

Running Head: RETROSPECTIVE OVERVIEW AND COMMENTARY

**This is a prepublication version of:**

(2009), Julian R. Betts, “The San Diego Blueprint for Student Success: A Retrospective Overview and Commentary”, *Journal of Education for Students Placed at Risk*, (14:1), pp. 120-129.

The San Diego Blueprint for Student Success: A Retrospective Overview and Commentary

Julian R. Betts  
Department of Economics  
University of California San Diego

Requests for reprints should be sent to: Julian R. Betts, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508. Email: [jbetts@ucsd.edu](mailto:jbetts@ucsd.edu).

## Abstract

In a series of reforms started in 1998 and formalized with the 2000 introduction of the Blueprint for Student Success, the San Diego Unified School District (SDUSD) implemented a wide-ranging series of reforms designed to boost the literacy skills of students. This paper provides an overview of the reforms, assesses the contributions made by the papers in this special issue of *JESPAR*, and highlights complementarities between this research and related quantitative research. Overall, the evidence suggests that the model of professional development in the 8<sup>th</sup> largest district in the nation took root and led to tangible changes in the classroom. Achievement analyses by Betts (2005), Betts, Zau, and King (2005), and Betts, Zau and Koedel (2008) suggest that the reading achievement of students improved in the lower grades as a result of the reforms, but dipped somewhat in high school. The paper concludes with thoughts on the political and financial challenges to sustaining such reforms and the practical difficulties of evaluating their impact.

Keywords: Literacy, literacy reform, professional development, value-added achievement models, instructional leadership, remediation, summer school

## **Overview of the San Diego Blueprint for Student Success and Outline of the Paper**

Motivated by large variations in achievement among students, senior administrators in the San Diego Unified School District (SDUSD) embarked in 1998 on a seven-year reform effort to boost student literacy. In 2000, the district officially expanded and formalized these literacy reforms. The program, known as the Blueprint for Student Success (SDUSD, 2000), quickly won nationwide attention, both for the broad and ambitious sweep of the reforms, and for the bitter political controversy that ensued. The Blueprint sought to boost literacy through three strategies: preventive measures for students at grade level in reading, interventions targeted at students identified as below grade level in reading, and widespread professional development for teachers. The main pedagogical focus of the Blueprint was “balanced literacy,” which emphasized active participation by students in reading, speaking, and writing, with the teacher initially actively supporting the students and then gradually demanding more of the students as they progressed.<sup>1</sup>

One of the most important forms of professional development for teachers was the provision of literacy peer coaches who worked with classroom teachers to improve their ability to teach reading and writing. But, as detailed in the accompanying papers in this special issue of *JESPAR*, professional development occurred on many levels and became a pervasive and constant feature in the lives of teachers in the district.

The student-based interventions included extra time on reading for students in the form of extra classes before and after school, double- and triple-length English classes, and summer school. Preventive measures included double-length English classes for students in their first

---

<sup>1</sup> For more details on the balanced literacy strategy, see Betts (2005), Stein, Hubbard, and Mehan (2004), and Betts, Zau, and King (2005).

grade of middle or junior high school who were also reading at adequate levels and, at the lowest-performing elementary schools, additional reading resources and in some cases a longer school year.

With the advent of the accountability provisions of No Child Left Behind (NCLB), stories of districts engaging in major reform efforts have become commonplace. What makes San Diego's story interesting? First, at its most prosaic, this reform predated the NCLB reforms by several years.

Second, and more fundamental, the sheer scope of the reforms went far beyond a more typical urban district reform effort. Professional development was deconstructed and rebuilt from the ground up, and spending on professional development rose sharply, in part financed by a variety of non-profit foundations. Hightower (2002) reported that spending on professional development mushroomed from about \$1 million per year before the reforms to about \$65 million at the peak of the Blueprint. Further, following on the heels of the Chicago Public Schools' efforts, SDUSD used diagnostic testing to identify students who were somewhat or dramatically behind grade level, and then targeted an ambitious array of interventions at these students. Two fundamental design elements provided the undergirding for all of these interventions. Professional development to help teachers provide better literacy supports to their students came first. Second, additional time on task for students identified as lagging behind, with support from teachers who had received the professional development, became the main conduit of assistance to students in the greatest academic need.

A third reason for interest in the San Diego reforms is that they generated intense controversy, and indeed were substantially dismantled after Superintendent Alan Bersin left the district in 2005. A constant criticism of the Blueprint, from the teachers' union (the San Diego

Education Association), and from individual teachers, centered on the complaint that the reforms were top-down and that their implementation snubbed the considerable working pedagogical knowledge already possessed by San Diego's more highly qualified teachers. Williams (2005) provided a vivid account of the tensions between district leadership and the teachers' union. In considering the history of the Blueprint years, the question becomes: Could the reforms have been implemented differently? Would a more consultative approach have led to a longer-lasting set of reforms, or would the need for compromise have led to a less coherent and therefore less effective program?

With such an ambitious and controversial reform, it is easy to imagine how diverse approaches to studying it could lead to quite different conclusions about the effects of the reform and its implementation. One cannot help but be reminded of the old tale of the six blind men who were asked to describe an elephant after feeling different parts of the elephant's body. Of course, in that story, each person came to a markedly different conclusion. Only in the end did the complete picture of the elephant emerge. With this concern in mind, the following sections of this paper compare the findings of the papers in this issue with a number of previous research efforts. Ultimately, a fairly clear picture of the San Diego reforms does emerge. Indeed, the analyses in this issue and elsewhere complement each other quite well, and suggest many strengths and weaknesses of the approach San Diego adopted.

### **Contributions in This Issue and Complementarities with Earlier Research**

The contributions in this issue are well summarized in the overview paper by O'Day and Quick and will not be itemized in a mechanical way here. Rather, my goal in reflecting on these papers was to think about what we have learned and which aspects of the reform we still have not explored sufficiently.

The links from leadership to professional development to instructional practice is studied in these papers in a way that would be impossible without considerable expenditures on qualitative and mixed-methods analysis. The papers summarize how professional development, including peer coaching, and the overall approach to instructional leadership, with seven instructional leaders working with individual principals, were all designed to provide supports—and oversight—for classroom teachers.

The papers in this issue combine an observational approach that tells us much about how teachers' classroom practice reflected the professional development. Some of the papers, quite uniquely, also address how these forms of classroom practice fed through to student achievement. None of the papers can provide a bird's eye view of the overall effect of the reforms, but that is not the goal of these papers.

One reason why it is so important to observe broad-based reform at the classroom level is to confirm that ideas promulgated from the central administration actually make their way into the classroom. Bitter, O'Day, Gubbins, and Socias (in this issue) confirm that teachers really did implement balanced literacy. This in some sense provides validation for any positive or negative overall effects on achievement that emerge from district-wide studies of the relation between the reforms and students' gains in achievement.

An especially important aspect of this paper is its attempt to find a link between teachers' classroom practice and student achievement gains. The authors establish that use of literacy teaching methods that encourage discussions of higher-level meaning of text are associated with larger gains in reading scores. A possibility is that teachers who adopted these practices might have been the more effective teachers to start with, and may have used these higher-level methods even without the professional development provided by the district. For this reason the

authors do not claim that the reforms themselves caused both teacher practice and test scores to change. But it is an important finding suggesting positive systemic effects of the literacy approach.

Also, the apparent effects from these higher-level practices are modest, which explains why those who have compared test score growth between SDUSD and other large urban California districts may not have found much. See for instance Bassok and Raymond (2005), who in essence ran a horse race between SDUSD and other large California districts, and found, very roughly, comparable growth in averages of test scores among the districts studied. Another reason why this approach may be too blunt to reveal causal effects of the reforms is that it implicitly assumes that no other district in California was implementing reforms of its own.

Quick, Holtzman, and Chaney (in this issue) report on a study of 100 teachers in nine elementary schools who agreed to keep logs on professional development activities in 2004-2005. This timing is noteworthy: It stands to reason that professional development takes time to work. Indeed, as detailed in the next section, the quantitative longitudinal analysis of student achievement in the district by Betts, Zau, and King (2005) and updated by Betts, Zau, and Koedel (2008) suggests that the addition of peer coaches across San Diego's schools had virtually no effect on student achievement gains in the first few years. But by the later years of the reform, the level of provision of peer coaching became strongly associated with achievement gains in reading.

The analysis by Quick, Holtzman and Chaney suggests that the peer coaching innovation had by 2004-2005 probably achieved some of the changes in teaching practice that reform authors (former Superintendent Alan Bersin and former Chancellor of Schools Anthony Alvarado) had initially envisaged. The teacher logs analyzed by Quick and her co-authors shows

that teachers' concept of good professional development aligns fairly closely with the standards outlined in the analysis of the Eisenhower Professional Development Program (Garet Porter, Desimone, Birman, & Yoon, 2001). Moreover, 84% of teachers reported that the actual professional development was consistent with their students' learning needs to a moderate or large extent. This near unanimity is quite striking given the anger and opposition that initially greeted the "top-down" implementation of the reforms beginning in 1998-1999 through the official introduction of the Blueprint in 2000-2001.

An emergent theme again displays itself: Initial reaction and impact can be quite different from longer-term reaction and impact.

A second important finding from this paper is that teachers viewed collaboration with other teachers as key, and reported less time spent with peer coaches, probably because with only one or two peer coaches per school, the peer coaches were spread rather thinly. One reasonable interpretation of this finding is that the professional development model, although implemented originally through formal classes for teachers, peer coaching, and observation and feedback from principals, gradually took on a life of its own. Further, this more collaborative and less hierarchical approach was probably essential given limits on the number of peer coaches provided to schools.

Graczewski, Knudson, and Holtzman (in this issue) provide valuable insights on the role played by principals as instructional leaders during the literacy reforms. This paper, which echoes many of the earlier findings by Stein, Hubbard, and Mehan (2004) and Quick et al. (2003), suggests that some principals did not have the content knowledge in literacy instruction to serve as effectively as possible as instructional leaders, and that teachers often found other teachers to be more helpful in designing their literacy instruction. The manager/worker tensions



inherent in the principal/teacher relationship sometimes further reduced principals' effectiveness as instructional leaders. Nothing in these results suggests that principals should not serve as instructional leaders. Rather, they imply that the reforms could have become more effective at the classroom level if these issues of subject-matter knowledge and tensions within schools had been more fully addressed.

In a similar vein, Hannaway and Stanislawski (2005) pointed out that in 2000-2001, only 5% of school administrators had received training from the Educational Leadership Development Academy (ELDA), which was the district's main conduit for acquainting principals and other administrators with the details of the literacy reforms, and by 2004-2005 this number had risen to just 47% of school administrators. Schnur and Gerson (2005) questioned whether principals had received sufficient training, and, more important, whether the high levels of endurance and engagement that they detected in San Diego principals could be sustained for long periods.

The paper by O'Day (in this issue) focuses on the experiences of nine elementary schools with English Learner (EL) students and how they fared with the literacy reforms. This question is critical for a district such as SDUSD, where in a typical year just under one third of students are ELs. O'Day characterizes the district's literacy interventions as initially not distinguishing between EL students and native English speakers who had reading difficulties. However, by 2004-2005, she reports, differentiation to respond to the more basic needs of English language development for EL students had started to gain ground. Her point is a sensible one. Again, it does not suggest that the literacy reforms failed to help EL students, but rather that a more nuanced distinction between student needs could have accelerated EL students' progress.

### **Complementarities with Quantitative Research on the San Diego Reforms**

The previous section has already pointed out some of the complementarities between the findings of the papers in this issue and the findings of earlier work. In particular, the papers reinforce some of the concerns raised by Stein et al. (2004) about how the reforms were implemented at the classroom level, and the difficulties in bridging the divide between teachers and instructional leaders such as peer coaches and principals. The papers also show how thinking gradually evolved towards the end of the reform period (1998-2005).

The papers do an exceptionally good job of portraying how professional development worked, and how it did and did not translate into changes in classroom practice. The precise focus of these papers on professional development is welcome. But it is also worth examining the other pillar of the Blueprint, which was to use student testing to identify students lagging behind, and then to intervene by placing those students into additional literacy classes during school, before and after school and in summer. This concept of “additional time on task” for struggling students was every bit as central to the Blueprint reforms as professional development itself.

Another way of looking at the elephant is to ask: “What was the overall effect of these reforms on student achievement”? A book by Betts et al. (2005) and a follow-up paper by Betts et al. (2008) provide the only longitudinal student-level analyses of the overall effects of the reforms on reading achievement. The findings from these studies are relevant because they address the overall questions “Did the literacy reforms boost student achievement?” and “Which elements of the reform were most helpful?” Additionally, the update provided by the latter paper bears directly on the intimation in many of the papers in this issue of *JESPAR* that over time the reforms evolved, attitudes changed, and short-run implementation issues were ironed out at least partly.

Conversely, the case studies on professional development cited in the previous section provide considerably more detail on the nature and extent of the various training activities for San Diego's teachers than an overall evaluation of student achievement could ever provide. Betts et al. (2005) and Betts et al. (2008) reported on detailed information on student interventions, but on the professional development side they focused mainly on the number of peer coaches at each school. (Teachers participated in many district-provided training sessions, but the district did not keep records of these sessions in a format amenable to research. The main problems, district administrators found, were that teachers logged themselves into training sessions and that many handwritten entries often did not match teacher names.)

Although these two achievement studies do not control for each specific instance of teacher training, they do control for variations in the provision of peer coaching across schools and time, and for variations among students in participation in a large number of reading interventions.

### ***Did the Literacy Reforms Boost Student Achievement?***

Betts et al. (2005) estimated value-added models of test score growth for individual students, which through the use of student-specific intercepts allowed for different growth trends among students. They then tested whether exposure to specific aspects of the literacy reforms once they started (e.g., whether students had a peer coach at their school, or whether they participated in the after-school reading program) were associated with deviations from the trend in the individual students' test-score trajectories.

Many elements of the reforms were associated with boosted reading scores for students. Because this study included test scores from Grades 2 through 11, it became possible to compare the relative effects across grade-spans. In order to assess the overall effects of the Blueprint, the

researchers focused on all students who had been exposed to any of the reform elements, and then asked, based on the regression estimates, what would these students' scores have been had the Blueprint not been put in place?

This exercise suggested major benefits accrued to low-scoring students in elementary schools, positive but small benefits accrued in middle schools, and small negative achievement effects resulted in high schools.

To obtain these estimates, Betts et al. (2005) followed over a two-year period all students who entered Grades 3, 6 and 9 in Fall 2000. Over the period from Summer 2000 to Spring 2002, they measured each student's participation in each aspect of the Blueprint. They calculated where these students finished in the district's own distribution of test scores in Spring 2002, and then asked the counterfactual question: "Where would these students have ranked if the Blueprint had not existed?"

---

Insert Figure 1 Here

---

Figure 1 shows the results for elementary school students. The authors divided students in these three grade cohorts into 10 equally sized groups, or deciles, based on their Spring 2002 reading test scores. The dark bars show where these Blueprint participants would have ranked without the Blueprint. In a world without the Blueprint, the vast majority of these students would have been in the bottom five deciles, which is not surprising given that the district's assessment tools had previously identified these students as lagging behind. The lighter colored bars show where these students actually ended up in the test score distribution after having participated in the Blueprint.

The results strikingly indicate that the Blueprint shifted well over 10% of these students out of the bottom two deciles of reading achievement and into higher deciles.

Results for middle schools were similar but smaller: just over 4% of participants appear to have been shifted out of the bottom two deciles of test-score performance. High school results displayed a perverse result: The Blueprint is predicted to have shifted just under 5% of participants *into* the two lowest deciles of test-score performance.

With the major exception of high school, then, the authors concluded that the Blueprint reforms meaningfully increased gains in reading.

### ***Which Elements of the Reform Were Most Helpful?***

Betts et al. (2005), who analyzed test scores from Spring 1998 through 2002, found that most students who participated in a given element of the Blueprint in a given year were predicted to gain more in reading scores than in years when they did not participate. After-school reading programs, literacy-based summer school, enriched curriculum materials and a longer school year in the worst-performing elementary schools all were predicted to boost reading achievement considerably.

Double- and triple-length English classes in middle and high school produced mixed results—strongly positive in middle school and distinctly negative in high school. It was not possible for the authors to state with certainty the reason for the divergent results from these interventions, but two anecdotal stories emerged locally. First, high school English teachers did not buy into the remediation activities to the same degree as did English teachers in lower grades, perhaps due to a greater focus by high school teachers on teaching literature relative to teaching basic reading skills. Second, several stories we heard supported the notion that high

school students who participated in the pull-out English classes were stigmatized far more than were middle school students who participated.

The degree to which a school received peer coaches for its teachers had a zero predicted effect overall during the first four years of the reforms.

### ***How Did Program Effectiveness Vary across Years?***

As mentioned, Betts et al. (2005) for the most part found zero effects of the number of peer coaches in the school (relative to enrollment) in the early years. However, they also found that this overall “zero” effect consisted of a small negative effect in 1999-2000 and small positive effects in 2000-2001 and 2001-2002. This pattern hints that this aspect of the teacher professional development program was, if anything, disruptive initially but over time led to improvements in teaching of literacy skills.

Betts et al. (2008) extended the analysis to Spring 2005. Their findings weakly corroborate the notion that peer coaching filtered down to student achievement increasingly as time progressed. For instance, in elementary schools, the estimated effect of increasing the number of peer coaches as a percentage of enrollment does not increase monotonically by year, but roughly speaking is two or three times as large by the last three years of the program as in the first year. High schools exhibited a similar pattern, while for middle schools, the estimated effect of peer coaching appeared to reach its zenith in 2001-2002 and 2002-2003.

These same authors found that double-and triple-length classes became more effective with each passing year, and even at the high-school level, where negative estimated effects of extended-length English classes persisted, for EL students the effects became significantly positive in 2004-2005.

Apart from these Blueprint measures, coefficients for most of the other Blueprint interventions exhibited an inverted U-shape when graphed against year. It is impossible to know without further information what these patterns mean. But they are consistent with the notion that with such a massive reform, with multiple interventions being introduced at once amid a new and wide-scale form of teacher training, it took several years for teachers to learn how best to make use of the extra class time afforded by the extended day reading program, the double- and triple-length English classes, and literacy summer school. Toward the end of the reform period, the effectiveness of some of these interventions may have fallen somewhat in part due to state budget cutbacks that caused a general reduction in the district's financial resources. For instance, in conversations with district officials, we learned that the number of minutes of after-school reading classes funded may have fallen slightly after 2001-2002, in part due to budgetary concerns.

**It's All about the Long Run: Political and Financial Challenges to Sustaining Reform and Related Challenges to Evaluating the Longer-Term Effects of Reform**

The general finding that the estimated effect of the interventions varied by year, and that effectiveness grew over time (while later dissipating in some cases), suggests that dynamics are important from both an educational and a political perspective. It stands to reason that it takes time for a school district to maximize the benefits for students of new ways of teaching. But can the various parties affected by reforms—administrators, teachers, school board members, and families of students themselves—wait long enough to see the proof of the pudding?

The financial underpinnings of school reform also appear to be critical to the long-term viability of major school reform. Non-profit foundations donated roughly \$100 million to the reforms in San Diego over a seven-year period. Program officers from these foundations clearly

hoped to see a transition plan whereby the district itself would eventually find a way to finance the professional development model and the student interventions purely from internal district funds. It was not inconceivable that the Blueprint, had it persisted, could have become internally funded. (To put the \$100 million in external assistance over seven years into perspective, in 2002-2003, SDUSD had revenues of \$1.15 billion.<sup>2</sup> Thus, expenditures of roughly \$25 to \$35 million a year represent about 2% to 3% of overall revenues.)

In practice, fate dealt California a bad hand, as a series of bleak state budgets led to retrenchment in school districts statewide. Indeed, SDUSD's proposed budget of \$1.2 billion for 2008-2009 is only about 5% higher than the actual 2002-2003 budget, so that after taking inflation during this six-year period into account, the district's budgets in recent years have gone down considerably in real terms.<sup>3</sup>

The time needed for reform to take root, contrasted with the uncertain long-term financial viability of ambitious reforms such as the Blueprint for Student Success, also poses difficult choices for researchers. There is a natural thirst for answers—Many ask “Is it working?” in the first year or two of reform, which is precisely the time when the reform is most likely to be undergoing teething pains. Further into the reform, as the program matures, more reliable measures of the long-term efficacy of the reforms can be made. But the chances of the reform surviving both political and financial pressures dwindle with each succeeding year.

Seen in this light, the collection of papers in this issue affords a rare glimpse, at the classroom level, of an ambitious reform at roughly the half-decade point. The papers suggest that

---

<sup>2</sup> Data downloaded from Ed-Data, <http://www.ed-data.k12.ca.us>, on December 22, 2008.

<sup>3</sup> For the proposed 2008-2009 budget, see <http://www.sandi.net/depts/budget/pdf/0809budgetbook.pdf>, as downloaded December 22, 2008.



the plan for disseminating skills to teachers had its flaws, as did the lack of student individualization, at least initially. These results in many senses corroborate some of the earlier qualitative analyses of how professional development was unfolding in the district. But the papers in this issue also make clear that much of the original vision of professional development really did take root. The complementary quantitative analyses by Betts et al. (2005) and Betts et al. (2008) suggested a similar arc of reform over time, with the reforms taking root and measurably improving student achievement, especially after a year or two of experience.

In the end, different approaches to understanding this particular “elephant” of reform, with its immense proportions, does not lead to wildly divergent views after all. The reform, imperfect as it may have been, tangibly shifted teacher practice in elementary schools, at least for a half decade, and, in elementary and middle schools, but not high schools, boosted reading achievement for the students exposed to the various literacy interventions.

## References

- Bassok, D., & Raymond, M. E. (2005). Performance trends and the Blueprint for Student Success. In F. M. Hess (Ed.), *Urban school reform: Lessons from San Diego* (pp. 299-323). Cambridge, MA: Harvard Education Press.
- Betts, J. R. (2005). The promise and challenge of accountability in public schooling. In F. M. Hess (Ed.), *Urban school reform: Lessons from San Diego* (pp. 157-176). Cambridge, MA: Harvard Education Press.
- Betts, J. R., Zau, A., & King, K. (2005). *From blueprint to reality: San Diego's education reforms*. San Francisco: Public Policy Institute of California. Retrieved on January 3, 2009, from <http://www.ppic.org/main/publication.asp?i=408>.
- Betts, J. R., Zau, A. C., & Koedel, C. R. (2008). *Blueprint for success? An evaluation of the direct and indirect effects of the San Diego reading reforms*. Unpublished manuscript. San Diego: Department of Economics, University of California San Diego.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915-945.
- Hannaway, J., & Stanislawski, M. (2005). Flip-flops in school reform: an evolutionary theory of decentralization. In F. M. Hess (Ed.), *Urban school reform: Lessons from San Diego* (pp. 53-70). Cambridge, MA: Harvard Education Press.
- Hightower, A. (2002). *San Diego's big boom: District bureaucracy supports culture of learning*. Seattle, WA: Center for the Study of Teaching and Policy.
- Quick, H. E., Birman, B. F., Gallagher, L. P., Wolman, J., Chaney, K., & Hikawa, H. (2003). *Evaluation of the blueprint for success in a standards-based system: Year 2 interim report*. Palo Alto, CA: American Institutes for Research.
- San Diego Unified School District. (2000, March). *Blueprint for student success in a standards-based system*. San Diego: Author.
- Schnur, J., & Gerson, K. (2005). Reforming the principalship. In F. M. Hess (Ed.), *Urban school reform: Lessons from San Diego* (pp. 93-113). Cambridge, MA: Harvard Education Press.
- Stein, M. K., Hubbard, L., & Mehan, H. (2004). Reform ideas that travel far afield: The two cultures of reform in New York City's District #2 and San Diego. *Journal of Educational Change*, 5, 161-197.
- Williams, J. (2005). The labor management showdown. In F.M. Hess (Ed.), *Urban school reform: Lessons from San Diego* (pp. 33-51). Cambridge, MA: Harvard Education Press.

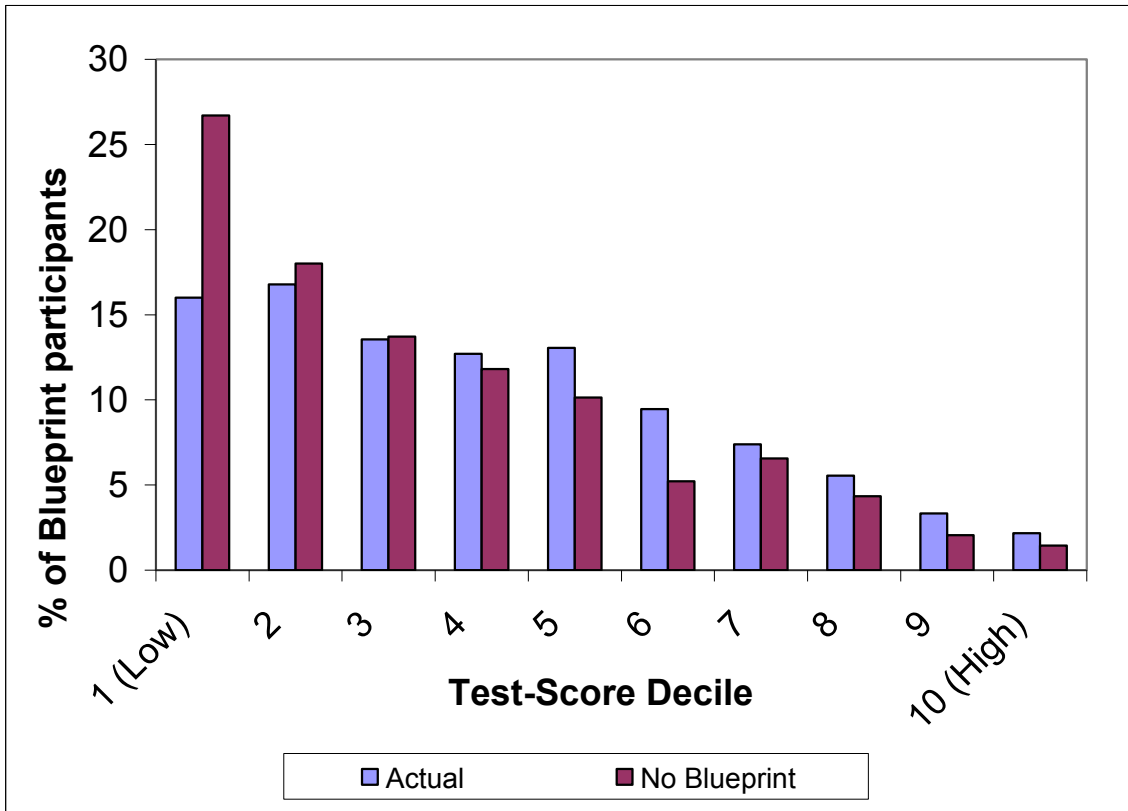


FIGURE 1 Distribution of Fall 2000 Grade 3 Blueprint participants by Spring 2002 test score decile: Actual and simulated distribution without Blueprint. *Source:* Betts et al. (2005).