

# Final Evaluation of the Reading Initiative



Report to the  
J. A. & Kathryn Albertson Foundation  
Board of Directors

Herbert J. Walberg

December 2001

# Contents

<b>I. Bulleted Executive Summary .....</b>	<b>3</b>
<b>II. Overview .....</b>	<b>5</b>
A. Context of the Reading Initiative .....	5
B. Service Delivery .....	6
C. Goal Attainment.....	7
<b>III. Relation of Programs to the Foundation’s Goals .....</b>	<b>7</b>
<b>IV. Program Costs .....</b>	<b>9</b>
<b>V. Waterford Outcome Results .....</b>	<b>10</b>
<b>VI. Accelerated Reader Outcome Results.....</b>	<b>13</b>
<b>VII. Service Delivery.....</b>	<b>16</b>
A. Pesky Center .....	16
B. Idaho State Library .....	16
C. Idaho Public Television .....	17
<b>VIII. Conclusions .....</b>	<b>17</b>
<b>IX. Appendices .....</b>	<b>20</b>
A. The Components of the Reading Initiative.....	21
B. Program Descriptions .....	22
C. Panel Recommendations .....	24
D. Evaluation Design Features.....	26
E. The Development of Vocabulary and Reading Skill .....	30
F. Bio for Herb Walberg.....	34

## I. Executive Summary

- The Reading Initiative programs supported by the Albertson Foundation provide a balanced approach for advancing four of the Foundation's Ultimate Critical Results: Readiness to Learn, Student Achievement, Educator Competence, and Stakeholder Satisfaction
- The Foundation invested nearly \$35,000,000 since March 1998 to five well-planned and continuously evaluated programs.
- Each program was evaluated by the present evaluator who was closely advised and informed by Albertson Foundation staff and outstanding Idaho educators and scholars with direct knowledge and experience with program operations in the state. They reviewed annual reports from each program vendor and raised questions that the vendors answered in orally and in writing.
- The Waterford Early Learning Program produced outstanding achievement results on young children's reading achievement, particularly those children who had initially poor prospects for learning to read well and those who completely finished the program.
- The Accelerated Reader program also produced substantial reading achievement gains particularly for grades 1 through 4 and particularly in schools where teachers were well trained in the use of the program.
- Independent of the first two programs, the Pesky Center provided sterling training in evidence-based techniques for teaching reading for substantial numbers of elementary school teachers throughout the state. It was very favorably received by training recipients.
- The Idaho State Library system reached out to sizable numbers of families involving children from birth to eight years of age to encourage them to read daily to their children and to encourage their children to use public libraries. The central Library system staff also carried out well regarded professional workshops on reading for li-

brarians, educators, and parents throughout the state and developed a special collection of early reading videos and other materials for statewide loans.

- The Idaho State Television programs provided well received workshops for many parents and guardians throughout the state and distributed approximately 25,000 books to needy families.
- Overall, the Reading Initiative is perhaps boldest, most comprehensive, and successful state program of its kind in the nation and provides an excellent model for citizen, education and legislative leaders in Idaho and other states.

## II. Overview

The J. A. and Kathryn Albertson Foundation Reading Initiative is a bold and well-conceived venture to help solve the problems of young children's poor reading abilities. Although being carried out in Idaho, its pioneering ideas and the evaluation findings have national significance.

### A. Context of the Reading Initiative

The Albertson Foundation concern and funding of the Reading Initiative arose during a time of heightened national concern about children's reading difficulties especially during the early school years and the consequences for their academic progress, their success in life, and for American society. Idaho was no exception. The Report to the Legislature on Reading Education in Idaho concluded:

- That 3rd graders who are reading a year or more below grade level and are poor and attending a school serving many other poor children have nearly no chance of graduating from high school.
- That effective classroom instruction in the early grades by well-prepared teachers is the most powerful method for preventing reading and learning problems.
- That a balanced and comprehensive approach to reading must offer an organized, explicit skills program that includes phonemic awareness, phonics and decoding skills, and a strong language, literature, and comprehension program that offers a balance of oral and written language.
- A study commissioned by the committee and completed in 1997 revealed that as many as 40 percent of Idaho 4th graders in our schools are reading below grade level.

The Foundation's Reading Initiative design includes several components, which constitute a comprehensive approach for solving these problems, and the distinctive program strengths complement one another. The programs are the Waterford Early Reading Program, the Accelerated Reader Program, the Pesky Center Program, the Idaho State Libraries Program, and the Idaho Public Television Program. As a set, the programs teach reading directly in the classroom, encourage and assess classroom and leisure reading, train teachers and librarians, and provide education for parents to encourage their children's reading at home and at school. Each program has a previous record of success within or outside the

state or cotanins proven components of success. This final evaluation report concerns on the results for the program from beginning to end, that is, June 1, 1998 through May 31, 2001.

## B. Service Delivery

It may be helpful to provide an overview of the results before turning to the specific findings. Overall, it can be concluded that the vendors provided the agreed upon amount and quality of services, provided a variety of staff training, and enhanced students' reading achievement in accord with the Foundation's Ultimate Critical Results.

To assess service quality, two panels of Idaho educators provided insights from their distinctive vantage points.<sup>1</sup> As potential and actual program Idaho consumers, the User Panel read and helped interpret written and oral reports from the five program vendors. A Review Panel of expert reading authorities from Idaho universities, the State Department of Education, and other institutions within Idaho also raised such questions and provided suggestions and advice to improve field operations.

During the second and third years of the Reading Initiative, the panels met together. The combined panels raised constructive questions that were answered by the vendors. The questions and answers suggested many program and implementation improvements as well as continuing issues and concerns.<sup>2</sup>

The vendor reports and answers to written and oral questions from experts and educators on the ground convinced the panels and the evaluator that all five programs provided high-quality services. In addition, the programs are well founded on reading research and exemplify effective practices likely to raise reading achievement. Vendor incorporation of panel suggestions improved field operations. The quantitative information on the amount and quality of services provided is presented in the main body of this report.

---

<sup>1</sup> In educational terminology, an evaluation is a report of the assessment of the worth of a program, product, service, or means of organization. As discussed further in appendices, a meta-evaluation is an evaluation of the evaluation. In the present case, this meta-evaluation consists of the commissioning and analysis of vendor reports and the solicitation of expert reviews of the evidence and user perspectives based on their experience.

<sup>2</sup> Evaluation for improving a program is referred to as formative evaluation. The present report is focused on summative evaluation, specifically, the amount and quality of service delivery and the impact of the Reading Initiative on achievement

### C. Goal Attainment

Needless to say, actual reading achievement outcomes are also important, perhaps the most important. To estimate achievement progress, the evaluator requested and received considerable numerical data on achievement outcomes from the programs, which were reviewed with the Idaho User and Review Panels.

Analyses of achievement data on two programs show that they both raised children's reading levels substantially. The Waterford Program produced substantial gains in reading achievement, especially for children who were at risk for reading failure. The Accelerated Reader Program also yielded large reading gains. Teacher training, student perseverance, and program completion further enhanced the outcomes. Without these, the effects are comparatively smaller.<sup>3</sup>

Subsequent sections of this report explain the five programs in relation to the Foundation's objectives and document the quantitative findings. This report is a concise, pointed, and interpretive summary of a substantial amount of material collected to evaluate the program. The original plan for the evaluation, the vendors' program reports, summaries of the panel deliberations, sample program materials, and other materials are available in Foundation files. In addition, some of the most relevant material may be found in several appendices to this report

### III. Relation of Programs to the Foundation's Goals

The programs bear upon several of the Albertson Foundation's goals. Based on program materials and descriptions, user and expert panel discussions, vendor reports, and Albertson staff views, I rated the programs according to their relevance to the Foundation's goals or Ultimate Critical Results as shown in the table. Each program's relevance is indicated with one asterisk. Strong, more direct relevance is marked with two asterisks.

---

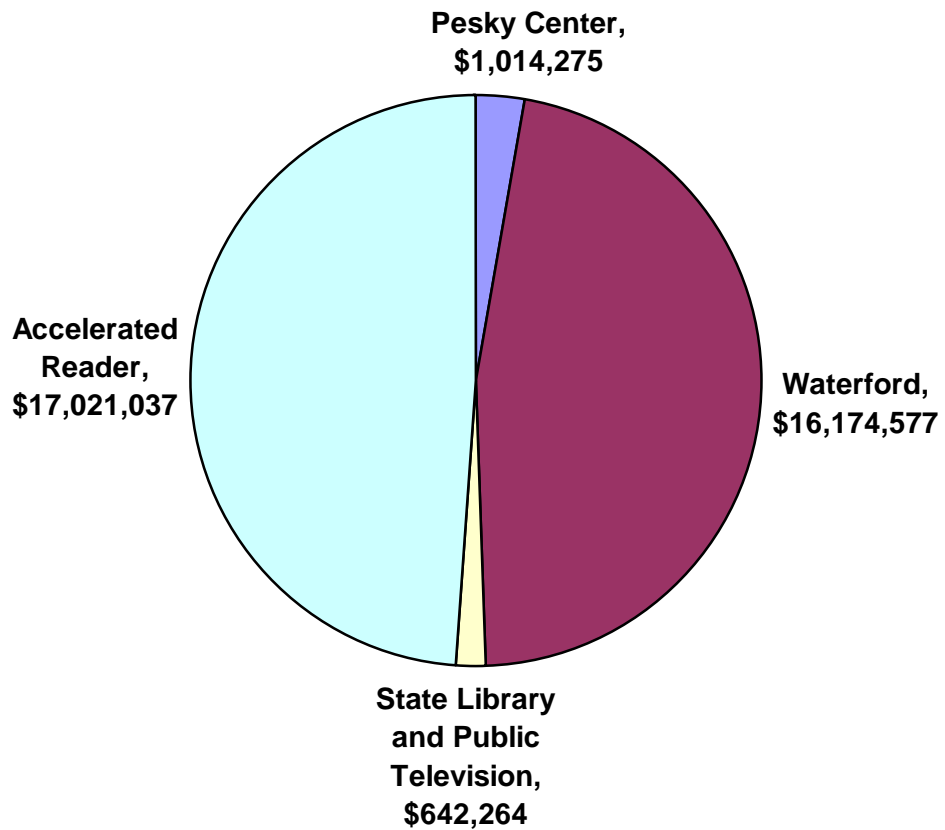
<sup>3</sup> Amount and quality of service information but not achievement outcome data were available for the Pesky Center, Idaho Public Television, and Idaho Public Library programs.

Program	Readiness to Learn	Student Achievement	Educator Competence	Stakeholder Satisfaction
Waterford Early Reading	**	**	*	*
Accelerated Reader	*	**	**	*
Pesky Center			**	
State Library	*			**
Public Television	*			**

As a set, the programs provide a balanced approach to the Foundation’s Ultimate Critical Results. Waterford and Accelerated Reader Programs both directly emphasize achievement. Waterford is more focused on beginning reading, Accelerated Reader on continuing achievement and teacher training, which enhance the program’s effects. The Pesky Center program focuses on educators’ competence rather than student achievement. The State Library and Public Television programs emphasize services to the ultimate stakeholders, parents, and other adult citizens.

#### IV. Program Costs

For the three years of the Reading Initiative, the Foundation's investment in the Reading Initiative programs was \$34,852,163. The proportional allocations of this expenditure are shown below in the chart.



## V. Waterford Outcome Results

The achievement outcome results for the Waterford Early Reading Program are striking. For the present evaluation, eight school districts were chosen to be representative of the state in terms of geographic location, size, and socioeconomic status. The eight districts were Blackfoot, Coeur d'Alene, Kuna, Marsing, Nampa, Twin Falls, West Jefferson, and Whitepine. Students tested in the same grades during the previous year that were unexposed to the Waterford program served as a "historical control group."

As shown in the graph, the effects of the Waterford program are impressive. Its effects on 2,414 students in the Idaho districts compare favorably with the most effective of some 274 teaching methods and conditions in research literature.<sup>4</sup> The program produced the best results for students that started out in the lower third of achievement and completed 100 percent of the program. Separate results for 1) all students starting in the lower third and 2) all students who completed the program gained about half as much, although still more on average than revealed by tutoring studies in the research literature. The effect on all students is about .20 or one-fifth the largest effect for lower-third, 100% completers; it is, however, twice as large as expensive class size reductions.

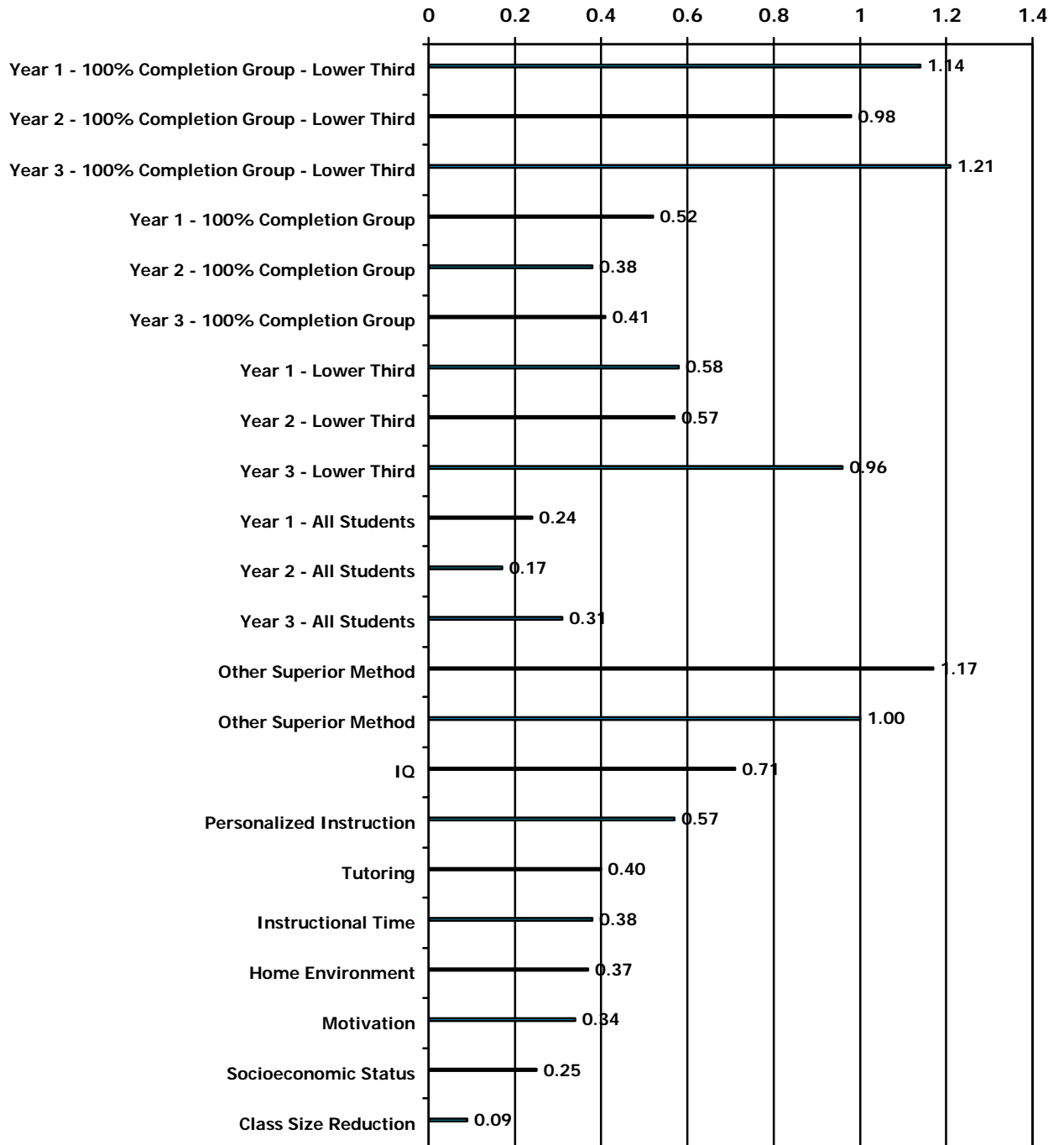
Thus, the Waterford program also appears spectacularly effective for beginning readers who initially scored in the lower third of the group when they began to learn to read. It is also remarkably effective for students either in the lower third or who completed the entire program. For these groups, Waterford effects, for example, were comparable or superior to tutoring, increasing instructional time, increasing the academically stimulating features of the home environment, and class size reduction.<sup>5</sup> When the scores of all the students including those who completed and did not complete the Waterford program were compared with a control group, the program also had highly worthwhile effects.

---

<sup>4</sup> Scale: The effects in the subsequent chart are expressed in "effect sizes" or standard deviation units, which are roughly the same as a grade equivalent or the normal progress an average student would be expected to make in one year. In the present case, the effect is the additional progress on Waterford test beyond what would be expected from normal progress.

<sup>5</sup> "Other superior methods" come from my coauthored chapter Herbert J. Walberg and Jin-Shei Lai, "Meta-Analytic Effects for Policy" in Gregory J. Cizek (Editor), *Handbook of Educational Policy* (San Diego, CA. Academic Press, 1999), pages. 419-454.

# Effects of the Waterford Program and Other Alternatives



Two minor testing ambiguities slightly cloud the Waterford results. First, the reading tests used to measure the effects may have been insufficiently difficult to show the full effects of the program on readers who began in the middle and upper thirds. Thus, the effects may be even larger than estimated. A second ambiguity is that Waterford employed its own tests, which probably biased the results in the other direction. Even so, other evaluations of Waterford show favorable effects on both Waterford and other tests. In this evaluator's view, reflecting also the Panel members, the program had truly excellent results.

As originally planned, it was important to study the program over a full three-year period. The graph shows fairly consistent results for all three years of the program in the sense that the lower-third, 100 % completers consistently made the best gains, the students without both these characteristics made least but still respectable progress, and the other groups were intermediate. Many faddish programs have a initial "Hawthorne" or "hothouse" effect but quickly fade in their effectiveness. Even with diligent, continuing effort, other programs may take a few years to attain their best results. The evaluation results show that Waterford maintained consistently excellent results during the three-year period.

## VI. Accelerated Reader Outcome Results

The effects of the Accelerated Reader Program are also impressive as revealed in analyses of some 21,534 students in 76 Idaho schools. Across all grades, more than half the students averaged better than 85 percent correct on objective tests about the books they read, as shown in the last row of the table. The average student read 38 books and spent 22 minutes a day reading, which was a 127 percent increase in the average time students at any given grade level spend reading. Still, the Accelerated Reader staff recommends 60 minutes of reading per day.

Grade	Average Number of Books Read	Estimated Minutes of Daily AR Reading	Percentage Gain in Reading Time	Percent of Students Averaging 85% Correct
1	43	17	240	73
2	67	23	191	66
3	56	24	140	57
4	54	29	127	52
5	39	24	88	52
6	18	22	108	50
7	13	16	93	49
8	12	14	84	45
9	8	8	122	44
All	38	22	127	54

The table shows a “grade effect.” Students in the lower grades read more books, gained more reading time, and achieved better on their tests. This grade fall off may be attributable to the tendency of early grade teachers to concentrate on reading in the early grades rather than, say, science and social studies. The effect may also be related to staff training concentrated in the early grades, changes in the reading skills required in advanced texts, or declining motivation to participate in the program as students grow older.

An interim evaluation showed that success in the program depends on the amount of staff training. The chart below is taken from that report. The results are expressed as extra units of estimated yearly growth.<sup>6</sup> The Accelerated Reader groups with various stages of implementation were compared with a national control-group sample. Using the program alone with little staff training raised achievement slightly (about .4 of one unit per year) on average.

The extra growth was substantial (about 1.6 units) if at least five school staff members were trained in Accelerated Reader use, and more than doubled (3.9) if the school had at least one classroom that was Renaissance certified (an indication of exemplary Accelerated Reader use). As in the case of the Waterford program, specific teacher training and student completion of the program are keys to large effects. The effects of the program were stronger in grades 1 through 4, and an annual growth of 4.0 extra units per year would enable poor children to read at the level of middle-class children in roughly five years.

Data from the last year of the project showed similar effects. Continuing training seems necessary for good effects.

**Effects of the Accelerated Reader Program**

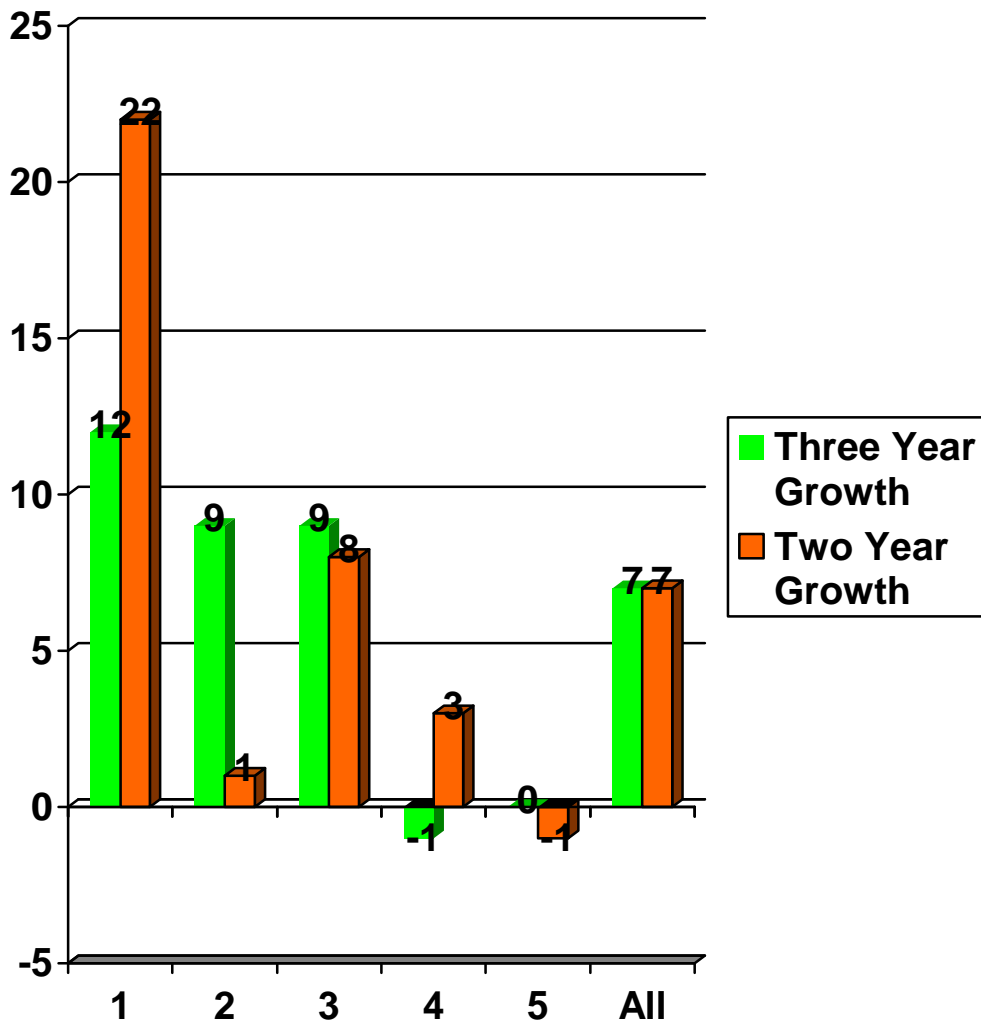


---

<sup>6</sup> The effect is expressed in the metric of normal curve equivalents, which are similar to percentiles in having a mean of 50 and a range from 0 to 100 but different in that the units between points on the scale are equal. The effect is the extra gain beyond expected normal progress made by the average students.

The continuing evaluation allowed estimates of the longer-term effects of the Accelerated Reader program. During last year of the project, data were available for 401 students in the program for three years and 1168 students for two years who started at various grade levels shown in the table. These are much smaller numbers of students than in the other analyses, but, the results, nonetheless, are striking. The program appears particularly effective when children begin in the early grades, especially the first grade. Those who enter after third grade merely make normal progress, that is, keep their place in national percentiles.

## Two and Three Year Growth Effects of Accelerated Reader



## VII. Service Delivery

### A. Pesky Center

At the request of the Pesky Center, Clearwater Research, Inc. of Boise interviewed and surveyed 1,100 recipients of Pesky Center training to determine the effectiveness of the mentor-training sessions. The program is focused on teaching techniques that solve the difficulties many children have in learning to read. Two-day training sessions provide mentors practice in teaching teachers, a tool kit of teaching resources, and a plan for communicating and mentoring when the participants return to their communities.

The evaluation showed that 78 percent of the teachers throughout the state found the training useful. Those who collaborated with colleagues and received administrative support or released time were more likely to incorporate new literacy teaching methods in their classroom. Trained teachers performed better than others without training on a knowledge test.

Overall, 92 percent of the mentor teachers found the training useful. The mentors were most likely to incorporate new reading practices in their classrooms, gather or develop additional reading materials, and participate in university coursework on reading instruction. The mentors performed better than teachers who had and did not have literacy training.

Pesky staff presentations, panel comments, and review of the course materials suggest that the Pesky program delivered valuable training that informed and stimulated teachers. The ideas presented to teachers were generally state-of-the-art, and the practices described and encouraged are known to promote reading achievement.

### B. Idaho State Library

The Library program provided three years of Begin with Books training, a two-day workshop on beginning literacy and on how librarians and their partners can promote early literacy activities. These professionals reached at least 2,160 parents and caregivers. The local libraries conducted a variety of outreach activities to help parents understand the importance of reading aloud to children and encouraging their leisure-time reading. Independent surveys of the program participants showed that 52 percent of the participating families read daily to their children. The program also encouraged library use. The survey showed that 75 percent of the families had obtained a library card, and that library usage increased. In in-

interviews, nearly all the parents agreed that reading aloud to their children will later help their reading and writing. Other Albertson-sponsored State Library training reached out to the growing Hispanic population and at-risk families.

The State Library also acquired one of the strongest early childhood video collections in the Pacific Northwest. It makes available some 300 video cassettes to any interested Idahoan including parent educators, teachers, trainers, librarians, and parents.

The grant also allowed the State Library to help five local libraries create literacy outreach programs for pre-schoolers. Twenty Idaho libraries have now followed their lead and acquired one-year federal and state funding to start similar programs. In extending its reach as a result of the grant, the State Library now works Idaho Public Television, the Idaho Commission on Hispanic Affairs, the Idaho Department of Education, the Head Start Association, and other organizations to advance reading and literacy.

### C. Idaho Public Television

With respect to quality of service and outreach, the program was well received by the User and Review Panels. The numbers on the program are also remarkable: During the three years of the project, Public Television presented 196 workshops, served 2368 parents and 9594 children, and distributed 26,229 books throughout the state. Even more impressively, the annual figures generally show substantial increases from year to year.

Public Television staff developed and distributed a monthly newsletter "Connections" for program participants. This publication links Ready-to-Learn television programming to specific books and literacy activities.

Public Television staffers carried out and commissioned questionnaire research to improve the program. Among other findings, the workshops were rated Excellent or Very Good by 93 percent of the participants; 77 percent indicated that they would change the way they use television in their homes.

## VIII. Conclusions

The Reading Initiative incorporates a balanced and comprehensive set of five innovative programs that either has been proven effective or that incorporate features that research shows to be effective. The Waterford and Accelerated Reader programs feature groundbreak-

ing computer technologies to directly encourage, guide, and assess reading progress. The Pesky, State Library, and Public Television programs exemplify excellent public-private statewide cooperation.

The program vendors delivered extensive, high-quality services. Although Idaho Review and User Panels raised constructive questions and suggestions about the programs, they view the design and operation of the programs in Idaho very favorably. The program vendors have carefully considered the suggestions and incorporated many of them into their Idaho operations.

The achievement results are sufficiently large to rule out the statistical possibility that they had a trivial or modest effect. In fact, the results of the Waterford program may have actually been larger than estimated since the test appears to have been insufficiently challenging to show larger effects.

Both the Waterford and Accelerated Reader success, however, depend in part on the degree of teacher expertise and the amount of time and effort that students put into the programs. The Waterford results are based on Waterford tests. Even so, previous studies by Waterford and independent groups show the program yields similar positive results on both national standardized examinations and Waterford tests.

Previous research on Accelerated Reader as well as Idaho results bear out its effectiveness in promoting achievement on a variety of tests. Since both the Accelerated Reader and Waterford programs depend on teacher expertise and student effort, it is reasonable to ask how these can be enhanced, an interesting question for future research and evaluation.

The balance of the programs in the Reading Initiative should be considered an asset. If a dollar value or other utility index could be assigned to each unit served, it would be possible to allocate funds to the programs partly or wholly based on their utility/cost ratios. Unfortunately, that is largely impossible in educational program evaluation. The larger funding allocations to the Waterford and Accelerated Reader programs, however, may be justified in that they directly serve students and provide more solid indications of outcome effects. Still, State Library, and Public Television programs serve complementary roles since they tend to reach “unreachable” families, and the Pesky program serves many students indirectly by affecting teachers.

Finally, the Albertson Foundation board and staff can take considerable pride in beginning and sustaining a bold experiment, one that addresses one of the most critical issues of our times—effectively teaching and encouraging children to read so that they may succeed in school and subsequent study and in their lives as citizens. The Reading Initiative, moreover, is fully consonant with and promotes Albertson Foundation

Ultimate Critical Results with respect to readiness to learn, achievement outcomes, educator competence, and stakeholder satisfaction.

In my view, foundations can uniquely carry out the initial activities but cannot or should not compete with government funding sources for routine support of proven programs for public institutions including schools. For this reason, the Foundation board and staff may wish consider what Reading Initiative activities remain to be funded and how the successes of the program might be expanded. Given the general success of the Reading Initiative, however, the legislature, school boards, and other public agencies may have chief responsibility for further support and expansion.

## IX. Appendices

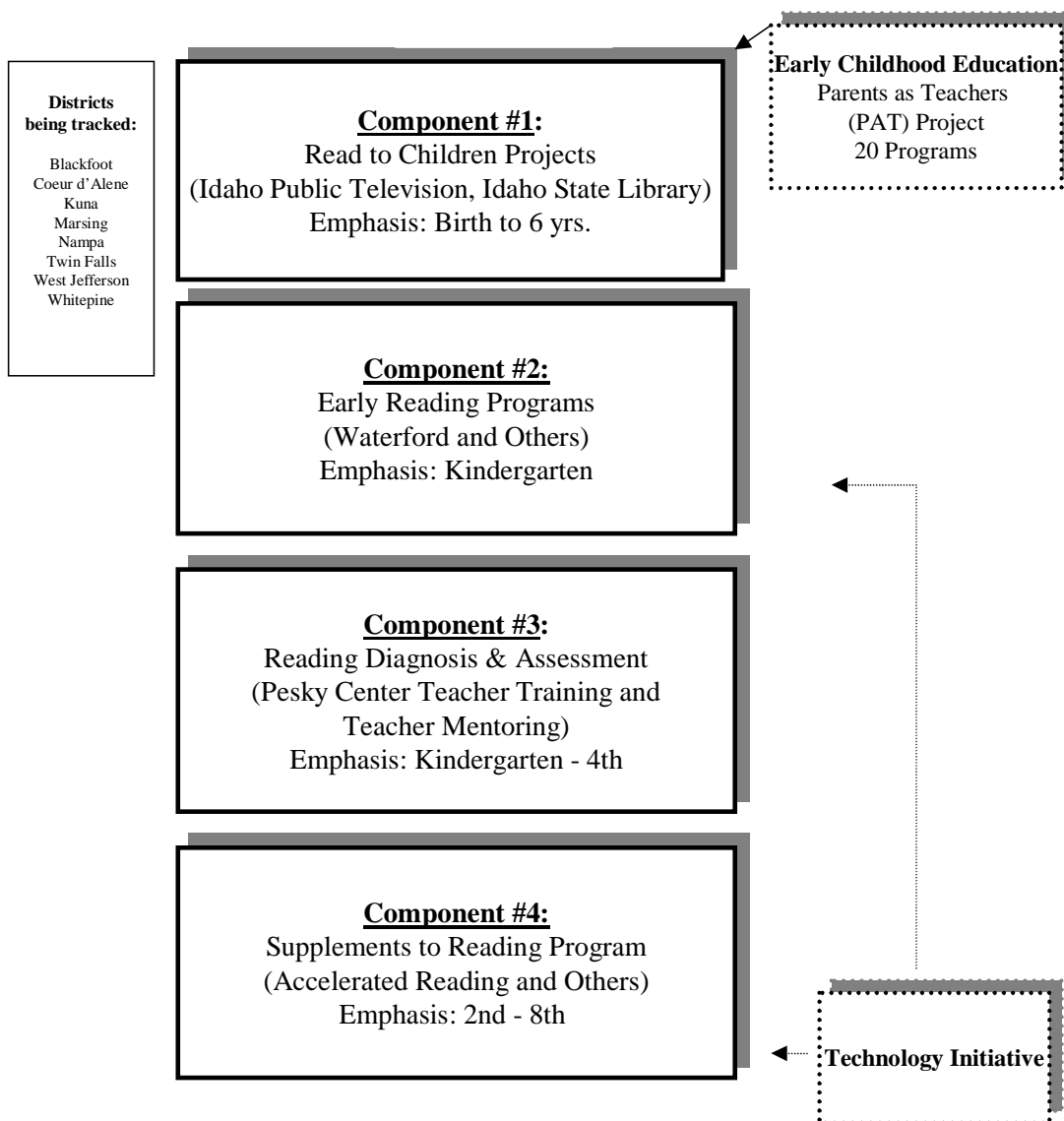
A. The Components of the Reading Initiative

J.A. & Kathryn Albertson Foundation

Reading Initiative

**Overall Purpose:** To improve reading performance of students by 4th grade.

**Timeline 1998 to 2001**



## B. Program Descriptions

### 1. The Waterford Early Reading Program

This program, though new, is based on sound research and advice from several of the nation's foremost authorities on reading. It relies chiefly on a sophisticated computer-assisted tutoring of individual children. The computer system employs sound, color, and animation to teach children to read. It begins with sound and letter recognition and, on this fundamental base, builds other reading skills including comprehension. The program captivates children's interest, and adults find its many features attractive. Children work in the same room but independently of one another. The computer program has many routes to learn the same essentials, and progress reports are readily available on the class or individual children so that teachers can monitor and guide their progress.

### 2. Accelerated Reader

Accelerated Reader assumes the well documented and obvious but often-unheeded idea that the more children read, the better they read. Reading material, moreover, that appeals to children and that appropriately challenges them promotes their reading ability most efficiently. The developers have categorized some 30,000 children's books according to their reading difficulty and children's interests, such as sports and history. Their computer program suggests to teachers which books would be the best choices for each student and then assesses with computer-administered questions each student's progress on books they have taken out of the school or community library. Teachers and students can trace progress on a point system based on the volume and difficulty of books read.

### 3. Pesky Center Program

Also based on sound research about best practice, the Pesky program Every Child Can Read is a course provided to teachers throughout Idaho that provides a foundation of understanding about how children learn to read. Teachers learn reading assessment and intervention tools to improve their reading instruction. In addition, mentors receive training to inspire and guide other teachers in best reading instruction practice.

### 4. Idaho State Library

This program enables local library staff to encourage family reading and library use. It provides children with a book a month for a year and gives literacy training to librarians and human service workers. It also distributes a catalog of early childhood videos to librarians

and social service workers and provides traveling displays and pamphlets about stimulating children's reading.

#### 5. Idaho Public Television

This program provides books for underprivileged children each month. The Program also presents workshops for parents and childcare workers, and developed and distributes "Connections," a monthly newsletter and other materials.

### C. Joint Panel Recommendations

The User Panel and Review Panel consisted of prominent university reading scholars and leading educators in Idaho. Based on written and oral reports by the vendors, extensive discussions with them, and their scholarship and field experience, the Panel members devoted much of their last two meetings in small groups to formulating conclusions and next-step recommendations. In summary, recommendations relevant to the present summative evaluation are as follows:

- The Reading Initiative has had far-reaching impact on Idaho public schools. Its components were generally effective, complemented one another, and provided a comprehensive approach to reading improvement in the early grades. The evaluation results support expansion of the components.
- The Reading Initiative provided an integrated network of collaborative partnerships focused on the critical reading needs of the early childhood years. It was based on and sustained by Albertson Foundation support. It encompassed parents, school administrators, boards, teachers, and librarians who shared knowledge, responsibility, and accountability. The evaluation showed the importance of training, partnerships, support, and leadership.
- The Waterford program and Accelerated Reader showed positive effects on reading achievement; they are both effective and efficient. The success of both programs depends on staff training, full implementation, and student completion of the programs. The results provide the evidentiary basis for increased funding, possibly by the legislature, which should be informed of the important Reading Initiative accomplishments.
- The Idaho Public Television and Idaho State Library programs as well as the Waterford program provided the means for children to develop their reading skills at home, although the Television and Library programs, given their budgets, were unable to supply the optimum number of books for the state nor sustain the involvement of all parents.

- , To improve reading achievement, the Pesky Center addressed the need for teacher training, not necessarily tied to a particular program such as Waterford. The evaluation suggested positive results of such training on teacher knowledge and attitudes.
- Three years of data and study sufficed to evaluate the programs, but, in some cases, the programs were compromised since the staff members were not fully trained and the programs were not fully implemented. Variations in training and implementation, however, allowed the evaluation to measure the substantial effects of training and implementation.
- One group recommended and hoped that the Albertson Foundation will maintain the focus on early literacy in Idaho. The Initiative set the targets of full implementation and training. The energies and resources of the state and districts should be employed to hit the targets. In addition, since the evaluation and other research show the importance of early verbal experience, pre-school and full-day kindergarten should be provided statewide.
- To sustain the gains requires continuing staff training, and the finance and support of such groups as the Idaho legislature, Department of Education, city leaders, and Parent Teacher Associations is required.

## D. Evaluation Design Features

In assessing the Albertson Foundation's Reading Initiative, we would like to have the most telling information on its success while at the same time avoiding extra testing of students, information burdens on schools and program providers, and other problems. We want the most rigorous studies possible but we also want the insights of university scholars and classroom and central-office educators. These desires mean we must make difficult trade-offs among various considerations of scientific validity and practical feasibility.

### 1. Design Features

My purpose here is to make the alternatives and trade-offs clear and concise. My intended readers are Albertson Foundation staff, their grantees, namely, the Academic Excellence Institute, Idaho Public Television, Idaho State Library Association, the Lee David Pesky Center, and the Waterford Institute, and representative educators in Idaho. Let us consider design alternatives that are particularly important for the evaluation of the Reading Initiative.

### 2. Experiment

Ideally but unrealistically, we would assign students with a flip of a coin to program and control groups. As in rigorous medical and agricultural research, this is called a "true experiment," and its strength is that there is no reason other than the program and small random variations that the groups differ. Of the thousands of education evaluations, however, only a few have had this feature because children, teachers, and schools cannot be easily assigned in this way. Experiments, moreover, are hardly foolproof. They may tell us how educators and students do in artificially contrived circumstances rather than in ordinary circumstances.

### 3. Quasi-Experiment

An alternative is a quasi-experiment (jokingly also called a queasy-experiment) in which non-randomly assigned program and control groups are compared. But they may have initially differed. Better-informed administrators and teachers or those in districts of higher socioeconomic status, for example, may have chosen to use the new program. Thus, apparent program superiority may be attributable to factors other than the program. For similar reasons, a high success rate of a new form of by-pass surgery on highly educated patients in university hospitals leaves open several questions.

#### 4. Statistical Control

Quasi-experiments may be strengthened substantially by employing statistical controls for initial status and conditions of the groups. Since achievement is usually the chief variable of interest in education evaluations, pretest achievement measures are the most valuable statistical controls. They allow comparisons of the gains from the start and to the end of the program or one of its phases. Sometimes, however, pretests are difficult to obtain since students move and classroom groups do not remain intact.

#### 5. Pre-Post Designs without Control Groups

Evaluators often confront situations in which a control group is lacking. Perhaps all available students are in the program group leaving no students for the control group; educators may have been reluctant to deny to some students what it is believed to be a superior program. At any rate, the gains may be measured over a given time period and compared with what might have been expected. For example, over one chronological year from, say, from May to May a one grade-equivalent gain might be expected.

Pre-Post Designs often raise additional questions having to do with how to measure achievement, skill, and other possible outcomes. They concern the breadth and representativeness of the measures. Program developers, for example, may have very specific objectives for their programs. Accordingly and reasonably, they design achievement tests and other assessments to measure the degree to which they have accomplished their particular objectives. The education consumer, however, may want to know how students do on tests other than the developer's, tests that may measure more conventional or general objectives. Two design features may help satisfy their interests.

#### 6. Nationally Standardized Tests

The Stanford, Iowa, and Metropolitan tests are examples of widely used tests employed to assess students, schools, districts, and programs. Though less than perfect, they are employed throughout the United States to assess achievement in reading, mathematics, science, and other subjects. Their manuals contain information on their scope, validity, and reliability. Their "norming" is based on national random samples, which allows expressions of scores such as percentiles in relation to all U.S. students rather than a local, convenience, or "grab" sample. Such scores are crude approximations to physical measures: A mile is the same distance in Idaho and Maine; a grade equivalent gain on a well-standardized test has similar universality.

## 7. Calibrated Developer Tests

We may face a situation for which scores on developer tests are available but not for nationally standardized tests. For example, it is my initial and unverified understanding that some Idaho districts employ ITBS tests in October; so, the accomplishments through the 1998-99 school year cannot be assessed until then if the evaluation were to be restricted to these tests. In such cases, we may have to rely on developer tests alone at least for the initial evaluation. In this case, it will be highly desirable to show that the test correlates substantially with nationally standardized tests. In addition, the tests should be calibrated so that the scores can be expressed on national scales such as grade equivalents.

## 8. Effect Sizes

For many years, evaluators sought to determine whether the program groups scored higher than control groups on outcome measures. It is more informative to know how much better. Effect sizes express this difference.<sup>7</sup> Effect sizes allow rough comparisons of program effects under different conditions and with various samples of students even though different tests may have been used. In addition to effect sizes for the Reading Initiative, either developer may calculate effect sizes from sites other than Idaho schools. These can provide context for the first year's findings. If some preliminary findings in Idaho are somewhat tentative or equivocal, they may be buttressed by knowing of good effects elsewhere.

## 9. Meta-Evaluation

Recognizing the advantages, disadvantages, and feasibility of the above design considerations, the developers can provide the best information possible. A ten-page summary and an appendix no more than 20 pages should best serve). The summary will be shared with non-technical educators as explained below. The appendix might provide more explicit evidence and its basis, summaries of reports of previous and related studies, and other material.

It will be chiefly my job to meta-evaluate or evaluate the evaluations. The chief technical and judgmental principles to be applied are covered in the seven points above. Depend-

---

<sup>7</sup> A conventional way of calculating effect sizes is to divide the difference between program and control group means by the standard deviation of the control group. For pre-post designs, the gain may be divided by the pretest standard deviation. The underlying scores should be raw scores, normal curve equivalents, or other equal interval expression.

ing on the data and other information provided in the evaluations, it will be possible to state some conclusions with confidence, and others with some uncertainty since they will be subject to alternate explanations (as in the by-pass example above). The effect sizes can be compared with others I have compiled to provide perspective on educational programs and methods.

#### 10. Review Panel

None of the design considerations above will yield direct, independent insights from reading scholarship, the classroom, and the central office. Nor will these considerations reveal how the two programs and other Albertson-sponsored programs are or could work together. In considering these and related matters, therefore, an Idaho review panel should add considerable value to the overall evaluation.

Preliminarily for discussion, I envision that each member of the panel would be supplied with ten pages about each of programs from each developer. To facilitate comparison and the possibility of mutually enhancements among programs, each reports should cover basic and evaluative information, specifically, the grade levels and numbers of students reached in Idaho and elsewhere, managerial and operational procedures, teacher preparation, number and length of sessions, and, most important, achievement results.

As implied above, the purpose of the panel is to help evaluate the progress of the Reading Initiative. The scholars can bring to bear their expertise in reading; the school staff, their direct experience with the programs. More specifically, the panel can be expected to yield summative and formative insights. How well are the Reading Initiative and its components working? How can they work better? They can also consider such questions as: How can the components complement rather than duplicate one another? What aspects of reading promotion may be left out of the Initiative or one of the components? Would university programs or school staff development bring considerable added value to the overall effort?

#### 11. User Panel

In addition to the Review Panel, a User Panel of representative educators (probably from eight districts) who are actively using several or all of the programs will be convened. Among other purposes, they will be informed about data needs for the evaluation. To complement the quantitative analysis, they will provide qualitative information from their experiences in employing the programs and insights about how the programs work together and possible gaps that remain to be filled in providing the very best reading instruction and optimal experience for Idaho children and youth.

## 12. Final Report

The Reading Initiative has the potential to be one of the most important and exciting developments in the promotion of reading skills in the history of American education. In addition, the Albertson board should have an overall report on what has been accomplished. Therefore, at the end of a second round of evaluation, an overall summary of its goals, operations, and results should be submitted.

### E. The Development of Vocabulary and Reading Skill

Small variations and changes in the verbally stimulating qualities of the children's home environments can make for large differences in their reading progress. During the first 18 years of life, children spend only about 13 percent of their waking hours in school and about 87 percent outside of school, particularly in their homes during the first six years of life. For these reasons, the origins of poor reading progress can often be traced to specific parental behaviors observed before and after children begin school that affect children's reading skills. Sticht and James (1984) pointed out that children first develop vocabulary and comprehension skills by listening, particularly to their parents before they begin school. As they gain experience with written language between the first and seventh grades, their reading ability gradually rises to the level of their listening ability. Highly skilled listeners in kindergarten make faster reading progress in the later grades, which leads to a growing ability gap between initially skilled and unskilled readers.

This growing gap between good and poor readers reflects race and social class differences. Writing in the authoritative *Handbook of Reading Research*, Wigfield and Asher concluded:

“The problems of race and socioeconomic status (SES) differences in achievement have been at center stage in educational research for nearly three decades. Research has clearly demonstrated that such differences exist; black children experience more difficulty with reading than white children, and the discrepancy increases across the school years. Similarly, children from lower SES homes perform less well than children from middle-class homes, and here too the difference increases over age” (p. 423).

As Wigfield and Asher point out, these differences stem from early childhood experience, especially with respect to parent behaviors that motivate children. They summarize studies that show that middle-class parents are more likely to hold high expectations for their children's achievement and to be more often engaged with them in promoting it. As they further conclude about learning situations:

“Lower-SES mothers provide their children with poorer problem-solving strategies, and they tend to take over for their children rather than letting them do the task . . . That lower-SES parents view school as a distant, rather formidable institution over which they have little control; engage in less effective teaching strategies; and lack confidence in their children’s ability does not bode well for their children’s school performance” (p. 429).

Home observations and interviews with parents reveal further SES differences among parents associated with higher achievement in reading, for example, “responsivity of the parent, the kinds of discipline techniques used, the organization of the physical environment, parental involvement, and provision of appropriate play materials” (pp. 431-432).

Such parent behaviors cause huge and growing gaps in children’s preparation for school and learning to read. Hart and Risley (1995) showed this graphically in preschool children’s vocabulary growth recorded during free play. Though vocabulary differences were miniscule at 12 to 14 months of age, by age three, sharp differences had emerged: African-American welfare children had vocabularies of about 500 words, middle/lower SES children about 700, and mostly white higher SES children had vocabularies of about 1100 words—more than twice that of welfare children (p. 47).

Such SES differences in vocabulary were strongly associated with parent behaviors exhibited in their homes. Higher SES parents spent more minutes per hour interacting with their children and spoke to them more frequently. On average, higher SES parents spoke about 2,000 words per hour to their children, African-American welfare parents, only about 500 (p. 68). By age four,

“An average child in a professional family would have accumulated experience with almost 45 million words, an average child in a working-class family would have accumulated experience with 26 million words, and an average child in a welfare family with 13 million words” (p. 198).

Higher SES parents, moreover, used “more different words, more multclause sentences, more past and future verb tenses, more declaratives, and more questions of all kinds. The professional parents also gave their children more affirmative feedback and responded to them more often each hour they were together” (pp. 123-124). Todd and Risley estimated that by age four, professional parents encouraged their children

with positive feedback 750,000 times, about six times as often as did African-American, welfare parents. The welfare parents, on the other hand, had discouraged their children with negative feedback about 275,000 times, about 2 1/2 times the amount employed by professional parents (p. 200). Such parenting behaviors predicted about 60 percent of the variation in vocabulary growth and use of three-year olds.”

Anderson, Wilson, and Fielding (1986) in a study of daily time logs of fifth graders showed large differences in amounts of daily reading among children. The number of minutes of reading varied from 0 to 71 minutes per day (2nd and 98th percentiles) with a median of 11.1 minutes; the number of words read per year at leisure varied from 8 to 4,700,000 with a median of 601,000. These enormous variations in reading, of course, lead to large differences in children’s’ vocabularies and comprehension abilities.

Time and practice, however, can compensate for, or remedy initial disadvantages. Given sufficient exposure or instruction, children and even adults can make rapid progress in learning difficult languages.

“Immersion’ methods, for example, can bring about striking rates of acquisition. The Army Language School set 1,300 hours of instruction as a standard for an adult to achieve near native competence in Vietnamese. A child who spends about ten hours a day in school, in play, and with media in English might gain comparable, though seemingly natural and effortless, experience in 130 days” (Walberg, Hase, and Rasher, 1978, p. 428)

## References

Anderson, R.C., Wilson, P.T., and Fielding, L.G. (1986) Growth in reading and how children spend their time outside of school. Urbana, IL.: University of Illinois Center for the Study of Reading.

Hart, B. and. Risley, T. R. (1995) Meaningful Differences in the Everyday Experience of Young American Children. Baltimore, MD.: Paul H. Brooks Publishing.

Sticht T. G. and James H. James, "Listening and Reading." In P. David Pearson (Ed.) Handbook of Reading Research. New York: Longman, 1984.

Walberg, H.J., Hase, K., & Rasher, S.P. (1978) "English mastery as a diminishing function of time, TESOL Quarterly, 12 (4), 427-437.

Wigfield, A. and. Asher, S. R. "Social and Motivational Influences on Reading. In P. David Pearson (Ed.) Handbook of Reading Research. New York: Longman, 1984.

## F. Bio for Herb Walberg

Herb is Emeritus Research Professor of Education and Psychology at the University of Illinois at Chicago. He is also Distinguished Visiting Fellow at the Hoover Institution at Stanford University. Holding a Ph.D. from the University of Chicago and formerly Assistant Professor at Harvard University, he has written and edited more than 55 books and written about 350 articles on such topics as educational effectiveness and exceptional human accomplishments. Among his latest books are the International Encyclopedia of Educational Evaluation, and Psychology and Educational Practice.

Herb served one the National Assessment Governing Board, referred to in the U.S. as "the national school board" given its mission to set education standards for U.S. students and measure progress in achieving them. The National Assessment provides information on changes in educational achievement in the U.S. as well as comparisons of individual states. A fellow of four academic organizations including the American Association for the Advancement of Science, American Psychological Association, and the Royal Statistical Society, Walberg is also a founding fellow and Vice-President of the International Academy of Education, headquartered in Brussels. He edits for the Academy a booklet series on effective educational practices, which is distributed by the International Bureau of Education to some 4,000 education officials in more than 150 developing countries

Walberg has given invited lectures in Australia, Belgium, China, England, France, Germany, Israel, Japan, the Netherlands South Africa, Sweden, Taiwan, Venezuela, and the U.S to educators and policy makers. He has frequently testified before U.S. Congressional committees, state legislators, and federal courts.

In his research, Walberg employs experiments and analyses of large national and international data sets to discover the factors in homes, schools, and communities that promote learning and other human accomplishments. He also employs research synthesis to summarize effects of various educational conditions and methods on learning and other outcomes, the results of which have important bearings on education policy and practice.

For the U.S. Department of Education and the National Science Foundation, he carried out comparative research in Japanese and American schools. For the U.S. Department of State and the White House, he organized a radio series and book about American education distributed in 74 countries. Walberg chaired the Scientific Advisory Group for the Paris-based Organization for Economic Cooperation and Development project on international educational indicators. He also advised UNESCO and the governments of Israel, Japan, Sweden, and the U.K. on education research and policy. He advised several states including

California and Virginia on standards and is a member of the National Academy of Sciences Committee on Educational Excellence and Testing Equity.

Currently Walberg chairs the board of the Chicago's Heartland Institute (web address: [www.heartland.org](http://www.heartland.org))--a think tank that provides policy analysis for the U.S. Congress, state legislators, talk show hosts, news reporters, and the public. He advises the Albertson Foundation—the second largest U.S. foundation concentrated on education. He serves on a committee that selects students with near perfect university admission scores for merit scholarships at top universities. He is a board member of one of the largest charter schools in the United States and a trustee of the California-based Foundation for Teaching Economics. Further details are available in “Who's Who in America,” “Who's Who in Medicine and Health Care,” and “Who's Who in the World.”

