Super WHY! Summer Camp

An Analysis of Participant Learning in 2008 and 2009 Summer Camps

James Marshall, Ph.D. | Diane Lapp, Ed.D. | Kimberly Cavoto, Ph.D.

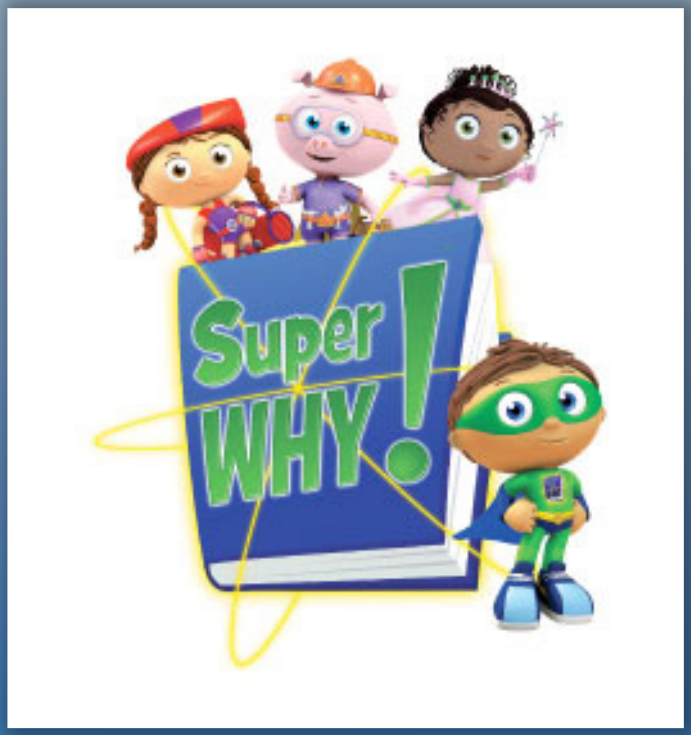

Table of Contents

Index of Tables

Index of Figures

Abstract

The Corporation for Public Broadcasting (CPB) funded *Super WHY!* Summer Camps during 2007, 2008, and 2009 as part of its Ready To Learn (RTL) initiative. The week-long camps were designed for preschool children from low-income families. Pre- and posttest results from summer camp participants indicate growth in literacy skills and abilities. On average, participants achieved an 11.7 percentage point gain as a result of the four-day, 12-hour intervention. All pre-to-posttest differences were statistically significant ($p = .000$). Of particular interest was the correlation between Near (*Super WHY!* episode–covered content) and Far (content that is similar, but not explicitly covered in the episode) scores. These scores shared a positive and significant correlation ($r = 0.589$, $p = .000$), suggesting transfer of *Super WHY!* skills to new and different content.

Executive Summary

The Corporation for Public Broadcasting (CPB) funded *Super WHY!* Summer Camps during 2007, 2008, and 2009 as part of its Ready To Learn (RTL) initiative. The week-long camps were designed for preschool children from low-income families. A carefully developed early literacy curriculum provided the content. In summer 2008, 453 children participated in 33 camps in 17 states. In summer 2009, 80 camps served 820 children in 16 states.

Each camp session served up to 20 children. Campers engaged in activities focused on specific reading skills brought to life by the *Super WHY!* characters. Each day began with participants watching the same episode of the *Super WHY!* series. Children also reviewed targeted segments of *Super WHY!* on classroom televisions and practiced their literacy skills through teacher-directed learning activities.

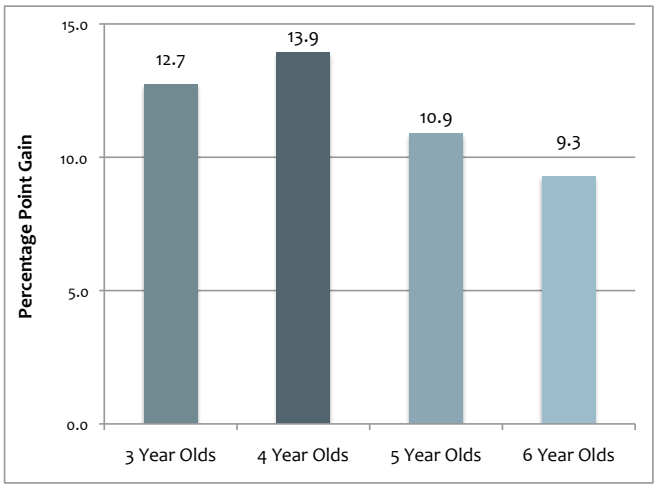

Figure 1. Pretest-to-Posttest Percentage Point Gains (n = 990)

Our analysis of *Super WHY!* Summer Camp data from the 2008 and 2009 sessions provides clear evidence of the program's impact. While only five days in duration, the program consistently produced measureable gains for participants (Figure 1). These gains proved statistically significant, and gains of various sizes were realized regardless of age or ethnicity. Results of this analysis reveal the following key findings:

1. Posttest mean scores were greater, relative to pretest scores, for each subtest and the full assessment. On average participants improved 11.7 percentage points. Four-year old participants achieved the greatest gains (13.9 percentage points) from pre-to-posttest. All pre-to-posttest differences were positive and statistically significant (p = .000).
2. There were no significant differences between the pretest scores of 2008 and 2009 participants. Each year's participants began the program with equivalent skills, as measured by the pretest.
3. There were no significant differences between the gain scores of 2008 and 2009 participants. Each year's participants made similar gains, indicating consistency in the intervention over time—and with the addition of twice as many participants in 2009.
4. Gain scores do differ based on the participant's age. Three- and four-year-old gain scores are reliably greater, relative to older (five- and six-year-old) participants.
5. Gain scores are greatest for participants pretesting in the second (25%–49%) quartile.
6. Near (*Super WHY!* episode–covered content) and Far (content that is similar, but not explicitly covered in the episode) scores shared a positive and significant correlation (r = 0.589, p = .000). This suggests transfer of *Super WHY!* skills to new and different content.

Overall, this analysis provides promising evidence of the summer camp program's impact. Additional inquiry that includes a matched comparison (control) group, as well as the addition of "Far" items never mentioned in any summer camp activity is recommended.

While opportunities to enhance and extend the findings reported here abound, this analysis provides convincing evidence of the *Super WHY!* Summer Camp program's impact on its young participants. Emergent literacy skills increase and those skills readily transfer to content not presented in the *Super WHY!* television program.

Background

The Corporation for Public Broadcasting (CPB) funded *Super WHY!* Summer Camps during 2007, 2008, and 2009 as part of its Ready To Learn (RTL) initiative.[1] The week-long camps were designed for preschool children from low-income families. A carefully developed early literacy curriculum provided the content.

Each camp session served up to 20 children. Teachers from local schools, childcare, and Head Start centers led the camps in partnership with 21 local PBS member stations—all participants in the PBS KIDS Raising Readers initiative. Summer camp participants engaged in activities focused on specific reading skills brought to life by each *Super WHY!* character. Each day of the summer camp began with participants watching the same episode of the *Super WHY!* series. Children also viewed targeted segments of *Super WHY!* on classroom televisions and practiced those skills through learning activities throughout the day. The range of skills addressed included:

- letter naming, based on letter symbol;
- letter identification, given a group of possible symbols;
- decoding;
- letter sounds;
- encoding; and
- reading words and opposites.

Preschool participants completed both pre- and posttests to assess learning gains from the five-day experience.

In summer 2008, 453 children participated in 33 camps in 17 states (based on assessment results provided for data analysis). Pre- and posttest data from these camps were analyzed, and a summary report was written. In summer 2009, 80 camps serving approximately 820 children were held in 16 states. This report summarizes results from the 2008 and 2009 *Super WHY!* Summer Camp programs.

[1]Ready To Learn is a five-year initiative funded by the U.S. Department of Education, in partnership with the Corporation for Public Broadcasting, PBS, and The Ready To Learn Partnership.

Analysis Plan

A complement of guiding questions helped scaffold the data analysis effort and results presented in this report. The data analysis team was formed and contracted post hoc—after the study's design, instrument construction, summer camp implementation, and data collection. This team was provided with 2008 assessment scores and raw data from the 2009 summer camp implementation (see Methodology section for a description of the data analysis process).

The data analysis effort commenced with the definition of guiding analysis questions. These questions, defined by the team and confirmed by CPB, focused the team's statistical analysis and guided its inquiry.

Table 1 presents the guiding questions. Each question is addressed, in turn, in the Results and Summary sections of this report.

Table 1. Guiding Questions for Data Analysis

Data Analysis Questions
1. Does pretest performance differ significantly from posttest performance?
2. Was pretest performance between 2008 and 2009 summer camp participants equivalent?
3. Do gain scores between 2008 and 2009 participants differ significantly, assuming pretest equivalency?
4. Do gain scores differ significantly based on the participant's age?
5. Are gain scores for low-performing participants (defined by pretest score) greater, when compared to higher-performing participants?
6. Do assessment scores specific to content covered in the *Super WHY!* television program (Near) correlate with assessment scores for similar content not covered in the television program (Far)?

These questions guided the analysis evaluation from start to finish, including the entry and arrangement of data, statistical analysis, and reporting presented here.

Methodology

CPB, through their contractor *Out of the Blue*, provided participant assessment data in various forms. This section describes the data analysis methodology used to present results contained in this report. A list of contextual factors that likely influenced the study results is also included.

Instrumentation

Out of the Blue, the production company that created *Super WHY!,* was responsible for developing the assessments used in this evaluation. The pre- and posttest assessments were developed in 2007 and employed for data collection in 2007, 2008, and 2009. The assessment instrument is included as an appendix to this report.

The same instrument was used for both pretesting and posttesting. This instrument included six distinct subtests:

1. Letter Naming, based on letter symbol
2. Letter Identification, given group of possible symbols
3. Decoding
4. Letter Sounds
5. Encoding
6. Reading Words and Opposites

Appendix II provides a copy of the assessment instrument and notes the specific day on which each subtest was administered as a posttest (all pretesting was accomplished on day one).

Instrument Reliability

Instrument development was a necessity of this study. As previously described, no off-the-shelf, validated instrument covering these specific concepts, relative to the *Super WHY!* program, existed. The instrument's reliability was of primary concern to the data analysis team. The team questioned whether the different sections of the *Super Why!* Summer Camp Assessment would prove to be statistically reliable measures. Therefore, the team's first analysis was for reliability.

Cronbach's alpha is a measure of consistency that estimates internal reliability. It takes a group of items and calculates a coefficient to indicate how well the items are aligned with one another. Generally, coefficients above .7 suggest reliability of the items as a group. Table 2 presents reliability statistics for each section of the assessment instrument.

Table 2. Reliability Analysis for Assessment Instrument

Assessment Component	Coefficient Alpha		Number of Scores in Calculation
	Pretest	Posttest	
1. Letter Identification (sounds and symbols)	.96	.96	4
2. Decoding	.83	.75	12
3. Letter Sounds	.87	.86	2
4. Encoding	.89	.92	8
5. Reading Words and Opposites	.85	.87	6
Total Assessment	**.91**	**.92**	**17**

This reliability analysis was largely based on 2009 data. The more detailed item and subtest scores were consistently available for each participant in the 2009 administration. Data from the 2008 summer camps supplemented the 2009 data when the team analyzed reliability for the full assessment (total test).

Overall, reliability is high for both subtest scores and the full test score. It ranged from .75 to .96. While the team did eventually reduce the sample (see following section), it should be noted that reliability remained constant and did not differ from figures presented in Table 2.

Data Collection

Children, ages 3–6, attended the five-day camp for three hours each day. Each day represented one of the four characters (i.e., Alpha Pig Day, Wonder Red Day, Princess Presto Day, Super Why Day) presented in every episode. The fifth day was Super You Day.

The pretest interview was conducted with each child at registration – prior to the start of the summer camp. Summer camp staff sat with each child and administered the assessment (termed "Interview" on the instrument itself) according to the stated instructions provided with the assessment. See Appendix 2 for the instrument and administration instructions.

Posttesting was conducted section-by-section, based on the content covered on a given day. Daily posttesting was necessary due to time constraints, staffing issues, and other logistics. For example, the "Day 1" subtest, which covered content presented during day 1 of the summer camp, was administered at the conclusion of Monday's camp session. This pattern followed for each of the summer camp days—Monday through Thursday.

Camp personnel recorded participant responses and scored performance according to instructions embedded in the assessment form. All pre- and posttest assessments were forwarded to *Out of the Blue* by the summer camp program implementation subcontractor. Forms were stored until the start of this data analysis effort, when copies of each assessment were provided in paper-based format to the data analysis team for entry and analysis.

Resulting Sample

The full data set included summer camp participants who ranged in age from 2 to 11 years. A total of twenty-nine 7–11 year olds, and two 2 year olds, participated in the 2009 summer camp programs. The remainder of the participants fell within the 3–6 year old age range. As one might expect, the majority (24 of 29) of the 7–11 year olds performed in the top 25% (upper quartile) based on pretest score. Because the summer camp program specifically targeted children of ages 3–6, the final sample discarded these 31 outliers (again, two 2-year old and twenty-nine 7–11 year old participants).

The resulting sample included summer camp participants of ages 3–6 years old. Table 3 presents a summary of demographics for this final sample. Note that totals vary across categories when descriptive information was missing from analyzed forms.

Table 3. Final Sample Demographics

Category	Subcategory	2008		2009		Total Sample	
		Count	Percent (%)	Count	Percent (%)	Count	Percent (%)
Gender	Male	177	44.8	423	52.1	600	49.7
	Female	218	55.2	389	47.9	607	50.3
	Total	395		812		1207	
Age	3	4	.9	48	6.0	52	4.2
	4	148	33.2	280	35.3	428	34.5
	5	251	56.3	367	46.2	618	49.8
	6	43	9.6	99	12.5	142	11.5
	Total	446		794		1240	
Ethnicity	African American	124	39.1	380	55.4	504	50.2
	Caucasian	92	29.0	174	25.4	266	26.5
	Hispanic	89	28.1	109	15.9	198	19.7
	Other/Asian/Mixed	12	3.8	23	3.4	35	3.5
	Total	317		686		1,003	
Total in Final Sample		**453**		**820**		**1,273**	

The original sample contained 1,305 cases (not all cases, however, were complete with both pre- and posttest scores). After reduction, the analyzed sample included 1,273 participants (n = 453 in 2008; n = 820 in 2009). Approximately half the participants were male, with the majority being 4 or 5 years of age, and African American. The 2008 study included slightly more males and Hispanics compared to the 2009 study.

Data Entry

Available data included 2008 and 2009 assessment results. Each was delivered to the data analysis team in a different format.

- 2008 data had been previously entered and summarized (i.e., descriptive statistics) by CPB. This data set included demographic information for each participant, as well as subtest and total scores. The data were delivered in Microsoft Excel spreadsheet format and subtotaled by station.
- 2009 data existed only in raw form. *Out of the Blue* delivered copies of each available pre- and posttest form. These data were shipped to the data analysis team in three boxes and organized, roughly, by state and station.

Data Processing

The data analysis team entered all 2009 pre- and posttest assessments. This process began by establishing a database to receive participant scores. With the 2008 data already existing in Excel, the team continued use of that platform for the 2009 data.

Two data entry clerks, using two separate computers and two separate Excel files, keyed each 2009 participant record into their respective files. This work necessarily involved raw data entry, as well as calculation of subtest totals. Each participant record received a consecutive number that was entered into both data files.

When both clerks had completed entry of all participant records, the separate spreadsheets were combined into a single file. Participant IDs were used to match up records. The team then employed Excel's MATCH formula (with the ISNA evaluation function) to evaluate and reconcile each data field (i.e., demographics, raw scores, and totals) between the two sets of data. Inconsistencies, where they existed, were reviewed by comparison to the original form and then adjusted by one data entry clerk. Each adjusted record was flagged and reviewed a second time by a second project staff member.

Data Analysis

The master Excel file was imported into the Predictive Analytics Software (PASW, formerly SPSS). With the dataset finalized and in the statistical analysis software, descriptive statistics were calculated. The team compared this output against similar statistics calculated on the raw file within Excel to confirm equivalency between the Excel and PASW datasets.

Dr. Marshall performed initial analysis on the dataset and informally reported topline results to CPB. Dr. Cavoto performed the detailed analysis on the dataset, guided by the analysis questions. The data analysis effort included the following categories and tests:

- Descriptive statistics—means, standard deviations, disaggregation based on demographics

- Inferential analyses
    1. Pre- and posttest comparisons were conducted using Repeated Measures Analysis of Variance (including calculations for effect size, power, etc.).
    2. Difference and gain scores comparisons were conducted using Univariate Analysis of Variance (ANOVA).
    3. Near and Far difference scores were compared using paired T-test, bivariate correlations, and simple linear regressions models.

A "list-wise approach" was applied to address missing values. This means that only cases with valid data on both the pretest and corresponding posttest were included in the analysis for a given subtest or the full assessment.

Gain Score Calculation

Gain scores were calculated for each set of pretest and posttest scores. Gain scores are the difference between pre- and posttest scores, where:

- negative values mean pretest scores are better than posttest scores;
- zero (0) is interpreted as no impact of instruction; and
- positive values mean posttest performance was better than pretest performance.

A frequency analysis of gain scores revealed a range of values with a notable number of negative values. In fact, 7.8%–16.8% of the overall sample (both years combined) scored better on the pretest than on posttest, depending on the specific subtest. Additionally, 80.6% of the full sample realized a positive gain from pretest to posttest. Table 4 summarizes the frequency of positive, negative and no differences in gain scores for the full sample.

Table 4. Percentage of Students with Positive, Negative and No Differences in Gain Scores (*n* = 990)

Test	Pretest > Posttest Gain Score Is Negative (%)	Neutral Gain Score Is Zero (%)	Posttest > Pretest Gain Score Is Positive (%)
Alpha Pig Letter Naming	11.1	57.2	31.7
Alpha Pig Letter Pointing	15.6	53.6	30.8
Wonder Red Decoding	7.8	37.7	54.5
Princess Presto Letter Sounds	16.8	30.7	52.5
Princess Presto Encoding	11.1	37.3	51.6
Super Why Day Reading Words	12.1	35.2	52.7
Full Assessment	13.3	6.1	80.6

Younger participants tended to provide more negative values when compared to older participants, although this was not a statistically significant pattern. Participants who pretested in the bottom two quartiles were more likely to produce negative gain scores than those in the top two quartiles (depending on subtest). While fully discussed in the following section, it must be noted that a ceiling effect exists for the Letter Naming and Letter Pointing

subtests. Here, one-half (50%) of participants pretested with a 100% correct score, providing no room for growth. This is reflected in the high neutral percentages, and relatively lower positive growth percentages in Table 4 specific to these subtests.

Contextual Factors

As with any study, contextual factors interact with quantitative results. This section provides a list of those factors that *may* have influenced the reported results. Contextual Factors are not necessarily negative. In the case of this data analysis, some of the following factors may have limited the positive results presented in this report.

- Instrumentation: Reliability and Validity—The *Super WHY!* assessment instrument does posses high levels of internal consistency (reliability) as discussed earlier in this section. With regard to validity, the data analysis effort was undertaken with the assumption that the involved instrument is a valid measure of the involved constructs. Instrumentation was created by the program developer and contains Near and Far measures of summer camp program content. It is recommended that the developer investigate both face and predictive validity as a part of future studies.
- Instrument Sensitivity—While developed to directly align with the *Super WHY!* programs it assesses, the instrument length is relatively short. The summer camp program constraints included staffing issues and limited time each summer camp day, necessitated a shorter instrument. However, this resulted in a limited number of items targeted to each skill. The relatively small number of items may limit the instrument's ability to measure fractional gains in skill or ability. This is especially relevant given the short intervention time. It is possible that a more sensitive instrument would improve the results reported here by capturing smaller gains in performance. More items, and a wider range of difficulty in those items, are two improvements to sensitivity the team recommends.
- Pre-to-Post Intervention Time—The elapsed time between pre- and posttesting varied from a matter of hours for the Day 1 assessment, to approximately 40 hours for the Day 4 assessment. By all accounts, this represents a very short intervention time. This contextual factor may have influenced assessment results. Participants practiced letter identification, which was posttested on the first day of the camp, throughout the entire summer camp experience. Conversely, Word Reading—tested on day four—was involved, at least indirectly, throughout the entire summer camp (*Super WHY!* episode and activities). Conducting the full posttest at the conclusion of the summer camp would capture effects of the intervention in **all** areas across the full experience.
- Assessment Administration—The team is unable to confirm the consistency with which assessments were administered. Data entry clerks did note the occasional inconsistency in recording an individual score. Given the large number of sites and assessment administrators, it is probable that provided directions varied to some extent when put into practice. Coaching of participants is an example of potential bias that could have influenced the reported results. While the team believes the instrument and implementation contractor did all within his or her ability to limit the possibility of variation, it is a contextual factor that must be accepted with any large-scale implementation.

It should also be noted that the data analysis team conducted this analysis effort post hoc. This arrangement is distinctly different when compared to our preferred approach for conducting an *evaluation* of the *Super WHY!* Summer Camp program. An evaluation would include visiting, and interacting with, implementation sites to (a) observe the assessment's administration; (b) understand the program and instruction; (c) quantify the implementation, including the amount and types of intervention and fidelity to the model; and (d) observe the participants' reactions, interactions, and engagement. Each is an important program evaluation touchstone.

This effort, however, was limited to data analysis—and was accomplished after the conclusion of all summer camp programs. The data analysis team never observed the program "in action." The team did not have the chance to watch assessment administration nor observe teachers and participants engage with the program. The absence of quantitative, reliable data to describe implementation fidelity—to the provided curriculum, and across the many participating stations—limits our ability to *interpret* the data presented in this report. We also cannot make data-driven recommendations for improvements to the program and its implementation with certainly. Where provided, such inferences must be considered with this lack of "ground truth" in mind.

While each of these situations requires consideration, none in and of itself provides rationale for dismissing study results. Rather, they are contextual factors about which the reader should be mindful when reviewing data analysis results.

Data Analysis and Findings

The Data Analysis and Results section presents findings from our review of both 2008 and 2009 data—and data from the full two-year dataset. This presentation is organized into the following six sections, guided by the key analysis questions.

1	Pretest-to-Posttest Performance
2	Comparison of Pretest Performance: 2008 vs. 2009
3	Comparison of Gains: 2008 vs. 2009
4	Gain Scores Based on Age
5	Gain Score Based on Pretest Performance Quartile
6	Near-to-Far Transfer

1. Pretest-to-Posttest Performance

Results Overview

Question: Does pretest performance differ significantly from posttest performance?

Finding: Posttest mean scores were greater, relative to pretest scores, for each subtest and the full assessment. All differences were statistically significant.

We began the analysis effort with a basic comparison of pre- to posttest performance. The full sample, which includes scores for both 2008 and 2009 summer camp participants, was analyzed by comparing pre-to-posttest scores (Figure 2). This comparison was accomplished using a Repeated Measures—ANOVA test. Table 5 presents a summary of these analyses and indicates all pre-to-posttest differences proved statistically significant (p = .000).

Effect sizes are also provided.

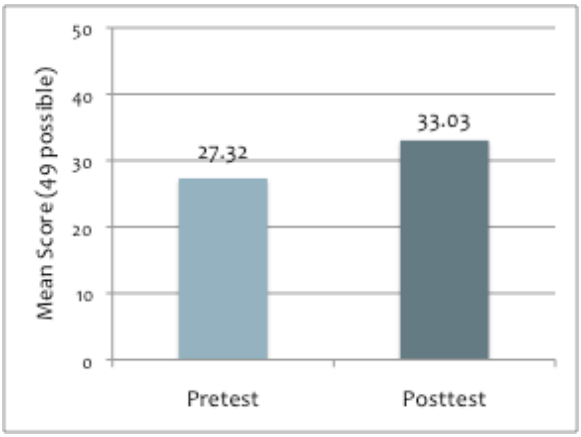

Figure 2. Pretest-to-Posttest Mean Scores (n = 990)

Table 5. Full Sample Pretest-to-Posttest Comparison (n = 909)

Test	Total Points	Pretest		Posttest		Gain		sig	Effect Size
		Mean	SD	Mean	SD	Mean	SD		
Alpha Pig Letter Naming	10	7.72	3.26	8.13	3.03	0.42	1.43	.000	0.08
Alpha Pig Letter Pointing	10	7.67	3.29	8.05	3.08	0.38	1.73	.000	0.05
Wonder Red Decoding	6	1.34	2.17	2.94	2.43	1.56	2.31	.000	0.32
Princess Presto Letter Sounds	11	5.99	3.91	7.06	3.79	1.09	2.51	.000	0.16
Princess Presto Encoding	8	2.23	2.60	3.71	3.13	1.52	2.65	.000	0.23
Super Why Day Reading Words	4	2.38	1.30	3.14	1.18	0.74	1.38	.000	0.24
Full Assessment	**49**	**27.32**	**13.19**	**33.03**	**13.71**	**5.71**	**6.70**	**.000**	**0.42**

As Table 4 describes, mean differences between pre- and posttests all proved positive, indicating (gains). Overall, this resulted in a 5.71 point average gain on the full assessment. This equates to an approximately 11.7% gain between the pre- and posttest. Figure 3 provides a histogram of the full gain score distribution and shows the greatest number of cases clustered around the 5.71 mean.

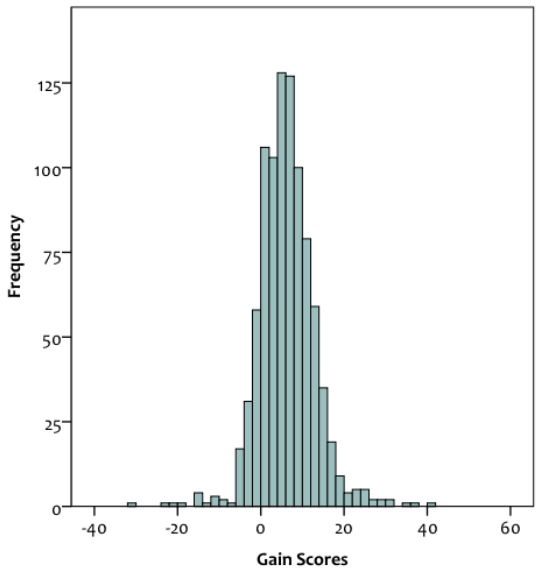

Figure 3. Gain Score Distribution (n = 990)

These differences between pre- and posttest proved statistically significant on each subtest, as well as the full assessment score. This suggests that the summer camp intervention was successful at improving each type of reading skill covered in the *Super WHY!* program.

For four of the six subtests, the standard deviation decreased from pretest to posttest. Based on posttest results for these subtests, not only are participants performing significantly higher, but more of the participants are clustered around that greater score. The lower standard deviation on these posttest scores indicates more consistent scores at the higher posttest level.

A more detailed explanation of the Letter Naming and Letter Pointing subtests is merited. More camp participants entered the program with mastery of these two skills than any other tested skill set. A full 50% of participants answered all Letter Naming pretest questions correctly. The same 50% statistic proved true for the Letter Pointing subtest. This effectively created a "ceiling" effect for half of the summer camp population: with 100% of the items answered correctly on the pretest, there was no room to grow on the posttest. This is reflected in the descriptive statistics, as well as the smaller effect sizes for these subtests (see Table 5).

Effect sizes ranged from .05 to .42, with Wonder Red Decoding (.32) and full assessment (.42) producing the largest effect sizes. Taking the full assessment score as an example, the .42 effect size indicates that knowing pretest scores explains or accounts for 42% of variance in posttest score. This means, given the pretest score we can predict a participant's posttest score 42% of the time. This leaves roughly 58% of unexplained variance. When one considers the many factors such as IQ, attention, and memory ability that can explain an individual's performance, the effect size is impressive.

We investigated whether demographic classifications of gender or ethnicity interacted with the pre- or posttest variables. No significant interaction was found either. Age, however, did interact in significant ways and is discussed in a later section of this report.

2. Comparison of Pretest Performance: 2008 vs. 2009

Results Overview

Question:	Was pretest performance between 2008 and 2009 summer camp participants equivalent?
Finding:	There were no significant differences between the pretest scores of 2008 and 2009 participants.

We analyzed pre-to-posttest performance by comparing 2008 and 2009 program participants. Figure 4 provides a graphical comparison of Full Assessment scores across the two program years. The overarching question was whether scores differed significantly between the two years. Before we could make a determination, we asked whether each year's participant group began the program with similar skills (as measured by the pretest).

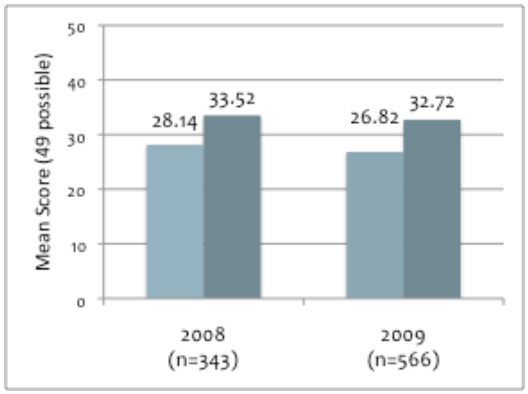

Figure 4. Pretest to Posttest Comparison by Year

Pretest Equivalency

The program year comparison began by investigating whether the two groups entered the summer camp program with equivalent scores. We wanted to confirm that the two groups of participants were of similar ability, on average, when entering the program. If equivalent, the need to make statistical adjustments before comparing gains in performance would be eliminated. Table 6 presents a comparison of pretest scores for the 2008 and 2009 subgroups.

Table 6. Pretest Score Comparison—2008 vs. 2009

Test	2008 Pretest (*n* = 343)		2009 Pretest (*n* = 566)		sig
	Mean	Standard Deviation	Mean	Standard Deviation	
Alpha Pig Letter Naming	7.91	3.00	7.60	3.41	ns
Alpha Pig Letter Pointing	7.89	3.08	7.54	3.41	ns
Wonder Red Decoding	1.25	2.08	1.39	2.22	ns
Princess Presto Letter Sounds	6.27	3.61	5.82	4.07	ns
Princess Presto Encoding	2.34	2.72	2.16	2.53	ns
Super Why Day Reading Words	2.49	1.23	2.31	1.34	0.05
Full Assessment	28.14	12.16	26.82	13.75	ns

Pretest performance in 2008 and 2009 was equivalent; differences in pretest scores were not significant for the full assessment and all but one subtest. The Reading Words assessment was the only subtest that showed a significant difference, revealing that 2008 pretest scores were, on average, better than 2009 pretest scores. Further review revealed that the effect size for this result was 0.004. This small effect size suggests this difference is minimal and, therefore, is not considered to be of practical value.

3. Comparison of Gains: 2008 vs. 2009

Results Overview

Question: Assuming pretest equivalency, do gain scores between 2008 and 2009 participants differ significantly?

Finding: There were no significant differences between the gain scores of 2008 and 2009 participants.

With the equivalency of pretest performance between years established, the analysis effort turned to a comparison of gain scores. We investigated whether gain scores differed between years and, if so, whether differences were statistically significant. Differences in gains scores would indicate differences in performance of the two groups: 2008 vs. 2009 camp participants. Figure 5 presents a graphical overview of gain scores. Table 7 summarizes gain scores for each year. Note that gain score means include all available gain scores for the given year—positive or negative (see the Data Analysis section in this report for more information about gain score calculation and classification).

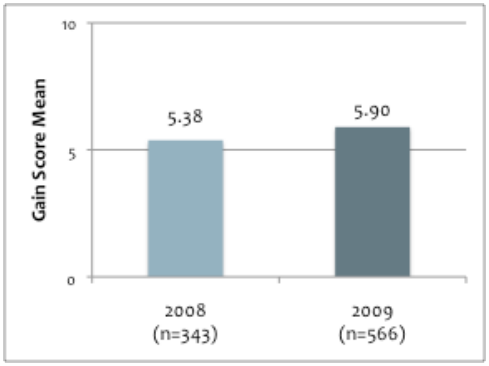

Figure 5. Gain Score Comparison by Year

Table 7. Gain Score Differences

Test	2008 Gains (*n* = 343)		2009 Gains (*n* = 566)		sig
	Mean	Standard Deviation	Mean	Standard Deviation	
Alpha Pig Letter Naming	0.39	1.17	0.44	1.57	ns
Alpha Pig Letter Pointing	0.28	1.62	0.44	1.80	ns
Wonder Red Decoding	1.56	2.21	1.56	2.37	ns
Princess Presto Letter Sounds	1.02	2.29	1.14	2.62	ns
Princess Presto Encoding	1.75	2.70	1.38	2.61	0.026
Super Why Day Reading Words	0.70	1.27	0.76	1.44	ns
Full Assessment	5.38	5.54	5.90	7.31	ns

Generally, there was no difference in the impact of the *Super WHY!* Intervention; the magnitude of the intervention-related impact was essentially the same between years. The one exception was 2008 participants, who showed a larger impact on the Princess Presto Encoding subtest when compared to 2009 participants. Yet, with a resulting effect size of .004, the value of this finding is fairly low. We conclude that performance was essentially the same for 2008 and 2009 summer camp participants across all subtests and the Full Assessment score.

4. Gain Scores Based on Age

Results Overview

Question: Do gain scores differ significantly based on participant age?

Finding: Gain scores do differ based on participant age—with younger participants gain scores being greater, relative to older participants.

We compared gain scores based on the age of the assessed participant. We were interested in knowing whether gains differed by age. Table 8 presents gain score means for each of the four age groups involved in the analysis. Shaded cells indicate the age group with the greatest gain for each subtest and the Full Assessment.

Table 8. Gain Score Differences by Age (*n* = 889)

Test	Age 3 (*n* = 39)		Age 4 (*n* = 289)		Age 5 (*n* = 455)		Age 6 (*n* = 106)	
	Gain Mean	SD	Gain Mean	SD	Gain Mean	SD	Gain Mean	SD
Alpha Pig Letter Naming	**0.77**	**1.97**	0.74	1.65	0.24	1.30	0.14	0.73
Alpha Pig Letter Pointing	**0.97**	**2.04**	0.75	2.03	0.17	1.55	0.08	0.83
Wonder Red Decoding	1.31	1.76	1.54	2.27	1.60	2.33	**2.08**	**2.60**
Princess Presto Letter Sounds	**1.38**	**2.12**	1.29	2.62	0.99	2.47	0.69	2.10
Princess Presto Encoding	0.67	2.09	**1.69**	**2.63**	1.57	2.77	1.05	2.83
Super Why Day Reading Words	**1.13**	**1.75**	0.80	1.44	0.75	1.35	0.52	1.01
Full Assessment	6.23	7.59	**6.82**	**7.14**	5.33	6.51	4.55	5.67

Overall, across years, older participants are impacted less than younger participants on the full assessment score and all subtests, with the exception of the decoding subtest. In other words, gain scores are smaller, in general, for older participants relative to their younger counterparts. Table 9 provides a summary of pre- and posttest scores, by age, and indicates effect sizes for each age group.

Table 9. Full Assessment Pretest-to-Posttest Comparison by Age

Age Group	*n*	Pretest		Posttest		Gain		sig	Effect Size
		Mean	SD	Mean	SD	Mean	SD		
3 Year Olds	39	10.10	9.67	16.33	12.76	6.23	7.59	.000	0.41
4 Year Olds	289	22.85	12.60	29.67	13.81	6.82	7.14	.000	0.48
5 Year Olds	455	29.17	11.42	34.50	12.33	5.33	6.51	.000	0.40
6 Year Olds	106	39.37	9.00	43.91	6.99	4.55	5.67	.000	0.39
Full Sample	990	**27.32**	**13.19**	**33.03**	**13.71**	**5.71**	**6.70**	**.000**	**0.42**

Differences in pre- and posttest scores were statistically significant for each age grouping. Tables 8 and 9 indicate the greatest average gain for four-year old participants—and the largest effect size (.48). From pre- to posttest, the 6.82 gain equates to a 13.9 percentage-point gain in performance.

5. Gain Scores Based on Pretest Performance

Results Overview

Question: Are gain scores for low-performing participants (defined by pretest score) greater, when compared to higher-performing participants?

Finding: Gain scores are greatest for participants pretesting in the second (25%–49%) quartile.

We also analyzed gain scores based on quartile distributions (established from the participants' pretest scores). We were interested in whether lower-scoring participants made greater gains as a result of the *Super WHY!* program. Table 10 summarizes gains based on pretest performance. Participants were divided into quartiles based on their pretest scores. Shaded cells indicate the quartile with the greatest gain for each subtest and the Full Assessment.

Table 10. Gain Score Differences by Pretest-Assigned Quartile (*n* = 909)

Test	Bottom 25% (*n* = 210)		25-49% (*n* = 221)		50-74% (*n* = 237)		Top 25% (*n* = 241)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Alpha Pig Letter Naming	**1.03**	**1.98**	0.70	1.75	0.09	0.70	–0.06	0.79
Alpha Pig Letter Pointing	**1.19**	**2.41**	0.70	2.17	0.14	0.85	–0.08	0.81
Wonder Red Decoding	0.88	1.65	1.52	2.09	**2.49**	**2.37**	1.41	2.65
Princess Presto Letter Sounds	1.18	2.44	**2.12**	**2.99**	0.93	2.26	0.15	1.60
Princess Presto Encoding	0.59	1.68	1.89	2.81	**2.38**	**2.93**	1.02	2.73
Super Why Day Reading Words	0.99	1.52	**1.00**	**1.52**	0.76	1.31	0.33	0.95
Full Assessment	5.79	7.66	**7.82**	**7.13**	6.60	5.30	2.81	5.60

Overall, top performers (participants who scored better on total pretest) showed the least impact from the *Super WHY!* Summer Camp. This was especially true for the basic skills of Letter Naming and Letter Pointing. Here, participants who pretested in the upper two quartiles showed virtually no impact as measured by average gain scores. The overall gain score mean for this group was 2.81.

Table 11 provides a summary of pre- to posttest performance, by quartile, for summer camp participants.

Table 11. Full Assessment Pretest-to-Posttest Comparison by Quartile (*n* = 909)

Age Group	*n*	Pretest		Posttest		Gain		sig	Effect Size
		Mean	SD	Mean	SD	Mean	SD		
Bottom 25%	210	7.92	5.00	13.71	9.58	5.79	7.66	.000	0.37
25%–49%	221	23.43	3.31	31.26	7.80	7.82	7.13	.000	0.55
50%–74%	237	32.59	2.36	39.19	5.42	6.60	5.30	.000	0.61
Upper 25%	241	42.61	3.89	45.41	5.09	2.81	5.60	.000	0.20
Full Sample	**909**	**27.32**	**13.19**	**33.03**	**13.71**	**5.71**	**6.70**	**.000**	**0.42**

The two middle performing groups showed the greatest impact on decoding and encoding subtests relative to the lowest and top performers, suggesting that perhaps the lowest performers are not ready for these skills and the top performers are already performing these skills and, therefore show less improvement overall. The second quartile group realized the greatest amount of growth. This group gained, on average, 7.82 points as a result of the intervention. This equates to an approximate 16 percentage point increase in performance from pre- to posttest.

6. Near and Far—Analyzing Transfer

Results Overview	
Question:	Do assessment scores specific to content covered in the *Super WHY!* television program (Near) correlate to assessment scores for similar content not covered in the television program (Far)?
Finding:	Near and Far scores shared a positive and significant correlation (r = 0. 589, p = .000), suggesting transfer of *Super WHY!* skills to new and different content.

While a child can learn a specific skill from a television program, the ultimate outcome is the child's understanding of the underlying rule or concept. For example, consider a program segment that covers the "-ig" word family, and provides practice with the initial consonant "p" (as in pig). The participant who recalls this specific content (p-ig) may be simply recalling (repeating) what he or she learned from the program. However, that same child may have a deeper understanding of the "ig" word family—one that transfers to other "ig" words (i.e., big, dig, jig).

The *Super WHY!* Summer Camp assessment instrument accommodated measurement of these "levels" of understanding. The instrument included four subtests (of six total subtests) with both "Near" and "Far" items. These classifications are defined as follows:

- Near items—content explicitly presented in the *Super WHY!* television program. These items target either recall of specific content, but *may* also indicate the application of a rule and/or conceptual understanding of the underlying skill.
- Far items—content not presented in the *Super WHY!* television program, but introduced through camp activities. These items arguably rely on application of a rule and/or conceptual understanding of the underlying skill.

The team questioned whether participants who pretested low on a given "Near" subtest and its corresponding "Far" subtest, would demonstrate:

- No difference on the posttest;
- Gains on only the **Near** posttest item—suggesting *recall* of the *Super WHY!* presented content; or
- Gains on both the **Near** and **Far** items—suggesting *transfer* of the *Super WHY!*-presented content.

To investigate this possibility, the team conducted comparisons of Near and Far interaction in three groups:

- Overall 2009 data (since Near and Far scores were only available for the 2009 data)
- Two age groups: 3–5 year olds and 3 and 4 year olds only
- Bottom quartile based on pretest Near totals

These analyses were based on difference scores. The team calculated difference scores for Near and Far items (i.e., subtracting pretest score from posttest score for all Near items, then all Far items), for each subtest and the full assessment. The full assessment calculation for this analysis excluded the two subtests that consisted entirely of Near items. Using this approach, the team would conclude that significant relationships between Near and Far (for participants who scored low on both variables on the pretest) suggest transfer of the *Super WHY!*-presented content.

Differences in Near and Far Improvement

The team began this investigation by determining whether significant differences existed between pre- and posttest scores for each assessment subtest that included both Near and Far items (as well as the full assessment). Two subtests consisted entirely of Near items, and therefore are excluded from the Total Assessment statistics. Table 12 presents the results of this Paired t-Test analysis.

Table 12. Pretest-to-Posttest Near vs. Far Comparison (*n* = 566)

Subtest/Full Assessment	Near or Far	Points Possible	Pretest		Posttest		sig
			Mean	SD	Mean	SD	
Alpha Pig Letter Naming	Near	5	3.78	1.76	4.01	1.64	.000
	Far	5	3.83	1.72	4.03	1.59	.000
Alpha Pig Letter Pointing	Near	4	2.98	1.41	3.15	1.29	.000
	Far	6	4.56	2.11	4.82	1.95	.000
Wonder Red Decoding	Near	3	0.74	1.17	1.73	1.29	.000
	Far	3	0.66	1.14	1.33	1.32	.000
Princess Presto Letter Sounds	Near	4	2.03	1.56	2.57	1.46	.000
	Far	7	3.78	2.62	4.41	2.58	.000
Full Assessment	Near	16	9.52	4.87	11.50	4.83	.000
	Far	21	12.83	6.42	14.59	6.36	.000

Positive growth was realized, on average, across all subtests for both Near and Far components. Each difference between pretest and posttest scores proved statistically significant.

Next, we wanted to determine if there was a significant difference between pre-to-posttest differences on Near items and pre-to-posttest differences on Far items. Table 13 presents the results of this Paired t-Test analysis. Again, the full assessment figures exclude the two additional subtests with Near items only.

Table 13. Pretest-to-Posttest Near vs. Far Gain Comparison (n=566)

Subtest/Full Assessment	Near Gain		Far Gain		sig
	Mean	SD	Mean	SD	
Alpha Pig Letter Naming	0.23	0.86	0.21	0.94	ns
Alpha Pig Letter Pointing	0.20	0.87	0.24	1.22	ns
Wonder Red Decoding	0.95	1.31	0.61	1.26	0.000
Princess Presto Letter Sounds	0.53	1.16	0.60	1.76	ns
Full Assessment	1.98	2.63	1.77	3.31	0.03

Results were mixed. For the full assessment, Near and Far totals differed significantly; revealing that Near performance (gain) was superior to Far performance. However, this seems to be derived from the Wonder Red Decoding subtest. All other subtests did not reveal a difference between Near and Far performance gains. Also, age likely plays a role as older participants show a much smaller gap in the Near and Far Wonder Red Decoding subset, relative to younger kids. This finding was the same for both our 3 and 4 year old and 3–5- year-old groups.

Near and Far Interaction

The final analysis investigated the key question of whether Near and Far scores share a reliable relationship. Based on the total assessment performance (excluding the two Near-only subtests), Near gain scores share a positive, strong and significant relationship with Far gain scores (r = 0.702, p = .000). This correlation indicates the greater the Near gain, the greater the Far gain score. Figure 6 presents a graphical representation of this relationship. This finding was true for the 2009 full group, and for both 3–4-year-old and for 3–5-year-old groups.

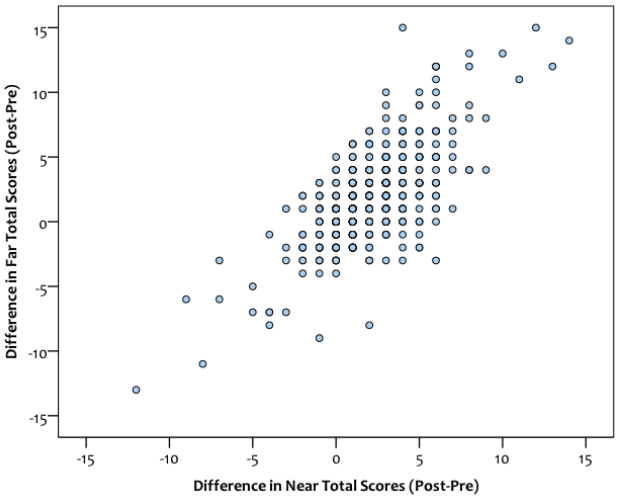

Figure 6. Scatter Plot of Near and Far Gain Score Interaction (n = 566)

We questioned whether two major factors had a significant influence on the Near–Far correlation. A regression model was used to test for any influence of the participant's age or the participant's pretest Near performance. Results indicated the relationship between Near and Far remains constant after controlling for both age and pretest Near totals. In other words, when the effects of age are removed, and pretest Near performance is equalized, the gains in Near performance continue to significantly predict the gains in Far performance. It should also be noted that neither age nor pretest Near totals significantly predict gains in Far performance.

Near and Far Correlations Based on Pretest Performance Quartiles

We investigated the Near–Far interaction based on quartile classifications. Quartiles were established by using 2009 participants' performance on pretest Near totals. There were significantly more 3 and 4 year olds in the bottom quartiles than in the top quartiles. Also, there were more females in the top quartile relative to the bottom quartiles. Table 14 presents a summary of Near–Far performance and correlation by quartile.

Table 14. Near–Far Means and Correlation Coefficients by Quartile

Quartile	*n*	Near Mean Difference Near Pre Score | Near Post Score	Far Mean Difference Far Pre Score | Far Post Score	*r*	sig
Bottom 25%	144	2.58 Pre: 2.56 | Post: 5.14	2.49 Pre: 4.00 | Post: 6.50	0.76	.000
26%–50%	144	2.67 Pre: 8.97 | Post: 11.63	2.61 Pre: 12.03 | Post: 14.16	0.54	.000
51%–75%	156	1.83 Pre: 12.35 | Post: 14.18	1.76 Pre: 16.44 | Post: 18.20	0.69	.000
Top 25%	122	.66 Pre: 14.76 | Post: 15.41	.48 Pre: 19.56 | Post: 20.04	0.81	.000

All quartile groups show a significant and positive relationship between Near gains and Far gains. These correlations ranged from 0.54 to 0.81 (0.76 for bottom quartile).

Results of this analysis further support the transfer from Near to Far items on the *Super WHY!* assessment. Of exceptional note were the greater gains for the lowest quartiles. The bottom quartile participants, who started with the lowest Near pretest scores, showed the strongest correlation between Near and Far.

A regression model conducted with the full 2009 sample revealed that Near gains significantly predict Far gains in performance—even after controlling for both age and pretest "Near" totals. Results for the bottom quartile mimic that of the full 2009 sample. Near and Far gain correlations for 3- and 4-year-old participants who pretested in the bottom quartile were about the same ($r = 0.74$) as all participants in lowest quartile. Near- and Far-gain correlations for 3-, 4-, and 5-year-old children who pretested in the lowest quartile were similar ($r = 0.75$) to all participants in the lowest quartile.

Summary of Key Findings

The *Super WHY!* Summer Camps provided children, ages 3–6, with a week-long experience centered on emergent literacy skills. Through interaction with television-delivered *Super WHY!* characters and supplemental exercises, camp participants gained new skills and abilities. An analysis of assessment data from 2008 and 2009 summer camp sessions established a statistically significant gain in participant performance from pretest to posttest. These gains were consistent year-over-year, indicating replicability of the summer camp program and its results.

The following is a summary of key findings from the *Super WHY!* Summer Camp data analysis effort.

1	Posttest mean scores were greater, relative to pretest scores, for each subtest and the full assessment. All differences were statistically significant.

Super WHY! Summer Camp participants realized gains in emergent literacy skills. On average, participants raised their pretest scores 5.70 points, which equates to an 11.7% increase in performance (total assessment points possible is 49). This growth is also described with an effect size of 0.42. Given the short intervention time (less than one week) and 49 point ceiling on the assessment form, the team considered the pre-to-posttest growth especially notable. Statistically significant differences between pre- and posttest existed regardless of age or ethnicity.

2	**There were no significant differences between the pretest scores of 2008 and 2009 participants.**

Samples consisting of 2008 and 2009 summer camp participants were considered equivalent based on pretest scores. Pretest means between the two groups differed by just 0.10 points (based on a 49-point possible assessment score). The team concludes that each group began the *Super WHY!* Summer Camp program with essentially the same level of performance.

3	**There were no significant differences between the gain scores of 2008 and 2009 participants.**

Participants made similar gains, regardless of whether they participated in a 2008 or 2009 summer camp session. There was only one subtest where 2009 participants scored lower (on average): the Princess Presto Encoding subtest. All other subtest scores, as well as the full assessment score, showed no significant difference based on program year. This finding leads to the conclusion that the *Super WHY!* Summer Camp program has a consistent impact— even when the population served doubled in 2009 to include more participants and more stations.

4	**Gain scores differ based on the participant's age—with younger participants' gain scores being greater, relative to older participants.**

The *Super WHY!* Summer Camp results in the greatest overall gains for 3 and 4 year olds. This finding is not surprising, given the emergent literacy skills covered and age range that the program targets. Gain score distributions may indicate that the youngest participants (three year olds), on average, are not developmentally ready for summer camp content. Conversely, six-year-old participants are, on average, more prepared. The fact that 14% of six-year-old participants pretested with a 100% correct score, compared to no three or four year olds and just 1.5% of five year olds, supports this possibility.

5	**Gain scores are greatest for participants pretesting in the second (25%–49%) quartile.**

Participants who pretested in the second quartile realized a 7.82 point gain, on average. The next closest gain was the third quartile's 6.60 point gain. The considerable difference in mean gain scores identified the second quartile participants as the optimal audience to target. In other words, the best-suited participants are those who enter the program with some level of knowledge. Indeed, participants in the second quartile pretested with a mean score of 23.43 and then realized the 7.82 gain. Participants in the first quartile pretested with a mean score of 7.91 and then realized a 6.60 gain.

Results of analysis by quartile likely reflect Vygotsky's Zone of Proximal Development (ZPD). Defined as the distance between a child's ability to (1) solve a problem with assistance and (2) solve the same problem independently, it is within the ZPD that learning occurs. Reading levels (independent, instructional, and frustration) are based on the ZPD concept. Subtest scores are consistent with this interpretation. Younger participants demonstrating growth in the more basic skills (letter naming and pointing), while 4 and 5 year old participants made the greatest gains with encoding and decoding.

Our analysis of performance by quartile suggests that that *Super WHY!* Summer Camp program aligns, on average, with the second quartile participants' ZPD.

6	**Near and Far scores shared a positive and significant correlation (r = 0.702, p = .000).**

The analysis of 2009 *Super WHY!* Summer Camp participants revealed an interesting interaction between Near and Far assessment content. Gains on Near content shared a positive, strong, and significant correlation to gains on Far content. This interaction indicates that as performance on Near items rises, so too does performance on Far items.

Near items were explicitly covered in the *Super WHY!* television; Far items were introduced only during camp activities. The strong interaction between Near and Far suggests that participants who master Near content are able to transfer those learnings to new and different instances of similar content. We take this promising evidence that learning goes beyond simple recall to the underlying concepts of presented literacy skills.

Conclusion

Analysis of *Super WHY!* Summer Camp data from the 2008 and 2009 sessions provides clear evidence of the program's impact. While only five days in duration, the program consistently produced measureable gains for participants. These gains proved statistically significant, and gains of various sizes were realized regardless of age or ethnicity.

Comparison of 2008 and 2009 results provided strong evidence of both consistency and scalability. While the 2009 summer camp almost doubled the number of participants, gains in participant performance remained consistent and did not differ from growth in the smaller 2008 summer camp implementation. These results also indicate scalability of the program: the addition of camps, supported with clear curriculum and implementation guidelines, did not lessen the gains achieved.

In particular, the program produced the greatest gains for participants who pretested in the second quartile (25%–49%). The second quartile finding provides an interesting indicator about when, and how, *Super Why!* presented reading skills should be taught. The noticeably lower scores for first quartile participants provide further evidence of this perceived "sweet spot" for summer camp intervention. Further investigation of this finding should be a component of future research.

Likewise, the Near–Far correlation is a promising finding of this analysis that may have implications for children's television. Further investigation of this relationship is merited. In particular, we suggest that the pre-and-post assessments be enhanced to include Far transfer items that are not covered in any part of the program. A correlation among these three measures (Near, Far—covered in camp, and Far—not covered in camp) would increase the evidence of television-taught skills transferring to new and novel content.

Such a study would also benefit from the inclusion of three groups: (1) summer camp program participants, (2) children who simply watch the episode each day at home, and (3) participants who do not ever watch the episode during the week-long intervention. This type of research cell arrangement would best isolate the impact of each summer camp element, as well as the television-instructional component alone. Conducting the study in this way increases the rigor and provides a quasi-experimental approach to isolate effect.

While opportunities to enhance and extend the findings reported here abound, this analysis provides convincing evidence of the *Super WHY!* Summer Camp program's impact on its young participants. Emergent literacy skills increase, and those skills readily transfer to content not presented in the *Super WHY!* television program.

Appendix I: IRB Approval for Data Analysis

Graduate and Research Affairs
Division of Research Affairs
San Diego State University
5500 Campanile Drive
San Diego CA 92182-8220
Phone: 619-594-5938
Fax: 619-594-4109

April 6, 2010

Faculty Researcher: *James Marshall, Ph.D.*
Department: *Educational Technology*

vIRB Number: *496054*
Title: *Analysis of Super WHY! Summer Camp Data: Understanding Relationship of Television-delivered Instruction and Literacy Outcomes*

Risk Level: *Minimal*
Exemption: *45 CFR 46.101(b)(4)*

Dear Professor Marshall:

The project referenced was reviewed and verified as exempt in accordance with SDSU's Assurance and federal requirements pertaining to human subjects protections within the Code of Federal Regulations (45 CFR 46.101). This review applies to the conditions and procedures described in your protocol.

The determination of exemption is final and requests for continuing review (Progress Reports) are not required for this study. However, **if any changes to your study are planned**, you must submit a modification request and receive either IRB approval (per 45 CFR 46.110 or 46.111) or IRB verification that the modification is exempt (per 45 CFR 46.101). To submit a modification request, access the protocol via the WebPortal, on the protocol Main Page, you will need to click on "Modifications" under Protocol Maintenance and enter a report. Once you have filled in your responses on the report form, click "submit". Additionally, notify the IRB office if your status as an SDSU-affiliate changes while conducting this research study (you are no longer an SDSU faculty member, staff member or student).

Please note the following for all exempt studies:

a) If this research involves the use of existing or secondary data sources, information obtained must be recorded so that subjects cannot be identified, either directly or through identifiers linked to the subjects.

b) If information will be obtained from individual medical records, please check with the organization authorized to provide access to these records to determine whether regulations relating to the Health Insurance Portability and Accountability Act (HIPAA) pertain to your research. Likewise, if academic records are accessed, Federal Education Rights and Privacy Act (FERPA) requirements must be respected. Notify the SDSU IRB office if protocol revisions are necessary to comply with HIPAA regulations.

c) If recruitment will take place through an outside agency or organization, confirm with that institution that you have permission to conduct the study prior to initiation of any study activities. If this research involves the use of existing or secondary data sources, confirm with

SAN DIEGO STATE UNIVERSITY

Graduate and Research Affairs
Division of Research Affairs
San Diego State University
5500 Campanile Drive
San Diego CA 92182-8220
Phone: 619-594-5938
Fax: 619-594-4109

the data owner that you have permission to access the data.

d) Approval is contingent upon the completion of the SDSU human subjects tutorial (found at: http://www-rohan.sdsu.edu/~gra/login.php) by all members of the research team. This certification must be renewed every 2 years.

For questions related to this correspondence, please contact the IRB office ((619) 594-6622 or e-mail irb@mail.sdsu.edu). To access IRB review application materials, SDSU's Assurance, the 45 CFR 46, the Belmont Report, and/or any other relevant policies and guidelines related to the involvement of human subjects in research, please visit the IRB web site at http://gra.sdsu.edu/research.php.

Graduate Students: This notification may be used as documentation to register in Thesis 799A. Attach a hard copy of this notice to your Appointment of Thesis/Project Committee form prior to submitting the completed form to Graduate and Research Affairs - Student Services Division.

Sincerely,

Jeanne Nichols
Chair, Institutional Review Board

Amy McDaniel
Regulatory Compliance Analyst

Choya Washington
Regulatory Compliance Analyst

Appendix II: *Super WHY!* Assessment Instrument

Camper's Name: _____

SUPER WHY & The Three Little Pigs
Reading Camp Pre-Interview Assessment

[INTERVIEWER] Hi, my name is _____. Today, we are going to play some games with letters and words. First, I need to ask you just a few questions. What is your name? How do you spell that?
Camper's Name: _____ Ethnicity: _____
Okay, now, how old are you? When is your birthday?
Age: _____ Birthday: _____
Great, now we can play.

To Get Ready:
• Sit on the floor or at a table, and ask the child to sit across from you, facing you.

MATERIALS:
• Letter Flash Cards: P, W, I, L, B, T, A, O, F, G
• A visual of the Alphabet presented in order.
• Word Flash Cards: WALL, TALL, BALL, FALL, HALL, CALL
• A visual of three words listed: good, red, small
• A visual of three words listed: small, good, red
Camper's Name: _____

Day 1: Alpha Pig Day: Letter Identification Assessment

[INTERVIEWER] We are going to look at some letters. I'm going to show you a letter, and you'll tell me its name, okay?

[HOLD OUT THE LETTER CARD IN FRONT OF YOU, SO THE CHILD CAN CLEARLY SEE THE LETTER.]

[INTERVIEWER] Ask, "What letter is this?"

[CODING: Write the letter that the child says on the line next to the letter asked about.]

___ P (Far)
___ W (Near)
___ I (Far)
___ L (Near)
___ B (Far)
___ T (Near)
___ A (Far)
___ O (Near)
___ F (Near)
___ G (Far)

Total: _____

[INTERVIEWER] Great Job!

Camper's Name: _____

Day 1: Alpha Pig Day: Letter Identification Assessment Continued
Now, I'm going to show you a picture with all the letters of the alphabet on it. I'm going to say a letter, and you find it on the picture, okay?

[PUT THE PICTURE OF THE WHOLE ALPHABET IN ORDER IN FRONT OF THE CHILD.]

[INTERVIEWER] Say, "Find the letter B." **[CONTINUE THROUGH THE WHOLE LIST BELOW.]**

[CODING: Write the letter that the child points to on the line next to the letter asked about.]

___ **B (Far)**
___ **L (Near)**
___ **O (Near)**
___ **I (Far)**
___ **A (Far)**
___ **T (Far)**
___ **F (Near)**
___ **W (Near)**
___ **P (Far)**
___ **G (Far)**

Total: _____

[INTERVIEWER] Great Letter Finding!

Camper's Name: _____

Day 2: Wonder Red Day: Decoding Assessment

[INTERVIEWER] We're going to read some words. I am going to put down a word, and you read it. Okay?

[PUT DOWN A WORD.]

[INTERVIEWER] Ask, "What word is this?" DO NOT TELL THEM THE WORD. Mark whether the child reads the word fully (as fluent readers would read it) or can only blend it (segmented reading). Move on to the next question.

[REPEAT FOR EACH WORD.]

[CODING: Write a "1" in the Read Word column if the child read the full word as a fluent reader would. If child reads the whole word as W-ALL, mark 1 in the Segments column. If the child could not read the word at all, mark both columns with a "0"]

Read Word **Read word in Segments** Transfer

___ WALL ___ W- -ALL (near)
___ TALL ___ T- -ALL (near)
___ BALL ___ B- -ALL (near)
___FALL ___ F- -ALL (far)
___ HALL ___ H- -ALL (far)
___ CALL ___ C- -ALL (far)

[INTERVIEWER] Great word reading!

Camper's Name: _____

Day 3: Princess Presto Day: Letter Sounds, Encoding Assessment

[INTERVIEWER] We're going to play a game with letter sounds. I'm going to say a sound, and you tell me what letter makes that sound, okay?

[INTERVIEWER] Say, "What letter makes the sound /T/?"

[CONTINUE THROUGH THE WHOLE LIST BELOW.]

[CODING: Write a "1" next to the sound if the child says the correct letter name. Write a "0" if the wrong letter is said.]

(Make the sound /t/ as in TABLE.) ___ **/T/ (Far)**
(Make the sound /i/ as in IGLOO.) ___ **/I/ (Near)**
(Make the sound /b/ as in BIG.) ___ **/B/ (Near)**
(Make the sound /f/ as in FISH.) ___ **/F/ (Far)**
(Make the sound /w/ as in WOLF.) ___ **/W/ (Far)**
(Make the sound /a/ as in APPLE.)
___ **/A/ (Far)**
(Make the sound /p/ as in PIG.) ___ **/P/ (Near)**
(Make the sound /g/ as in GOOSE.) ___ **/G/ (Near)**
(Make the sound /o/ as in OCTOPUS.) ___ **/O/ (Far)**
(Make the sound /o/ as in OPEN.) ___ **/O/ (Far)**
(Make the sound /L/ as in LADDER.) ___ **/L/ (Far)**

Total:

PART TWO: ENCODING
[INTERVIEWER] Say, "Okay, now I'm going to ask you to spell some words. How do you spell BIG? How do you spell PIG?
[CODING: Mark a "1" for each correct letter said in any order, and add a "1" if the whole word is spelled correctly *and* in the right order.]
____ BIG (Near) ____ PIG (Near)
____ B ____ P
____ I ____ I
____ G ____ G

Camper's Name: _____

Day 4: Super Why Day: Reading Words and Opposites ALL NEAR

[INTERVIEWER] We're going to play a game with words. I'm going to ask you a question, and you pick the word that answers my question, okay?

[SHOW THREE WORD CHOICES.]

[CODING: Write a mark next to the word the child picks. The child needs to pick one word.]

[INTERVIEWER] Which word is the opposite of big? Is it small, red, or good?
___ small ___ red ___ good

[INTERVIEWER] Which is the word *small*?
___ small ___ red ___ good

If no response or wrong response to previous question:
[INTERVIEWER] Which is the word *small*? I think it has the "all" in it.
___ small ___ red ___ good

[SHOW THREE WORD CHOICES.]

[INTERVIEWER] What is the opposite of bad? Is it red, good, or small?
___ red ___ good ___ small

[INTERVIEWER] Which one is the word *good*?
___ red ___ good ___ small

If no response or wrong response to previous question:
[INTERVIEWER] Which one is the word *good*? I think it starts with the letter g?
___ red ___ good ___ small

Total: _____

Appendix III: Overall Performance by Station

Station	Camps Conducted		*n*	Pretest		Posttest		Gain		sig
	08	09		Mean	SD	Mean	SD	Mean	SD	
1. APT	✓	✓	34	32.35	13.62	38.32	10.68	5.97	8.87	sig
2. IPTV	✓	✓	41	26.09	10.00	34.60	13.25	8.51	5.40	sig
3. KAET	✓	✓	13	26.30	9.34	33.46	9.58	7.15	4.16	sig
4. KAET (Phoenix)	✓	✓	11	18.81	10.52	24.90	16.17	6.09	6.69	sig
5. KLRN	✓	✓	19	24.31	13.77	26.78	14.89	2.47	6.76	ns
6. KPBS (Calexico)	✓		20	31.40	10.89	36.80	10.23	5.40	4.57	sig
7. KPBS (National City)	✓		19	29.37	12.69	32.89	12.44	3.53	2.86	sig
8. KQED	✓		5	20.80	9.26	21.60	8.64	0.80	6.30	ns
9. LPB	✓	✓	99	18.79	13.96	26.66	15.58	7.87	9.10	sig
10. MPB	✓		26	32.61	9.96	37.65	9.89	5.04	4.34	sig
11. MPT	✓	✓	37	31.32	11.47	35.81	12.07	4.49	4.98	sig
12. NJN		✓	13	32.38	6.04	39.31	7.75	6.92	4.87	sig
13. WFSU	✓	✓	52	32.82	10.85	35.42	11.53	2.60	5.02	sig
14. WGTE	✓	✓	19	23.00	10.35	27.84	15.23	4.84	7.06	sig
15. WHRO	✓	✓	23	27.30	11.65	33.56	11.27	6.26	4.16	sig
16. WHUT	✓	✓	26	41.34	5.49	44.31	4.83	2.96	3.45	sig
17. WHUT (Howard)	✓	✓	12	42.58	4.18	47.50	3.09	4.92	2.97	sig
18. WHUT (Ross)	✓	✓	14	43.00	3.76	47.71	1.77	4.71	3.07	sig
19. WJCT		✓	1	2.00	NA	5.00	NA	3.00	NA	—
20. WLJT	✓		22	20.73	10.65	24.27	13.13	3.55	6.23	sig
21. WNED	✓	✓	89	30.68	10.71	35.61	10.84	4.93	5.73	sig
22. WNPT	✓	✓	158	26.26	14.41	33.68	14.78	7.42	7.22	sig
23. WPSU	✓	✓	93	23.33	12.03	28.64	13.52	5.31	5.95	sig
24. WSIU	✓	✓	23	24.21	10.23	28.61	11.92	4.39	4.38	sig
25. WVPT	✓	✓	40	26.22	14.31	31.78	14.66	5.55	9.81	sig

The stations on this list, and subsets of data from the same station where noted, represent the organization of raw score "packets" received by the data analysis team.

About the Data Analysis Team

James Marshall, Ph.D.

Dr. Marshall is a faculty member in the Department of Educational Technology at San Diego State University and an independent consultant to corporate business entities and school systems. He teaches graduate-level courses in instructional design, human and organizational performance, and evaluation. His large-scale research studies for federal and state government agencies have evaluated over $20M in funded projects. His client list includes Bank of America, Anheuser Busch, Court TV, McGraw Hill Companies, The Princeton Review, The Transportation Security Administration, TIAA-CREF, The Corporation for Public Broadcasting, and the U.S. Department of Education.

Dr. Marshall can be reached at marshall@mail.sdsu.edu.

Diane Lapp, Ed.D.

Dr. Lapp is a Distinguished Professor at San Diego State University and an internationally recognized authority in language acquisition and development, reading and language arts, and first and second language assessment. Co-Chair of the International Reading Association's (IRA) Commission on Literacy, she has also been recognized by IRA with induction into its Reading Hall of Fame in 2005.

Kimberly Cavoto, Ph.D.

Dr. Cavoto specializes in applying and integrating solid research design with multivariate statistical analytics to guide academic, commercial, and governmental organizations in making sound data-driven decisions. In addition to 20 years' experience as a researcher and data analyst, she has taught college level statistics and, after 11 years, continues to be a statistical trainer for SPSS, a top statistical software company.