DOCUMENT RESUME

ED 373 456 EA 026 071

AUTHOR Ross, Steven M.; And Others
TITLE Using "Success for All" To Restructure Elementary Schools: A Tale of Four Cities.
PUB DATE Apr 94
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; At Risk Persons; Early Childhood Education; *Early Intervention; Grade 1; Grade 2; Grade 3; Kindergarten; *Low Achievement; Prevention; Primary Education; *Program Evaluation; *Reading Ability; *Reading Achievement
IDENTIFIERS *Success for All Program

ABSTRACT

The Success for All (SAF) program was designed as a K-3 preventative and intensive intervention for students at risk of failing to learn to read for the purpose of bringing every student to grade level in reading by third grade. Key program elements include individualized tutoring, regrouping across grades into homogeneous reading classes, smaller reading classes, family support, and a comprehensive reading program. This paper presents findings of a study that examined the effects of the SFA program in four cities on student reading achievement. The sites included Memphis, Tennessee; Montgomery, Alabama; Fort Wayne, Indiana; and Caldwell, Idaho. Specifically, the study sought to assess student achievement when the SFA program is introduced in settings geographically removed from program developers. A comparison of individual reading test results with those of matched control groups indicated advantages at three of the sites, with particular advantages for the lowest achieving 25 percent of students relative to their control counterparts at all four sites. Results also indicate that the SFA program can be replicated at distant locations with limited monitoring by program developers. Although there was variability in implementation quality and achievement outcomes, SFA has the following advantages: systematic procedures, comprehensive components, facilitator leadership, monitoring components, and extensive and ongoing staff development. Six figures and one table are included. (LMI)

Using "Success For All" to Restructure Elementary Schools: A Tale of Four Cities

Steven M. Ross
Lana J. Smith
Jason Casey
Brenda Johnson
Carole Bond

Memphis State University

Running head: FOUR CITIES

A paper presented at the annual meeting of the American Educational Research Association
April, 1994
New Orleans, LA

Abstract

The Success for All (SFA) program is designed to bring every student to grade level in reading by third grade. Key elements include individualized tutoring, regrouping across grades into homogeneous reading classes, smaller reading classes, family support, and a comprehensive reading program. In the present study, influences of the program on reading achievement of elementary students from schools in four cities were examined. Individual reading test results indicated advantages in three of the sites, with particular advantages for the lowest-achieving 25% relative to their control counterparts at all four sites. Results also indicate that the SFA program can be implemented in sites geographically removed from the developers and apart from their direct supervision.

Using "Success For All" to Restructure Elementary Schools: A Tale of Four Cities

Charles Dickens introduced his well-known tale of two cities with an oxymoron informing readers that the 1700's were the "best of times" and the "worst of times" (Dickens, 1962). As it was in the 1770's, so it is in the 1990's for research and knowledge of how to prevent the academic failure of many of today's students. It is the best of times in that we know more than we have ever known about how children learn, the basic skills they will need to keep up, and when to intervene to achieve maximum success. But in the truest sense of an oxymoron, we remain in many ways "wisely foolish" about translating this knowledge into practices and programs that capitalize on the research base. It was precisely this challenge of taking what was known from research and translating it into effective practice that led Robert Slavin and his colleagues at Johns Hopkins University to develop a program called Success for All (SFA), implement it in diverse contexts, and study its effects over time (Slavin, Madden, Karweit, Dolan, & Wasik, 1992).

SFA was designed to be a K-3 preventative and intensive intervention for students at risk of failing to learn to read. Key elements include individualized tutoring, regrouping across grades

into homogeneous reading classes, smaller reading classes, family support, and a comprehensive reading program incorporating phonics and whole-language methods. Early evaluations of SFA demonstrated that the program could be highly successful in increasing reading achievement among very disadvantaged students (Madden, Slavin, Karweit, Dolan, & Wasik, 1993). Developers at Johns Hopkins have continued to monitor the progress of schools and districts that have implemented the program since its inception in 1987 at Abbottston Elementary School in Baltimore, and results continue to support its robustness in a variety of contexts (Slavin, Madden, Dolan, Wasik, Ross & Smith, in press). The viability and acceptability of SFA as a prototype model for at-risk prevention, however, depends to a great extent, especially among the research community, on scrutiny by independent investigators and the extent to which it can be replicated in sites apart from its home base. This paper presents results and conclusions from two to three years of independent evaluations conducted in four cities and states beyond the original home of SFA in Baltimore.

The Success for All Model

The SFA program was developed on the basis of several assumptions regarding effective interventions for disadvantaged students (Slavin, Madden, Karweit, Livermon, & Dolan, 1990). One assumption was that early intervention to prevent learning problems from occurring would be more influential than remediating already

established deficiencies. Another was that reading and language arts are the most critical curriculum areas for applying special interventions, because the skills taught in these areas provide the foundation for learning in all other subjects. Finally, interventions need to be comprehensive to include factors outside of the classroom that affect students' readiness, motivation, and opportunity to learn. Such factors include parental involvement, school attendance, and health needs.

Based on these assumptions, the SFA model was designed to include the following major elements: (a) one-on-one tutoring in reading by certified teachers; (b) regrouping of students in grade 1-3 from heterogeneous, age-grouped classes into homogeneous, cross-grade ability groups during the language arts period (approximately 90 minutes); (c) reduced class sizes in language arts as a result of using the reading tutors as reading teachers during that period; (d) frequent (6 to 8 weeks) assessments in which students are tested to determine reading progress and needs for tutoring and/or group changes; (e) a systematic reading program integrating story telling and retelling (STaR), phonics in beginning reading, vocabulary building, comprehension skills, and cooperative learning; (f) a Family Support Team to provide parenting education and assistance for students experiencing personal or health problems; and (g) a program facilitator who works in the school to coordinate and oversee the program. A brief summary of the SFA program elements is provided

below (for more comprehensive description see the monograph by Slavin et al., 1992).

In kindergarten the SFA program usually begins with the Peabody Language Development Kits and with story telling and retelling (STaR) in selected children's literature to help build language concepts essential to reading. During the second semester, kindergarten students begin with sequenced sound-to-symbol instruction and then move to reading "shared stories," in which part of the story is read by the teacher and part by students. First graders start the year with the "Beginning Reading" program, which emphasizes the development of comprehension and Word Attack skills through a combination of whole-language and phonics methods. Partner reading and initial writing activities are also main components. Students then progress to the "Beyond the Basics" curriculum which can continue through the fifth or sixth grade. Its focus is building comprehension, thinking skills, fluency, and positive reading attitudes by integrating use of the school system's basal reader with cooperative learning, partner reading, process writing, and other components (see Slavin et al., 1992).

First to third-grade students are regrouped according to ability during the 90-minute reading period. Class sizes are reduced by having the SFA tutors serve as reading teachers during that time. Cross-grade grouping is used where appropriate for particular students. Students identified by assessment, the facilitator, and

teacher as most in need of additional help beyond the regular 90 minute reading class receive 20 minutes of individual tutoring each day during a support subject (e.g., art or social studies), with priority given to first graders. Each tutor can generally work with up to 11 students per day. Every 6 to 8 weeks, brief reading and language tests are administered by teachers to determine the students' progress and assignments to reading classes and/or tutoring.

Purpose of the Study

The present research was designed to address the issues of student achievement when the SFA program is introduced into new settings outside the developers' home city and monitored over time by outside evaluators. In addition, we also wanted to extend previous studies by examining some of the processes involved in implementation as well as teachers' experiences and attitudes. Programs in four cities were examined: Memphis, TN (1 SFA school), Ft. Wayne, IN (2 schools), Montgomery, AL (4 schools), and Caldwell, ID (2 schools). The data presented here include three years of implementation at the Memphis site (1990-93) and two years at the other three sites (1991-93).

Evaluation Design

A common evaluation design was used for all four project sites. Procedures and conditions specific to individual sites are described in the separate sections to follow. Every SFA school was matched with a control school similar in poverty level (percent of students qualifying

for free lunch), historical achievement level, ethnicity, and other factors. Student cohorts were formed by matching SFA subjects to control counterparts based on pretest scores on the Peabody Picture Vocabulary Test (PPVT). In some instances where scores for kindergarten students were not available, random samples were selected from the kindergarten populations at the SFA and control schools.

Except where otherwise noted, the basic method used for the test comparisons was Multivariate Analysis of Variance (MANOVA). Overall significance across all test dimensions was evaluated via the multivariate F, and, where multivariate significance was found, individual test significances were evaluated via univariate F's. Reading achievement was measured at all four sites using the following instruments:

Woodcock reading mastery test-revised. Four scales, Letter Identification, Word Identification, Word Attack, and Passage Comprehension (Woodcock, 1987), were individually administered to students by trained testers. The Letter Identification test measures student knowledge of the alphabet; Word Identification measures students' recognition of common sight words; Word Attack assesses decoding and phonics skills; and Passage Comprehension determines comprehension in context.

Durrell analysis of reading difficulty. The Oral Reading scale presents a series of graded reading passages which students read

aloud, followed by comprehension questions (Durrell & Catterson, 1980). It was also individually administered. Two additional language measures, the Merrill Language Screening (Mumm, Secord, & Dykstra, 1980) and the Test of Language Development (TOLD) (Newcomer & Hammill, 1988) were used in kindergarten in 1991-92 only.

The data presented in this paper summarize student performance using effect size (ES), computed by dividing the difference between the SFA and control group mean by the control group standard deviation. All analyses used raw scores. Results are also separately presented for students in each grade who scored in the lowest 25% on the pretest.

Results

Memphis, Tennessee

Summary results for Memphis and the other sites are presented in Table 1. The SFA program was first implemented in grades K-2 at an inner-city school in Memphis, TN in the 1990-91 school year (Ross & Smith, 1991; Ross & Smith, in press). Thus, it has the longest history of the four cities and an additional year of data. To assess first- and second-grade outcomes, SFA students were individually matched to control students from a school in the same inner-city neighborhood. The two schools were nearly identical in the percentage of students receiving free lunch and in standardized achievement scores obtained during previous years. Both schools also serve student populations

that are nearly 100% African American. All SFA-control student matches were based on student performance on total language skills scores and total reading scores on the California Achievement Test (CAT), a state-mandated assessment administered to district students each spring semester. At the kindergarten level, no prior achievement test scores were available for use in matching. Therefore, random samples were selected from the kindergarten populations at the SFA and control schools.

Year 1

Results for kindergarten (SFA n = 20, control n = 23) were the most consistent and striking during the first year (1990-91) with significant SFA advantages ($p < .05$) favoring SFA for Word Identification (ES = .72), Word Attack (ES = 2.00) and Oral Reading (ES = 1.67). Also, a significantly higher percentage of SFA than control students (35% vs. 3%) successfully read the initial paragraph on the Oral Reading test. At the first grade level (n = 45 pairs), SFA students were directionally superior to the control group on all tests; however, significant advantage ($p = <.05$) was indicated on Oral Reading only (ES = .41). Advantages for students who had scored in the lowest 25% in reading on the previous year CAT test were much more pronounced, with positive, if not significant, differences attained on Word Identification (ES = .74), Word Attack (ES = 1.25) and Oral Reading (ES = .61). At the second grade level, the total SFA and control samples performed similarly, but for the lowest 25% of

students, SFA had noticeable advantage in Word Identification (ES = .56, p = .07). As Table 1 indicates, the average ES's for kindergarten and grade 1 during the first year were .97 and .17, respectively. For the lowest performing first grade students, the average ES was .90.

Insert Table 1 about here

Year 2

At the kindergarten level, SFA and control students were matched (n = 35) on the basis of pretest scores on the Peabody Picture Vocabulary Test (PPVT) given in the fall semester. Although all comparisons favored SFA students over the control students, the ES's were small in magnitude, averaging only .08 overall. Parallel analyses conducted on the scores for the pairs scoring in the lowest 25% yielded low effects (Mean ES = .00).

At first grade (second year in SFA), however, ES's were substantially larger than those obtained for kindergarten, averaging .90 across all tests (SFA n = 26, control n = 68). The two largest effects were 1.10 for Word Attack and .93 for Passage Comprehension. Univariate comparisons using t-tests for independent samples, showed significant advantages for SFA on all tests. Parallel analyses conducted for the lowest 25% subsamples (SFA n = 7, control n = 18) also indicated extremely high ES's, averaging 3.15 overall. Despite the

small n's, all univariate tests significantly favored the lowest-25% SFA group.

Results for the second grade sample (n=17 pairs; matched on CAT scores in first grade) indicated the SFA group surpassed the control group on all but the Durrell Oral Reading test. However, effects for the overall sample were non-significant (average ES = .14). Only 4 SFA and 5 control students remained in the low-25% subsample. Informal comparisons showed that the SFA group was favored on all tests, with mean ES = 1.35.

Year 3

Kindergarten was not evaluated during the third year. Because the samples of surviving pairs were so small at both first and second grades, the design was altered and the SFA and control program groups were treated as independent samples. Results for first grade (SFA n = 30, control n = 33) indicated SFA program advantage on all of the tests (mean ES = .38). The overall MANOVA, however, was not significant. Effect sizes for the individual tests were: Word Identification = .32, Word Attack = .32, Passage Comprehension = .35, and Oral Reading = .53. The lowest 25% subsample of first graders (SFA n = 7, control n = 8) performed significantly better than the control subsample on Word Attack only (ES = .18). The overall mean ES was .11.

The MANOVA effect for second graders (SFA n = 15, control n = 43) was stronger than for first grade but only approached significance.

$F(4,53) = 2.43$, $p = .06$. Second grade SFA students did significantly better than control students on Word Identification (ES = .56) and Passage Comprehension (ES = .70). Results for Oral Reading (ES = .52) and Word Attack (ES = .27) were positive but not significant. The overall mean ES for second grade was .51. Results for lower-ability second grade students were more dramatic. With the MANOVA yielding $F(4,11) = 9.53$, $p < .001$, significant group differences favoring SFA occurred on all tests: Word Identification (ES = 2.62), Word Attack (ES = 2.59), Passage Comprehension (ES = 2.44), and Oral Reading (ES = 2.98). The mean ES was 2.66.

Outcomes of the three years of implementation in Memphis showed moderate advantages for SFA over the control school in first grade and relatively strong advantages in second grade for students who had begun the program in kindergarten. The most clear and convincing support for the benefits of SFA were for the lowest-performing students, those most at risk of school failure. Congruent with previous SFA findings, the longer the children were in the program, the greater were the benefits.

Montgomery, Alabama

SFA was implemented at four elementary schools in Montgomery in 1991-92. Both the program and evaluation procedures were fundamentally the same as those described for the Memphis program. All schools were inner-city schools with an African American population of from 91-99%. Each school was matched to a comparable

school having similar demographic characteristics and prior achievement on standardized tests. During the first year, the evaluation encompassed all four schools at Grade K but only two schools at Grade 1. The reason for excluding two schools from the Grade 1 evaluation was that their first graders had been in SFA as kindergartners in the previous year and thus could no longer be matched to control counterparts without possible contamination from SFA effects.

During year two (1992-93), due to the unavailability of suitable matching control schools, only two of the four SFA schools were evaluated. First grade students at one of the two schools were individually matched with comparable control students on the basis of PPVT pretest scores obtained in the previous (kindergarten) year; however, first graders at the second school were not matched for the 1992-93 school year, as the control school was new and the students had not been pretested. For this school pair, an independent samples MANOVA design was used.

Year 1

The matched MANOVA design at the kindergarten level consisted of four school pairs (Pair 1 n = 39, Pair 2 n = 34, Pair 3 n = 19, Pair 4 n = 48). There was only one significant contrast for the overall kindergarten sample: SFA students were superior to the control students on the PPVT Posttest (ES = .24). Overall, SFA and control performances were almost identical (mean ES = 0.00). In addition, no

significant program differences were found in any of the comparisons between the control and SFA lowest-25% subsamples.

The results were also similar for both samples of first grade students (Pair 1 n = 29, Pair 2 n = 69). Examination of the means and effect sizes (mean ES = .11) showed that SFA students did slightly better than control students on the Word Identification, Durrell Oral Reading, and the Word Attack tests, but had nearly identical performance to the control group on Passage Comprehension. For the low-25%, SFA students were directionally higher than the control group on all tests; however, none of the effects was significant, and the mean ES was only .15 (see Table 1).

Because the outcomes in Montgomery were rather weak overall and were not consistent to other program sites in the first year of implementation, the researchers conducted a followup observation study (see Bond, Ross, & Smith, 1993) of the control schools. Findings from this study produced a likely explanation of the results in that the control schools in Montgomery contained many of the essential program elements of SFA, such as tutoring, regrouping, reduced class sizes, and language development programs such as STaR and Peabody in beginning reading.

Year 2

Only two schools were evaluated during the second year. Examination of individual school pairs showed for Pair 1 (SFA n = 57, control = 81) significant differences and quite large effect sizes

favoring SFA first grade students on all of the tests (mean ES = 1.28). For School Pair 2 (14 pairs) MANOVA revealed no significance $F(4,10) = 1.52$, $p = .27$ (mean ES = 1.29). Due to small sample size and large observed mean effects, univariate tests were examined. Univariate tests revealed significant SFA effects on Word Identification (ES = 1.43), Word Attack (ES = 1.54), and Oral Reading (ES = 1.25). Results for Passage Comprehension were positive but not significant (ES = .93). The low-25% samples from both schools (SFA $n = 17$, control $n = 24$) also significantly outperformed the control school on all tests $F(2,29) = 6.39$, $p < 0.01$. The mean ES was 2.28 (see Table 1).

Although results for the SFA implementation from the first year were rather weak for the Montgomery schools, the second year yielded much greater benefits. More appropriate matching of experimental schools to schools with traditional types of programs is the most probable explanation for the findings. Further study of these students as they move to upper grades will continue.

Ft. Wayne, Indiana

Two elementary schools in Ft. Wayne, IN implemented SFA in grades K-6. Unlike Memphis and Montgomery, these schools were racially mixed and located in mid- to lower-income neighborhoods. Kindergarten and first grades were pre- and post-tested in the first year and first and second grades were tested during the second year using the same instruments as the Memphis and Montgomery sites.

Year 1

The MANOVA performed on the kindergarten overall sample (n = 73 pairs) yielded a significant program effect, F(5,67) = 9.34, p < .001. Univariate tests showed that SFA students were superior to control students on Letter Identification (p = .003), TOLD (p = .04), and Word Identification (p < .001). The overall mean ES across all tests was .38 (see Table 1). For the low-25% subsample, the MANOVA result only approached significance, F (5,14) = 2.65, p = .069. Inspection of univariate comparisons showed significant differences on Word Identification only (p < .01), with SFA students having directionally higher means on all measures except TOLD. The overall ES was .56.

The first grade MANOVA results (n = 68 pairs) also indicated a significant program effect: F (4, 63) = 6.45, p < .001 (mean ES = .47). Univariate tests showed the SFA schools to be superior to the control schools on Word Identification (p = .002), Word Attack (p < .001), and Durrell Oral Reading (p = .002). The SFA advantage on Passage Comprehension approached significance (p = .052). Inspection of the low-25% sample (n = 19 pairs) revealed significant effects favoring SFA: F(4, 15) = 4.45, p = .014 (mean ES = .60). Univariate tests significantly favored SFA on Word Identification (p = .01), Word Attack (p = .03), and Durrell Oral Reading (p = .02)

Year 2

First grade SFA students (n = 25 pairs) outperformed controls on all four reading test measures [F(4, 20) = 3.37, p = .029] with SFA students doing significantly better on Word Identification (ES = .80) and Word Attack (ES = .66). Results for Passage Comprehension (ES = .25) and Oral Reading (ES = .33) were not significant. The overall ES for first grade was .51. MANOVA revealed no low-25% sample effects (p = .118) due to small sample size (n = 7 pairs). SFA students did significantly better than control students on Word Attack (ES = 1.24) only. The overall mean ES was .79.

MANOVA comparing second grade combined results for both pairs of schools revealed a significant effect favoring SFA: F(4,31) = 3.40, p = .02 (mean ES = .44). Again, SFA students did better than control students with significant effects on Word Identification (ES = .62). Results for Word Attack (ES = .53), Passage Comprehension (ES = .36) and Oral Reading (ES = .24) were not significant.

Overall effect sizes for both the total and lowest-25% first graders increased slightly from .47 in year 1 to .51 in year 2, while the lowest-25% increased in effect size from .60 to .79. (No first year comparisons were made for second grade). The Ft. Wayne results indicate benefits for SFA consistently increased from Year 1 to Year 2. In addition, as in Memphis and Montgomery, the low-25% students realized the greatest benefits.

Caldwell, Idaho

SFA was implemented in kindergarten and first grade of one elementary school during 1991-92 in the Caldwell, Idaho community. In 1992-93, the Caldwell district opened a new school and moved the principal and part of the staff from the original SFA implementation, making both schools SFA schools. These two sites were the first rural SFA schools to be evaluated. All evaluation procedures and instruments were consistent with the other three sites.

Year 1

Kindergarten outcomes were generally positive with slight negative effects on the Letter and Word Identification measures. The MANOVA yielded $F(5, 47) = 2.63$, $p < .05$, with an average ES of .10. Univariate analyses revealed a significant difference only on the PPVT posttest, $F(1,51) = 8.52$, $p < .01$ (ES = .41). The analyses of the low-25% sample of grade K (n = 14 pairs) revealed no significant differences on either the overall MANOVA or the univariate analyses due to the small number of pairs. However, there were large effect sizes favoring SFA (mean ES = 1.59).

For Grade 1 (n = 64 pairs), MANOVA was significant, $F(5,59) = 2.69$, $p < .05$ (mean ES = .01). Follow-up univariate analyses, however, yielded no significant effects on any test. Results from the low-25% sample (n = 19 pairs) were negative with a mean ES of -.21, $F(5,14) = 4.53$, $p < .05$. Univariate tests favored the control sample on Word Identification only (p = .05, ES = -.83)

Year 2

The MANOVA for first grade students (n = 31 pairs) was highly significant, $F(4,27) = 12.50$, $p < .01$ (mean ES = -.21). Univariate analyses indicated SFA students were comparable to control students on Word Attack (ES = .07) and on Passage Comprehension (ES = -.07). Control students did significantly better than SFA students on Word Identification (ES = -.51), and directionally better on Oral Reading (ES = -.34). For low-25% first grade students (n = 9 pairs) the overall MANOVA was also significant, $F(4, 5) = 9.80$, $p = .01$ (mean ES = .-52). Univariate results indicated significant advantages for the control students on Word Identification (ES = -.79) and Oral Reading (ES = -.58). Results for Word Attack (ES = -.41) and Passage Comprehension (ES = -.29) were not significant.

MANOVA for second grade was highly significant, $F (4, 28) = 3.24$, $p = .03$ revealing SFA advantages (mean ES = .12). SFA students did significantly better than control students on Passage Comprehension (ES = .39), and had a nonsignificant advantage on Word Attack (ES = .30). Results for Word Identification (ES = -.07) and Oral Reading (ES = -.16) slightly favored the control group. MANOVA revealed no significant results for the low-25% second grade subsample, $F(4, 4) = 2.18$, $p = .24$ (mean ES = .33). Univariate results revealed directional advantages favoring SFA on all of the tests: Word Identification (ES = .07), Word Attack (ES = .44), Passage Comprehension (ES = .65), and Oral Reading (ES = .15).

Outcomes at the Idaho site have been the most inconsistent of all site comparisons (Slavin, et al. in press) and are particularly puzzling. Observations of the SFA implementation have indicated that the school is implementing the program with high fidelity. Given that the control school is a high-quality school that provided special reading programs, and that it is located in a more suburban, less rural area of the region, it is reasonable to question the validity of the SFA-control comparison. The results for the overall second grade sample, and especially the second grade low-25% subsample, however, do reflect the general pattern of findings at other sites and suggest the cumulative benefits of SFA over time. That is, the longer a lower achiever is in SFA, the greater the effects.

Overall Comparison of Four Sites

Kindergarten

Overall kindergarten results for the four sites are presented in Figure 1. Effect sizes for each site's first year of implementation ranged from .97 in Memphis to -.00 in Montgomery. (See description and explanation of the Montgomery results above.) It is interesting to note that for Memphis, the only site where kindergarten results were collected for two years, there was great disparity in results from the first year to the second (.97 ES to .08 ES). During the first year of implementation, the SFA program's more academic emphasis represented a fairly substantial deviation from the traditional kindergarten curriculum taught at the control school. However,

during the second year, the control school had shifted its emphasis to include more direct teaching of alphabetic and sound principles in addition to acquiring a Write to Read computer program in which all kindergarten children participated.

Data for the lowest performing kindergarteners across all four sites are presented in Figure 2. Effect sizes for this group are generally higher than those for the overall sample and ranged from a low of -.08 at the Montgomery site to a substantial 1.59 ES in Caldwell.

First Grade

First grade results are presented in Figures 3 and 4 and are the most key comparisons when considering the program's primary focus of giving all students a strong start in reading. Of particular note are the second year's results at each site. It is during the second year of implementation, when teachers have had a year's experience with the program and have received children from kindergarten with the early SFA experiences in language, storytelling, and graphophonic principles, that one can begin to see the relative impact of the program. With the exception of Caldwell, all sites realized increased benefits of program during the second year of implementation, ranging from .51 to 1.27. Similar to the findings for kindergarten, for the lowest 25% first graders the effect sizes were even stronger, ranging from .79 to 3.15.

As can be seen in Figures 5 and 6, results for the overall second grade samples were fairly strong in Memphis (ES = .51) and Ft. Wayne

(ES = .44) but weaker in Caldwell (ES = .12). For the lowest 25% sample, the strongest effects occurred in Memphis (ES = 2.66) and Ft. Wayne (ES = .79), with more moderate results in Caldwell (ES = .33).

Conclusions

Based on the present results in Memphis, Montgomery, and Ft. Wayne, achievement benefits for SFA students clearly occurred in all grade levels with even stronger effects for the lowest performing students. The inconsistency of results at two sites with results from previous studies suggests that it may be unrealistic to expect outcomes to be identical across all applications. Programs differ in teacher training, school conditions, and also in the appropriateness of the "control" programs used for evaluation comparisons. Some of the key issues evolving from this study of replications of SFA center around these program differences.

Achievement results were most consistent in Memphis and Ft. Wayne with previous studies of SFA (Madden et al., 1993) and less so in Montgomery and Caldwell. Outcomes at the latter two sites may have been the result of matches with control schools that are also changing and improving their curricula to more closely reflect effective practices identified by research so that they have become more similar to the "program" schools.

Other types of naturally occurring dynamics from school to school also probably had effects that were not measured. In some instances, a clear competitive spirit naturally arose between school

staffs and principals of the matched schools as well as between schools who were starting up new programs at the same time, as in Ft. Wayne and Montgomery. Principal and major staff transfers, as at Caldwell and to some degree in Ft. Wayne, and facilitator changes (Memphis' first year; Caldwell, Year 2), could conceivably have had major impact on program implementations. Other noted differences from one site to another included quality and extent of training of teachers; level of commitment and fidelity to the program by teachers, principals, and facilitators; urban vs. rural environments; and background characteristics of students.

Based on these considerations and from the achievement results of this study, it seems fair to infer that Success for All can be replicated at distant locations from, and with limited monitoring by, the program developers. Although there was variability in implementation quality and achievement outcome within and across sites, SFA has the advantage over many other programs of having systematic procedures, comprehensive components, facilitator leadership, monitoring components, and extensive and ongoing staff development. These elements give it robustness and increased ability to ensure that disadvantaged children will have opportunity to learn and that, with relentless effort, can be successful in school.

References

Bond, C.L., Ross, S.M., & Smith, L.J. (1993). Montgomery, Alabama Success for All program control school visitation study. Memphis: Memphis State University.

Dickens, C. (1962). A tale of two cities. New York: Macmillan Co.

Durrell, D., & Catterson, J. (1980). Durrell Analysis of Reading Difficulty. New York: Psychological Corporation.

Madden, N. A., Slavin, R.E., Karweit, N. L., Dolan, L.J., & Wasik, B.A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. American Educational Research Journal, 30, 123-148.

Mumm, N., Secord, W., & Dykstra, K. (1980). Merrill language screening test. NY: Psychological Corp.

Newcomer, P. & Hammill, D. (1988).Test of language development. Austin, TX: Pro-Ed.

Ross, S.M. & Smith, L. J. (1991). Final report: 1991 success for all program in Memphis. Memphis: Memphis State University, Center for Research in Educational Policy.

Ross, S. M. & Smith, L. J. (in press). The impact of using the Success for All model on reading achievement, teacher adjustment, and restructuring at an inner-city school. Elementary School Journal.

Slavin, R. E., Madden, N. A., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1992). Success for All: A relentless approach to prevention and

early intervention in elementary schools. Arlington, VA: Educational Research Service.

Slavin, R. E., Madden, N. A., Karweit, N. L., Dolan, L. J., & Wasik, B. A., Ross, S., & Smith, L. (in press). Whenever and wherever we choose...: The replication of Success for All. Phi Delta Kappan.

Slavin, R. E., Madden, N.A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. American Educational Research Journal, 27, 255-278.

Woodcock, R. W. (1987). Woodcock reading mastery tests-revised. Circle Pines, MN: American Guidance Service.

Table 1
Average Effect Sizes for the Overall and Lowest 25% Samples.

Site	Year and Grade							
	Year 1			Year 2			Year 3	
	K	1	2	K	1	2	1	2
Memphis								
Overall								
SFA n	20	45	47	35	26	17	30	15
Control n	23	45	47	35	68	17	33	43
Mean ES	0.97	0.17	0.00	0.08	0.90	0.14	0.38	0.51
Low - 25%								
SFA n	-	11	11	9	7	4	7	5
Control n	-	11	11	9	18	5	8	11
Mean ES	-[a]	0.90	0.42	0.00	3.15	1.35	0.11	2.66
Montgomery								
Overall								
SFA n	129	94	-	-	71	-	-	-
Control n	129	94	-	-	96	-	-	-
Mean ES	0.00	0.11	-	-	1.27	-[a]	-	-
Low - 25%								
SFA n	33	27	-	-	17	-	-	-
Control n	33	27	-	-	24	-	-	-
Mean ES	-0.08	0.15	-	-	2.28	-[a]	-	-
Ft. Wayne								
Overall								
SFA n	73	68	-	-	25	36	-	-
Control n	73	68	-	-	25	36	-	-
Mean ES	0.38	0.47	-	-	0.51	0.44	-	-
Low - 25%								
SFA n	19	19	-	-	7	12	-	-
Control n	19	19	-	-	7	12	-	-
Mean ES	0.56	0.60	-	-	0.79	0.79	-	-
Caldwell								
Overall								
SFA n	52	64	-	-	31	32	-	-
Control n	52	64	-	-	31	32	-	-
Mean ES	0.10	0.01	-	-	-0.21	0.12	-	-
Low - 25%								
SFA n	14	19	-	-	9	8	-	-
Control n	14	19	-	-	9	8	-	-
Mean ES	1.59	-0.21	-	-	-0.52	0.33	-	-

[a] No sample tested for this population in this year.

Figure Captions

Figure 1. Kindergarten: Mean Effect sizes for overall samples.

Figure 2. Kindergarten: Mean Effect sizes for lowest 25% samples.

Figure 3. First grade: Mean Effect sizes for overall samples.

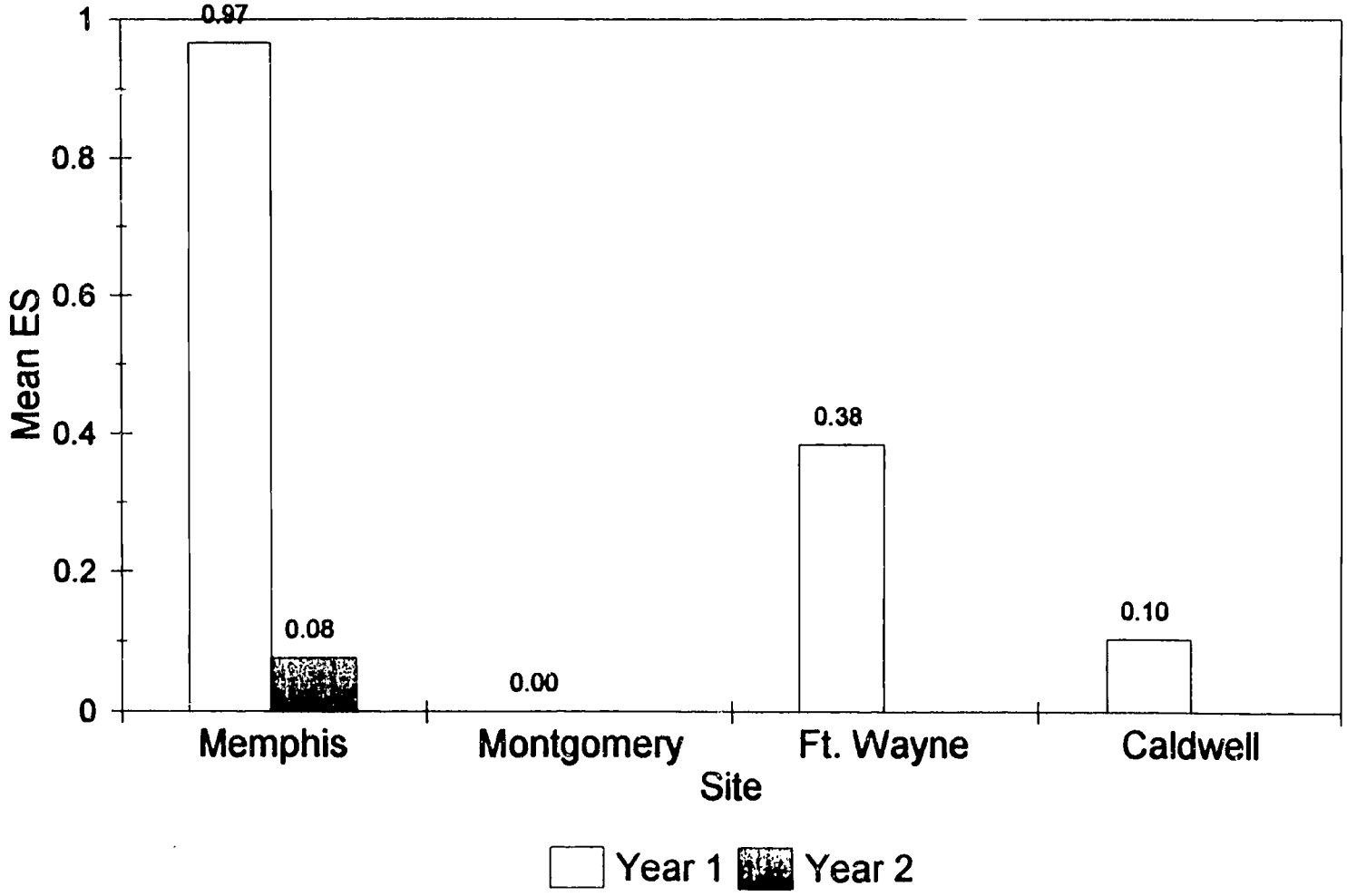
Figure 4. First grade: Mean Effect sizes for lowest 25% samples.

Figure 5. Second grade: Mean Effect sizes for overall samples.

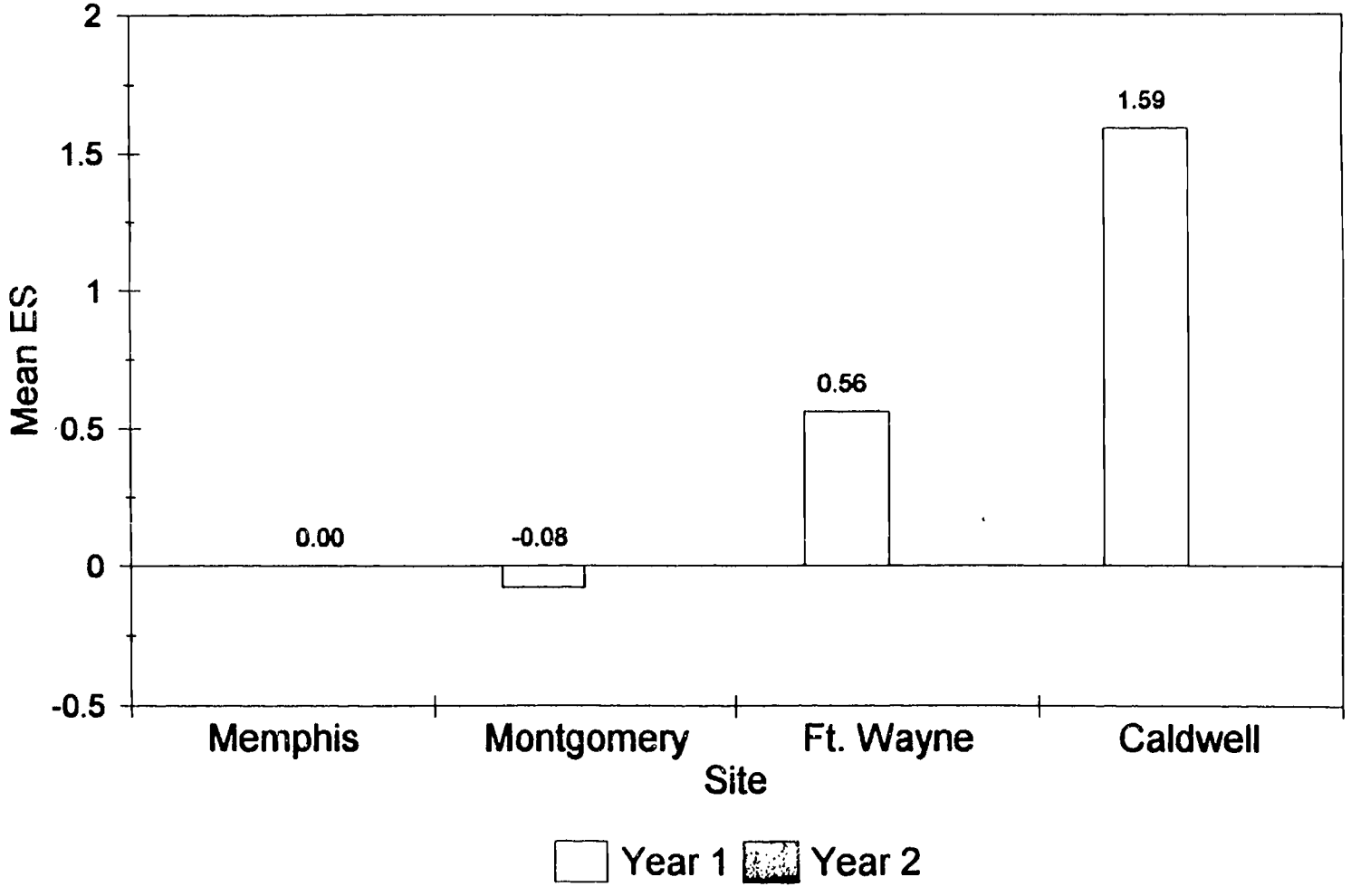
Figure 6. Second grade: Mean Effect sizes for lowest 25% samples.

Grade K Comparisons

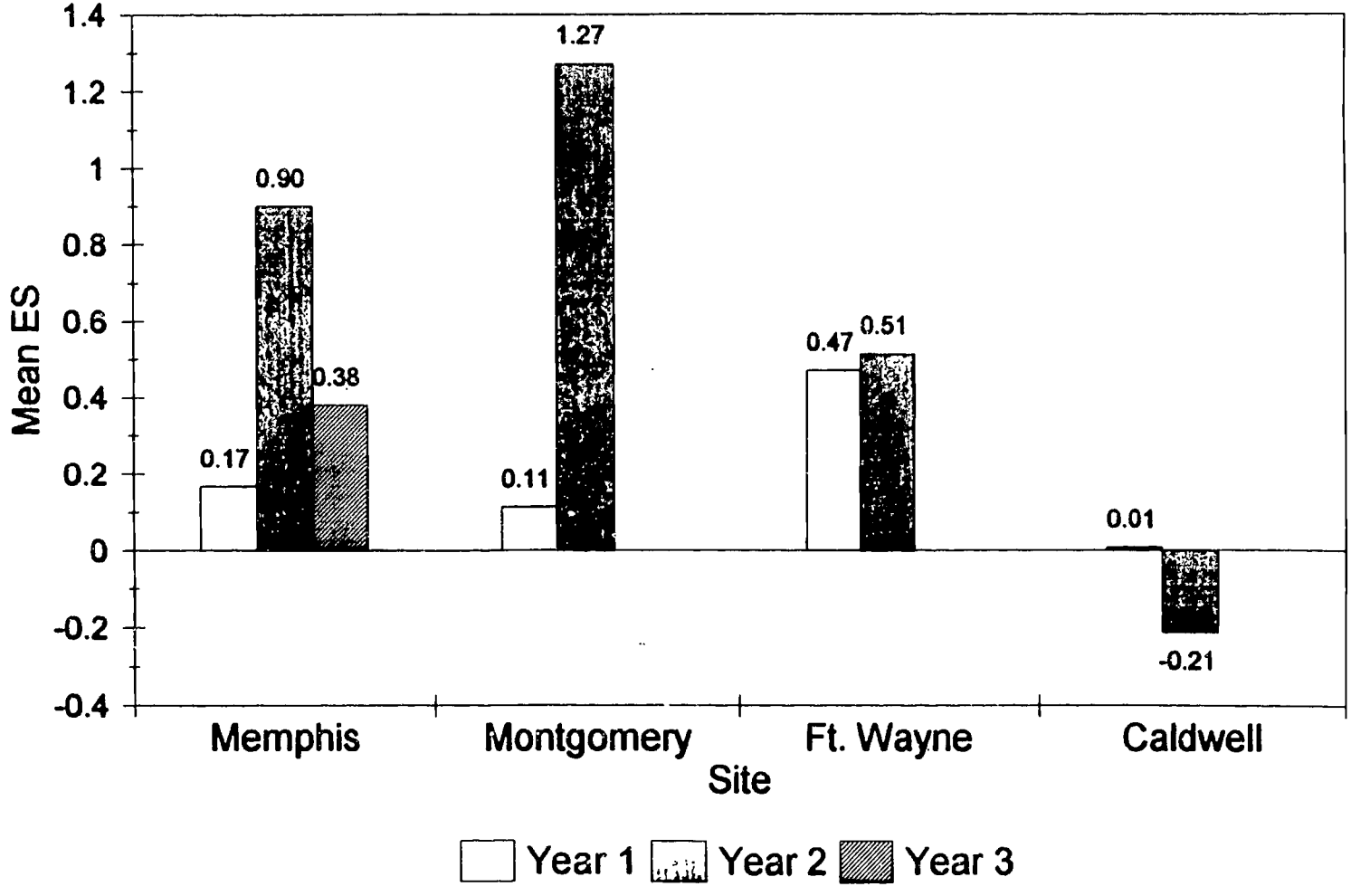
Overall Samples



Grade K Comparisons
Lowest 25%

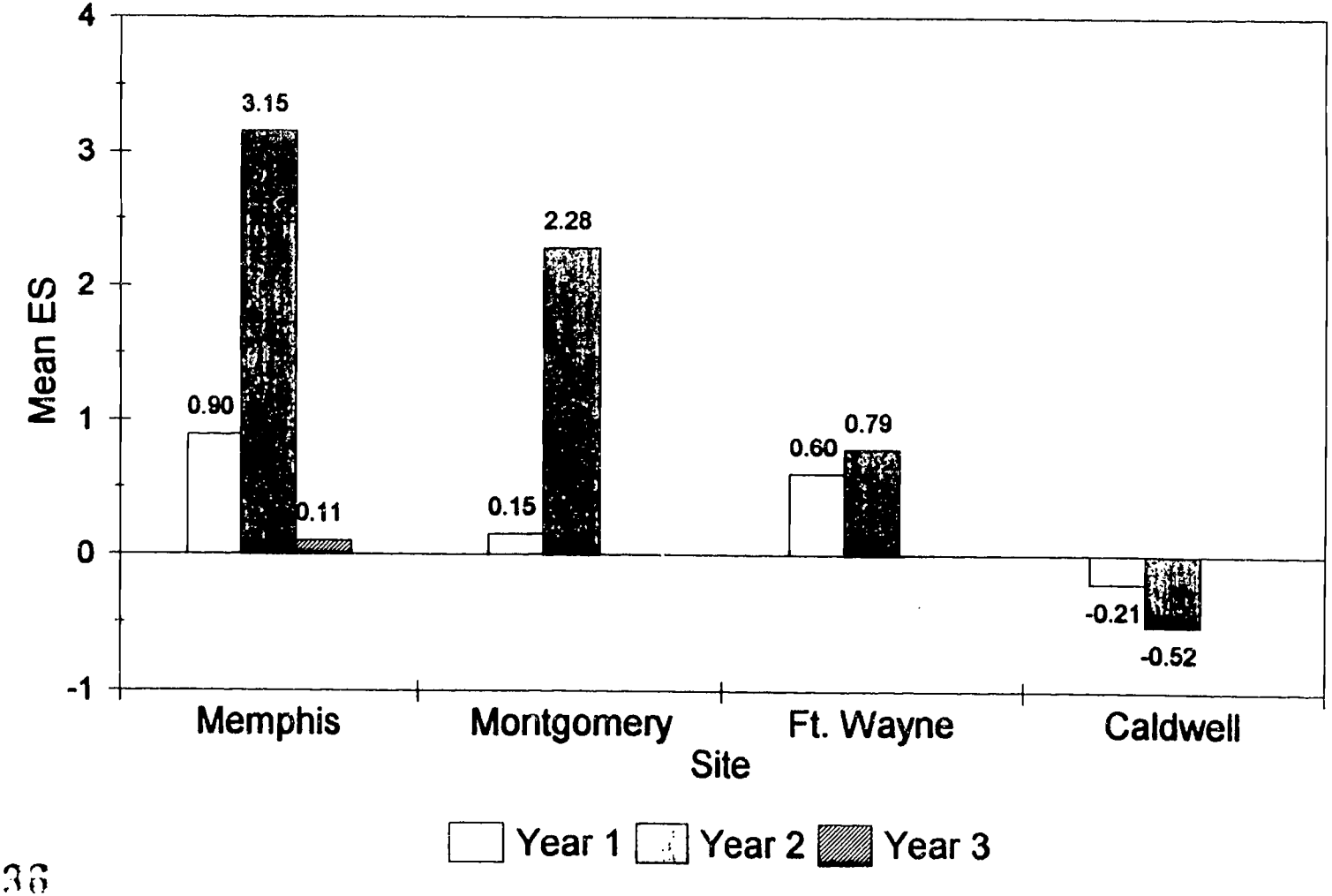


Grade 1 Comparisons
Overall Samples

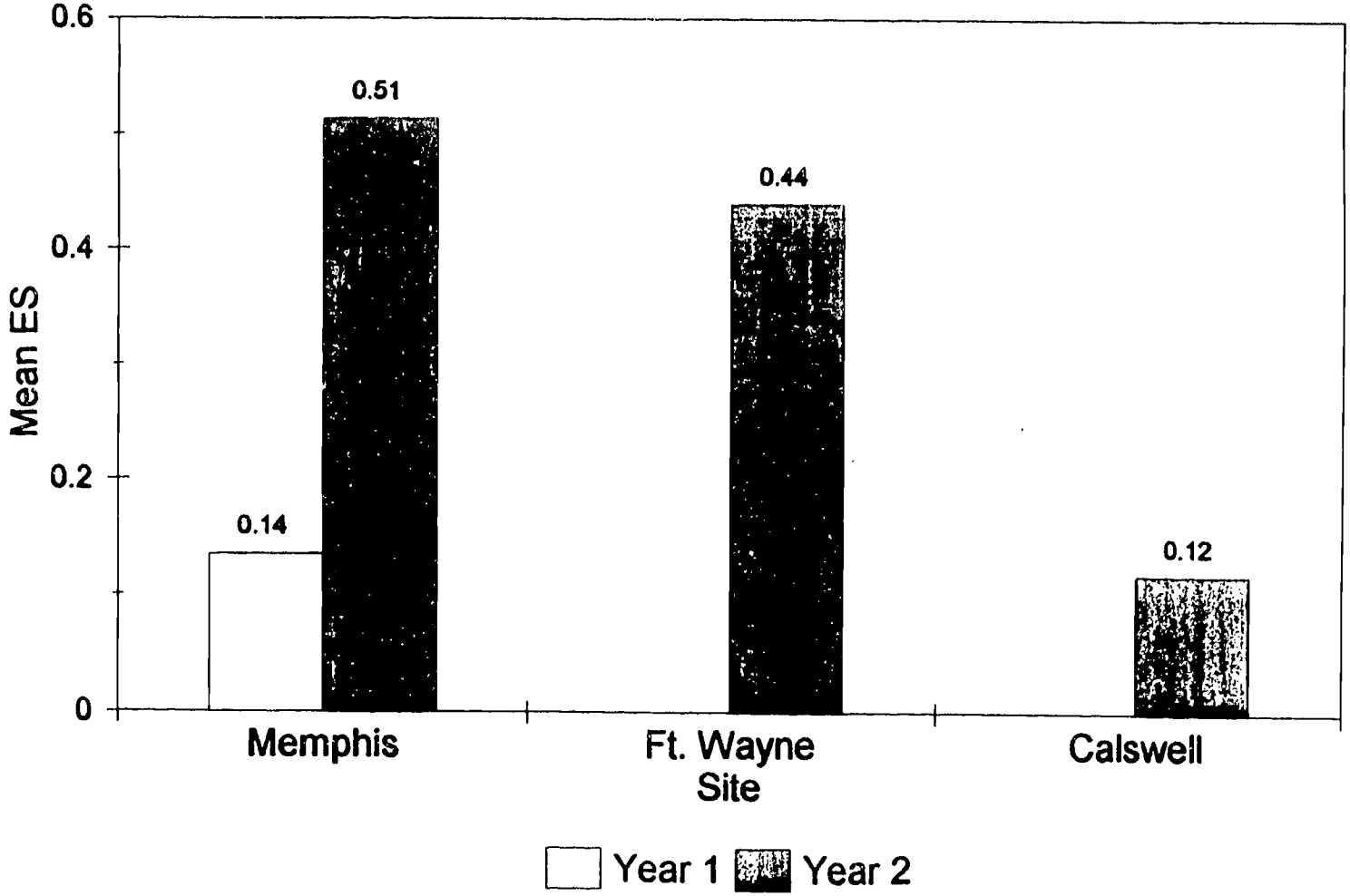


Grade 1 Comparisons

Lowest 25%



Grade 2 Comparisons
Overall Samples



33

Grade 2 Comparisons
Lowest 25%