University of HUDDERSFIELD

University of Huddersfield Repository

Tracey, Louise, Chambers, Bette, Slavin, Robert, Hanley, Pam and Cheung, Alan

Success for All in England: Results From the Third Year of a National Evaluation

Original Citation

This version is available at http://eprints.hud.ac.uk/29075/

http://eprints.hud.ac.uk/

SAGE Open

Success for All in England: Results From the Third Year of a National Evaluation
Louise Tracey, Bette Chambers, Robert E. Slavin, Pam Hanley and Alan Cheung

The online version of this article can be found at:

Published by:

⑤SAGE

⑤SAGE open

Additional services and information for *SAGE Open* can be found at:

Email Alerts: /cgi/alerts

Subscriptions: /subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Success for All in England: Results From the Third Year of a National Evaluation

Louise Tracey[1], Bette Chambers[1,2], Robert E. Slavin[1,2], Pam Hanley[1], and Alan Cheung[3]

Abstract

This article reports the third-year findings of a longitudinal evaluation in England of Success for All (SFA), a comprehensive literacy program. Eighteen SFA schools across England and 18 control schools, matched on prior achievement and demographics, were included in this quasi-experimental study. The results of hierarchical linear modeling analysis reveal a statistically significant positive school-level effect for SFA schools compared with control schools on standardized reading measures of word-level and decoding skills, and there were directionally positive but nonsignificant school-level effects on measures of comprehension and fluency. Practical and policy implications of these findings are discussed, particularly as they relate to recent English government policies encouraging schools to implement research-proven approaches.

Introduction

Currently, 25% of children in the United Kingdom live in poverty, and the percentage is growing (Department for Work and Pensions, 2009; Wilkinson & Pickett, 2010). The chances of these children succeeding in school and in life are much smaller than are those of their more advantaged peers (Mongon & Chapman, 2008; Strand, 2008). Inequality starts early, and although research suggests that it is what parents *do*, not who parents *are*, for example, in terms of occupation and income, for parents in poverty and challenging circumstances it is harder to provide children with the skills they need for school readiness (Duncan & Murnane, 2011; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004). Consequently, every year thousands of children enter school with educational differences already apparent (Hills et al., 2010; Lee & Burkham, 2002). This educational gap often continues to increase during a child's school career, and the longer they persist, the harder they are to close (Goodman, Sibieta, & Washbrook, 2009). Consequently, there is a need for proven-programs designed to address these inequalities.

Early Literacy

One of the keys to later educational attainment is early literacy acquisition. Strong literacy achievement can lead to children being able to access the whole curriculum, and the earlier they do so the more beneficial it is. In contrast, early literacy problems can hinder children's knowledge and development, with long-term consequences for their educational outcomes (Lesnick, George, Smithgall, & Gwynne, 2010).

For many years, research on beginning reading has supported the use of explicit, systematic phonics instruction (National Reading Panel, 2000). In particular, the Rose (2006) Review called for phonics to be taught systematically in schools. Synthetic or systematic phonics involves teaching the discrete sounds that letters make, to help children blend or "sound out" new words.

The importance placed on synthetic phonics in England is reflected in current policy. For example, following the Rose Review, the English Primary Framework and the associated phonics resources *Letters and Sounds* (Department for Education and Skills, 2007) emphasize synthetic phonics practices. *Letters and Sounds* outlines systematic phonics teaching in daily, briskly paced lessons to facilitate early literacy acquisition. In 2011, a national phonics assessment to be administered at the end of Year 1 (when children are typically 6 years old) was piloted and then rolled out nationwide in 2012. The government's intention was to incentivize schools to better support pupils to master basic phonics skills in Year 1. Only 40% of children tested achieved the

[1]University of York, UK
[2]Johns Hopkins University, Baltimore, MD, USA
[3]The Chinese University of Hong Kong, People's Republic of China

Corresponding Author:
Louise Tracey, Institute for Effective Education, University of York, Heslington, York YO10 5DD, UK.
Email: louise.tracey@york.ac.uk

recommended level on the first assessment administered in summer 2012.

Alongside this increased emphasis on phonics in teaching literacy, education policymakers in England (and elsewhere) are beginning to encourage schools to implement programs and practices with strong evidence of effectiveness in helping children living in poverty to succeed in school (Allen, 2011; Department for Children, Schools and Families [DCSF], 2009). For example, they have instituted a policy of allocating a "pupil premium" to schools based on the number of children who are eligible for free school meals due to poverty. Schools are encouraged to adopt proven practices with this additional funding. A systematic review of reading programs suggested that the most successful programs were those that included a broad curriculum that included systematic phonics instruction as well as teaching in vocabulary and comprehension (Slavin, Lake, Chambers, Cheung, & Davis, 2009). The Allen (2011) report listed programs with a record of effectiveness. One of the few programs listed in the top category was the Success for All (SFA) program, based mainly on evidence from evaluations conducted in the United States.

SFA

SFA is a whole-school reform program, started in the United States in 1987. The program's design is based on a model that posits that substantially enhancing success in high-poverty schools depends on a multidimensional intervention approach. This includes providing extensive professional development, effective teaching strategies, emphasizing cooperative learning, and school-wide structures focusing on school leadership, parent involvement, and attendance, which are expected to jointly enhance reading performance and other outcomes (see Slavin, Madden, Chambers, & Haxby, 2009, for a complete description). SFA began to be used in the United Kingdom in 1997, and it now serves more than 100 schools in England, Scotland, and Wales. One previous small-scale study found positive effects of SFA in the United Kingdom (Hopkins, Youngman, Harris, & Wordsworth, 1999), although another study found positive outcomes in Year 1, but mixed outcomes in Year 2 (Tymms & Merrell, 2001). However, in a policy context demanding more rigorous evidence for interventions intended for use in the United Kingdom, there remained a need for a large-scale evaluation of the approach in the schools as they exist today.

The theory of action behind SFA emphasizes *prevention* and *early, intensive intervention*, to keep pupils on the path to success throughout their time in primary school and beyond. *Prevention* includes approaches used in Nursery and Reception intended to build children's background knowledge, vocabulary, and phonemic awareness using strategies emphasizing cooperative learning, theme-based activities, and stories. It then focuses on proven teaching methods for beginning reading in Reception and Year 1 and for reading and writing in Year 2 and beyond. *Early intervention* includes tutoring and outreach to parents to solve children's problems before they become serious.

A strong emphasis within teaching in SFA is the use of cooperative learning. The cooperative learning strategies employed have children learning in mixed-ability pairs or groups of four, in which positive interdependence and individual accountability are key. This means that teams are rewarded for each team member's learning. Each pupil is held accountable for his or her own learning and for helping group mates learn as well. Research has long suggested that cooperative learning has many benefits across the curriculum, including in literacy development (Law, 2008; Slavin, 1995, 2009; Stevens, 2003). Cooperative learning has been shown to increase student motivation, and the brisk pacing and in-built routines ensure that time is used effectively by both students and teachers. It is designed to give students opportunities to try out their understandings in a safe environment, to receive immediate feedback, and to "learn by teaching" in describing their current state of knowledge to a peer. From Year 1 onward, pupils are taught in classes regrouped so that all are at one reading level, though they may be from different year levels. Groupings are changed each term in light of pupils' performance. A daily 90-min beginning reading program called Reading Roots introduces synthetic phonics and builds further vocabulary, fluency, and comprehension skills, in line with the recommendations of the Rose (2006) Review and the U.S. National Reading Panel (2000). Reading Roots uses phonetic readers, which children read to each other in pairs, as well as real children's literature, vocabulary development, and comprehension activities. Embedded video introduces letter sounds, sound blending, vocabulary comprehension skills, and writing. Children work in pairs and small groups to help each other master the content.

After the Year 1 reading level, children enter Reading Wings, where they work in four-member cooperative groups to help each other apply and extend their phonics skills and build comprehension, fluency, and vocabulary. Pupils learn and apply metacognitive skills such as clarification, summarization, and graphic organizers to gain skill in comprehending texts of increasing sophistication. Pupils also learn a writing process approach in which they help each other plan, draft, revise, edit, and "publish" compositions in various genres.

The SFA teaching elements are sufficient to enable most children to succeed in reading, but some children need more than this. The schools provide tutoring for children who are struggling in reading, and reach out to parents to help with home literacy, attendance, behavior, and other issues of importance in children's overall development.

Each SFA school has a facilitator on its staff, an experienced teacher who works with the staff members to help ensure that all program elements are being implemented with high quality and coordinated with each other. Extensive

Continuing Professional Development (CPD) is provided to each school by SFA-UK, a U.K. registered charity. This includes initial training and coaching visits to the school over time.

SFA rose to prominence through the whole-school reform movement in the United States. Research on whole-school reform has highlighted the difficulties involved in the scale-up of programs across schools, and in particular, their replicability, Rowan, Camburn, and Barnes (2004) have indicated that SFA has been particularly successful in replicating its program due to its "bureaucratic approach" to program implementation. However, other studies have also indicated that a more pragmatic approach facilitated SFA's rapid expansion across schools, not least due to the practice of "collaborating with schools in a knowledge-producing enterprise" (Peurach & Glazer, 2012, p. 170). The adaptability of SFA is also in evidence in its introduction and implementation in the United Kingdom.

Although the United Kingdom has not traditionally followed the whole-school reform movement begun in the United States, the scale-up and replication of programs is, as indicated above, increasingly becoming an issue. This is partly a response to the Allen (2011) Report but has been reinforced by the establishment of the Education Endowment Foundation (http://educationendowment-foundation.org.uk/) and the establishment of the What Works evidence centers for social policy.

Although its basic structure is the same as its U.S. version, SFA has been substantially adapted to the language, culture, and standards of England, Scotland, and Wales. It is aligned with the *Letters and Sounds* requirements of England's Department for Education and emphasizes the same curricular elements, focusing on systematic instruction in phonemic awareness and phonics, as well as vocabulary and comprehension. It has a fast-paced and structured approach to teaching, intended to ensure that pupils have solid reading skills by the end of Key Stage 1 (when they are 7 years old).

The adaptation has, however, also brought its own concessions to adjust to local circumstances. For example, the family services aspect of SFA has proved to be underutilized in the United Kingdom, with the emphasis being on within-school practices. Consequently, this evaluation focuses on the school-based components of the program.

Previous Research on SFA

More than 40 empirical studies have shown positive effects of SFA on a variety of measures of student attainment (Quint et al., 2013; Slavin et al., 2009b; Slavin, Lake, Davis, & Madden, 2011). The cumulative evidence from these studies shows positive effects of SFA on a variety of measures of student achievement, as well as on assignments to special education, retentions, and other outcomes (Borman, Hewes, Overman, & Brown, 2003).

In particular, a large cluster randomized controlled trial in the United States involved a 3-year longitudinal sample of children who participated in the SFA or control condition from kindergarten through the second grade (Borman et al., 2007). Hierarchical linear model (HLM) analysis revealed statistically significant school-level effects of assignment to SFA on literacy outcomes. In addition, a synthesis of 23 studies of SFA found a mean effect size of +0.29 for students in general and +0.52 for students in the lowest 25% of their classes at pretest (Slavin et al., 2011). These studies of SFA tend to focus on the early years of learning to read, building on the evidence that early literacy acquisition can reduce educational inequalities (Lesnick et al., 2010). Studies on the later grades in primary/elementary schools by Hanselman and Borman (2013) suggest that these gains may not be sustained. Although Hanselman and Borman suggest that early exposure may be the key to later gains and these findings may be more a factor of the high levels of student turnover in high-poverty schools.

As noted earlier, the first major evaluation of the adapted program in the United Kingdom found positive impacts on literacy outcomes (Hopkins et al., 1999). Tymms and Merrell (2001) found positive effects in Year 1 and mixed effects at the end of Year 2. Using outcome estimates from the most rigorous studies with U.K. cost estimates, the Dartington Social Research Unit estimated that each pound invested in SFA yields £14.78 pounds of benefits to the individuals and society (http://dartington.org.uk/projects/investing-in-children). However, earlier U.K. studies involved a small number of schools, allowing for the possibility that school characteristics could explain observed differences. In addition, these studies began in the 1990s, whereas this study places schools within the more recent and changing educational context in England.

The study reported here is the first large-scale, comprehensive evaluation of SFA in England in the early years of schooling.

Method

Research Design

This quasi-experimental study involved 20 schools already implementing SFA and 20 schools matched to the SFA schools in terms of prior attainment and demographics, to provide a comparison cohort. As the intervention affects the whole school, a long-term evaluation was considered to be the most appropriate approach to gain an assessment of the impact on students' reading achievement over time. The large number of schools involved in the study enables the use of appropriate statistical methods for clustered (school-level) designs, with adequate statistical power to detect true differences. It also allows for an evaluation of the program as it is actually used in England as opposed to a small study that

might provide more implementation support than a typical scenario.

Sample

Schools were recruited in Spring 2008 by researchers from the Institute for Effective Education (IEE) at the University of York, using lists provided by SFA-UK. Once 20 SFA schools were recruited, researchers started to recruit control schools whose overall characteristics matched those of the SFA schools. Key matching characteristics were

- percentage of children achieving Level 4 or above on the Key Stage 2 (KS2) literacy SATs, for the 3 years prior to the intervention school adopting SFA;
- percentage of pupils in receipt of (or eligible for) free school meals (FSM); and
- percentage of pupils with English as an Additional Language (EAL).

All schools (control and intervention) agreed to allow for individual testing of their children and to allow observers access to SFA/literacy classes of the appropriate year group. In fall 2008, these were the children entering Reception classes. Children were pretested in September 2008 and then were posttested in June–July 2009 at the end of Reception, and then at the end of Year 1 and Year 2. Control schools were provided with a financial reward of £2,000 per year for participating in the study to compensate for the potential disruption to the school during the assessment period.

In addition to receiving head teacher consent, parental information and opt-out forms were sent to the schools to be distributed to all children entering Reception that September. All children in Reception who had permission to participate were then individually pre- and posttested. Each year, teachers were to continue with their normal classroom practices—whether that was SFA or any other method for teaching literacy (e.g., *Letters and Sounds*, Jolly Phonics, Read Write Inc.). The SFA schools had been involved with SFA for between 1 and 8 years. Trainers from SFA-UK made their normal implementation visits to each school throughout the year.

Table 1 shows the characteristics of the original 40 schools recruited and of the final sample of (36) schools by the end of the academic year 2010-2011. This includes the original variables on which they were matched and

- the percentage of children achieving Level 4 or above on the KS2 maths and science SATs, for the 3 years prior to the intervention school adopting SFA;
- the percentage of pupils with statements of Special Educational Needs (SEN);
- the enrollment figures; and
- the overall level of absences reported within the control compared with the intervention schools.

Over the years, four schools dropped out of the study, two control and two experimental. The experimental and control schools involved in the original baseline were well matched on all characteristics except (significantly) the percentage of children with EAL and in receipt of FSM, despite these being key matching criteria. This can be explained by the fact that, during recruitment, researchers used school-provided data as the matching criteria, which was inevitably flawed as recruitment occurred prior to children entering Reception classes and schools appear to have varied in providing school-level data and the anticipated profile of their September 2008 intake. Researchers therefore obtained official data regarding from the (then) DCSF (now the Department for Education [DfE]). In both cases, however, the intervention schools had higher levels of pupils in each category (45% vs. 22% and 44% vs. 33%, respectively). In addition, nonstatistically significant differences can be seen in the pretest measure of receptive vocabulary on starting school (the British Picture Vocabulary Scale [BPVS]), the proportion of pupils within schools with Special Educational Needs supported by outside agencies, and in the KS2 SATs results for the 3 years before take-up of SFA, again, with the advantage being in favor of the control schools.

However, because of attrition, by the end of the study, there remained a significantly higher percentage of children with EAL in the SFA schools (46% vs. 24% for the control schools). SFA schools also had nonsignificantly higher levels of children eligible for FSM (43% vs. 36%).

The sample includes schools from a range of regional contexts throughout England representing the national reach of the program, with a relatively high percentage of children eligible for FSM (approximately 40%).

Measures

The pretest, undertaken on entry to Reception (September 2008), was the British Picture Vocabulary Scale–Second Edition (BPVS-II). This is a measure of receptive vocabulary and is an English adaptation of the Peabody Picture Vocabulary Scale. Children are told a word and then asked to point to one of four pictures that represents that word. The BPVS-II was normed on a national sample of children in the United Kingdom and has a Cronbach's alpha of .93 and a split-half reliability of .86 (Dunn, Dunn, Whetton, & Burley, 1997).

During June/July 2011, the same children at the end of Year 2 were posttested using the Word Identification and Word Attack scales of the Woodcock Reading Mastery Tests–Revised (WRMT). The Word Identification scale measures the child's ability to read isolated words and the Word Attack scale assesses children's ability to decode and "sound out" nonsense words. The WRMT was normed on a U.S. national sample of children and the internal reliability coefficients for the two scales used were .97 and .87, respectively (Woodcock, McGrew, & Mather, 2001).

Table 1. Comparison of Baseline Characteristics for Success for All (Intervention) Schools and Control Schools.

Variable	Condition	n	M	SD	t	n	M	SD	t
Pretest BPVS[a]	Intervention	20	36.80	5.77	−1.531	18	37.88	5.75	0.607
	Control	20	39.58	5.69		18	39.00	5.03	
KS2 English SATs[b]	Intervention	19	51.58	15.12	0.730	17	50.28	15.11	0.944
	Control	17	55.10	13.65		15	55.24	14.59	
KS2 maths SATs[b]	Intervention	19	57.42	11.87	−1.484	17	57.12	12.25	−1.293
	Control	17	51.65	11.40		15	51.64	11.97	
KS2 science SATs[b]	Intervention	19	72.70	18.17	−0.725	17	71.392	18.70	−0.715
	Control	17	68.57	15.75		15	66.96	16.05	
Enrollment[c]	Intervention	19	346.32	132.03	−0.660	17	326.41	120.78	−0.050
	Control	18	319.28	116.23		16	324.31	122.35	
% pupils eligible for free school meals	Intervention	20	44.08	18.62	−2.060*	18	43.41	19.05	1.403
	Control	20	33.12	14.82		18	35.74	13.17	
% of pupils with SEN, supported at School Action[c,d]	Intervention	19	22.46	12.17	0.097	17	15.94	8.31	0.197
	Control	18	22.84	11.23		16	16.48	7.56	
% of pupils with SEN, supported at School Action Plus[c,e]	Intervention	19	14.94	10.77	−1.022	17	11.23	5.29	−0.905
	Control	18	11.84	7.22		16	9.74	4.05	
% of pupils with EAL	Intervention	20	45.21	34.44	−2.294*	18	45.87	34.26	2.052*
	Control	20	22.19	28.75		18	23.91	29.79	
Overall absences[c]	Intervention	19	6.937	1.58	−1.345	17	6.77	1.56	−0.483
	Control	18	6.36	0.93		16	6.56	0.71	

Source. Department for Education, including published Performance Tables, 1994-2012 (http://www.education.gov.uk/schools/performance/archive/index.shtml).
Note. BPVS = British Picture Vocabulary Scale; KS2 = Key Stage 2; SEN = Special Educational Needs; EAL = English as an Additional Language.
[a]Authors' own data.
[b]Percentage of pupils achieving Level 4 or above for the 3 years prior to the intervention school in each matched pair adopting the SFA program. One intervention and three control schools in the original cohort of 40 schools were lower schools (i.e., took pupils up to age 9) during the 3-year period when the average KS2 English SATs were calculated for this study so we do not have results for the specified time frame. During the posttest period, this was one intervention and two control schools.
[c]One intervention and two control schools were lower schools at the beginning of the study, so enrollment figures, percentage of SEN pupils at School Action or School Action Plus, and overall absence data were not available from the published Performance Tables.
[d]A pupil identified by the teacher as having special educational needs with additional resources and interventions to address those provided within the school.
[e]As with School Action (detailed above) but where additional advice or support is provided by outside specialists.
*p < .05.

In addition, testers administered the York Assessment of Reading Comprehension (YARC), a standardized measure of accuracy, reading rate, and comprehension. The YARC was normed on a U.K. national sample of children. The reliability coefficients for the three subtests for Year 2 children were .87, .95, and .62, respectively. All assessors were hired, trained, and supervised by the researchers at the IEE. Assessors were not made aware of schools' treatment conditions.

Intervention schools were visited by SFA consultants and researchers to assess implementation fidelity and control schools were visited by researchers to observe general literacy practices and assess to what extent key elements of the SFA program (in particular cooperative learning) were being practiced.

As noted previously, four schools left the study, two intervention and two control schools, making assessment of those children impossible. This was primarily due to changes in head teacher or in the direction of the school. Given the longitudinal nature of the study, a number of children left due to the high rate of movement between schools often found in vulnerable communities. As this was a longitudinal study and the analyses completed at the school level, subjects were not replaced with new students.

Analyses

This quasi-experimental cluster evaluation was analyzed using a HLM with school as the unit of analysis (Raudenbush & Bryk, 2002). All multilevel models were estimated using the HLM software's restricted maximum likelihood (REML) estimation procedure (Raudenbush, Bryk, Cheong, & Congdon, 2000). Pretests (the BPVS) were used as covariates. This multilevel approach is the optimal design for school-based interventions. It addresses the effects of students being clustered within schools, and it is well aligned with the theory of how this educational intervention works: as a coordinated, systemic initiative delivered by school-level elements acting in concert. Multilevel analysis greatly reduces statistical power, requiring many more schools than

Table 2. Multilevel Models Predicting Student and School-Level Literacy Outcomes: Woodcock Reading Mastery Test.

Level 1 model : $Y_{ij} = \beta_{0j} + r_{ij}$.

Level 2 model : $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Mean BPVS})_j + \gamma_{02}(\text{SFA})_j + u_{0j}$.

Fixed effect	Word identification (I = 415, C = 471)			Word attack (I = 415, C = 471)		
	Effect	*SE*	*t*	Effect	*SE*	*t*
School mean achievement						
Intercept	0.01	0.04	0.19	0.01	0.04	0.11
Mean BPVS pretest	0.02	0.10	0.16	−0.07	0.08	−0.90
SFA assignment	+0.20*	0.09	2.33	+0.25**	0.08	3.16
Random effect	Estimate	χ^2	*df*	Estimate	χ^2	*df*
School mean achievement	0.18	59.15	33	0.15	49.68	33
Within-school variation	0.98			0.98		

Note. BPVS = British Picture Vocabulary Scale; SFA = Success for All; I = Intervention; C = Control.
*p < .05. **p < .01.

would be needed in individual-level analysis, but individual-level tests of statistical significance, which assume that the outcome for an individual is independent of that for any other student, are inappropriate for evaluations of school-level interventions.

Using HLM, one may simultaneously model both student- and school-level sources of variability in the outcome (Raudenbush & Bryk, 2002). Specifically, we developed two-level hierarchical models that nested students within schools. The fully specified Level 1, or within-school model, nested students within schools. The linear model for this level of the analysis is written as

$$Y_{ij} = \beta_{0j} + r_{ij},$$

which represents the summer posttest achievement for student i in school j regressed on the Level 1 residual variance, r_{ij}.

At Level 2 of the model, we estimated SFA treatment effects on the mean posttest achievement outcome in school j. We included a school-level covariate, the school mean BPVS pretest score, to help reduce the unexplained variance in the outcome and to improve the power and precision of our treatment effect estimates. The fully specified Level 2 model is written as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Mean BPVS Pretest})_j + \gamma_{02}(\text{SFA})_j + u_{0j},$$

where the mean posttest intercept for school j, β_{0j} is regressed on the school-level mean BPVS score, the SFA treatment indicator, plus a residual, u_{0j}.

Results

Empirical Analyses of Reading Achievement

For each of the achievement posttest outcomes, a series of multilevel models were specified. The analyses for the

WRMT subtests are shown in Table 2 and those for the YARC are shown in Table 3. These show a school-level significant effect on the WRMT Word Identification (Effect Size (ES) = +0.20, p < .03) and Word Attack (ES = +0.25, p < .01). On the three YARC scales, effects were directionally positive but not statistically significant, with effect sizes as follows: Rate (ES = +0.11, *ns*), Comprehension (ES = +0.06, *ns*), and Accuracy (ES = +0.12, *ns*). In addition, we tested for the random effects of the BPVS pretest at the student level across schools but found no significant effects.

Implementation Observations

Observations were designed to enable researchers to assess the fidelity with which the SFA program was being implemented and to determine whether critical components of the SFA instructional process were being implemented in control schools. They included factors relating to literacy instruction, cooperative learning, and assessment. These observations were in addition to the regular, routine visits by SFA advisors. There were 19 items on the implementation checklist in total: 14 related to teacher behaviors (e.g., "Teacher models phonics skills correctly") and 5 to pupil behaviors (e.g., "Children read the texts fairly accurately, self-correcting if errors are made"). Each item was rated on a scale from 0 (observed none of the time) to 3 (present virtually all of the time), with an option for "Unobserved." During the 2010-2011 academic year, 10 of the intervention schools were visited once and 8 were visited twice during the school year. Nine out of 18 control schools were visited once each during the school year. The visits were carried out by two researchers who initially co-observed lessons at two schools to enhance the consistency of evaluator ratings.

For both the intervention and control schools, where possible, observers attempted to observe more than one class or literacy group within each school, covering a range of ability levels within Year 2. At most SFA schools, this involved two

Table 3. SFA Versus Control Multilevel Models Predicting School-Level Literacy Outcomes: YARC.

Fixed effect	YARC rate ability (I = 356, C = 381)			YARC comprehension (I = 407, C = 461)			YARC accuracy (I = 412, C = 468)		
	Effect	SE	t	Effect	SE	t	Effect	SE	t
School mean achievement									
Intercept	−0.03	0.05	−0.55	0.03	0.04	0.73	0.01	0.04	0.35
Mean BPVS pretest	−0.08	0.12	−0.70	0.45**	0.07	6.41	0.16	0.08	1.89
SFA assignment	+0.11	0.14	1.09	+0.06	0.08	0.69	+0.12	0.08	1.46
Random effect	Estimate	χ^2	df	Estimate	χ^2	df	Estimate	χ^2	df
School mean achievement	0.24	74.99	33	0.17	57.47	33	0.16	52.67	33
Within-school variation	0.97			0.97			0.98		

Note. SFA = Success for All; YARC = York Assessment of Reading Comprehension; BPVS = British Picture Vocabulary Scale; I = Intervention; C = Control.
**p < .01.

or three classes, whereas control school visits were more likely to involve one or two classes (usually depending on the size of school). Within SFA schools, teachers usually taught groups of students from different year levels but all at one reading level. Students changed classes at reading time to make this possible as the SFA model prescribes. In control schools, teachers generally taught several reading groups within the year-level class they taught all day.

Phonics. In all the schools visited who had a phonics lesson observed, researchers saw some form of instruction in synthetic phonics in use—teaching letter sounds and then blending these sounds together to read whole words. *Letters and Sounds* (Department for Education and Skills, 2007) was the most widely used system in control schools, although other programs were also mentioned by teachers or recognized by observers, including Jolly Phonics and Read Write Inc.

Cooperative learning. A key feature of the SFA program is cooperative learning. Partner work was observed in many control as well as intervention schools. At its most limited, this involved sharing of resources within pairs. At its most sophisticated, it involved partner reading—a strategy observed in most SFA schools, whereby children work together and take turns reading and summarizing text. The most popular form of partner work in both control and intervention schools involved partner talk, the sharing of ideas, formalized in the SFA schools by such strategies as "Think-Pair-Share." In most SFA schools and some control schools, the partnerships were formally assigned by the teacher, who directed children to work together in specific pairs.

Cooperative teaching strategies and pupil behaviors were much less frequently observed in control than SFA schools. For instance, almost a half of SFA classes scored 3 (i.e., "present almost all the time") on the rating "Teacher has pupils working in heterogeneous partners or teams"

compared with only one instance in the control schools. Almost a third of classes in SFA schools scored 3 on "Pupils display cooperative behaviors in group work," but no control classes were rated 3. Although control classes might be organized to sit in groups, or sometimes even perform tasks in groups, cooperative learning behaviors were rare.

SFA implementation fidelity. It is important to note that there was variation in implementation of the SFA program among the intervention schools. Project researchers rated the implementation of the SFA Year 2 teachers on a scale of 0 to 3, with 0 being "No fidelity to the program or cooperative learning in place" and a score of 3 meaning "materials and routines are followed with fidelity and cooperative learning is embedded within the school culture." Given the nature of SFA, with its cross-class, cross-year reading classes grouped by reading level, this is an important indicator of overall fidelity to the program. An average fidelity score was calculated across reading classes within schools and, where more than one visit was made, an average was then obtained across visits. This resulted in an implementation fidelity score for each school on a scale of 0 to 3, with 0 being very weak implementation fidelity and 3 being very high. Scores were calculated at a school level because the nature of the SFA program means that children are likely to be moved through a number of groups during the year. In most, although not all cases, implementation fidelity was very similar within any one school.

Of the 18 intervention schools, 10 schools received a rating of 3, 7 were rated 2, and 1 was rated 1. Where more than one class was observed, scores have been averaged and rounded to the nearest whole number to enable comparison. A similar pattern of ratings was recorded by SFA advisors, with the caveat that they are not directly comparable because the advisors were rating SFA implementation throughout the school whereas the research study focused on classes

containing Year 2 children. The key element that varied among schools was the extent to which cooperative learning was followed with consistency. Additional issues elicited from informal interviews after observations in SFA schools included the following:

- Lessons in some schools were reduced from 90 min per day, 5 days a week to less than 90 min a day and/or less than 5 days a week.
- Other reading schemes were used within the class-room within the designated literacy time in a few schools, thereby diluting the "whole program" approach advocated by SFA.
- Sometimes the organization of the schools did not facilitate the mixing of year groups within a school.

This latter point was compounded by the fact that the Year 2 children were due to take the end of Key Stage 1 SATs during the summer of 2011. This meant that sometimes the mixed year group approach advocated by SFA was not followed through at this stage because of the importance to schools of ensuring that children reached their required "competences" and the felt need to provide additional coaching to Year 2 children who were less likely to meet those requirements.

SFA implementation outcomes. As indicated above, the major-ity of the SFA schools had a medium or high implementation rating (17 schools). Additional analysis was therefore con-ducted to compare the outcomes of these two groups. How-ever, no significant differences on both pre- and posttests were found between the two groups.

Discussion

These results indicate educationally significant results for the word-level reading assessments. There are positive effect sizes for all the assessments undertaken but only the effects for the Word Identification and Word Attack subtests of the WRMT were statistically significant. There may have been additional beneficial outcomes of the SFA program that we did not measure. For example, one teacher indicated that behavior tended to improve with the cooperative learning framework.

These outcomes of this evaluation of SFA are noteworthy for several reasons. First, we were able to obtain the coopera-tion of a sufficient number of SFA and control schools to provide an acceptable level of statistical power to detect school-level effects within a multilevel model framework. SFA and control samples were reasonably well matched on a variety of baseline characteristics, including many demo-graphics and the BPVS pretest scores.

Second, there were positive effect sizes demonstrated for the intervention schools on the beginning reading skills of word identification and decoding. Despite the strong focus on phonics and decoding in English schools in recent years, SFA schools still greatly improved phonics and word-level outcomes in comparison with control schools.

Third, the treatment fidelity and SFA implementation quality seemed reasonably good. There were similarities between the intervention and control schools in that they all were using some form of synthetic phonics instruction, yet cooperative learning strategies were generally absent within control schools.

The pattern of the Year 2 treatment effects appears to be consistent with the previous U.S. studies, the SFA program theory, and more general research and theory on the develop-ment of young children's emergent literacy skills. We found effects of both statistical and educational significance. The logic model behind the program is consistent with more gen-eral theories of how young children develop as emergent readers (Snow, Burns, & Griffin, 1998). Specifically, power-ful decoding strategies and phonemic awareness, as stressed by the SFA Reception and Year 1 program, are key building blocks upon which children can develop a broader range of skills. However, comprehension results, although generally positive, were not statistically significant and this is worthy of further research. The largest cluster randomized evalua-tion of SFA in the United States, by Borman et al. (2007), found positive effects of SFA in word-level outcomes in kin-dergarten and first grade, but comprehension effects did not appear until second and third grades. If the current evalua-tion were continued for another year, it might also find sig-nificant comprehension effects as Borman et al. (2007) did.

This aside, implementation was not perfect and this has always been an issue with large-scale program evaluation. There was variability in the level of implementation among SFA schools, although the majority of schools implemented the program with at least moderate fidelity (10 out of 18). Yet, they did not always do so for 90 min, 5 days a week. This has implications for the transferability of programs from U.S. to U.K. settings and for the implementation of pro-grams per se. Yet, the nature of the SFA program, involving as it does small group teaching, with these groups regularly changing during the school year, presented its own research problems.

In addition, although on many variables the imple-menta-tion and control schools demonstrated similar profiles, the school matches were not ideal. Researchers were originally reliant on largely school-provided data but official data col-lected later, and not available at the time of initial recruit-ment, suggests that on key variables, such as the proportion of children with EAL and the proportion in receipt of FSM, the intervention and control schools had significant differ-ences. This raises issues around the difficulties of recruiting schools that are already struggling to research studies. In each case, these differences did not favor the SFA schools, yet the fact that SFA schools had produced less favorable KS2 SATs results in English in the 3 years prior to starting the SFA program than their matched control schools suggests

that they were more open to change than perhaps their similarly performing peers at the time (i.e., not the control school group).

The findings of this experiment are important for policy and practice. Overall, this study responds to the doubts that have been raised about the viability and appropriateness of large-scale evaluations in school settings (Cook & Payne, 2002). As a large, lengthy field evaluation, rather than a relatively small, brief experiment, the results of this study have strong external validity and relevance for policy and practice. This project ties together two central themes of educational research and policy: the scale-up, or replication, of school-based interventions and the development of high-quality evidence of their causal effects. These outcomes have established that large-scale quasi-experiments involving replicable school-based interventions are possible in the United Kingdom, while highlighting some of the difficulties in doing so, in particular with regard to recruitment and implementation. It also suggests that there is a need for more randomized controlled trials to assess the impact of programs within U.K. schools.

As government policies now provide schools in England more autonomy and less top-down prescription, more schools should have opportunities to choose among effective and replicable interventions to provide educators with tools capable of narrowing the achievement gap between high and low income populations. With the movement toward evidence-based education in England, hopefully more programs such as SFA will be created, evaluated, and disseminated to help them do so, alongside more research on the barriers to implementation and collaboration with schools to provide more effective strategies to reduce inequalities.

Authors' Note

Acknowledgment

Declaration of Conflicting Interests

Funding

References

Allen, G. (2011). *Early intervention: Smart investment, massive savings* (Independent review on early intervention delivery). London, England: Cabinet Office.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*, 125-230.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*, 701-731.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. F. Boruch (Eds.), *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution, 150-178.

Department for Children, Schools and Families. (2009). *Analysis and evidence strategy 2009-10: Delivering evidence based policy*. London, England: Author.

Department for Education and Skills. (2007). *Letters and sounds: Principles and practice of high quality phonics six-phase teaching programme*. London, England: Author.

Department for Work and Pensions. (2009). *Households below average income: An analysis of the income distribution 1994/95–2007/08*. London, England: Author. Retrieved from https://www.gov.uk/government/collections/households-below-average-income-hbai–2

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. New York, NY: Russell Sage Foundation.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale* (2nd ed.). London, England: GL Assessment.

Goodman, A., Sibieta, L., & Washbrook, E. (2009). *Inequalities in educational outcomes among children aged 3 to 16: Final report for the National Equality Panel*. London, England: Institute for Fiscal Studies.

Hanselman, P., & Borman, G. D. (2013). The impacts of Success for All on reading achievement in grades 3–5: Does intervening during the later elementary grades produce the same benefits as intervening early? *Educational Evaluation and Policy Analysis*, *35*, 237-251.

Hills, J., Brewer, M., Jenkins, S., Lister, R., Lupton, R., Machin, S., & Riddell, S. (2010). *An anatomy of economic inequality in the UK: Report of the National Equality Panel*. London, England: Government Equalities Office.

Hopkins, D., Youngman, M., Harris, A., & Wordsworth, J. (1999). Evaluation of the initial effects and implementation of Success for All in England. *Journal of Research in Reading*, *22*, 257-270.

Law, Y. K. (2008). Effects of cooperative learning on second graders' learning from text. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *28*, 567-582.

Lee, V. E., & Burkham, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Lesnick, J., George, R., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment?* Baltimore, MD: Annie E. Casey Foundation.

Mongon, D., & Chapman, C. (2008). *Successful leadership for promoting the achievement of white working class pupils*. Nottingham, UK: National College of School Leadership. Retrieved from www.nationalcollege.org.uk/index/docinfo.htm?id=17406

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development.

Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, *13*, 155-190.

Quint, J. C., Balu, R., DeLaurentis, M., Rappaport, S., Smith, T. J., & Zhu, P. (2013). *The Success for All model of school reform: Early findings from the investing in innovation (i3) scale-up*. MDRC. Retrieved from http://www.mdrc.org/publication/success-all-model-school-reform

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: SAGE.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. T. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Rose, J. (2006). *Independent review of the teaching of early reading*. London, England: Department of Education and Skills.

Rowan, B., Camburn, E., & Barnes, C. (2004). Benefiting from comprehensive school reform: A review of research on CSR implementation. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research*. Washington, DC: National Clearinghouse for Comprehensive School Reform, 1-52.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston, MA: Allyn & Bacon.

Slavin, R. E. (2009a). Cooperative learning. In G. McCulloch & D. Crook (Eds.), *International encyclopedia of education*. Abington, UK: Routledge, 177-183.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009b). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, *79*, 1391-1465.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*, 1-26.

Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009c). *Two million children: Success for All*. Thousand Oaks, CA: Corwin Press.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: Committee on the Prevention of Reading Difficulties in Young Children, National Research Council.

Stevens, R. J. (2003). Student Team Reading and Writing: A cooperative learning approach to middle school literacy instruction. *Educational Research and Evaluation*, *9*, 137-160.

Strand, S. (2008). *Minority ethnic pupils in the Longitudinal Study of Young People in England: Extension report on performance in public examinations at age 16* (DCSF Research Report 029). London, England: Department for Children, Schools and Families. Retrieved from www.education.gov.uk/research/data/uploadfiles/dcsf-rr029.pdf

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The Effective Provision of Pre-School Education (EPPE) project* (Technical Paper 12—The Final Report: Effective Pre-School Education). London, England: Department for Education and Skills, Institute of Education, University of London.

Tymms, P., & Merrell, C. (2001). *Success for All quantitative evaluation report: Second year*. Durham, UK: Curriculum Evaluation and Management Centre, Durham University.

Wilkinson, R., & Pickett, K. (2010). *The spirit level: Why more equal societies almost always do better*. London, England: Penguin Books.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Author Biographies

Louise Tracey is a Research Fellow in the Institute for Effective Education. Her research interests include, research methods, professional development, early years education, and literacy.

Bette Chambers is is Director of the Institute for Effective Education. Her current work focuses on evaluating replicable programmes that use co-operative learning and technology-embedded instruction, particularly in early childhood education and early literacy. She actively promotes the use of evidence-based programmes and practices in education.

Robert E. Slavin is Director of the Center for Research and Reform in Education at the Johns Hopkins University School of Education in Baltimore, and Professor at the Institute for Effective Education at the University of York. His areas of interest include whole-school reform, cooperative learning, and research review.

Pam Hanley is a Research Fellow in the Institute for Effective Education. She has been involved in evaluations of science, maths and literacy interventions and CPD programmes at both primary and secondary school level.

Alan Cheung is currently associate professor in the Department of Educational Administration and Policy at the Chinese University of Hong Kong. His research areas include school reform, educational technology and research reviews.