

WHITE PAPER

Validity and Reliability

Let's Go Learn's Educational Foundation

Purpose of this paper

Today educators are increasingly being asked to evaluate and justify the actions they undertake in the process of educating students. This increase in accountability has placed new demands on educators as they seek to evaluate and scrutinize outside goods and services offered to them by many educational publishers or service providers.

The purpose of this paper is to clarify the many terms that surround this evaluation process and to shed additional light on where and how Let's Go Learn's own products fit into the educational process as tools for teachers and schools. In this white paper, we will define terms and describe how Let's Go Learn's online reading assessments meet today's educational accountability requirements.

Contents

Terms	2
Foreword by Dr. Richard McCallum	3
Educational Expertise	3
Let's Go Learn's Validity	4
State Standards Alignment	5
Studies, Subtest Specifications, and Correlations	5-7
Appendix A - Study on Targeted Instruction	8
Appendix B - SBIR U.S. Department of Education Pilot Results	28

Terms

Validity - An assessment instrument is valid to the extent that it actually assesses the underlying skill or construct it is designed to assess. A properly calibrated postage scale, for example, is a valid means of assessing how much an envelope weighs. But assessing the component skills underlying a complex phenomenon like reading is much more difficult. The difference is that weight is a directly observable feature of physical reality, whereas reading skills are latent (not directly observable) traits within a person's mind. The validity of an instrument designed to assess such latent traits includes (1) Construct Validity: The theoretical connection between the instrument and the skill to be assessed – provided by the experts in the field who create the instrument, and (2) Criterion Validity: The empirical connection between performance on the instrument and other outcomes recognized as correlates of the skill to be assessed – such as correlation with other assessment instruments or relevant outcomes. The Let's Go Learn reading assessments derive their construct validity from the same techniques and content as the Cal Reads diagnostic assessments. LGL's criterion validity is established by its high correlation with scores on Cal Reads assessments and other nationally-normed commercial assessments.

Reliability - An assessment is reliable to the extent that its results are consistent over repeated administrations. Reliability is a necessary condition for an instrument to be valid. A perfectly valid and reliable instrument will give the same score over and over when assessing the same person in the same skill state. In reality, however, repeated assessments of a single individual do not result in the same score, as the person's score can be expected to increase with practice over time. The reliability of an instrument is therefore established by other means, such as comparing one part of the instrument to another part (split-half reliability) or by the internal consistency of test items, computed as Cronbach's "alpha" reliability coefficient. The reliability of the *LGL Reading Assessment* is consistently high.

Nationally Normed - Nationally norming a test means giving a particular test to a large pool of test-takers across the nation. Because it is given across a nationally representative sample of test-takers, scores can be compared to a national norm. Unfortunately, national norming by itself says nothing about the accuracy or validity of a test. A test can be nationally normed and still be a terrible test. Therefore, it is important to note the test's validity first and foremost. National norming then becomes relevant if a percentage comparison is needed against the national norm. National norming is necessary for accountability tests that rank individual programs according to a national average.

Criterion-Referenced - Often tests fall into two categories: accountability testing and diagnostic testing. Diagnostic tests generally use criterion referencing. In other words, these tests compare specific abilities to detailed measures or standards. For instance, reading specialists may state that by the early second grade students should have mastered certain phonological rules. If a student has not mastered those particular "criteria," he or she is considered below second grade level in that skill. The way that criteria are defined can vary according to the experts who defined them. However, for diagnostic purposes what is more important is that the same measurement is used to plot progress. For instance, a child may grow in height over the years. Whether one measures the child in inches or centimeters does not matter. What is important is that the measurer uses the same system so that when comparing measurements, growth or lack of growth can be recognized.

Research-Based - Something is research-based if it was developed by recognized experts in a particular field. If these experts have statistical data or studies to support their findings the claim is stronger. Often ties to universities or other research organizations help verify claims of being "research-based." Companies with no ties to public research institutions often lack third-party verification of claims and thus have less clout.

Foreword by Dr. Richard McCallum

Let's Go Learn was founded on the belief that timely and accurate assessment data is a key component of successful learning. This fact is especially true in reading: parents and teachers need both diagnostic and on-going assessment data to make effective instructional decisions for students. The goal of the *LGL Reading Assessment* is to bring the best practices in literacy assessment into an intelligent online application. To achieve this goal we began to assemble a set of reading instruments, delivered online, that will provide: 1) individualized assessment data in reading, and 2) a management system for the reporting and analysis of students' scores.

The most recent version of the *LGL Reading Assessment* provides an online tool for collecting information that might normally be collected by a teacher or specialist using informal reading inventories, word lists, reading passages, and other classroom-based diagnostic measures of reading ability. Our goal is to utilize the strengths of online technology to get individualized diagnostic assessment data into the hands of educators. Our first attempts at developing such tools have been warmly greeted by parents, educators, and administrators in schools.

There are several distinct advantages for teachers using the *LGL Reading Assessment*. First, teachers save time using this tool. Collecting individualized assessment data is time consuming; teachers will tell you that the time commitment alone is enough to mitigate against collecting such data. Second, when students are assessed in an online environment, no data is lost. That is, when a teacher or specialist assesses a child, subtle patterns in his or her behavior may be lost if the assessor is not highly trained and aware of the many nuances involved. In the *LGL Reading Assessment*, however, the thoughtful design of the test items and the database structure allow for all test data to be captured. For example, in the word analysis subtests in the *LGL Reading Assessment*, distracters were chosen with several key variables in mind: the nature of the sound pattern and its position in the word. Over the course of a subtest this information can be used to identify subtle patterns in the student's response within the measure.

Educational Expertise of Dr. Richard McCallum, Co-Founder of Let's Go Learn

Let's Go Learn was co-founded by Richard McCallum, Ph.D. For the past eight years, Dr. McCallum has been the Academic Coordinator for the Advanced Reading and Language Program in the Graduate School of Education at the University of California, Berkeley. In Dr. McCallum's program, graduate students earn Master's degrees in Reading Education and California teaching credentials as reading specialists. In addition to the course work required for the degree, Dr. McCallum's graduate students receive extensive field training through Cal Reads, a nationally recognized school-site intervention program in reading.

Cal Reads provides individualized one-to-one tutoring for low achieving intermediate and middle school students. As is the case with all effective interventions programs, Cal Reads administers individualized diagnostic reading assessments for all children served by the program. Based on these measures, an individual literacy profile is developed for every child. This profile provides the instructional roadmap for individualized reading remediation.

Cal Reads succeeds, in part, because the program collects both diagnostic and on-going assessment data on students. This detailed information is essential if we are to bring students' reading abilities back up to grade level. Unfortunately, parents and classroom teachers are not in a position to collect the type of assessment data a reading specialist or intervention program might utilize. For this reason, Richard McCallum and a small group of other experts in education and web-based business technology founded Let's Go Learn.

LGL Reading Assessment: Construct Validity

The validity of an assessment instrument refers to its ability to support valid assessment inferences. That is, do test results support a valid conclusion about a student's level of knowledge or skill? Building a valid test begins with accurate definitions of the constructs (i.e., the knowledge domains and skills) to be assessed. If the assessment activities in a test (i.e., the test items) tap into the constructs that the test is designed to assess, then the test has construct validity. Although additional factors affect overall test validity, construct validity is the basic logical bedrock of any test.

The construct validity of Let's Go Learn's reading assessments derives from the most current research-based and classroom-proven models of reading sub-skill acquisition and diagnostic reading assessment. These models are based on the work of Richard McCallum, Ph.D., a recognized expert in the use of diagnostic assessment to tailor instructional interventions precisely to students' individual profiles of reading sub-skill deficiencies. Dr. McCallum is one of the founders of Let's Go Learn and is the company's Chief Educational Architect.

Richard McCallum has many years' experience training graduate students as reading specialists at Saint Mary's College and at UC Berkeley. During this period, he and his teams of classroom intervention specialists have also trained hundreds of elementary and middle-school teachers to diagnose sub-skill deficiencies and to turn their at-risk kids into real readers. He has conducted lectures and training seminars on reading instruction and assessment for district administrators and principals. Dr. McCallum combines superb scholarly and academic credentials with many years of practical experience in implementing extremely effective classroom instructional intervention. He is uniquely suited to guide and integrate Let's Go Learn's pioneering efforts in creating and expanding an online educational environment based on valid assessment and effective instruction.

LGL Reading Assessment: Criterion Validity

The nationally recognized Cal Reads program started by Dr. Richard McCallum utilized a battery of diagnostic reading assessments as an essential part of reading remediation. It has proven that diagnostic assessment guiding targeted reading instruction can result in dramatic gains for students across differences in school attended, age, ethnicity, and gender. See Appendix A for the latest year's results. It shows a 3 and 1.8 year gain for two schools' respective student-served groups, compared to control groups in both schools that made an average of only a 0.5 year gain.

Let's Go Learn designed its assessment to measure the same reading outcomes that Cal Reads measures in its reading assessments. Taking the Cal Reads reading intervention model created by Dr. McCallum and the need to make reading assessment and instruction more available to all students, Let's Go Learn's goal was to create an online assessment that would approximate the one-to-one assessments traditionally administered by reading experts like those of the Cal Reads program. Let's Go Learn succeeded in doing this by developing an intelligent and innovative custom-web application that mimics the decisions a reading specialist would make when administering traditional one-on-one reading assessments. During the pilot performed in early 2002, funded in part by the U.S. Department of Education, Let's Go Learn demonstrated that its online assessment was highly correlated to similar reading assessments administered by reading specialists of the Cal Reads program. See Appendix B for the final pilot report submitted to the U.S. Department of Education.

Standards Alignment and Criterion Referencing

During its initial development, the *LGL Reading Assessment* based its criteria on the California Standards for grade-level achievement in reading. In addition, it drew from the best practices of current literacy assessment. As a result, the *LGL Reading Assessment* compares students' individual performances in six reading measures against a criterion of mastery for each sub-skill, resulting in an individual reading profile for every student. Today the assessment is aligned to all 50 states. Teachers can instantly compare any student's individual assessment to any state's standards.

Correlation to State Language Arts Standards

By virtue of being a diagnostic reading assessment that collects individual assessment data across six measures in reading, Let's Go Learn was able to correlate its assessment to many of the most rigorous State Language Art Content Standards. Immediately after a student completes an assessment, teachers or parents are able to view the student's performance against the State standards of their choice. All 50 states are available.

Accuracy and Refinement

Assessment, by its very nature, attempts to measure a skill with test items that sample across a given domain. Common sense dictates that the shorter the test, the less accurate it is. For this reason, the *LGL Reading Assessment* sees itself as in a class separate from the 10-15 minute reading assessments offered by multiple other educational publishers. Our assessment takes approximately 45 to 60 minutes and adapts to each test-taker as he or she undertakes an assessment. The detailed report that we produce for each student is far more accurate and diagnostic than what a "short" test can offer. Of course, assessment improvement is an ongoing process. Through detailed item analysis and sub-test refinement we are continuing to improve our assessment. By improving our assessment's adaptation to the test-taker, we expect to be able to reduce the total test time by 25% in 2004. By performing ongoing item analysis with specific pools of students, we are improving our accuracy even more.

Test-Retest Study: Q3 2003

Test-Retest is the ability of a test to be taken once and then immediately again and have similar results. Let's Go Learn undertook a second test-retest study in Q3 of 2003. This data was added to an earlier study performed in Q1 of 2003 in order to provide a larger sampling and to reduce the margin of error. The combined results were once again excellent. Variability was low, meaning that the *LGL Reading Assessment* is very precise and can be re-administered with low bias.

Grade level delta is the mean measured ranges between test and retest. SE is standard error for mean delta in grades.

• High-Frequency Words	Grade level delta: 0.07	SE=0.11
• Word Recognition	Grade level delta: -0.03	SE=0.12
• Word Analysis	Grade level delta: -0.02	SE=0.07
• Word Meaning	Grade level delta: -0.17	SE=0.19
• Spelling	Grade level delta: -0.07	SE=0.10
• Silent Reading	Grade level delta: 0.07	SE=0.13

Subtest Specifications

High-Frequency Words: 72 criterion-referenced words. 24 words per grade from 1st to 3rd grade.

Word Recognition: 120 criterion-referenced words. 10 words per grade from 1st to 12th grade.

Word Analysis: 80 criterion-referenced words. 20 words per grade from 1st to 4th grade.

Word Meaning: 60 criterion-referenced words. 5 words per grade from 1st to 12th grade.

Spelling: 60 criterion-referenced words. 5 words per grade from 1st to 12th grade.

Silent Reading: 12 Flesch-Kincaid leveled passages with 6 questions per passage. 1 passage per grade.

Reading Level Calculation Adherence

In the silent reading subtest of the *LGL Reading Assessment*, the following method for reading level calculation was chosen. The sixth subtest, silent reading, is made up of 12 reading passages for grades 1 through 12. These passages were systematically constructed to adhere to the Flesch-Kincaid Reading Grade Level Index. This index calculates the number of words, syllables, and sentences in a given passage. It then utilizes the average syllables per word and words per sentence to articulate a readability formula. After researching the methods used by textbook publishers and children's book publishers, our research found this to be the most reliable and widely accepted system for leveling reading material.

Let's Go Learn Correlates Significantly to the Nationally Recognized Cal Reads Program

In its FIRST comparison to one-on-one paper-and-pencil assessments performed by Cal Reads reading specialists, Let's Go Learn achieved high correlations with statistical significance beyond the $\alpha=.01$ level. Cal Reads reading specialists used the following assessments in their face-to-face assessments with the students in the study:

Sight-words familiarity: Fry's high-frequency word list

Word recognition: *Diagnostic Assessments of Reading* (DAR) published by Riverside Publishing

Word meaning: *Diagnostic Assessments of Reading* (DAR) published by Riverside Publishing

Spelling: *Diagnostic Assessments of Reading* (DAR) published by Riverside Publishing

Silent Reading: *Qualitative Reading Inventory* (QRI) published by Pearson Allyn & Bacon.

- Sight-word familiarity $r=.89$ $(n=17)$
- Word recognition $r=.81$ $(n=20)$
- Word meaning $r=.60^*$ $(n=20)$
- Spelling $r=.78$ $(n=20)$
- Silent reading $r=.89$ $(n=19)$

This study was conducted in the Tahoe/Truckee Unified School District in California in 2/02. Students were tested both by Cal Reads reading specialists and online using Let's Go Learn within a three-week time period. The Word Analysis subtests were not compared because of incompatible methods with which Cal Reads and Let's Go Learn reported their final results. See Appendix B for more information surrounding this pilot.

* Lower correlation results in Word Meanings were traced to two confusing items in this initial study. Shortly after this initial study, the Word Meaning component of the *LGL Reading Assessment* was modified to improve its correlation.

Item Analysis Major Revision: Q1, 2003

Item analysis is performed across all LGL assessment systems. A pool of 1000 students was used.

Items across all six subtests with more than 75% of students answering correctly or fewer than 25% answering correctly were flagged. These extreme values represent errors that are outside the range the items were designed for.

Example 1: “and” received an unusually high error rate. The word “an” was a distracter that was selected with a high percentage. Conclusion: The audios of “and” and “an” are too similar. Students might not be hearing the /d/ sound and thus may think the target word is “an.”

Example 2: Target vocabulary word: “Caravan.” Too often students chose a picture of one car and one van. Overall the error rate was too high for this particular word. Conclusion: Many students define “caravan” as the Dodge Caravan vehicle and not as a line of camels walking through the desert.

LGL Reading Assessment Comparison to Nationally-Normed Paper-and-Pencil Assessments:

April 2003

Tested students (Grade range: 2-6) within one week’s time on the *LGL Reading Assessment* and the following paper-and-pencil tests.

LGL HFW subtest and the Slosson Oral Reading Test
LGL WR subtest and the Woodcock Word Identification Test
LGL WA subtest and the Woodcock Word Attack

Correlation (HFW & SORT):	r=0.95	SE=0.073	(n=21)
Correlation (WR & WI-W):	r=0.92	SE=0.088	(n=21)
Correlation (WA & W-WA):	r=0.91	SE=0.097	(n=21)

High correlation demonstrates criterion validity of the *LGL Reading Assessment*.

Recalculation of April SP and SR results + New Tests October 2003

LGL SP subtest and the WRAT
LGL SR subtest and the Gray Oral Reading Test

Correlation (SP & WRAT):	r=0.85	SE=0.210	(n=21)
Correlation (SR & GORT):*	r=0.65	SE=0.250	(n=21)

Medium to high correlations demonstrate concurrent validity of the *LGL Reading Assessment*.

* Lower correlations to the GORT are attributed to a high variability observed in the GORT results. Students inconsistently tested well above their grade levels on the GORT. Subsequent SR comparisons with more consistent paper-and-pencil assessments are recommended.

Research Awards

In September 2001, the U.S. Department of Education awarded Let’s Go Learn a prestigious grant reserved for private companies. This SBIR grant seeks to support innovative new technologies that bring with them the ability to realize scalable solutions in education. The U.S. Department of Education recognized Let’s Go Learn’s unique team of experts and product plan and gave Let’s Go Learn extremely high marks across the entire spectrum of review criteria.

Appendix A

Final Report to School, University Partnership Program

CAL Reads

For Academic Year 2001-2002

Richard D. McCallum, Ph.D.

**University of California, Berkeley
Graduate School of Education**

Background

CAL Reads: Literacy Intervention for Low Achieving Urban Youth. The overall goal of Cal Reads is to increase the literacy achievement of compensatory education intermediate and middle school students. Participating students receiving services through CAL Reads have typically made average gains of two grade levels or more in a number of reading measures. During the academic year 2001-02, CAL Reads served students at two schools within two local districts: Sixth grade students at Lowell Middle School in Oakland Unified School District, and 4th and 5th grade students at Malcolm X School in Berkeley Unified School District. Over the course of the year 51 total students participated in the program. These students were served by approximately 35 volunteer undergraduate tutors from UCB, and 12 graduate students in the Graduate School of Education at UCB.

The heart of CAL Reads is an individualized tutorial that is tailored to each child's particular needs. Participating students are given a diagnostic assessment battery, and based upon these assessments, lessons are designed to teach these students the range of strategic behaviors that they need to read texts assigned in the school curriculum. Students work with their tutors three times a week for 50 minutes. CAL Reads completed its fifth year of operation at the close of Spring 2002; during the academic year 2001-02 approximately 51 low achieving students in grades 4-6 at two participating schools received CAL Reads' services.

CAL Reads is a collaboration that brings to bear resources from several areas: urban San Francisco Bay Area School Districts, the Graduate School of Education (GSE) at U.C. Berkeley, and AmeriCorps National Service program. The GSE provides expertise in the area of literacy assessment and instruction through the participation of faculty and graduate students. These individuals provide oversight to the program, coordinate scheduling, conduct literacy assessments, plan individualized instructional lessons, coordinate the evaluation and data-gathering associated with research efforts, and provide professional development training to tutors and teachers in the project. Tutors are drawn from the undergraduate population at UCB. The school district and school sites provide a room on-site for the housing of the administrative components of the project, a .4 FTE ("full – time equivalent") at each site to assist in program coordination, and various materials utilized in instruction and administration of the program.

CAL Reads is operationally- defined by five main areas: assessment, instruction, materials, training and administration.

1. Assessment. All students involved in the program complete an assessment battery composed of ten sections drawn from a variety of commercially-available informal reading inventories and commonly-used reading methods textbooks. Components of these informal measures are adapted so as to insure that sufficient information is being collected to identify the strategies that students are employing when approaching text.

The assessment battery contains the following measures: an Interest Inventory (adapted from Ruddell, 1998); Sight-Word Recognition (drawn from Gillet & Temple, 1994); Word Recognition (in isolation) and Word Analysis (both drawn from The Diagnostic Assessment of Reading, Rosewall & Chall, 1992); Echo Reading (drawn from Gillet & Temple, 1996, and used with passages adapted from The Qualitative Reading Inventory, Leslie & Caldwell 1995); and Oral Reading, Silent Reading, Spelling, and Word Meaning (all drawn from Rosewall & Chall, 1992).

Assessment measures are administered and interpreted by assessment teams composed of project staff. The teams administer and interpret data, and present it in individual write-ups. Copies of all assessment write-ups are given to the classroom teachers and are discussed in teacher-training sessions. All measures are administered at three points during the academic year: in the fall, winter and spring.

2. Instruction. The instructional component of the intervention is based on a modified Reading Recovery lesson format (Clay, 1993). The tutor's role in these settings is to provide the supportive teaching necessary to 1) develop the student's awareness of the processes employed while reading; 2) help the student develop new strategies for attacking text; and 3) create opportunities for the student to experience success. The lessons are structured such that the attention of the tutor and student is on strategies or operations that lead to successful comprehension.

Each fifty-minute lesson is personalized and prepared by the coordinators. Each is composed of the following activities: 1) check-in (both an opportunity for the student to transition into the lesson and for the tutor to ascertain the student's mood, focus, and attitude at the moment); 2) skill-work (word segmenting and sorting activities or various "word – bank" activities, for example); 3) guided reading (with a focus on the application of word recognition and comprehension strategies); and 4) interactive writing.(with a focus on spelling strategies and basic writing skills). Following each lesson the tutor spends five to ten minutes responding, in writing, to various questions on a "lesson evaluation form."

3. Materials. The materials employed are all drawn from commercially - available leveled reading materials. Care is taken to provide material at students' instructional levels which make a visual impact, attract interest and generally motivate students to engage further in the text. Texts represent a range of subject matters including science, social science, history and literature. Students have opportunities to appraise and then choose the materials employed in the tutorials.

4. Tutor Training. As noted above, the tutors are undergraduate and graduate students from the University of California, Berkeley (UCB). The purpose of training component of the program is to help tutors develop the skills that are required to accelerate their tutees' literacy development. Training sessions are designed and conducted by the project coordinators and are delivered through a credit bearing university course offered by the Graduate School of Education and taught by the Project Coordinator. The course meets once a week for 3 hours, and attendance is a key requirement for passing the course.

Within the course, participating tutors are introduced to the goals of the program, the underlying model of reading employed, the lesson format, activities employed within the lesson, tips for working with middle school students, and the administration of the program. Specific presentations are made covering various approaches to comprehension and word – strategy instruction, and specific issues and questions that tutors might have about their particular tutee. Tutors are asked to bring their questions to the sessions, and considerable time is spent problem-solving and discussing directions for instruction. These discussions give coordinators opportunities to reinforce material presented earlier as well as to provide on-going monitoring of progress, encouragement and support.

Tutors also receive individualized feedback from the CAL Reads coordinators who write the lesson plans for their tutees. After each lesson, the tutor evaluates the session by responding to various questions on a post – lesson feedback form. Coordinators review these comments and in turn respond, both in writing on the subsequent plan and in face-to-face conversations with the tutor. Coordinators also observe tutoring sessions and follow up with verbal feedback to tutors. This on-site training reinforces and extends the course work by linking the broader issues to the specific context of the school and student.

Additional evaluative feedback is gained by examining the written reflections that tutors are asked to complete during the course and tutoring experience. Tutors are asked to reflect on their views of low achieving students, their role in instruction and the nature of the language they employ, the nature of schooling and its impact on students, and the nature of the change in their

own views as they take part in the program.

5. Administration. The day to day aspects of the project rest with the site coordinators. For this proposal, the school provided a .5 FTE to cover the coordination of the project. This individual had the responsibility for arranging space and materials, acting as a liaison with teachers, coordinating tutors, maintaining the files, and producing individualized lesson plans. A second coordinator from UC Berkeley, had responsibility for the overall management of the program. This person trained and managed the graduate student researchers (GSR's) who, in turn, had responsibility for conducting assessments, writing three lesson plans a week for each of ten tutoring pairs, training tutors, and meeting with teachers and other project members.

Design and Methodology

Participants. Participating students were chosen for inclusion in the program based on a two step process. First, all students in the targeted grade levels who scored two or more years below grade level on the SAT9 were identified at each site. Second, participating students were chosen at random from the overall pool. At Malcolm X, approximately half of the total pool was chosen to participate, and the remaining students were designated as the control group for the purposes of this study. At Lowell, approximately 30 students were chosen from the larger pool of 150 students, and another 30 students were selected to form a control group.

All students involved in the study were given the same assessment battery at three times during the school year: fall, winter and spring. The primary assessment used to determine general grade level gains in reading ability was the Qualitative Reading Inventory. The Qualitative Reading Inventory assesses both oral reading (oral fluency and decoding) and comprehension of level passage. Scores from both oral and silent reading are taken into account when determining an informal reading level (IRL). The highest level of mastery on the Qualitative Reading Inventory determines, in part, their informal reading level. Both comparison and intervention students were administered pre, mid and post assessments. In addition, a range of covariates were collected for all participating students. These variables included status (treatment or control), school, year of birth, gender and ethnicity.

Key Research Questions. Five key questions were tested in this report. These key questions are listed below.

1. Question: Are the treatment and control groups comparable?
2. Question: Which set of variables appears to be the most strongly related to gains from pre-post test?
3. Question: Is program participation associated with higher mid assessment scores?
4. Question: Is program participation associated with an increase in IRL scores from time 1 (mid) to time 2 (post)?
5. Question: Is program participation associated with higher post assessment scores?

Results

Question 1: Are the treatment and control groups comparable?

All subsequent analyses are contingent upon the results of the tests associated with question #1. If the groups (treatment and control) are not comparable, the assumptions necessary for employing the regression/ANOVA model will not be met.

A logistic regression was employed to answer the question. The groups appear to be comparable. (see Figure #1). No relationship was found between STATUS (placement in treatment/control group) and SCHOOL, YOB, GENDER, and EHTNICIY (checked all possible reference groups).

FIGURE 1: Binary Logistic Regression: Test of Comparable Groups

Link Function: Logit

Response Information

Variable	Value	Count
Status	1	52
	0	40
	Total	92

92 cases were used
2 cases contained missing values

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	0.0848	0.9423	0.09	0.928			
IRL0	-0.0571	0.1286	-0.44	0.657	0.94	0.73	1.22
School	0.0038	0.1498	0.03	0.980	1.00	0.75	1.35
Gender	0.5669	0.4498	1.26	0.208	1.76	0.73	4.26
AA	0.0136	0.6687	0.02	0.984	1.01	0.27	3.76
MidEast	10002	500011	0.02	0.984	*	0.00	*
API	1.487	1.267	1.17	0.241	4.42	0.37	52.98
Other	-1.480	1.335	-1.11	0.267	0.23	0.02	3.11

Log-Likelihood = -57.328

Test that all slopes are zero: G = 11.313, DF = 7, P-Value = 0.125

Question 2: Which set of variables appears to be the most strongly related to IRL2 (Total IRL from post test)?

A best subsets regression was used to address this question. (see Figure #2). Results indicate that IRL0 (Total IRL from pre test) and STATUS are the two most strongly related variables. The fact that IRL0 came up is expected; the fact that STATUS came up is encouraging. This is exploratory, just to get a handle on what to expect with the other regression models.

Figure 2: Best Subsets Regression

Response is IRL2

80 cases used 14 cases contain missing values.

Vars	R-Sq	R-Sq (adj)	C-p	S	S	S	G	M	O	A	A	P	E	I	R
					t	a	h	I	n	E	t	o	Y	R	d
					u	o	O	L	e	A	s	P	e		
					s	l	B	0	r	A	t	I	r		
1	34.9	34.0	25.7	1.5700				X							
1	9.2	8.1	65.8	1.8535	X										
2	52.0	50.8	1.0	1.3563	X	X									
2	36.4	34.7	25.4	1.5616		X	X								
3	53.7	51.9	0.2	1.3403	X	X	X								
3	53.3	51.5	0.9	1.3466	X		X			X					
4	54.9	52.5	0.5	1.3328	X	X	X			X					
4	54.0	51.5	1.9	1.3458	X		X	X		X					
5	54.9	51.9	2.4	1.3407	X	X	X			X	X				
5	54.9	51.9	2.4	1.3409	X	X	X			X	X				
6	55.0	51.3	4.3	1.3489	X	X	X			X	X	X			
6	55.0	51.3	4.3	1.3491	X	X	X	X		X	X	X			
7	55.1	50.8	6.1	1.3563	X	X	X			X	X	X	X		
7	55.0	50.7	6.2	1.3576	X	X	X	X		X	X	X	X		
8	55.2	50.1	8.0	1.3651	X	X	X	X	X	X	X	X	X		
8	55.1	50.1	8.1	1.3655	X	X	X	X	X	X	X	X	X		
9	55.2	49.4	10.0	1.3747	X	X	X	X	X	X	X	X	X	X	

Question #3: Is program participation associated with higher mid assessment scores?

The answer appears to be yes. (see Figure #3). Holding IRL0, SCHOOL, YOB, GENDER, and ETHNICITY constant, program participants scored 0.87 points higher on the IRL1 (mid test) than non program participants. This is about 1 school year's (9 month's) worth of reading gain in about half the time. To the extent that the randomization of treatment and control groups was effective (i.e. that groups are comparable), the results support the claim that this is a causal effect.

There are two very strong caveats. First, the residual plots indicate a problem with the model. Second, the analysis is based on only 65 of 94 cases. Apparently, not every student has a score for IRL1, so these results may not represent what participants as a whole experienced. To determine how representative these results are, we need to determine if missing cases are missing at random. Also note that SCHOOL, YOB, GENDER, and ETHNICITY were not significant, meaning that once the other variables were held constant (especially IRL0), these covariates were not associated with different IRL1 scores.

Figure 3: Regression Analysis

The regression equation is

$$\text{IRL1} = - 1.0 + 0.903 \text{ Status} + 0.874 \text{ IRL0} + 0.136 \text{ School} + 0.011 \text{ YOB} \\ - 0.115 \text{ Gender} + 0.385 \text{ AA} - 0.711 \text{ MidEast} + 1.03 \text{ API} + 0.063 \text{ Other}$$

65 cases used 29 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	-0.96	22.05	-0.04	0.965	
Status	0.9032	0.3514	2.57	0.013	1.2
IRL0	0.87417	0.09753	8.96	0.000	1.2
School	0.1362	0.1553	0.88	0.384	2.2
YOB	0.0111	0.2404	0.05	0.963	2.0
Gender	-0.1154	0.3292	-0.35	0.727	1.1
AA	0.3850	0.4986	0.77	0.443	2.0
MidEast	-0.7108	0.9181	-0.77	0.442	1.5
API	1.0300	0.8109	1.27	0.209	1.5
Other	0.0625	0.9428	0.07	0.947	1.6

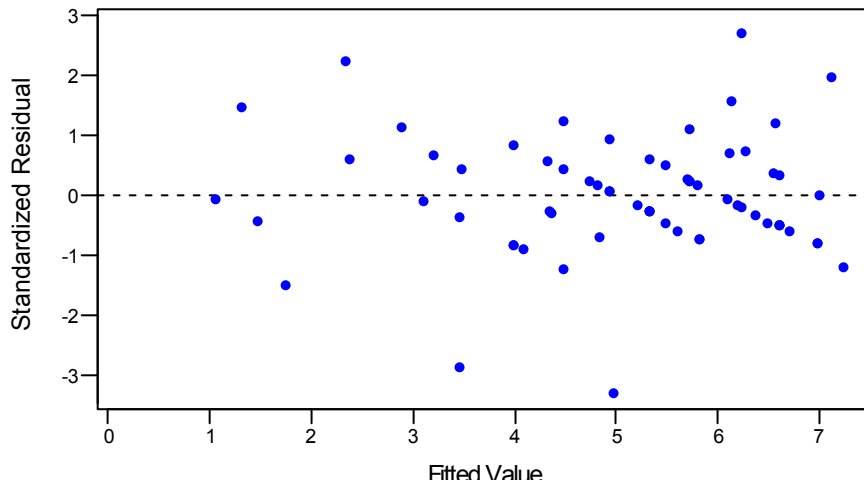
S = 1.275 R-Sq = 62.5% R-Sq(adj) = 56.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	148.950	16.550	10.18	0.000
Error	55	89.412	1.626		
Total	64	238.362			

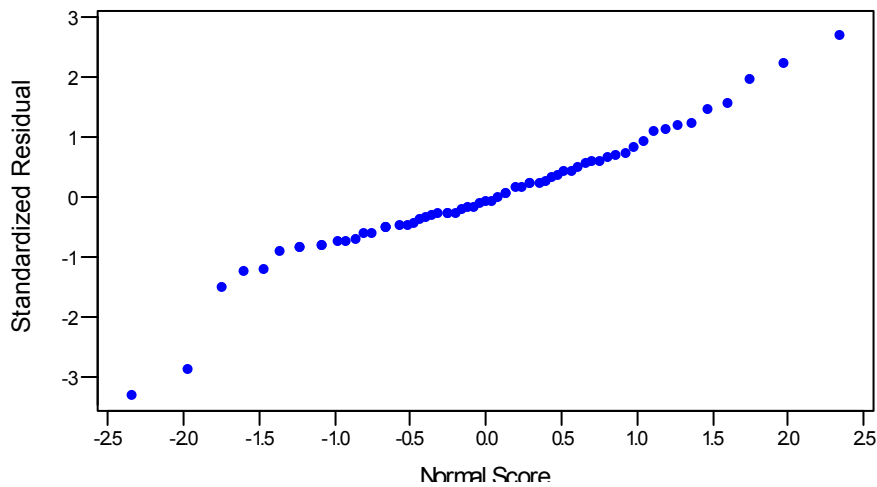
Residuals Versus the Fitted Values

(response is IRL1)



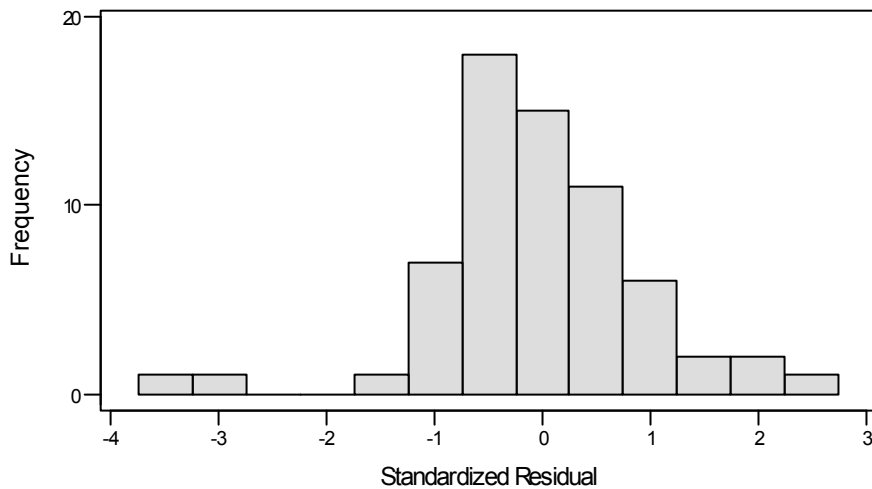
Normal Probability Plot of the Residuals

(response is IRL1)



Histogram of the Residuals

(response is IRL1)



Question #4: Is program participation associated with an increase in IRL scores from time 1 (mid) to time 2 (post)?

Again the answer appears to be yes. (see Figure #4). Students exhibited a similar gain from mid to post (0.83 points on the IRL) holding the other variables constant. Again, SCHOOL, YOB, GENDER, and ETHNICITY were not significant. The same two strong caveats apply to this model as the previous model.

Figure 4: Regression Analysis

The regression equation is

$$\text{IRL2} = -13.2 + 0.833 \text{ Status} + 0.772 \text{ IRL1} + 0.090 \text{ School} + 0.160 \text{ YOB} \\ + 0.211 \text{ Gender} + 0.243 \text{ AA} - 0.10 \text{ MidEast} - 0.287 \text{ API} + 0.689 \text{ Other}$$

63 cases used 31 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	-13.16	22.47	-0.59	0.561	
Status	0.8325	0.3517	2.37	0.022	1.2
IRL1	0.77168	0.08389	9.20	0.000	1.1
School	0.0901	0.1525	0.59	0.557	2.1
YOB	0.1597	0.2454	0.65	0.518	2.1
Gender	0.2109	0.3278	0.64	0.523	1.1
AA	0.2429	0.4872	0.50	0.620	1.9
MidEast	-0.099	1.052	-0.09	0.925	1.4
API	-0.2867	0.7884	-0.36	0.718	1.5
Other	0.6887	0.9071	0.76	0.451	1.5

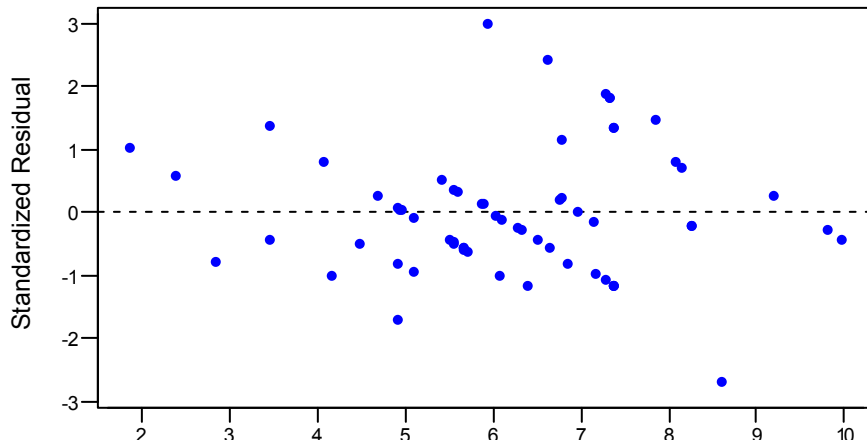
S = 1.239 R-Sq = 66.9% R-Sq(adj) = 61.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	164.378	18.264	11.90	0.000
Error	53	81.336	1.535		
Total	62	245.714			

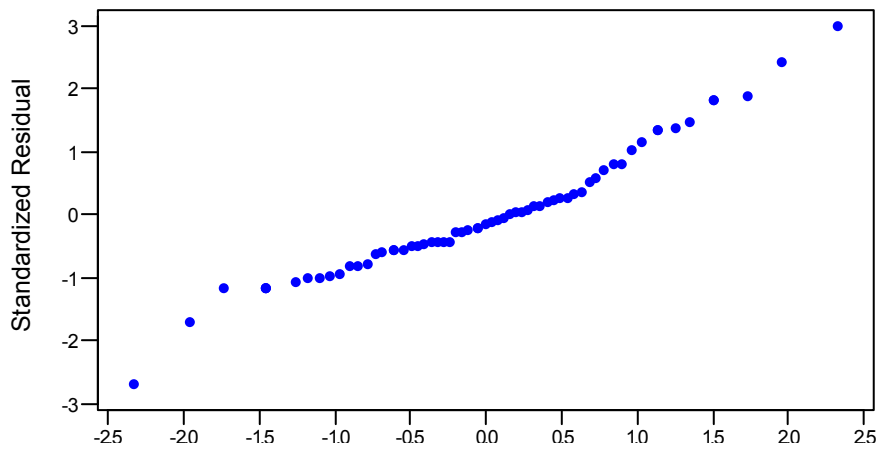
Residuals Versus the Fitted Values

(response is IRL2)



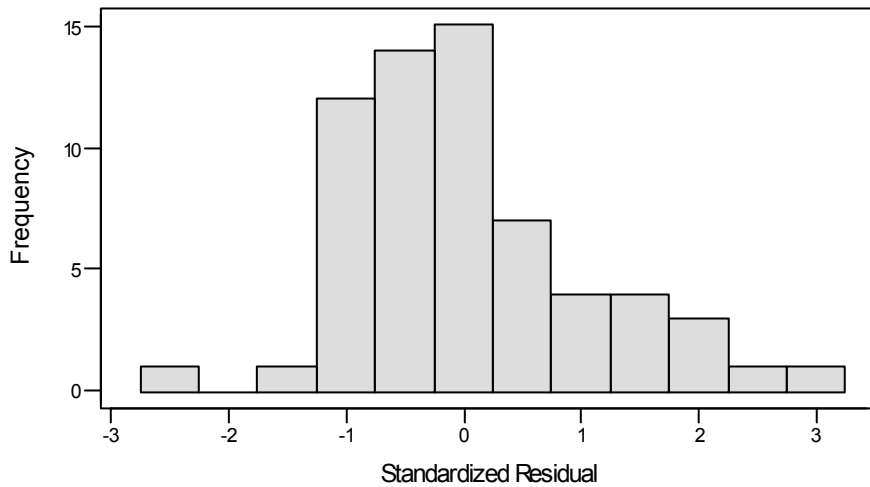
Normal Probability Plot of the Residuals

(response is IRL2)



Histogram of the Residuals

(response is IRL2)



Question #5: Is program participation associated with higher post assessment scores?

This is the key question for the analysis: Are students better off for having participated in the program? Again the answer appears to be yes. (see Figure #5). Participants scored 1.72 points higher on the post test than non-participants, holding all the other variables in the model constant. This is very large program effect -- a year and three-quarters reading gain in less than a school year. And given that randomization looks good (at least with an initial test), the data support the causal claim that this effect is due to the students participation in the program.

Again, SCHOOL, YOB, GENDER, and ETHNICITY were not significant. And, the caveats mentioned in previous tests are toned down a bit for this model. First, the residual plots look much better. There is a slight positive skew, but that could be adjusted for pretty easily. And there are a few outliers (students who improved a much more/less than expected) that are perhaps of interest for qualitative follow up. Second, the model is based on 80 of the 94 cases. 80 out of 94 cases provide a much better starting place than 65 or 63 out of 94.

Figure 5: Regression Analysis

The regression equation is

$$\text{IRL2} = -1.4 + 1.72 \text{ Status} + 0.723 \text{ IRL0} + 0.189 \text{ School} + 0.030 \text{ YOB} - 0.076 \text{ Gender} + 0.226 \text{ AA} - 0.832 \text{ MidEast} + 0.382 \text{ API} + 0.463 \text{ Other}$$

80 cases used 14 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	-1.45	23.22	-0.06	0.950	
Status	1.7218	0.3280	5.25	0.000	1.1
IRL0	0.72294	0.09291	7.78	0.000	1.2
School	0.1893	0.1554	1.22	0.227	2.3
YOB	0.0298	0.2527	0.12	0.906	2.3
Gender	-0.0763	0.3187	-0.24	0.811	1.1
AA	0.2256	0.4963	0.45	0.651	2.0
MidEast	-0.8320	0.9617	-0.87	0.390	1.4
API	0.3822	0.7787	0.49	0.625	1.5
Other	0.4630	0.8798	0.53	0.600	1.6

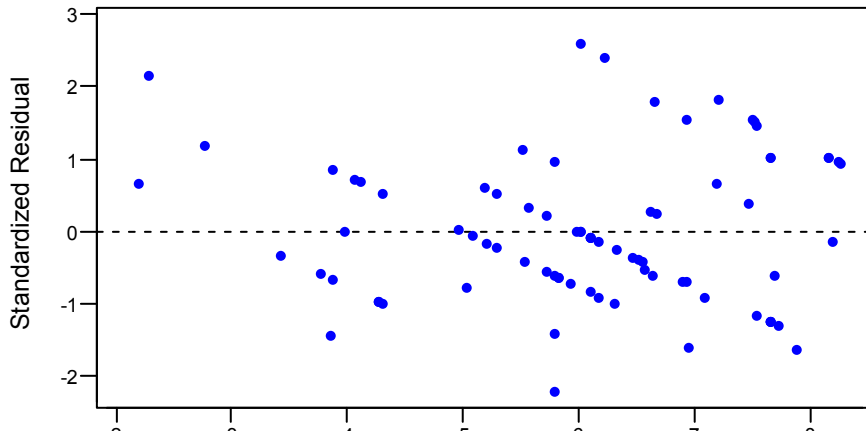
S = 1.375 R-Sq = 55.2% R-Sq(adj) = 49.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	162.888	18.099	9.58	0.000
Error	70	132.284	1.890		
Total	79	295.172			

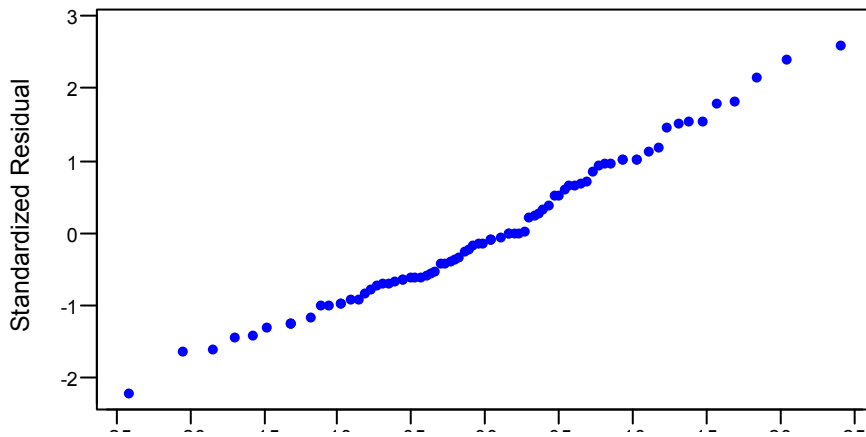
Residuals Versus the Fitted Values

(response is IRL2)



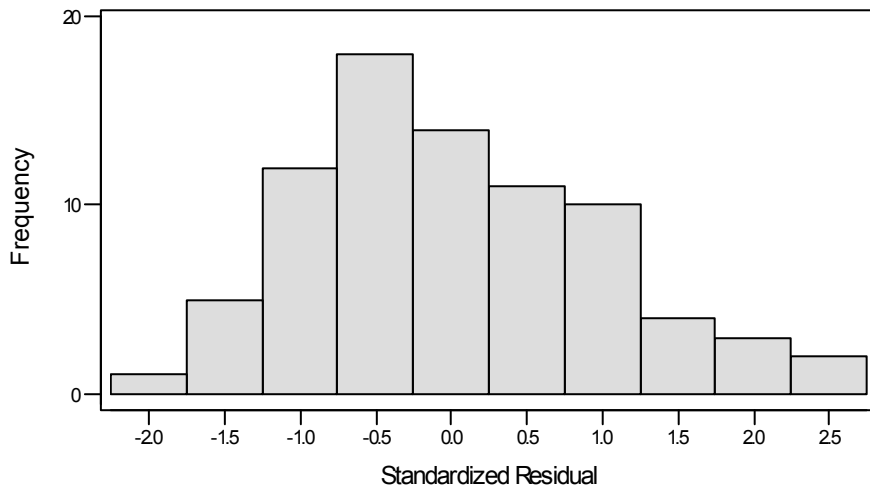
Normal Probability Plot of the Residuals

(response is IRL2)



Histogram of the Residuals

(response is IRL2)



Discussion

Results of the regression/ANOVA analyses suggest that the program had a significant effect on the reading ability of students who took part in the program. The finding that students in the program scored, on average, 1.84 and 3.0 grade levels higher at the end of instruction (for Malcom X Elementary and Lowell Middle schools respectively) than students who were not in the program (0.5 grade gain for both control groups), suggests that the program had a powerful impact on participating students. Further, this effect held even when other covariates such as ethnicity, date of birth, and school were considered. In other words it appears that the program is effective both for elementary students and middle school students regardless of their ethnicity.

This finding is consistent with results from previous analyses of CAL Reads. Over the course of the last five years we have been able to show that low achieving students, when given individualized strategy based instruction can make significant gains in reading – even though instruction occurs twice a week for 50 minutes and is conducted by undergraduate volunteer tutors. There are two key issues that are implicit in this finding that must be acknowledged, even though they were not the focus of this report.

The instruction that tutors provided to these low achieving students was supported by two very key components of the program. First, the tutors were involved in a semester long training course taught by members of the CAL Reads team. In this course, tutors were introduced to the nature of the reading process, their role in instruction, and attuned to the key dimensions of the program that are designed to foster independent strategic reading. These issues were addressed via video-taped analysis of instruction, activities, and simulations in class. In addition to the issues addressed in the course, tutors received specific on-site training from CAL Reads staff. Tutors were supported on-site with assistance adjusting and adapting instruction (is the text too hard?), changing their focus and language employed (what are you asking your student to attend to?), and building interest and motivation for students. The focused and specific nature of the training is a key element in the success of the program.

The second key dimension of the program that must be acknowledged is the role that the Graduate Student Researchers (GSR's) played in the implementation of the program. As mentioned, tutors were not asked to conduct assessment, write lesson plans or in other ways carry

the responsibility associated with planning for instruction. These tasks and responsibilities fell to the GSR's. As former teachers and reading specialists in training, these individuals were key to the effective implementation of the program. Their contribution can not be underestimated. In effect, these GSR's were the tutors for the tutors, and such a design make sense given the model of learning implicit in the program (tutors provide scaffolding and support as students learn to read. GSR's play the same role for the tutors).

CAL Reads is based on the belief that professional development in reading (for whatever constituent group – tutors, teachers, specialists, administrators) must be anchored upon direct interaction and work with students. The knowledge that tutors and GSR's take away from their experience with CAL Reads (and I believe that this knowledge is considerable) is directly related to the fact that we did not just TALK about reading (the normal course of events for professional development) we TAUGHT real kids, with real problems how to be strategic readers.

Collaborators

1. Graduate School of Education

Provides structure, graduate students, and training course for tutors. CAL Reads is also housed in the GSE.

2. AmeriCorps

Provided each site with two AmeriCorps members who recruited tutors, managed tutor attendance and tutor appreciation. Each AmeriCorps member also tutored one student each. AmeriCorps also organized an event called “Cubs for a Day” for tutors and tutees.

3. America Reads

Provided work study tutors.

4. Lowell Middle School (OUSD) and Malcolm X Elementary School (BUSD)

Provided space, materials and supplies on site, support from teachers, and .4FTE who assisted in the coordination of the program and acted as a liaison between UC Berkeley and the school.

5. Patten College

Provided undergraduate tutors who were pursuing a career in teaching. Tutors participated in training course and received credit for it at Patten College.

**CAL Reads
Malcolm X Elementary School
Participants
2001-2002**

Name	OUSE ID	Grade	Gender	Ethnicity	Date of Birth	WR-Pre	WR-Post	Gain	WM-Pre	WM-Post	Gain	Overall IRL-Pre	Overall IRL-Post	Gain
XXXXXXXX		4	2	1	51692	1	2	1	1	3	2	2	3	1
XXXXXXXX		5	1	1	1991	4	5	1	1	4	3	4	7	3
XXXXXXXX		5	1	6	111691	2	2	0	4	4	0	4	4	0
XXXXXXXX		4	2	5	103191	5	7	2	2	2	0	1	4	3
XXXXXXXX		4	2	5	1292	4	7	3	3	4	1	3	6	3
XXXXXXXX		5	1	1	12591	4	7	3	3	4	1	4	4	0
XXXXXXXX		4	2	1	82091	2	2	0	2	3	1	-1	3	4
XXXXXXXX		5	1	1	1191	5	8	3	3	5	2	4	7	3
XXXXXXXX		5	1	1	63090	3	4	1	2	4	2	3	4	1
XXXXXXXX		4	2	1	12992	2	3	1	1	2	1	-2	1	3
XXXXXXXX		5	2	4	11590	5	5	0	3	3	0	4	4	0
XXXXXXXX		5	1	2		2	3	1	1	2	1	-1	3	4
XXXXXXXX		5	1	5	10991	6	7	1	3	3	0	3	4	1
XXXXXXXX		5	1	1	11391	3	5	2	3	4	1	4	4	0
XXXXXXXX		4	1	1	2892	3	4	1	3	5	2	1	5	4
XXXXXXXX		5	1	1	2891	5	7	2	4	4	0	3	7	4
XXXXXXXX		4	1	4	31892	2	5	3	1	3	2	1		
XXXXXXXX		5	1	1	62091	3	5	2	3	3	0	4	4	0
XXXXXXXX		4	1	1	101392	3	4	1	1	1	0	-1	3	4
XXXXXXXX		5	1	2	123190	4	5	1	3	3	0	3	5	2
XXXXXXXX		5	2	1	4490	1	3	2	2	4	2	1	3	2
XXXXXXXX		5	1	1	61991	4	6	2	3	4	1	3	4	1
XXXXXXXX		5	2	1	21891	3	4	1	4	5	1	4	4	0
XXXXXXXX		5	1	1	22690	2	3	1	2	2	0	2	3	1
XXXXXXXX		5	1	4	41190	2	3	1	2	2	0	4	4	0
XXXXXXXX		5	2	1	103089	3	5	2	3	5	2	3	3	0
AVERAGE GAIN								1.5			1.0			1.8

**CAL Reads
LOWELL MIDDLE SCHOOL
Participants
2001-2002**

Name	Grade	Gender	Ethnicity	Date of Birth	WR-Pre	WR-Post	Gain	WM-Pre	WM-Post	Gain	Overall IRL-Pre	Overall IRL-Post	Gain
XXXXXXXXXX	6	1	4	121389	1			1			-2		
XXXXXXXXXX	6	1	1	10191	5			1			3		
XXXXXXXXXX	6	2	1	32090	1.5			4			3		
XXXXXXXXXX	6	1	1	72590	5	5	0	5	5	0	4	6	2
XXXXXXXXXX	6	2	1	70190	3	3	0	3	3	0	1	4	3
XXXXXXXXXX	6	2	1	51790	2	2	0	2	2	0	1	4	3
XXXXXXXXXX	6	1	1	52789	4	7	3	5	5	0	4	7.5	3.5
XXXXXXXXXX	6	1	1	71690	6			4			4		
XXXXXXXXXX	6	2	1	32290	4	6	2	5	5	0	3	7.5	4.5
XXXXXXXXXX	6	1	1	82890	2	3	1	4	4	0	1	7.5	6.5
XXXXXXXXXX	6	1	2	20889	7	11.5	4.5	1	3	2	3	7.5	4.5
XXXXXXXXXX	6	1	1	71990	1	1.5	0.5	2	3	1	-2	0	2
XXXXXXXXXX	6	2	1	61989	4			4			4		
XXXXXXXXXX	6	1	2	32990	1.5	1.5	0	2	3	1	1	1	0
XXXXXXXXXX	6	2	1	82390	3	5	2	3	3	0	3	4	1
XXXXXXXXXX	6	2	1	62690	4	7	3	2	2	0	1	3	2
XXXXXXXXXX	6	1	1	62689	5	9.5	4.5	5	8	3	4	7.5	3.5
XXXXXXXXXX	6	2	1	41090	6	6	0	3	4	1	1	4	3
XXXXXXXXXX	6	1	5	91889	1.5	1.5	0	6	6	0	-2	2	4
XXXXXXXXXX	6	1	2	32690	2	5	3	1	1	0	1	2	1
XXXXXXXXXX	6	1	1	52390	5	7	2	4	6	2	3	6	3
XXXXXXXXXX	6	1	2	12490	3	5	2	3	4	1	0	3	3
XXXXXXXXXX	6	1	5	70589	7			4			3		
XXXXXXXXXX	6	2	1	122689	6	7	1	4	5	1	3	7.5	4.5
XXXXXXXXXX	6	2	1	21190	5		1.6	4		0.7	4		3.0
AGERAGE GAIN													

Appendix B

LetsGoLearn Final Report #3, Period 1/1/02 to 2/28/02

Small Business Innovation Research (SBIR)
U.S. Department of Education

**Accelerating the Literacy Achievement of Students through Technology:
The Design and Development of an Individualized,
Assessment-Driven, Internet-Based Literacy Intervention**

Richard Capone, Chief Executive Officer & Principle Investigator

LetsGoLearn, Inc.
705 Wellesley Ave.
Kensington, CA 94708
(510) 558-8844

**Topic 11. Development or Adaptation of Emerging and Cutting-Edge Technologies for
Improving Teaching and Learning in Pre-Kindergarten to Postsecondary Education, or for
Supporting School/Classroom Management and Comprehensive School Reform**

ABSTRACT (FROM GRANT PROPOSAL)

Technical Abstract

Schools around the nation have the challenge of raising the literacy achievement of many struggling readers. It has been shown that targeted, consistent, individualized instruction is the ideal way to accelerate the literacy development of struggling students. However, current models of individualized instruction are traditionally limited to serving only a fraction of the students in need while the pool of trained professionals necessary to implement these programs is scarce. *LetsGoLearn, Inc.* proposes a way to extend individualized literacy instruction to all students and a way for teachers to manage each student's reading progress and apply individualized computer-based activities to authentic classroom practices. It is our intention to 1) complete the design and development of a computer-based diagnostic assessment which would drive individualized instructional activities delivered over the Internet, 2) pilot this initial version of the program at one school, 3) plan the development of an expandable instructional platform, and 4) analyze all data collected to inform future development plans.

Summary of Anticipated Results and Implications

Through the proposed research and development plan for Phase I, we anticipate creating a viable model of a dynamic assessment and instruction engine that would deliver, via the internet, targeted, individualized literacy instruction to all students. By Phase II, we anticipate generating more instructional activities to be added to the instruction engine and the development of more features associated with the system. These will include 1) multimedia demonstrations of instructional techniques that teachers could implement to extend individualized instruction into classroom practices, 2) features which would easily allow parents to participate in their child's instruction at home, and 3) the development of features that would allow the product to be implemented in other settings outside of school (e.g., afterschool centers, correctional facilities, etc.). We anticipate that the development of this product could greatly increase the capacity to serve more people nationwide. While literacy problems arise in many institutions and many contexts across the nation, it is our contention that struggling learners everywhere can achieve success through individualized instruction. It is our goal to help the country become "a nation of readers" through the development of this product.

LetsGoLearn Final Report #3, Period 1/1/02 to 2/28/02
Small Business Innovation Research (SBIR)
U.S. Department of Education

FINAL REPORT

Note:

The content in this third and final report will cover progress made since our last report submitted two months ago in addition to a summary of the results of our pilot study.

Recent Progress: Overview of the Final Stage of Pilot (Jan & Feb. 2002)

Over the past two months, we have concentrated our efforts on three areas: 1) administering our reading assessment product to students in the Tahoe/Truckee Unified School District (TTUSD) and measuring the results of these assessments against similar pen and paper assessments 2) developing the framework for the instructional component to our product and 3) tapping into the home-school market as a source of revenue. In the process of implementing the assessment engine in schools we have collaborated with the Tahoe/Truckee Unified school district which has a diverse population of students in need. As we work on the implementation of the assessment engine, we started to develop a process (both technically and pedagogically) for integrating an instructional component to our product. The activities we have conducted the past two months have led to many lessons learned.

The process of implementing our assessment product to schools in TTUSD has involved making several trips to 2 primary sites in Lake Tahoe and Truckee, CA. We initially delivered an overview of our assessment to the teachers and principals and then followed up with training sessions on how it works. Once the on-line assessment was implemented, we encountered several glitches that negatively affected the implementation of the on-line assessment in the schools. One problem involved a bug in Macromedia's Flash plug-in for MacIntosh versions of Internet Explorer. The bug in the Flash plug-in made our assessments terminate prematurely on MacIntosh computers. This problem incapacitated the use of our system by Tahoe and Truckee teachers for one and a half weeks as we developed a solution. Eventually, we were able to edit the assessment engine in a way which took into account the bug present in the Flash plug-in.

In addition to modifying our system to work with Flash, we learned, through first-hand experience, a lesson on Internet security. Mid-February, hackers penetrated our server and forced us to lose one week of use as we rebuilt our system. While our data was protected, the nature of their attack made us discover faults in our first layer of defenses. In this situation we learned that we need to allocate more of our resources now and in the future towards security to keep similar breeches and shut-downs from occurring. This means maintaining our security patches more diligently, upgrading our servers, and hiring an information technology company to regularly maintain our server's security. This additional cost is necessary and will become a key point in our financial planning.

In February, we integrated into our assessment product, our initial instructional activity module tied to the first subtest of High Frequency Words. We determined the starting point of the students in the instructional activity based on where they last demonstrated mastery in our leveled high frequency words test list. Our instructional activity then pulled from a larger complete list of high frequency words to drill and test the student. The students only moved forward with instruction as they demonstrated mastery of the current high frequency words. This insured that the pace of instruction was in line with the pace of the student's ability to learn. We currently will

use this initial module as a basis to learn and shape the development of the remaining instructional activities for the balance of our assessments.

As we completed the assessment engine in December 2001, we began to market our product to home-school parents. In the home-school market, we discovered that we will need many more home-schooling web site partners to get significant use by home-schooling parents. This was apparent when our lead partner site, HomeSchool.Com exhibited fewer website hits than expected. Naturally, our web-site also received fewer than expected hits. However, the parents that did purchase assessments for their children have given us an overwhelming amount of positive remarks about our assessment. In general, these parents found the product to be useful because without it, they lacked the technical understanding of the complexity of learning to read and they generally had children with very unique reading profiles. Many of these children exhibited extreme peaks (or areas of high achievement) and valleys (areas of weakness). While our current resources and time preclude us from performing further formal statistical analyses of these home-school children, we will certainly consider this for future planning.

Overall Summary of Results of Pilot Study

1. Development of Assessment and Instruction Engine

As mentioned in our second progress report, we completed our assessment engine development in December. Since the launch of the assessment engine we have applied many preventative modifications and enhancements. Because our product is a web server application we have been able to update it as needed on a daily basis, making the current version of the assessment tool immediately available to users. For instance, a teacher at TTUSD informed us one day that a student's test suddenly terminated. Upon examining that particular child's profile of use on the server, we discovered that we had a bug that halted that particular student's test because our system had moved that student down a grade on a particular subtest and assumed the test was over. That afternoon we corrected the problem and the next morning the student resumed taking the assessment.

Piloting the assessment engine in TTUSD has helped us in many ways. We have received many useful tips suggestions from teachers concerning our assessment. Some of these suggestions which were already considered in the development process helped us reinforce the need for these features. For instance, teachers have instinctively asked for a feature that would tie the assessment to some sort of instructional activity. To compensate for this developing feature of our assessment, we have included instructional suggestions with each individual profile that is generated for the teachers. Some teachers suggested changing the script of the mascot who guides the student through the assessment. Others have asked us to build in a function that would help the teachers group their students into various profiles to help with their grouping decisions for small group instruction. Some suggestions have been adopted while others are on waiting lists to be reviewed and possibly implemented.

In late December 2001, we finalized the instructional paradigm that the instructional engine would follow. This paradigm moves a student's learning through three domains. These domains are: 1) discrete knowledge: Memory Activities (i.e., letter-sound patterns, sight words, word analysis items, vocabulary words, affixes, etc.) 2) application of knowledge on a word level (i.e., decoding strategies in word recognition, encoding strategies or spelling, vocabulary usage) and 3) application of knowledge on a sentence or passage level (i.e., comprehension, fluency, writing/composition). The computer program would start the instructional engine at the appropriate place given the results of the assessment engine. This instructional paradigm

provides the initial logic which we will use to build the interaction of the assessment engine to instruction and different instructional components with each other.

As we summarized above in the 'Recent Progress' section, we have completed the development of one instructional activity related to the first domain of the paradigm. This instructional activity involves practicing the student's recognition of sight words. We have plans of creating more activities that interact with each other which speak to the individual profiles of each student

2. Pilot Beta Version of Assessment and Instruction Engines

As mentioned earlier Tahoe Truckee Unified School District is extensively testing many of their students using our assessment. The teachers are excited to gain detailed information on how their students are performing in the subskills necessary for fluent reading.

In our pilot testing, we compared our online automated assessment system to the best diagnostic assessment methodology currently available for early readers, which has been developed by the CalReads program in the UC Berkeley Graduate School of Education. But whereas the CalReads assessments are administered to children one-on-one by highly trained specialists with advanced degrees in education, the LetsGoLearn system is an automated tool that requires no staff to administer.

For this test, 20 kindergarten through grade 5 boys and girls were identified by the principal at Glenshire Elementary School, in the Truckee Unified School District, near Lake Tahoe in California. These 20 children were chosen because they all were manifesting difficulties in mastering reading skills appropriate for their grade level. All 20 kids were thoroughly assessed by the LGL system and by the CalReads experts. Both systems assess children of 5 diagnostic outcome variables: 1. Sight-word familiarity 2. Word recognition 3. Word meaning 4. Spelling and 5. Silent reading. The diagnostic outcome variables for both systems were placed on a common scale and correlation coefficients were computed. The correlations between the assessments made by the two systems for the same kids are as follows:

- Sight-word familiarity $r=.89$ (n=17)
- Word recognition $r=.81$ (n=20)
- Word meaning $r=.60$ (n=20)
- Spelling $r=.78$ (n=20)
- Silent reading $r=.89$ (n=19)

All correlations are statistically significant beyond the $\alpha=.01$ level. Note that some of the correlations above are based on fewer than the full sample of 20 children. Observers noted that three sight-word subjects and one silent reading subject did not actually participate in the assessment activity.

Although these correlations are quite good, we are currently evaluating the detailed qualitative observations of our system, provided as part of the pilot study by the CalReads professionals, to improve the accuracy of LGL's assessment algorithms and performance criteria. The LGL system will be continuously refined with more testing, across a broader range of kids and classroom contexts.

Some examples of the kind of refinements we will examine include adjusting our scoring to improve alignment with CalReads. For example, we did not put in a time limit for a correct response for Word Recognition. We will try applying this limit to see if it will help compensate for our reduced correlations. In Word Meaning, the sub test with the lowest correlation, we may

try to compensate by adjusting the word difficulty. The Let's Go Learn Word Meaning sub test has 4 images for students to choose from which provides additional contextual clues.

Qualitatively, we found that the majority of students piloting our on-line assessment enjoyed using it. They usually saw it as a game and not an assessment. As the children moved up in age from the elementary to the middle school level, they did find our graphics less interesting. The entire pilot was performed using our animal theme that was designed for younger children. Due to financial constraints we were not able to build parallel graphic "skins" for older children and adults yet. We do, however, have the hooks in our system to grab different graphics files based on the age of the learner and hope to put these in sometime in April 2002.

Additionally we noticed that younger children generally had more problems initially getting started. Teachers had to explain more thoroughly how the "game" worked and what they were supposed to do. In one of our first versions of the High Frequency Words subtest, our instructions said, "Click on the fly that says the word correctly..." However, when the mouse rolled over the fly, the fly would rise up a centimeter and flap its wings. Younger kids would then try to move their mouse up before clicking. As they moved up, they moved out of the click zone and thus many became confused. We fixed this problem by increasing the click zone. This illustrates the challenge of designing on-line interfaces with younger children in mind. Moving forward we are working on fixing some similar sources of confusions for younger children and have already begun to reword some of our audio instructions that sometimes are too confusing for young children. When younger children are confused we concluded that less words are better than trying to explain more. We will also have a practice area where teachers can have students try the "games" without actually starting a real assessment.

While we have designed the first of many instructional activities we have not been able to pilot this feature of our product in time for this final report. As mentioned above this instructional activity practices the recognition of sight words. The instructional component functions in the following way: Once an assessment is completed our system automatically starts the student with instruction. Instruction begins at the exact level where the student was tested in our assessment. The instructional activity is based on the same Flash module which the assessment module was built. We added some additional user interfaces into the existing module. One module instructs the student by presenting them with a series of words and matching tasks. When they get an answer wrong another module, a correction interface signals the system to continue to instruct the child. The idea is that the student only advances when they have mastered the current set of high frequency words that they are learning.

We do not directly teach the 300 high frequency words present in our assessment engine. We created a separate list of 1000 high frequency words commonly accepted and used by reading specialists. While some of the three hundred words are part of this list, the premise involved is move away from teaching to the test and moving towards accelerating the literacy of the children interacting with our product.

3. Prepare Product and Establish Partnerships for Future Developments Beyond Phase I

We have been very successful at creating effective and innovative partnerships with other companies. We have signed an agreement with a company called Inventive Communication. Inventive Communication is an educational portal that has a URL library that it has currently sold to 3450 schools nationwide. Let's Go Learn, Inc. has had several meetings with Inventive Communication's founder. Currently Inventive Communication is planning on selling Let's Go Learn's assessment along side its product. Our estimates is that come September 2002, the Let's Go Learn assessment can possibly be in use by 175 to 225 schools. This achievement is

extremely important given the costly nature of our development and the need to become operationally profitable as soon as possible.

A public educational company called Scientific Learning has also signed some initial agreements with us to allow them to use our assessment in up to 10 pilots. They have 180,000 students using their computer based reading instruction system but have no assessment component. They are in some extremely large school districts like Los Angeles Unified in Southern California. This will potentially turn into a great opportunity for revenue. Given their long track record in computerized learning, they have already offered us much feedback that we expect to implement this summer.

4. Human Subjects Protection

Federal Wide Assurances completed and being followed.

5. Financial Update

Currently, operating the company has been costly and funding has been difficult. Let's Go Learn is in the process of raising a new series B funding round. We hope to raise \$100K in April that will allow us to continue to improve our assessment engine and to continue where our SBIR pilot study has left off. Expanding the supplemental instructional activities to be complete enough for school teachers to use by fall 2002 is our goal. However, if we do not acquire funding we will have to slow down development and wait for revenues to grow. Otherwise will wait and hope for a phase II SBIR grant to be awarded before being able to resume development. Fortunately, the signing of the Inventive Communications partnership will help convince potential investors that revenue is in the pipeline.

6. Future Outlook

Our future outlook remains strong. Our partnership deals will insure that we can start generating significant revenues in 2002. This is absolutely essential for our survival and growth. Free flowing VC funding is virtually non-existent in the education sector. This means that we need to become financially independent. We have already begun to plan well beyond 2002 and into 2003 and 2004. Our hopes are to provide an effective total solution for schools in the area of reading. This means providing reading assessment, reading instruction tied to assessment that adapts to the learner's pace of development and finally professional development for teachers.

Dr. McCallum has become more and more focused on K-4 teachers. His belief is that good tools and teacher training is necessary for any system-wide literacy program to succeed. We are working towards this goal as well. And our initial foray into this market is the posting of "e-seminars" on teaching reading. We have already posted one 30-minute interactive e-seminar that is free to all teachers and parents that explains a basic introduction to reading theory. Our goal is to create full online courses that teachers can take for credit. The professional development director at Tahoe is extremely excited about this option given their remote location and the difficulty teachers experience when trying to attend courses at university or colleges.

In conclusion, everyone at Let's Go Learn, Inc. is extremely excited about our success to date. Teachers and parents using our site have been extremely happy. Our "new features" list is becoming extremely long as our users get excited about our system and eagerly submit new ideas to us. Our plan is to remain focused but ambitious. By focusing on what is absolutely necessary we can expand our system and add great new tools that will benefit schools, teachers and students. This focus will also allow us to be realistic and make our deadlines. While we do not want to create ideas too grandiose to be implemented, we do need to be ambitious and do what no one else has done.

Thank you for your support in the phase I SBIR grant. In addition to the direct financial support it provided, the prestige and validation that it has brought to us has been invaluable. We look forward to applying to the phase II competition and hope to be given a phase II award to help us accomplish our goal of helping every child become successful readers. Once again thank you for your support and we hope to be working with the U.S. Department of Education in the future.

Appendix A: "Work Plan Activities" as submitted in original grant proposal

WORK PLAN ACTIVITIES	Sept. 2001	Oct. 2001	Nov. 2001	Dec. 2001	Jan. 2002	Feb. 2002
Development of Assessment and Instruction Engine						
Adapt assessments for delivery on web						
Development of database to analyze patterns of error						
Development of instruction engine framework: align to assessment database						
Development of instructional objectives						
Development of instructional activities						
Development of embedded instruction engine matrix to analyze and adjust instruction assessment according to student progress						
Pilot Beta Version of Assessment and Instruction Engines						
Obtain consent from teachers to pilot beta version of program						
Orient teachers to the program: conduct needs assessment						
Administer assessment measures to students						
Administer beta version of instructional activities						
Collect qualitative data and quantitative data						
Analyze qualitative and quantitative data						
Prepare Product and Establish Partnerships for Future Developments Beyond Phase I						
Develop platform logic to easily add other features for post Phase I						
Commence planning for future developments and features						
Develop current partnerships						
Seek new partnerships						
Disseminate results of study						