even larger gains, suggesting that "more is more." These findings suggest that the recommendation of 15 minutes of usage per day might be more profitably positioned as a minimum.

When added to the regression model, Compliance appeared to overlap heavily with Level1Usage: both were significant predictors of Level1Gain when all children were included in the sample, but Compliance supplanted Level1Usage when only the Pure sample was used, suggesting that the dose-response relationship was potentially more nuanced than just "usage to gain."

Progress Findings. Progress through the structured reading curriculum, as quantified by the number of unique Level 1 learning objectives mastered by the child, was found to significantly enhance the dose-response model. This suggests that it is not just the amount of time that a child spends using the program that matters, but rather how that time is used as well.

In some ways, progress-monitoring could be seen to function as an indicator of the integrity of the ALT modeling itself. Assuming Berliner's claim that learning cannot occur without ALT, it is possible that mastery of objectives over time could be seen as a necessary albeit insufficient indicator of ALT. Thus, if children were not receiving early-reading ALT, they could not master novel early-reading material, and no progress would be realized. Progress, however, does not necessarily translate into ALT being successfully applied. For example, if I were to take Level 1 of the reading curriculum, I would succeed, not because it is providing me with early-reading ALT – according to Berliner's definition, it is not providing me with ALT because it is far below my instructional level – but rather because I had already mastered the skills necessary for success.

Best Model Fit Findings. Best-fitting models were constructed separately for the whole sample of children and for the Pure sample. This dichotomy was required by the discovery of significant interactions between the Pure sampling variable and both the Compliance and L1Obj variables, the effect of which was to effectively double the UBC depending on sample membership.

For the full sample, the best-fitting model included L1Obj with either Level1Usage or Compliance. This model explained just under 10% of the overall variance. For the Pure sample, the best-fitting model contained L1Obj by itself and accounted for 18.6% of the overall variance.

The strong presence of L1Obj in both models underscores the importance of objective-based curricular progress. Assuming that the Pure sampling variable did not unduly bias the overall findings, the best-fitting model was one in which each unique Level 1 objective that was mastered added approximately 3.3 points to a child's WACS score on average. Based on a total of 380 Level 1 objectives, the achievement gain due to objective mastery could have ranged to 1,254 points, or roughly one-and-a-third years of growth in early-reading skills. Using the average efficiency of 17.3 minutes per objective measured for the Pure sample, this would require just under 6,600 minutes or 110 hours of usage of the curriculum to effect.

In a Broader Context

Despite its limitations, this dissertation has advanced the formal study of ALT. It is the first study that has attempted to measure ALT on a large scale, in an informal environment, and with preschoolers. This was enabled primarily by moving the burden of instruction and measurement from a teacher to a computer.

The ability to conduct serious educational research over an extended period of time in diverse home environments is a sizable achievement in its own right. Traditionally, a researcher had to be onsite, a constraint that was neither cost-effective nor scalable. A computer-based instructional approach overcomes these barriers of venue and human presence. The children who participated in this study came from many walks of life – from the Salt Lake City metro area to Native American reservations. And participate they did – averaging over 35% more usage than required, and over the summer months, no less.

This study in many ways is an early example of what Woolf (2009) describes as the coming inflection point in educational research: a point when artificial intelligence, the internet, and cognitive science are jointly brought to bear on persistent educational problems. That this point is rapidly approaching is beyond doubt – even *Science* has recently jumped on board, devoting an entire issue to technology and education, with the Editor-In-Chief penning an article entitled "Making a Science of Education" and proclaiming that "we will much more emphasis on both science and the 'science of education'" (Alberts, 2009, p. 15).

Learning efficiency. Many children grew substantially in their mastery of early-reading skills across the study. On average, children who participated in the UPSTART program gained 191 points in early-reading skills as measured by WACS, while those in the Pure sample gained 313 points. Interpreting these results in the context of WACS, where 1,000 points is equal to a year of achievement, the average child gained one-fifth of a year in early-reading achievement, while the average child in the Pure sample gained approximately one-third of a year in early-reading achievement.

The efficiency of these gains is less than what would have been expected based on the amount of time in the curriculum for the previously normed samples; on average, the children should have gained 1,000 / 365.25 * 134.4 = 368 points. A potential explanation for this disparity is that the normed sample was composed entirely of students in preschool programs, implying that they would have received more early-reading ALT than those in the pilot study, where less than half of the children were enrolled in preschool programs.

The manner in which children used the reading curriculum was not efficient. Almost all of the children in the study did not take the placement test and consequently started the program at the beginning regardless of their preexisting ability. As mentioned in the Introduction, RRLWM's sequencing algorithm has no mechanism for accelerating a child's progress by skipping content. Therefore, children who knew most of the letters of the alphabet potentially spent a large portion of their time working on activities that were designed to teach them what they already knew. This was evidenced by the average pre-test score of 2,338 on the combined measure. In essence, this meant that the preschoolers scored at the Kindergarten Winter benchmark level on the pre-test. It is possible, therefore, that the learning would increase (and the required time to achieve norm-based expected growth would decrease) if the initial placement of children were to be improved (e.g., through the use of the placement test).

In addition, children spent roughly 20% of their overall usage time in Free Play activities, which ultimately were not found to be predictive of gain. The rechanneling of these minutes into the structured curriculum might have enhanced efficiency, although the impact on overall engagement would need to be weighed in the balance.

Effect sizes. In order to understand the effect-size findings, it is important to situate them in a broader educational context. The gains children achieved, as measured by effect size, were notable. In his interpretation of effect size (ES), Cohen (1992) proposed three different designations – small, medium, and large – which correspond to values of 0.2, 0.5, and 0.8, respectively. According to Cohen:

> My intent was that medium ES represent an effect likely to be visible to the naked eye of a careful observer. I set small ES to be noticeably smaller than medium but not so small as to be trivial, and I set large ES to be the same distance above medium as small was below it. (p. 156)

Effect sizes are linear with respect to one another, so the magnitude of a small effect size is 40% of a medium effect size and 25% of a large one. Using this parlance, the early-reading gains were medium or large, indicating that they should be readily noticeable.

Effect sizes are often used in meta-analyses to compare variables across a large number of studies. Table 52 draws from Walberg's (1984) summary table. I have modified the table to include Bloom's (1984) findings of the impact of expert one-on-one tutoring, as well as some of the repeated-measures effect sizes from the pilot study.

Table 52
*Effect Size Comparisons (Pilot Study Findings in **Bold**)*

Method	Effect size	Size ('X' = 0.1 effect size)
Bloom's Instruction	2.0	XXXXXXXXXXXXXXXXXXXX
Pure, Overcompliant	**1.67**	**XXXXXXXXXXXXXXXXX**
Pure	**1.26**	**XXXXXXXXXXXXX**
All Children	**0.71**	**XXXXXXX**
IQ	0.71	XXXXXXX
Personalized Instruction	0.57	XXXXXX
Tutoring	0.40	XXXX
Instructional Time	0.38	XXXX
Home Environment	0.37	XXXX
Motivation	0.34	XXX
Socioeconomic Status	0.25	XXX
Treatment Group	0.21	XX
Class Size	0.09	X

Bloom's results represent an ideal of sorts (Pon-Barry, 2004) and, in many ways, could be seen as the upper limit of what is possible. As can be seen from Table 52, the effect sizes the children achieved across an average of 134 days were impressively large.

One of the most encouraging facets of these findings is the modest amount of time that was required to achieve them. In formal environments, elementary students receive very small amounts of individualized instructional time – on the order of 1 – 2 minutes per day out of a 6-hour school day (Conant, 1973). In the pilot study, the average curricular usage for the most effective sample (Pure Overcompliant), was only 2,250 minutes across the 134 days or just under 17 minutes per calendar day or roughly 23.5 minutes per weekday. This is shorter than the length of an average children's cartoon show. Or, assuming a 4-year-old is awake for 14 hours a day, it is roughly 2% of that child's waking time.

This suggests that there is ample time for additional learning to take place. Using the best-fitting model for the Pure sample as a guide, 110 hours of usage should result in roughly 16 months of gain. If the blocks of usage time could either be lengthened or occur more frequently without adversely impacting the child or the efficacy of the learning, it is possible that children could learn much more quickly. For instance, if children had two half-hour sessions on each weekday, that would result in 5 hours of usage per week, or roughly 20 hours per month. This is a fraction of the time that a child would spend attending a preschool, but could have outsized results on the child's rate of learning – 20 hours / 110 hours * 16 months = roughly 2.9 months of progress per month. While such extrapolations are prone to error for a variety of reasons, there is reason to believe that such an approach is promising.

It is noteworthy that these gains were realized in a non-formal learning environment. Although the majority of research on learning occurs in formal, school-based environments, Walberg (1984) estimated that only 13% of a child's waking hours before the age of 18 are spent in formal instructional environments. Home environments provide an alternative setting in which to situate educational reform. Traditionally-cited weaknesses of the home environment are a lack of structure and expertise. The computer-based approach used in UPSTART supplies both and makes them available without the additional costs of traditional learning environments.

It is also notable that the program was in place over the summer months. Previous researchers found that the retention of school learning from the end of one school year to the beginning of the next is impacted by opportunities to learn during the summer months (Alexander, Entwisle, & Olson, 2007). Based on the gains achieved by these children in the pilot study, the UPSTART program could potentially serve as a model for how to avoid the dreaded summer slump.

Summary and Future Directions

Both Carroll and Berliner called attention to the importance of looking more closely at the process of learning. Their collective consensus, namely, that instructional time is at the heart of academic learning, has been echoed ardently by researchers in the intervening years. This ardor has not been matched by quantitative rigor to this point. This dissertation has:

a. Provided an operational definition of Berliner's notion of ALT;
b. Identified the necessary tools for a pilot study of ALT in the area of early-reading; and
c. Conducted the study and analyzed the results.

Its findings provided support for all three hypotheses. Additional post-hoc investigations suggested the important of looking at how the curriculum was utilized in comparison with the recommended guidelines for its use (compliance) and at the trajectory of the children through the curriculum (progress monitoring).

Still, over 80% of the variance in the linear regression model of achievement gain was not accounted for, suggesting that much of the story remains untold. This might be partly an artifact of attempting to model the relation linearly when research suggests nonlinear models might be more accurate (Fredrick & Walberg, 1980, p. 191). Longitudinal data analysis might help address this problem (see Singer & Willett, 2003). Additional data points should be available from the same initial cohort of UPSTART children in the future, which would allow for further tests of the linearity of the dose-response relation.

Another question for future research is whether the gains demonstrated by the children will have a long-term impact. Alexander, et al. (2007) reported that the initial differences in scores on a test in 1st-grade (with an approximate standard deviation of 0.7) continued to account for roughly a third of the gap between low- and high-SES students in high school (p. 21; see their footnote #34 on p. 30 for the details of the initial gap). This gap is roughly equivalent to the effect size of early-reading skill gains produced by children in the pilot study (using the repeated-measures method). Although researchers have demonstrated the importance of getting off on the right foot in reading (Cunningham & Stanovich, 1997), other promising interventions for preschool children (Whitehurst et al., 1999) have shown an attenuation in gains over time.

Although there are several ways to improve the study of the effects of computer-based ALT on early-reading achievement moving forward, I will focus on two. First, there was a level of precision that was not achieved in the pilot study because the amount of time spent on individual skills was not available. Consequently, usage time and early-reading skills were evaluated at a coarser level, which left questions about individual early-reading skills and their relationships to specific portions of the overall curriculum unanswered. Second, most children included in the study were not initially placed within the reading curriculum and therefore may have spent significant portions of time with material they had already mastered. Although accommodations for this limitation were attempted (e.g., the creation of the Pure sample), it would be both theoretically and methodologically superior to avoid such complications in the first place.

The hypothesized dose-response relation was supported, but two moderating variables also were discovered. The first was compliance, which suggests that it is not just important to look at how much the curriculum is used, but also to consider the specific context of recommended use. The second moderating variable was curricular progress, which suggests that curricular usage is not predictive of gain by itself, but rather it is only predictive insofar as it results in the mastery of the specific learning objectives of the program.

Appendix A – WACS 2.0 Sequence Logic for Prekindergarteners

Overview

***All** kids see* Blending (BL), Initial Sound (IS), Letter Sound (LS), Letter Recognition (LR), Vocabulary (VO)
LC will be seen if the first or last gate is FAILED.
RW, SW will only be seen if the first gate is PASSED.
NW will only be seen if the second gate is PASSED.
RC will only be seen if the last gate is PASSED.
No child will ever see SG

Detail

Gate 1 = LS
IF the child fails LS

- Vocabulary (VO)
- Listening Comprehension (LC)
- DONE

IF the child passes LS he goes on to the next gate, Real Words (RW)

Gate 2 = RW
IF the child fails RW

- Sight Words (SW)
- Vocabulary (VO)
- Listening Comprehension (LC)
- DONE

IF the child passes RW

- Nonwords (NW)
- Sight Words (SW)
- Next gate, Vocabulary (VO)

Gate 3 = VO
IF the child fails VO, he gets LC and then is DONE.

IF the child passes VO he goes on to the next gate, Reading Comprehension (RC)

Gate 4 = RC
If the child fails RC, he gets LC and then is DONE.

If the child passes RC, he is DONE.

Appendix B – WACS 2.0 Skill Difficulties and Item Count for Prekindergarteners

Note: © Waterford Research Institute. Modified and reprinted with permission.

Assessment	# Questions In Item Bank	Raw Score Range	4 & Out	Maximum Growth
Letter Recognition	118	1334 – 2333	1001	1,332
Letter Sound	86	1334 – 3333	1001	2,332
Initial Sound	57	2001 – 3000	1001	1,999
Blending	181	1334 – 4000	1001	2,999
Real Words	233	2334 – 4667	1001	3,666
Nonwords	182	2334 – 5000	1001	3,999
Sight Words	226	2334 – 4000	1001	2,999
Vocabulary	335	1334 – 5000	1001	3,999
Reading Comprehension*	35 Passages*	2001 – 6000	2001	3,999
Listening Comprehension*	43 Passages*	1001 – 5000	1001	N/A

* Passages have 8 associated questions on average.

Appendix C – WACS 2.0 Correlation With RRLWM 1.3

Note: © Waterford Research Institute. Modified and reprinted with permission.

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
BLENDING *Objective:* Combine spoken sounds to form a word.	**Phonological Awareness** 12 Blend Individual Phonemes Instruction 1 and 2* 10 Blending Riddles 2—Individual Phonemes* 8 Blending Every Sound Practice* Blending Assessment* 8 Stick and Spell* 6 Find the Picture* 8 Change One Sound* 8 One-Two-Three—Sounds* Word Traveler	**Phonological Awareness** 6 Blend Onset/Rime Instruction 1 6 Blend Onset/Rime Instruction 2* 8 Blending Dragon* 8 Blending Riddles: Onset/Rime 10 Where is the Sound* 5 Barnyard Bash 5 Circus Clown Climbers—Substitution*	NRP, pp. 2-41 PRD, pp. 80–81, 280 Adams, pp. 75–76
	Level One 9 Blend Decodable Words Instruction* 9 Blend Decodable Words Practice	**Level One** —	
	Level Two Spell and Blend Word Blending	**Level Two** —	
	Level Three Spell and Blend	**Level Three** —	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION

WACS Assessments	Reading Correlation (Explicit)	Reading Correlation (Implicit)	National Standards / Research
LETTER RECOGNITION *Objective:* Demonstrate recognition of lowercase and capital letter names.	**Level One** 19 ABC Songs What's Your Name Assessment* 5 What's Your Name? 78 Letter Checker—Capital* 78 Letter Checker—Lowercase* 52 Hidden Letter 52 Hidden Picture 52 Letter Hunt (Menu Choice) Ant, Plumber, Spider, Tractor, Trash 52 Letter Picture 52 Letter Picture Writing 52 Make a Scene 26 Find the Letter—Capital* 26 Find the Letter—Lowercase* 26 Fast Letter Fun Assessment—Jungle Gorilla* 52 Fast Letter Fun Automaticity Distinguish Letters—Introduction, Instruction Distinguish Letters Assessment* Similarities and Differences in Letters—Introduction, Instruction Similarities and Differences in Letters Assessment* 5 Name that Letter Assessments*—Aa–Ee, Ff–Jj, Kk–Oo, Pp–Tt, Uu–Zz 5 Alphabet Review—Aa–Ee, Ff–Jj, Kk–Oo, Pp–Tt, Uu–Zz	**Level One** 6 Letter Sound Songs 26 Sing a Rhyme Songs 26 Sing around the World Songs Treasure Hunt: Matching Capital Letters Capital to Lowercase Letter to Picture Catch a Match: Letters All Concepts Coloring Box: Letters Letters Make Words—Assessment* Letters Make Words—Introduction, Instruction, Practice Distinguish Letters Introduction, Instruction, Practice, Assessment* Similarities and Differences in Letters Introduction, Instruction, Practice, Assessment* Patterns—Introduction, Instruction, Practice Patterns—Assessment*	NRP, p. 2-41 PRD, pp. 80, 113–115, 184–185, 322 Adams, pp. 55, 351–352, 359–364

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
LETTER RECOGNITION CONT. *Objective:* Demonstrate recognition of lowercase and capital letter names.	**Level Two** —	**Level Two** Letter Sound Screening Sound Hunt Name that Sound 63 Readable Books 22 Read-along Books 16 Traditional Tales	
	Level Three —	**Level Three** ABC Order (1st letter) Song ABC Order (2nd letter) Song ABC Order (1st Letter) Instruction ABC Order (2nd Letter) Instruction Skill Builder Assessment—Alphabet Action* Spelling Exploration 30 Readable Books 54 Read-along Books	
	Writing —	**Writing** ABC Book Card Maker Free Choice Me by Me Sign Studio Things to Do	
LETTER SOUND *Objective:* Demonstrate identification of letter sounds.	**Level One** Find the Letter *Picture Sound Song* *Scientist Sound Song* *Show and Tell Sound Song*	**Level One** Choose a Sound 26 Sing a Rhyme 26 Read with Me Books Treasure Hunt: Matching	NRP, p. 2-132 PRD, pp. 81- 83 Adams, pp. 31, 49-50, 251-252,

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
	Apples and Bananas Vowel Song *Tongue Twister Sound Song* *Old MacDonald's Vowels Song* *S Steals the Z Song* *C-K Rap Song* *T-H Has Two Sounds Song* *Chip Chop Song* *Where Is a Whale Song* *C and G Song* Make a Scene 12 Letter Sound Instruction 12 Letter Sound Practice 9 Choose a Sound 10 Where is the Sound? Instruction* **Level Two** Sound Room Sound Hunt *Letter Sound Screening (feedback) Spell and Blend Word Blending Say and Trace **Level Three** *P-H and G-H Say Fff* Song	Letter to Picture 10 Blend Decodable Words 10 Blend Decodable Words Practice 10 Decodable Books 10 Decodable Book Practice **Level Two** Word Construction Word Pattern Introduction Word Pattern Spelling **Level Three** —	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
INITIAL SOUND *Objective:* Identify the first sound in a spoken word.	**Phonological** 8 Initial Sound Instruction 16 Initial Sound Practice 16 Right Initial Sound*	**Phonological** 3 Phoneme Eliminator (initial sound) Blend Onset/Rime Instruction 1 Blend Onset/Rime Instruction 2* Blending Dragon*	NRP, p. 2-41 PRD, pp. 152–154 Adams, pp. 80, 248, 249, 331
	Level One 12 Letter Sound Instruction 12 Letter Sound Practice 9 Choose a Sound books 10 Where is the Sound? Instruction*	**Level One** 26 Sing a Rhyme 7 Letter and Sound Songs 26 Read with Me Books Treasure Hunt: Matching Letter to Picture	
	Level Two Letter Sound Screening Sound Hunt Name that Sound	**Level Two** 8 Stick and Spell 6 Find the Picture 8 Change One Sound Blending Riddles 3 Word Traveler 5 Barnyard Bash Sound Room	
	Level Three —	**Level Three** Spelling Exploration	
SIGHT WORDS *Objective:* Identify sight words (high-frequency words).	**Level One** 20 Power Word Instruction 20 Power Word Practice	**Level One** 26 Sing a Rhyme Songs 26 Read with Me Books	PRD, pp. 182, 194, 322 Adams, pp. 160–161
	Level Two Power Word Introduction*	**Level Two** Power Word Progress	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
	Power Word Screening* Identify Power Words—Rascal Presents a Word* Spell Power Words—Spelling Scramble* Word Mastery Practice Word Assessment* Lesson Screening (Preassessment)*	Word Mastery Games 3 Power Word Readables 60 Readable Books 22 Read-along Books 16 Traditional Tales Sentence Dictation Readable Walk-through Readable Jump-through	
	Level Three Make and Spell Preassessment* Spell and Blend Spelling Instruction Automatic Word Recognition Spelling Postassessment	**Level Three** Spelling Exploration 30 Readable Books 54 Read-along Books Word Recognition Assessment	
	Writing —	**Writing** ABC Book Card Maker Free Choice Me by Me Sign Studio Things to Do	
REAL AND NONWORDS *Objective:* Combine letter sounds to form real (hat) and nonsense (vup) words.	**Level One** 9 Blend Decodable Words Instruction 9 Blend Decodable Words Practice 10 Decodable Books 10 Decodable Book Practice	**Level One** 12 Letter Sound Instruction 12 Letter Sound Practice Word Traveler 7 Letter and Sound Songs 26 Sing a Rhyme Songs 26 Sing around the World Songs	NRP, pp. 2-131–2-138 PRD, pp. 182, 184–185, 194, 322 Adams, pp. 210, 211 (pseudo words) Adams, pp. 107–135

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
		26 Read with Me Books 26 Picture Story Treasure Hunt: Matching Word to Picture Choose a Friend: Put It Together	
	Level Two Pattern Word Blending Key Word Screening: Spell and Blend* Word Blending Word Pattern Introduction Word Pattern Spelling Say and Trace Word Mastery Practice Key Word Instruction Key Word Match Pattern Hunt*	**Level Two** Lesson Screening (Preassessment)* Letter Sound Screening* Sound Hunt Name that Sound* Word Assessment* Sound Room Key Words Song 63 Readable Books (record) Watch Me Read 22 Read-along Books 16 Traditional Tales Sentence Dictation Readable Jump-Through	
	Level Three Spell and Blend Readable Word Play Songs: Bossy Mr. R Compound Words Double the Fun Drop Magic E	**Level Three** Spelling Exploration 30 Readable Books 54 Read-along Books Make and Spell Preassessment* Spelling Instruction Spelling Games*	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
	P-H and G-H Say Fff Schwa Sound Silent Letters (G and H) Silent Letters (K and G) Silent Letters (W)	Word Recognition Assessment—Create a Face* Automatic Word Recognition All-star Spelling Postassessment*	
SEGMENTING *Objective:* Divide a spoken word into separate sounds.	**Phonological Awareness** 12 Phoneme Segmentation Practice* 10 Where is the Sound—Instruction* 8 Phoneme Eliminator* 8 Change One Sound* 8 Initial Sound Instruction* 16 Right Initial Sound* 8 Final Sound Instruction* 8 Right Final Sound*	**Phonological Awareness** 5 Barnyard Bash 8 One-Two-Three—Sounds* 6 Find the Picture*	NRP, p. 2-41 PRD, pp. 81, 280, Adams, pp. 67–71
SEGMENTING CONT. *Objective:* Divide a spoken word into separate sounds.	**Phonological Awareness cont.** 5 Circus Clown Climbers 1—Substitution* 5 Circus Clown Climbers 2—Assessment*		
VOCABULARY *Objective:* To obtain a broad vocabulary in order to communicate effectively—both orally (words spoken or recognized while listening) and visually (words used in print).	**Level One** Vocabulary Introduction Vocabulary Instruction—Secret Picture Game* Vocabulary Instruction—Word Journey* Read with Me Vocabulary Instruction Vocabulary Assessment	**Level One** 52 Make a Scene 26 Sing a Rhyme Songs 26 Sing around the World Songs 26 Read with Me Books 26 Picture Story Treasure Hunt: Matching Word to Picture	NRP, pp. 4-24–4-27 PRD, pp. 63, 280 Adams, pp. 146–150

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
		Choose a Friend: Put It Together Over, Under, and Through—Introduction, Instruction, Practice Over, Under, and Through—Assessment* Top, Beside, Bottom—Introduction, Instruction, Practice Top, Beside, Bottom—Assessment* Identify Parts of the Face—Introduction, Instruction, Practice Identify Parts of the Face—Assessment* Simple Shapes—Introduction, Instruction, Practice Simple Shapes—Assessment* Make Comparisons—Introduction, Instruction, Practice Make Comparisons—Assessment*	
VOCABULARY CONT. *Objective:* To obtain a broad vocabulary in order to communicate effectively—both orally (words spoken or recognized while listening) and visually (words used in print).	**Level Two** 5 Use a Clue (Strategy)* Look for a Clue song Rusty and Rosy's Clues 63 Walk-through 8 Build Knowledge for TT 5 Build Knowledge for RA	**Level Two** 63 Readable Books 22 Read-along Books 16 Traditional Tales 8 Pre-comprehension strategies TT 16 Post-comprehension strategies TT 17 Read-Along pre-comprehension 22 Read-Along post-comprehension	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
TT = Traditional Tales RA = Read Alongs	**Level Three** Song: Reading Detective (Build Vocabulary) 12 Build Knowledge Pre-comprehension activities 6 Assessment: Build Vocabulary	**Level Three** 54 Read-along Books 30 Readable Books Pre- and Post-comprehension activities Revision—Using Interesting Words 1 Songs: Antonym Ant Homophone Monkey Put It at the Front Put It at the End Let's Compare Homophone Instruction Skill Builder Practice—Homophone Adventure* Antonyms Instruction Skill Builder Practice—Antonyms* Synonym Tree Synonym Instruction Skill Builder Practice—Synonyms* Prefixes Instruction Prefix Story* Suffixes Instruction	
VOCABULARY CONT. *Objective:* To obtain a broad vocabulary in order to communicate effectively—both orally		**Level Three cont.** Skill Builder Practice: Suffix Machine Comparing Adjectives Instruction Skill Builder Assessment: Comparing Adjective Story*	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
(words spoken or recognized while listening) and visually (words used in print).	**Writing** —	**Writing** ABC Books Card Maker Me by Me Sign Studio	
TEXT COMPREHENSION *Objective:* To obtain and construct meaning from written language.	**Level One** —	**Level One** —	NRP, pp. 4-42–4-52 PRD, pp. 6–7, 62–65, 210–211 Adams, pp. 99–103. 140–142, 146
	Level Two Comprehension Strategies Introductions for 22 Read-along and 16 Traditional Tales Books: Step into the Story Peek at the Story Ask a Question Build Knowledge Predict Compare Characters* Comprehension Strategies Assessments: Find an Answer* Sum Up: Remember Order* Describe Characters* Connect to Me* Recall Details* Sum Up: Five Ws* 5 Use a Clue (Strategy)*	**Level Two** 63 Readable Books What is a Sentence?—Song Sentences Practice Skill Builder Practice—Sentences Assessment* Sentence Marks—Song Sentence Marks—Practice Skill Builder Practice—Sentence Marks Assessment* Look for a Clue Song Context Clues: Rusty and Rosy's Clues	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
TEXT COMPREHENSION CONT. *Objective:* To obtain and construct meaning from written language.	**Level Three** Comprehension Strategies Introduction for 30 Read-along Books: Build Knowledge Peek at the Story Step into the Story Comprehension Strategies Assessment: Build Vocabulary* Check My Guess* Map the Story* Sum Up—Remember Order* Compare Characters*	**Level Three** 24 Read-along Books 30 Readable Books Reading Detective song	
LISTENING COMPREHENSION *Objective:* Demonstrate understanding of literal meaning of a story being told	**Level One** 6 Recall Details— Missing Pictures* The Apple Tree* All the Pretty Little Horses* Mother, Mother, I Am Ill* Three Little Kittens* Wee Willie Winkie* 4 Sequencing— What Comes Next?*	**Level One** 26 Sing with Me Songs 26 Sing around the World Songs 7 Letter and Sound Songs 26 Picture Story Real and Make Believe in Stories—Introduction, Instruction, Practice Real and Make Believe in Stories—Assessment*	PRD, pp. 61, 64–65, 75, 80, 214–224, 280, 332 Adams, p. 358

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
	Hey, Diddle, Diddle* Ten Little Goldfish* Little Miss Muffet* Eensy, Weensy Spider 1 Predicting Assessment—This Little Pig* Predicting Practice Picture Clues Introduction, Instruction, Practice Picture Clues Assessment*	Memorization Skills—Introduction, Instruction, Practice Memorization Skills—Assessment* Words Tell about the Pictures—Introduction, Instruction, Practice Words Tell about the Pictures—Assessment* First, Next, Last—Introduction, Instruction, Practice First, Next, Last—Assessment	
LISTENING COMPREHENSION CONT. *Objective:* Demonstrate understanding of literal meaning of a story being told		**Level One cont.** Opposites—Introduction, Instruction, Practice Opposites—Assessment* Looking at Details—Introduction, Instruction, Practice Looking at Details—Assessment*	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
	Level Two Comprehension Strategies Introductions for 22 Read-along and 16 Traditional Tales Books: Step into the Story Peek at the Story Ask a Question Build Knowledge Predict Compare Characters* Comprehension Strategies Assessments: Find an Answer* Sum Up: Remember Order* Describe Characters* Connect to Me* Recall Details* Sum Up: Five Ws* 5 Use a Clue (Strategy)*	**Level Two** 63 Readable Books What is a Sentence?—Song Sentences Practice Skill Builder Practice—Sentences Assessment* Sentence Marks—Song Sentence Marks—Practice Skill Builder Practice—Sentence Marks Assessment* Look for a Clue Song Context Clues: Rusty and Rosy's Clues	

WACS / READING CURRICULUM / NATIONAL STANDARDS CORRELATION			
WACS Assessments	**Reading Correlation (Explicit)**	**Reading Correlation (Implicit)**	**National Standards / Research**
LISTENING COMPREHENSION CONT. *Objective:* Demonstrate understanding of literal meaning of a story being told	**Level Three** Comprehension Strategies Introduction for 30 Read-along Books: Build Knowledge Peek at the Story Step into the Story Comprehension Strategies Assessment: Build Vocabulary* Check My Guess* Map the Story* Sum Up—Remember Order* Compare Characters*	**Level Three** 24 Read-along Books 30 Readable Books Reading Detective song	

Bibliography

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

National Reading Panel (2000). *Report of the National Reading Panel: Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Reports of the Subgroups.* Washington D.C.: National Institute of Child Health and Human Development

Snow, Catherine E., Burns, M. Susan, & Griffin, Peg (Eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington DC: National Academy Press.

Appendix D – Reading Activities Not Correlated With WACS

Note: © Waterford Research Institute. Modified and reprinted with permission.

WACS / READING /NATIONAL STARNDARDS CORRELATION			
Additional Reading Activities Not Correlated with WACS			
PHONOLOGICAL AWARENESS	**Level One** Get Started with Sounds What Do You Hear? Sounds in Order Rhyme Preassessment Syllable Preassessment Blending Preassessment Initial Sound Preassessment Phoneme Segmentation Preassessment Rhyme Instruction 1 2 Rhyming Words Song 6 Rhyme Instruction 2 8 Finish the Picture 10 Make It Rhyme 4 Rhyme Match 5 Rhyme Postassessment Identify Nonrhyming word (6 One Doesn't Rhyme) 2 Syllable Instruction 1 4 Syllable Instruction 2 12 Syllable Safari 5 Syllable Postassessment	**Level Two** —	**Level Three** —

WACS / READING /NATIONAL STARNDARDS CORRELATION			
Additional Reading Activities Not Correlated with WACS			
	Syllable Deletion (6 Take away Syllables) 5 Blending Postassessment 5 Initial Sound Postassesement 5 Phoneme Segmentation Postassessment		
PHONICS	**Level One** Capital Letter Assessments Lowercase Letter Assessments 26 Letter Sound Assessment 12 Individual Letter Sound Assessments 9 Blend Decodable Word Assessment 20 Power Word Assessments Play and Practice (Menu) 10 Decodable Book Assessments	**Level Two** Choose a Readable (Menu) Meet the Readable Mystery Words Write about It Reader's Choice (Menu) Play and Practice (Menu) Word Mastery Games 27 Name that Sound Assessments	**Level Three** Choose a Readable (Menu) Play and Practice (Menu) Word Recognition Assessment (Create a Face)
COMPREHENSION	**Level One** All activities listed in "Phonics: Level One"	**Level Two** All activities listed in "Phonics: Level Two"	**Level Three** All activities listed in "Phonics: Level Three"
VOCABULARY	**Level One** All activities listed in "Phonics: Level One"	**Level Two** All activities listed in "Phonics: Level Two"	**Level Three** Antonyms Assessment
LANGUAGE CONCEPTS	**Level One** Look, Listen, Match 1 Look, Listen, Match 2 Look, Listen, Match 3 Look, Listen, Match 4	**Level Two** Double the Fun Song Double the Fun Practice Adjectives Describe Song Adjectives Assessment	**Level Three** Adjectives Instruction Adjectives Practice Adjectives Assessment Adverbs Song

WACS / READING /NATIONAL STARNDARDS CORRELATION			
Additional Reading Activities Not Correlated with WACS			

	Look, Listen, Match 5
	Print Conventions
	    Introduction
	    Instruction
	    Practice
	    Assessment
	Print Directionality
	    Introduction
	    Instruction
	    Practice
	    Assessment
		Compound Words Song
		Compound Words Assessment
		Contraction Action 1 Song
		Contractions Practice
		Contractions Assessment
		Nouns Song
		Nouns Practice
		Nouns Assessment
		More Than One Song
		Plural Nouns Practice
		Plural Nouns Assessment
		Apostrophe Pig Song
			Adverbs Instruction
			Adverbs Practice
			Adverbs Assessment
			Let's Compare Song
			Let's Compare Practice
			Let's Compare Assessment
			Verbs Song
			Verbs Practice
			Verbs Assessment
			Strange Spelling Song
			Irregular Plurals Instruction
			Irregular Plurals Practice

LANGUAGE CONCEPTS CONT.	**Level One cont.**
	Common Sounds
	    Introduction
	    Instruction
	    Practice
	    Assessment
	Dots, Lines, and Circles
	    Introduction
	    Instruction
	    Practice
	    Assessment
	Look at Details
	    Introduction
	    Instruction
	    Practice
	    Assessment
	Make Comparisons
	    Introduction
		Level Two cont.
		Possessive Nouns Assessment
		Sentences Practice
		Sentences Assessment
			Level Three cont.
			Irregular Plurals Assessment
			Pronouns Song
			Pronouns Instruction
			Pronouns Practice
			Pronouns Assessment
			Irregular Verbs Song
			Irregular Verbs Instruction
			Irregular Verbs Practice
			Irregular Verbs Assessment
			Tricky Y to I Song
			Change Y to I Instruction
			Change Y to I Practice
			Change Y to I Assessment
			Word Processor Tutorial
			Word Processor

WACS / READING /NATIONAL STARNDARDS CORRELATION			
Additional Reading Activities Not Correlated with WACS			
	Instruction Practice Assessment Match Numbers 1–5 Introduction Instruction Practice Assessment Numbers 1–5 Introduction Instruction Practice Assessment		Writing Process—Edit Edit Calendar Capitals Edit Capitals Edit Commas Edit End Punctuation Edit Punctuation Edit Spelling Edit Tricky Spelling Writing Process—First Draft First Draft 1 and 2 First Draft 3 First Draft 4 First Draft 5 First Draft 6 First Draft 7
LANGUAGE CONCEPTS CONT.	**Level One cont.** One-to-one Correspondence Introduction Practice Assessment Patterns Introduction Instruction Practice Assessment Red, Yellow, and Blue Introduction Instruction Practice Assessment Sort Introduction		**Level Three cont.** Writing Process—Prewrite Writing Introduction Prewrite Mapping 1 Prewrite Mapping 2 Prewrite Mapping 3 Prewrite Mapping 4 Word Bank 1 Word Bank 2 Word Bank 3 Writing Process—Revise Revise: Add Details 1 Revise: Add Details 2 Revise: Delete Extra Words Revise: Start Sentences Differently Revise: Stick to the Topic 1 Revise: Stick to the Topic 2

WACS / READING /NATIONAL STARNDARDS CORRELATION			
Additional Reading Activities Not Correlated with WACS			
	Instruction Practice Assessment		Revise: Use Interesting Words 1 Revise: Use Interesting Words 2
FLUENCY	**Level One** —	**Level Two** —	**Level Three** 30 Fluency Speed Passage 30 Fluency Comprehension 3 Fluency Expression Instruction—Exclamations 3 Fluency Expression Instruction—Pauses 3 Fluency Expression Instruction—Phrases 3 Fluency Expression Instruction—Questions 3 Fluency Expression Instruction – Quotations The Mighty Sparrow: Expression Recordings 30 Fluency Speed Passages

Appendix E – WACS 2.0 Cross-Validation (Spring, 2009)

Note: © Waterford Research Institute. Modified and reprinted with permission.

	WACS Assessments		
	Kindergarten Skills	**1st Grade Skills**	**2nd Grade Skills**
DIBELS, Beginning Kindergarten	$r = .69$, $p < .001$ r = .68		
DIBELS, Beginning 1st Grade		$r = .68$, $p < .001$ r = .69	
DIBELS, Beginning 2nd Grade			$r = .59$, $p < .001$ r = .62
TPRI, 1st Grade		$r = .58$, $p < .001$ r = .60	
TPRI, 2nd Grade			$r = .45$, $p < .001$ r = .48
ITBS Kindergarten	$r = .44$, $p < .001$ r = .53		
ITBS 1st Grade		$r = .70$, $p < .001$ r = .74	
ITBS 2nd Grade			$r = .61$, $p < .001$ r = .67
SAT 10 1st Grade		r = .755, p < .001 r = .76	
SAT 10 2nd Grade			$r = .65$, $p < .001$ $r = .65$

Appendix F – Recruitment Flier (Front and Back)

Note: © Waterford Research Institute. Modified and reprinted with permission.


UPSTART
Utah Preparing Students Today for A Rewarding Tomorrow
POWERED BY Waterford Institute
Programa preescolar gratis para usar en su casa
UPSTART es un programa preescolar por computadora financiado por el estado de Utah que preparará a su niño de 4 años para la escuela con un comienzo divertido en lectura, matemáticas y ciencias en solo 15 minutos por día.
Gratis
Este programa gratis ofrece lecciones COMPLETAS en lectura, matemáticas y ciencias para principiantes.
Muy divertido
Su niño apreciará los juegos, libros, canciones y actividades FÁCILES DE USAR.
La calidad de Waterford
Millones de niños en todo el país han aprendido exitosamente con el programa Waterford y ahora su niño también puede hacerlo.
Podría cumplir los requisitos para usar una computadora y servicio de Internet gratis.
¡Llame ahora mismo!
Vacantes limitadas
¡Llame ahora mismo!
1-800-669-4533
(¡Hablamos español!)
www.UtahUpstart.org
FLY.UPST.001.0309

Appendix G – Demographic Coding Information

Schooling (InPreK)

Question: "Will this child attend any other preschool while participating in UPSTART?"

Coding: 1 = Yes; 0 = No.

Ethnicity (NonWhite)

Coding: 1= Non-White and/or Hispanic; 0 = Non-Hispanic White.

Socioeconomic Status (LowIncome)

Coding: 1 = Less than 200% of Federal Poverty Guidelines; 0 = At or above 200%.

Gender

Coding: 1 = Male; 0 = Female.

Primary Language (English)

Coding: 1 = English; 0 = Other.

Appendix H – Training Overview

Training Sessions

Invitations:
Participants will be sent snail-mail and e-mail invitations that list the times and locations of training sessions and will be asked to RSVP either by calling a toll-free number or e-mailing their response. There will also be an option to RSVP via the website. Families will be encouraged to bring friends, neighbors, or other people wishing to get more information about the program.

Schedule:
Dates for training sessions will begin in late June and run through August 1.

There will be two training sessions per day: A morning one for stay-at-home parents (10 AM) and an early evening for working parents (7 PM).

Locations:
Each district will have at least one day of training sessions. Areas with high concentrations of Spanish speakers will also have sessions in Spanish.

Meeting locations will include libraries, community centers, assembly halls or meeting rooms in local lodging facilities. Some meetings will be visits to participants' homes in districts where there are only one or two participants.

Session topics:

- Discussion of the importance of Early Childhood Education
- Review Program Timeline: What we've done so far and what is yet to come in the months ahead
- Town meeting-style forum—Question and answer sessions. Questions that are repeated will be collected and posted under the "Common Questions" page on the website so that those not able to attend sessions will still benefit.
- Review of resources on the website.

Staff attending will include some combination of the following:

- Training personnel
- Project leader
- Technical support
- Spanish-speaking User Support personnel

Appendix I – Parent Motivational Material

Note: © Waterford Research Institute. Modified and reprinted with permission.

For parents whose children were below usage for the week:

Dear Parent:

Here is the weekly usage chart for your child's Rusty and Rosy Learn with Me™ /UPSTART program participation. This week, your child has not used the program as much as required for UPSTART. It is very important that your child use the software at least 15 minutes a day, 5 days a week in order to experience the greatest learning gains.

If you are having any kind of technical issue or if your child is having trouble using the program, please call User Support line at XXX-XXX-XXXX. We are here to make sure you and your child have the best experience possible, and there are many ways we can help.

We look forward to hearing from you soon!

User Support
Waterford Institute
XXX-XXX-XXXX

For parents whose children exceeded usage for the week:

Dear Parent:

Here is the weekly usage chart for your child's Rusty and Rosy Learn with Me™ /UPSTART program participation. This week, your child used Rusty and Rosy Learn with Me™ even more than the recommended time per week. We are so glad to see that your child is getting the most out of this wonderful program! We hope your child is enjoying the math and science portions of the software, which are accessed after the required 15 minutes of reading instruction are completed. Keep up the great work!

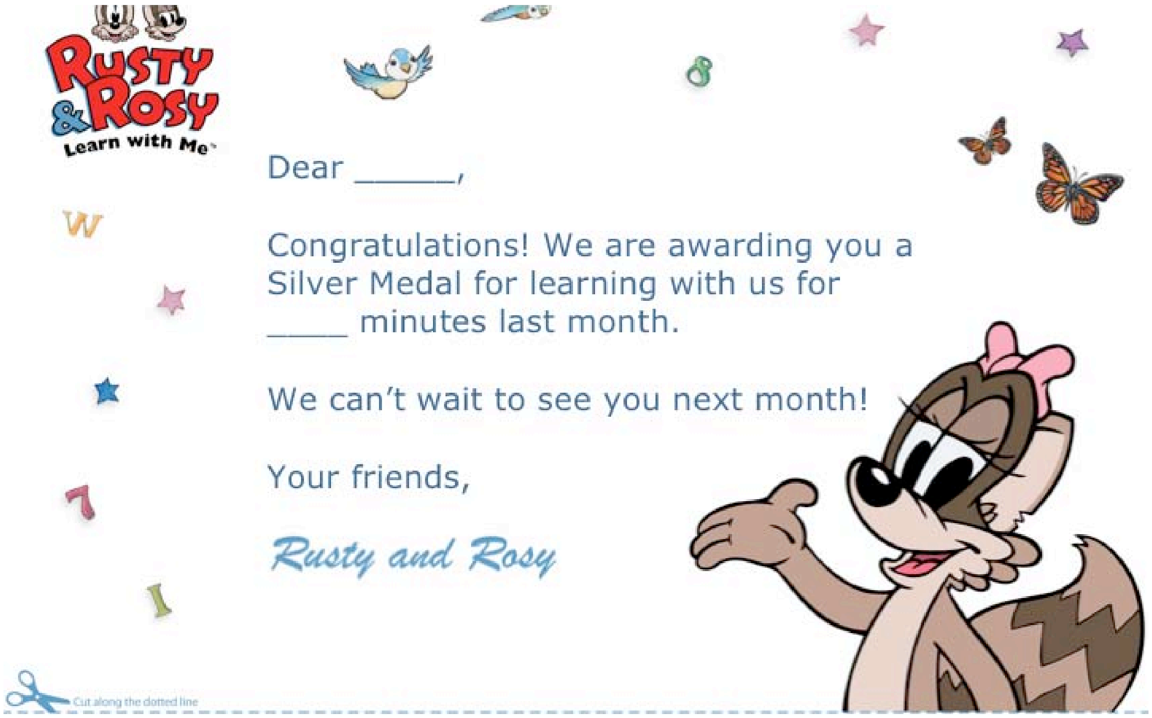
Please call XXX-XXX-XXXX if we can assist you in any way. We are here to help.

Warmest regards,

User Support
Waterford Institute
XXX-XXX-XXXX

Appendix J – Child Motivational Materials

Note: © Waterford Research Institute. Modified and reprinted with permission.



Appendix K – Selected Survey Question Results

Child Motivation

(CURRENT) Overall, does your child enjoy getting the monthly certificates?
(EXITED) Overall, did your child enjoy getting the monthly certificates?

	Total
Definitely	82%
Probably	12%
Probably not	5%
Definitely not	2%
Depends (SPECIFY)	0%
Don't know (DO NOT READ)	0%

(CURRENT) How helpful do you think monthly certificates are in keeping your child motivated to use the program ?
(EXITED) How helpful did you think the monthly certificates were in keeping your child motivated to use the program?

	Total
Very helpful	45%
Somewhat helpful	34%
Not very helpful	14%
Not at all helpful	6%
Don't know (DO NOT READ)	0%

Parent Motivation

(CURRENT) How effective are these calls in encouraging you to have your child use the program?
(EXITED) How effective were these calls in encouraging you to have your child use the program?

	Total
Very effective	48%
Somewhat effective	35%
Not very effective	7%
Not at all effective	3%
Don't know (DO NOT READ)	7%

Child Enjoyment
(CURRENT) Overall, would you say your child enjoys using the program?
(EXITED) Overall, would you say your child enjoyed using the program?

	Total
Definitely	73%
Probably	24%
Probably not	3%
Definitely not	0%
Don't know (DO NOT READ)	0%

Parent Enjoyment
If you had another child eligible for the UPSTART program, how likely would you be to enroll him or her?

	Total
Very likely	90%
Somewhat likely	7%
Somewhat unlikely	1%
Very unlikely	1%
Depends (SPECIFY)	1%
Don't know (DO NOT READ)	0%

How likely would you be to recommend the UPSTART program to a friend or relative?

	Total
Very likely	91%
Somewhat likely	7%
Somewhat unlikely	1%
Very unlikely	1%
Depends (SPECIFY)	0%
Don't know (DO NOT READ)	0%

References

Alberts, B. (2009). Making a science of education. *Science*, *323*, 15.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167-180.

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.

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

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.

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.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

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

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.

Department of Education (DOE). (2008). *Reading First.* Retrieved 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.

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

Fredrick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *The Journal of Educational Research, 73*(4), 183-194.

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.

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.

Heuston, D. H. (1997). Response to the evaluation and meta-evaluation. *International Journal of Educational Research 27*(2), 175-181.

Heuston, E. B. H. (2008). The promise of academic learning time in a dose-response model of early-reading achievement (Master's Thesis). Retrieved from http://contentdm.lib.byu.edu/ETD/image/etd2688.pdf

Hutchens, T. A., Hamilton, S. E., Town, P. A., Gaddis, L. R., & Presley, R. J. (1991, April). The stability of I.Q. in preschool years: A review. Talk presented at the meeting of the American Association for Counseling and Development, Reno, NV. ERIC #: ED353068.

Johnson, E. P., Perry, J., and Shamir, H. (in press). Variability in reading ability gains as a function of computer assisted instruction method of presentation. *Computers and Education*.

Kazdin, A. E. (2003). *Research design in clinical psychology*. Boston: Allyn and Bacon.

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

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.

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

Pon-Barry, H. (2004). In search of Bloom's missing sigma: Adding the conversational intelligence of human tutors to an intelligent tutoring system (Master's Thesis). Retrieved from http://godel.stanford.edu/old/muri/papers/HeatherPonBarryThesis.pdf

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.

Shamir, H., Johnson E. P., & Brown, K. (2009, July) Validity and reliability of the Waterford Assessment of Core Skills. Talk presented at the 74th annual meeting of the Psychometric Society, University of Cambridge, Cambridge, England.

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

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.

Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Farley & N. J. Gordon (Eds.), *Psychology and Education: The state of the union.* (pp. 81-108). Berkeley, CA: McCutchan Pub. Corp.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19-27.

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

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology*, *91*(2), 261-272.

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Boston: Morgan Kaufmann Publishers/Elsevier.