The *Between the Lions* American Indian Literacy Initiative
Research Component:
Report Prepared for the United States Department of Education

Deborah L. Linebarger, PhD
Annenberg School for Communication
University of Pennsylvania

Introduction to *Between the Lions* Research

In an initial examination of *Between the Lions*, kindergarten children from moderately at-risk backgrounds (i.e., determined by initial reading skill) who watched the program performed significantly better on almost all outcome measures of reading achievement when compared to similar children who did not watch the program. Specifically, the kindergarten children who watched 17 episodes from the first season of BTL were able to identify more letters and read more words than those who did not watch. Their understanding of concepts of print was enhanced and their phonemic awareness and letter-sound correspondence skills, skills highly predictive of later, fluent reading, were significantly accelerated over those who did not watch.

Between the Lions was also the focus of a research study conducted in Mississippi (Prince, Grace, Atkinson, Linebarger, & Huffman, 2002). More than 50 teachers and nearly 1,000 preschool, kindergarten, and 1st grade students from the Choctaw Indian Reservation and the Delta region (children who were low income and/or ESL students) were assigned to either a treatment group (i.e., received videotapes and ancillary materials) or a control group (i.e., did not receive the materials; however, they might see the program on their local PBS station). This investigation differed from the initial evaluation in two important ways. First, the evaluation period lasted 9 months (versus just over 1 month). In addition, this project combined extensive ancillary intervention materials, training, and follow-up (e.g., teachers' guides, handbooks, posters, a CD-ROM with content form the Web site, sets of classroom books, and take-home books, stickers, bookmarks, and information for parents) with videotapes of the episodes.

Researchers found children who watched half-hour episodes of *Between the Lions* regularly, and whose teachers carried out related activities, significantly outperformed control groups on several key reading skills (Prince et al., 2002). For instance, on the TERA-3, a standardized test that measures reading ability at the earliest stages, all the Choctaw viewers and the Indianola kindergarten viewers significantly outperformed the control groups on the conventions subtest, a measure of what children know about books and other basic concepts of print, such as reading from top to bottom and left to right. In additions, Indianola viewers, who started significantly below the control group at the outset of the project on the test measuring the ability to identify initial word sounds, significantly outperformed their non-viewing peers and gained skills at a faster rate by the end of the project. On a test of letter/sound correspondence, a high-level skill associated with knowledge of the alphabetic principle and predictive of later fluent reading, *Between the Lions* viewers on the Delta performed significantly better and gained skills at a faster rate than students in the control group. On a test of receptive, or listening, vocabulary and knowledge word meanings, an important predictor of reading achievement in later grades, delta kindergartners who viewed *Between the Lions* significantly outperformed the control group.

Project Purpose

The overall purpose of this project was to increase the early English literacy skills of American Indian children in New Mexico Head Start programs using the PBS children's program *Between the Lions*, a television series designed with the intent of supporting language and literacy acquisition, and related resources adapted specifically for the American Indian communities.

Embedded within the larger project goals was an evaluation designed to assess the impact of this intervention on the early literacy skills of participating American Indian Head Start children.

METHOD

Research Design

We used a varied program of evaluation that included extensive-evaluation of 5 tribal communities and limited-evaluation of the remaining 6 tribal communities. A priority was to develop a design that not only ensured valid, meaningful and replicable results but also considered the needs and concerns of the tribal communities involved. Prior to installation of the curriculum, a series of discussions and a succession of revisions to the outreach and research materials transpired resulting in a classroom syllabus, research design, and assessment procedures agreed upon by all parties.

The chosen design used a comprehensive and multi-method approach that allowed all sites to be fully eligible for participation and enabled the installation to be flexible to the needs of each site while still maintaining the necessity of convergence.

The research took place in two phases over a two year period.

Year 1: Multiple Baseline Design

In Phase One of the initiative, we used a modified single-subject methodology in order to allow all tribal communities equal opportunity to participate in the evaluation. Because some of the locations contained just one classroom, we were unable to use typical group design methods (e.g., experimental/control designs). If we excluded locations because they did not fit into a typical group design framework, then our research would be artificially constrained to only those locations with multiple classrooms or sites. Further, in our discussions with tribal representatives, we believed that the uniqueness of each location necessitated a design that would examine allow us to examine the effectiveness of the intervention across multiple settings. With these criteria and constraints, a single-subject methodology was chosen as most appropriate.

 Single-subject designs are based on repeated and controlled application of a particular intervention (Kazdin, 1982). The goal is to demonstrate the efficacy of an intervention using appropriate control conditions. Then, we can evaluate the degree of change in behavior as compared to pre-intervention levels. Repeated measurements of the target behaviors (i.e., in this case, children's English literacy skills) are tracked and graphed. Next, we visually examine these graphs to determine whether changes in behavior co-occurred with the introduction of the intervention in question.

There are a variety of single-subject designs (e.g., A-B-A-B; alternating or simultaneous treatments; multiple baseline).We opted to use a multiple baseline design across settings. That is, we examined the acquisition of children's English literacy skills across multiple settings (i.e., each setting was either a tribal community Head Start or multiple classrooms within that Head Start location). Each of the multiple baseline settings started the

intervention approximately 1-2 months apart, while all the children engaged in repeated assessments across a baseline period and into the treatment period. See Table 1.

Analysis of a multiple baseline design includes visually inspecting children's graphed data (e.g., their ability to identify alphabet letter names) as well as fitting separate trend lines to the baseline data and to the intervention data. Trend lines provide information about the mean level of performance and the slope of the line (i.e., how much the behavior is growing within baseline when there was no intervention and within intervention). Typically, single-subject designs focus on the behavior of a single child. We modified the traditional single-subject design by tracking multiple children at each 'setting' in the study. Therefore, we could calculate a general trend line that used all children's data at each setting across baseline and intervention. For this report, we have fitted the trend lines using a hierarchical linear modeling (HLM) technique. HLM allows us to calculate the average performance at a given point in the intervention (e.g., during baseline, after 3 intervention lessons, after 6 intervention lessons) as well as average growth in a particular skill.

Table 1. Multiple Baseline Design and Assessment Schedule

Site	NOV.	DEC	JAN	FEB/MAR	APR/MAY
Site A Taos	X_{wave1} I			X_{wave2} I	X_{wave3} I
Site B/SF Class 1, 2	X_{wave1} I			X_{wave2} I	X_{wave3} I
Site C Laguna	X_{wave1} I			X_{wave2} I	X_{wave3} I
Site B Class 3, 4	X_{wave1} B	COMMUNITY HOLIDAYS		X_{wave2} I	X_{wave3} I
Site D	X_{wave1} B			X_{wave2} B	X_{wave3} I
Site B Class 5, 6	X_{wave1} B			X_{wave2} B	X_{wave3} I

Note. X = Wave of Assessment; B = Baseline; I = Start Intervention

Year 2: Standard (Control/Experimental) Design

During the second year of the initiative (i.e., Phase Two), we were able to introduce an experimental research design as our tribal community had multiple locations that were similar demographically to one another. An experimental design is one in which children are randomly assigned to different groups, the groups receive different manipulations (e.g., some children receive the intervention and other children do not), and then we evaluate differences in outcome measures. Because we randomly assigned classrooms to groups (rather than individuals), we adopted a quasi-experimental framework. The major disadvantage of this design is that the original groupings (by classroom) may have differed from one another in substantial or meaningful ways. Therefore, we have less confidence that the manipulation caused the changes in the outcomes. Instead, there may have been differences in the groups that resulted in the changes (e.g., different instructors, curricula, or ability levels of the children). We did attempt to reduce this problem by selecting Head Start classrooms geographically near one another. We also collected detailed information from each family and Head Start location to evaluate any initial differences. If there were none, we will have begun to establish that the differences between groups are truly a result of the manipulation and not any initial group differences. Next, we would examine the extent of the changes in the outcome variables. For instance, if we find the same pattern of results across all outcomes favoring our manipulation, our confidence in the effectiveness of the manipulation is further enhanced.

Our Phase Two tribal community allowed us to implement the experimental design in four classrooms. Therefore, we were able to implement the intervention in two of the four classrooms. The other two classrooms continued with their regular literacy curriculum. two control and two experimental sites (randomly assigned by classroom). As with Phase One, there were three waves of assessment during Phase Two. Each wave took place at the same time in all sites.

Participant Recruitment and Selection

Sites

Eleven different tribal communities located in New Mexico comprising a diversity of culture and literacy priorities were approached to participate. The intervention took place during the regular academic year (i.e., September to May over 2 years). In year one, four of the eleven tribal communities were randomly selected to participate in the intensive evaluation, serving as independent intervention settings. That is, each of the sites served as a separate replication of the intervention. Six tribal communities participated in a limited evaluation, using all the intervention materials while participating in limited assessment activities that included the administration of teacher measures, parent measures, a brief literacy screening tool (i.e., administered by caregivers or education specialists at each location), and weekly post-test evaluations on 4 students per classroom.

In year two, one tribal community participated with 2 sites randomly selected to incorporate the curriculum into their classroom and 2 sites acting as control classrooms (i.e., these sites did not use the intervention).

Child Participants

Year 1 –
4 Tribal Communities participated in intensive assessment activities with University of Pennsylvania project staff (15 classrooms). Of the 250 children available across the 4 sites, parental consent was received for 135 children.
N= 135 (53% girls / 47% boys)

Year 2 –
1 Tribal Community (4 sites / 4 classrooms: 2 control / 2 experimental)
Consent forms were received for 70 children
n=39 experimental; n=31 control (42% girls / 57% boys)

Teacher Participants

Year 1 – Phase One

5 tribal communities returned teacher surveys. These surveys were used to examine how differences in teacher qualifications and classroom activities might be impacting the effectiveness of the implementation. In Phase One of the research the average number of years teaching was 11.6 (with a range from 2 to 30 years in the classroom). These teachers had an average of 14.8 years of education (approximately that of an Associate's Degree). Years of education ranged from 12 to 16.

Teachers were asked to allocate the time spent each day to a range of activities from health activities (i.e., meals, hand washing, brushing teeth) to transitional activities (i.e., coming in from outdoor play time, moving from one lesson/activity to another) to learning activities (i.e., academic curriculum). Teachers reported an approximate 55 hour per week were spent in learning activities in the classroom (this includes time in which more than one activity is happening simultaneously).

Phase Two

Teacher surveys were not returned for the second phase of research.

Research Objectives

√   Objective 1:   To describe the teachers involved in the intervention, document the amount and type of literacy strategies used by these teachers, and evaluate the role that teacher characteristics and behavior played in relation to children's outcomes

√   Objective 2:   To describe children and families who participated in this project, including the home media and literacy environments of these children and families

√   Objective 3:   To evaluate child English literacy skills

Constructs and Measures

It was important to the researchers that children's areas of strengths rather than areas of weaknesses were examined for this project. A series of measures were compiled, piloted and altered in order to ensure they were valid and culturally appropriate measures for the American Indian population in this initiative. Where available, we have listed benchmarks for various measures to contextualize the results.

Educator Demographic Questionnaire

Education levels, attitudes toward media and literacy instruction, experience, and other demographic information were collected and examined in relationship to implementation of the intervention and child outcomes.

Classroom Literacy Environment

Detailed questionnaires were completed by teachers in the 46 classrooms across the 12 tribal communities. We planned to document the amount and type of strategies and interactions associated with literacy instruction in the classroom (e.g., children create stories; teachers discuss alphabet; the environment is labeled with print and signs, teachers talk about and using rhyming and alliteration).

Children and Families Demographic ad Home Media Environment

Home literacy practices and demographic characteristics were be assessed by parent report using an adaptation of the Stony Brook Family Reading Survey, a 52-item questionnaire designed to measure the literacy environment, parental expectations, and parental characteristics of each family. In addition, we included other questions to examine attitudes toward media and media use and access.

Storytelling and Narrative Abilities

Storytelling and narrative competence were measured using the *Narrative Comprehension of Picture Books* (Paris & Paris, 2003). This task examined young children's comprehension of wordless picture books to assesses their thinking and comprehension of narrative sequences independent of any decoding skills (an essential feature for preschool children who would not be expected to be reading). The measure is composed of 3 separate tasks (i.e., storybook picture walk; retelling; and prompted comprehension) that yield five different composite scores including spontaneous reactions to the story; retelling of the

story; comprehension of explicit story information; comprehension of implicit story information; and total storybook comprehension.

Oral Language and Vocabulary

Oral language and vocabulary knowledge was evaluated using a tool that measures their expressive language, the *Picture Naming Task* (PNT, Missall & McConnell, 2004). The child is shown a series of pictures and asked to name as many as he/she can in one minute. Categories of objects include animals, food, people, household objects, games and sports materials, vehicles, tools, and clothing.

Letter Name Knowledge

Children were given a sheet of displaying all 26 upper-case letters out-of-order and were asked to identify as many as possible (moving in order from left to right and top to bottom). There was no time limit.

Environmental Print Awareness

Children were asked to identify a series of examples of environmental print: road signs, product labels, corporate logos, etc. Example: Child is asked to describe a picture of the McDonald's golden arches, a seat belt roadside sign, a Wal-Mart logo.

Phonemic Awareness

We used 4 different tasks to measure phonemic awareness: blending, rhyming, alliteration, and sound isolation. All four tasks have pictures with them to support children's cognitive capacity (i.e., the child can concretely see the item and not have to think about it in any way, rather he or she can concentrate on the actual task at hand).

- *Sound Isolation Example:* "Do you hear 'v' in van?"; "Do you hear 'b' in cat?"
- *Blending Example:* "Find what you get when you put dog and house together. Find dog (pause) house."
- *Rhyming Example:* "Rhyming means words end with the same sound, for example Cat rhymes with Hat – do you hear the sound /at/at the end of each word? Here is a frog. Here is a doll and a dog. Does frog sound like doll? Or sound like dog?"
- *Alliteration Example:* "We're going to look at pictures. Two of them start with the same sound, the other one starts with a different sound. Find the one that starts with a different sound than the other two words – This is bug, bag, sun…".

General Literacy Abilities

General literacy abilities were evaluated using the *Get Ready to Read! Screener.* This screener assesses print knowledge (i.e., knowledge of the letters of the alphabet); book knowledge (recognition of how books work including the difference between words and images); phonological awareness (i.e., understanding that spoken words are composed of individual sounds); phonics (i.e., recognition of the sounds letters make); and writing (i.e., understanding how text should look: letters grouped together into words). The measure

consists of twenty items. Each item requires the child to select a response from a group of four pictures (or four letters, words, etc.). Example: "These are pictures of a book. Find the one that shows the back of the book." Example: "Find the letter that makes a *tuh* sound." Example: "Some children wrote their name. Find the one that is written the best."

Note about Results Presented Below

Included in this report are results related only to Phase One of the research. We plan to create a formal report that includes all data from both Phases; however, as part of an agreement that allowed us to work with one of the 11 tribal communities, we are not currently able to present any findings related to the community involved in Phase Two research until this community's institutional review board has had an opportunity to examine, discuss, and approve the release of these findings. At this time, we do not have an expected date of release. However, we will be sure to keep all parties informed as our analyses and writing progress.

Analytical Approach

We used a Hierarchical Linear Modeling (HLM) analytic technique to analyze child outcomes. Because the study design involved baseline and intervention time periods and because these time periods occurred at different assessment points for children in different settings (i.e., for Site D, intervention began at assessment 3), we were interested in evaluating the effects of the amount of intervention occurring at each wave for each setting. To accomplish this analysis within HLM, we included number of lessons of intervention at a particular wave of assessment in the Level-1 model as a time-varying covariate and evaluated whether slope differences were significant. A significant difference indicated a linear model where each subsequent intervention lesson had an additive effect on the outcome of interest. We also tested whether gender, entered as a Level-2 predictor, moderated any of these relationships.

Two statistical controls correcting for number of hours of learning activities presented in the classroom (as reported by teachers) and teacher's years of education were included in all analyses. Control variables regularly account for many of the individual differences in the outcome measures. Therefore, these analyses represent a stringent test of the effects associated with viewing this program.

RESULTS

Children's General Literacy Abilities at the Beginning of the Intervention Period (Across All Locations, Limited and Intensive)

At the beginning of the intervention,

- 39% of all participating children were below the anticipated developmental level and were therefore considered to be "at risk"
- 38% of children were considered "not-at-risk" and were developing language and literacy skills at the expected rate

- 23% of all children were performing at an "above-average" level for language and literacy development

Child Outcomes from Intensive Locations

Storytelling and Narrative Abilities

Pilot studies in Philadelphia and New Mexico indicated that American Indian children had strong storytelling skills. See Table 2. The final results have not yet been analyzed.

Table 2. Pilot Results Collected During September and October 2004

Measure	Location		Age	
	New Mexico	**Pennsylvania**	**3 yo**	**4 yo**
Get Ready to Read (out of 20)	9.4	9.2	7.5	10.5
Picture Naming (most in 1 minute)	17.4	16.67	15.0	19.25
Environmental Print (out of 10)	8.5	7.0	7.0	8.5
Letter Naming (out of 26)	9.0	10.5	5.33	13.25
Narrative Comprehension (out of 10)	7.5	4.8	4.25	7.43
Blending	5.67	NA	NA	5.67

NA = not administered

Oral Language and Vocabulary

Results. After 14-16 *Between the Lions* intervention lessons, children in this study were correctly able to name 30.6 pictures and were growing at 2.09 words per every 3 *Between the Lions* intervention lessons completed.

Benchmarks. Typically developing 66-month old children (5.5 years) have approximately 27 words with a growth of .44 words/month. Children from high poverty backgrounds at the same age have approximately 19 words, with a growth of .36 words/month.

Letter Name Knowledge

Results. After 14-16 *Between the Lions* intervention lessons, children in this population were correctly able to name 21.5 upper-case letters, growing at 2.6 letter names per 3 intervention lessons.

Environmental Print Awareness

Results. There was no improvement in this measure associated with intervention lessons. Children averaged 7.8 items correct (out of 13) with no significant growth (i.e., -

.13 items per 3 intervention lessons); however, this is not reason for concern. There is no research which indicated levels of environmental print awareness impact later academic success.

Phonemic Awareness

Sound Isolation Results. After 14-16 *Between the Lions* intervention lessons, children in this study were correctly able to identify 9.6 (out of 10) sounds in isolation and were growing at .79 sounds per every 3 *Between the Lions* intervention lessons completed.

Blending Results. After 14-16 *Between the Lions* intervention lessons, children in this study were correctly able to blend 8.8 items (out of 10) and were growing at .46 blends per every 3 *Between the Lions* intervention lessons completed.

Alliteration Oddity Results. No improvement was found in this measure. Children averaged 4.1 items correct (out of 10) with no significant growth (i.e., .12 items per 3 intervention lessons). It is likely this is due to the instructions provided by researchers to the children. Researchers suggest the children were not clear on the task.

Rhyming Results. No improvement was seen in this measure. Children averaged 5.3 rhymes correct (out of 10) with no significant growth (i.e., -.02 items per 3 intervention lessons). Children may have had difficulty with the directions on this task; alternatively, teachers reported less instructional time was devoted to rhyming tasks compared with letter naming and sound identification.

General Literacy Abilities

Results – Intensive Only. After 14-16 *Between the Lions* intervention lessons, children in this population answered 14.5 items correctly, growing at 2.2 items correct per 3 intervention lessons.

Results – Full Sample. At the end of the intervention period the number of children who were considered "at risk" pre-intervention decreased from 39% to 12%. Those in the "not-at-risk" category decreased from 38% to 24%. A significant increase was seen in children who were performing at an "above-average" literacy level. The number of children in this category increased from 23% pre-initiative to 64% at project end.

Benchmarks. The national Head Start average was 8.52 during the fall assessment period. Children in our project scored 8.5 items correct. Scores greater than 11 on this measure are predictive of reading success by 2nd grade. The standard developmental gain from fall to spring (without the *Between the Lions* curriculum) is approximately 15% in the general population.

Anecdotal Results

Teachers and Head Start Directors offered many insights into the impact of the initiative that could not be gathered through simple "data gathering waves". These anecdotal reports suggest the program is having a much broader reach than can be measured in a research methodological way.

Teachers receiving children into Kindergarten and Grade 1 from the Head Start programs that had, in the previous year, been incorporating the Between the Lions program into the classroom curriculum found the children to be different academically and socially than children who had not had the curriculum. They felt the children were more prepared academically and were similarly more confident in their new classroom environment.

Other teachers mention parental reports of children bringing their lessons home with them. After watching the program "Shooting Stars" and working with the "Shooting Stars" curriculum, one parent asked, "what were you watching on TV in class yesterday? [My child] insisted we go out side and look at the stars!"

Another parent mentioned that, while walking with his child the child suddenly announced, "Dad you're big and I'm little" following a viewing of the "Little Big Mouse" episode in which these concepts are addressed.

CONCLUSIONS

- Children's skills improved in direct relation to the amount of intervention experienced for oral language knowledge, letter names, phonemic awareness (i.e., isolating sounds in words and blending words) and general literacy abilities.

- Get Ready to Read! benchmarks at the final wave (across all locations) indicated that, on average, children's scores would predict 2nd grade success in reading (i.e., 14.5 items correct; 11 correct predict 2nd grade success in reading).
  - Average national gain fall to spring = 15%
  - This project gain fall to spring = 25%

- Anecdotally, teachers and parents noted spontaneous comments about learned curriculum components. Teachers also reported that children were especially engaged in project activities and seemed to really enjoy these activities.

References

Kazdin, A. E., (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford, England: Oxford University Press.