Occasional Paper No. 41

National Center for the Study of Privatization in Education
Teachers College, Columbia University

Student Academic Achievement in Charter Schools: What We Know and Why We Know So Little [1]

Gary Miron and Christopher Nelson

The Evaluation Center
401B Ellsworth Hall
Western Michigan University
Kalamazoo, MI 49008-5237
<www.wmich.edu/evalctr/>

December 2001

Abstract. Student achievement is central to the policy debate over charter schools. This paper reviews what is currently known about charter schools' impact on student achievement. After documenting the surprising dearth of systematic empirical studies on the topic, the paper combines the findings of existing research into an overall impact rating, weighted by the studies' methodological quality. We find that the existing body of research reveals a mixed picture, with studies from some states suggesting a positive impact, studies from other states suggesting negative impact, and some providing evidence of positive and negative impacts. Overall, the charter impact on student achievement appears to be mixed or very slightly positive. However, this conclusion is tempered by the fact that there are, as yet, no systematic studies of charter achievement in several states that have large numbers of charter schools. The paper concludes by offering some preliminary explanations for variations in charter performance across states and for the paucity of empirical evidence in many states.

---

[1] The authors intend to update their analyses in 2002 to include new studies that have since become available. This paper benefited from the capable research assistance provided by Carla Howe and Kim Reynolds.

The Occasional Paper Series of the National Center for the Study of Privatization in Education (NCSPE) is designed to promote dialogue about the many facets of privatization in education. The subject matter of the papers is diverse, including research reviews and original research on vouchers, charter schools, home schooling, and educational management organizations. The papers are grounded in a range of disciplinary and methodological approaches. The views presented in these papers are those of the authors and do not necessarily represent the official views of the NCSPE.

If you are interested in submitting a paper, or wish to learn more about the NCSPE, please contact us at:

NCSPE, Box 181, Teachers College, Columbia University,
525 W. 120th Street, New York, NY 10027

tel: (212) 678-3259 fax: (212) 678-3474
ncspe@columbia.edu www.ncspe.org

1. Introduction

At the heart of the charter school concept lies a "bargain": schools will receive more autonomy in operations in exchange for being held more accountable than other public schools for student outcomes. Yet, in spite of the central role that achievement plays in this formulation of the concept, few studies or evaluations have analyzed student achievement in charter schools. Citing both the relative newness and small size of charter schools, many researchers have avoided the question of achievement altogether. More recently, and with the maturing of the movement, a number of studies have taken up the question of charter schools' impact on student learning and achievement. In this paper we attempt to summarize what is currently known about the impact of charter school attendance on student achievement.

In presenting this synthesis of research on student achievement in charter schools, we are mindful of four important limitations. First, it is important to state unequivocally that student achievement is not the only relevant "output" of charter schools or any other school. A full assessment of charter schools' effectiveness and overall desirability must examine other outcomes, such as equity; customer satisfaction and market accountability; the schools' legitimacy in the eyes of key stakeholders; and so on. Moreover, even if student achievement were the only goal of charter schools, standardized test results are only one of many ways to assess it. Few, if any, measurement experts would endorse evaluating a school or student on the basis of standardized tests alone. Second, readers should bear in mind that the research and evaluation literature has not yet produced clear and unambiguous statements of fact about achievement in charter schools. While perhaps frustrating, the cause of educational reform is best served by careful and honest consideration of the evidence. This, in turn, requires stakeholders to weigh the strengths and weaknesses of that evidence. We have attempted to derive a fair accounting of these factors and arrive at some tentative conclusions. Third, like any review or meta-analysis, our portrayal of the existing literature is colored by the selection of studies for examination. While we have attempted to make explicit the rules that guided our search for and selection of studies, readers should bear in mind that a less restrictive list of studies might generate a different set of conclusions. Where possible we sought to test our

findings' sensitivity to these assumptions. Generally, we have been somewhat surprised at the paucity of studies of student achievement in charter schools and the difficulty we had in obtaining some of the studies. Finally, we must point out, in the spirit of full disclosure, that we have authored and contributed to three of the studies covered in this paper. We have attempted to set out clear evaluative criteria and apply them fairly. However, it is inevitable that a certain amount of bias will remain in any attempt to evaluate one's own work.

In the following (second) section we outline expectations for charter school accountability across the states with a review of testing, reporting, and evaluation requirements for charter schools. Since the results of any literature review or meta-analysis are sensitive to choice of method, we devote a considerable amount of space in Section 3 to such issues as how one should measure student achievement, which studies should be included, how one should evaluate the quality of studies, and how one should combine findings into a "bottom line" statement of overall charter school impact. While we encourage all readers to carefully consider our methodological choices, some readers may wish to skip Section 3 and move directly to Section 4's presentation of findings regarding the current knowledge base about student achievement in charter schools. The final sections of the paper consider some preliminary explanations for variations in charter schools' achievement impacts across states and for the dearth of knowledge about student achievement in charter schools.

2. Expectations for Accountability in Charter Schools

Before summarizing what we know about student achievement in charter schools, we address questions pertaining to how researchers, citizens, and stakeholders learn about student achievement in this section. We first examine the extent to which charter schools are required to meet state standards and administer state assessments. Next, we examine whether state laws require independent evaluations of charter school performance. Finally, we report on how many states' charter schools have been subject to independent evaluations or other research studies that include information about student achievement. Answers to these questions provide the necessary context for the summary of existing research that follows in Sections 3 and 4.

The analysis in this section is based on a database we created that includes information about testing, evaluation, and accountability requirements in the 38 currently existing charter school laws (i.e., 37 states and the District of Columbia). Most of the data was obtained from state departments of education Web sites as well as the U.S. Department of Education-sponsored Web site <www.uscharterschools.org> and the Web site for the Center for Education Reform: <www.edreform.com>. These latter two Internet resources contain their own directories of information organized by states.

2.1 Required State Standards and State Assessment Tests

The charter school laws in 34 of 38 states clearly require charter schools to meet the same standards as noncharter public schools. Likewise, these 34 states require charter schools to administer some sort of standardized test. While the specific test was mentioned in legal and Web site materials from 31 of the 34 states, similar material for 3 other states did not mention which test must be taken. In all cases, charter schools were required to use the same test required by noncharter public schools in the state. Sixteen percent of the states required both criterion-referenced tests and norm-referenced tests. In 77.5 percent of the states, only a criterion-referenced test was required; in 6.5 percent of the states only a norm-referenced test was required. Figure 1 illustrates these findings.

The information on required testing suggests that extensive standardized test results should be available, which would facilitate evaluations of student achievement in charter schools. However, this was not the case, since many states commissioned evaluations but were not able to address student achievement due to limited availability of student achievement results for charter schools.

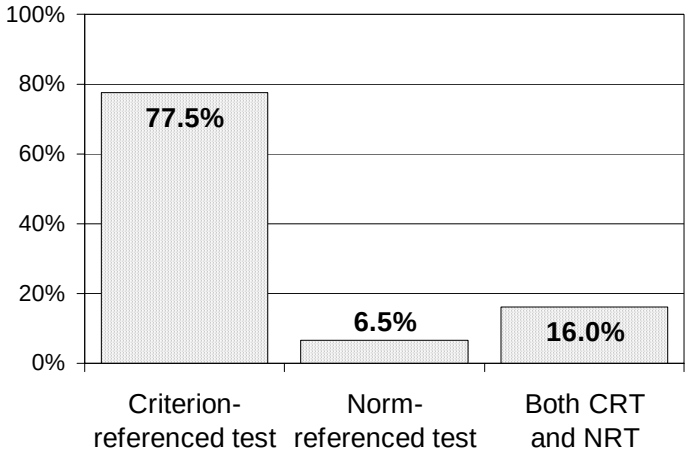

Figure 1. Required Testing in Charter Schools

2.2 Independent Evaluations of Charter Schools

State level policymakers might request evaluations of charter schools for a number of reasons. First and foremost, charter schools have been granted greater autonomy in exchange for accountability. One of the best ways to determine whether or not charter schools are living up to their end of the bargain is to have an independent evaluation of the schools. Another important reason is that evaluations with formative components provide feedback information to help improve the implementation of the reform.[2]

While many states require some form of formal reporting on charter schools or other state bodies, it is important to have an independent evaluation. Independence is particularly important in view of the highly polarized context in which many of the reforms are being implemented. "Independence" is difficult to observe in practice. However, we used two general rules of thumb in determining whether a given evaluation is independent. First, independence requires that the evaluation is conducted by a third party–that is, by some organization other than the charter schools and the agency or bureau that administers the state's charter school law. Second, independent evaluations must include data other than those that are self-reported by the schools.[3]

Our review of documentation and literature turned up some surprising findings. First of all, very few states actually require independent evaluations. In total, we found reference to required independent evaluations in only 3 states. In 15 states, however, the state board of education or the state department of education is expected to submit an annual descriptive report to the state's legislature regarding the status of the charter schools. In some states such reports are completed only

[2] Pennsylvania provides a good example of using evaluation for improvement purposes. The Commonwealth of Pennsylvania passed a charter school law in 1997 that requires an independent evaluation of the reform after five years. After only two years, however, the state's department of education commissioned an initial evaluation (see Miron & Nelson, 2000) to provide improvement-oriented information to the charter schools. At the same time, the study provided feedback information to policymakers and oversight agencies so that adjustments in the reform could be made before the five-year mandatory evaluation.

[3] We emphasize that these are merely rules of thumb. Indeed, we have included a small number of apparently nonindependent studies in our later analysis where we were convinced that the data reported were subject to audits.

irregularly; in Michigan, for instance, the report has been completed only once, and even then it was submitted well after the deadline. In Illinois, by contrast, the reports have been prepared every year, as required. However, most descriptive reports prepared by state education agencies contain only brief profiles of the schools with little or no reference to the academic progress of students.

Legal requirements notwithstanding, we found that 18 states have conducted what we deem to be independent evaluations (see Figure 2). At least five additional states have evaluations under way. Seven of the 18 states with independent evaluations had evaluations that were quite dated and/or of limited scope. The remainder of this section describes these 7 states and the evaluations that were conducted.

California: SRI International conducted an evaluation in 1997. Due to "a lack of statewide data" it was not possible to evaluate the impact of the charter schools on student achievement. The following year two new studies were released (Wells et al, 1998; and WestEd, 1998). These two studies were limited to a relatively small number of schools and districts. The Wells et al. study examined 17 schools in 10 school districts. This was largely qualitative in nature and did not directly analyze student achievement results. One finding from this study indicated that "charter schools in

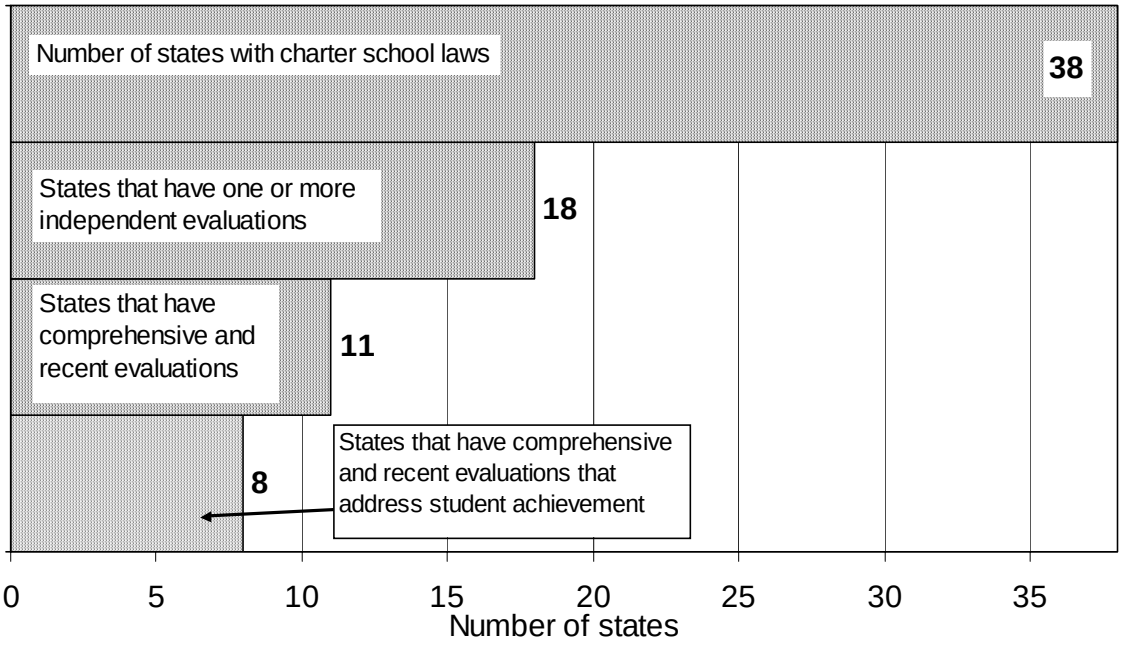

Figure 2. Evaluations of Charter Schools by Nature and Type

California are, in most instances, not yet being held accountable for enhanced academic achievement of their students" (Wells et al., 1998, p. 4). The WestEd study (1998) used a variety of approaches to examine student achievement in five Los Angeles schools that were coming up for renewal. The analyses focused on change in average achievement results over time, change in district ranking over time, and change in the performance of "stayers" over time. The study found that the five charter schools generally maintained or slightly improved their students' performance over time as compared with similar noncharter public schools.

Louisiana. An evaluation of the 11 charter schools in operation in 1998/99 was completed in October 1999. Particular attention was given to the 3 schools that had been in operation for 3 years. The nature and type of the test results varied by school, which made cross-school reporting impossible. The evaluation found that 65 percent of the interviewed students reported that they were doing better at the charter school than at their previous school. Only 1 school reported above average ITBS scores, while the other schools were not performing at levels consistent with the state average or with the parish school systems. Since many of the schools catered to at-risk students, comparisons were not made with similar schools.

Massachusetts: In 1997, the Massachusetts Department of Education (1997) conducted a preliminary study of student achievement in the state's charter schools. At most, only two years of test data could be traced and then only for a handful of schools. The report interpreted high enrollments in charter schools as evidence of success, but was cautious in making any claims about gain scores. The following year the Massachusetts Department of Education (1998) completed another report about the charter schools. This report did not consider student achievement but did note that a more elaborate study would be released in March 1999. To the best of our knowledge, this latter study was not made public if it were completed.

Minnesota: The Minnesota State Board of Education commissioned the Center for Applied Research and Educational Improvement (CAREI), University of Minnesota, to conduct an evaluation of charter schools in 1996 (CAREI, 1998). One of the key questions addressed was whether or not charter schools were improving student achievement. The study described some of the measures and

methods schools used and also considered other indicators of performance besides results on standardized tests. Nevertheless, very limited data from standardized tests were available, and then only for a fraction of the schools. From a contemporary perspective, the study is quite dated, since it covers only 1996 and 1997.

New Jersey: The state commissioned an evaluation that was completed in 1998 (Kane, 1998). Since the evaluation was conducted at an early stage in the reform, limited test data were available. The state has since commissioned a second evaluation, but the results were not available in time to consider it in this paper.

North Carolina. In 1998, the evaluation section of the North Carolina Department of Public Instruction conducted an initial evaluation of the state's charter schools. The evaluation found that the charter schools did not perform as well as other public schools on the state's ABCs accountability program. However, the student achievement results were clearly tentative given the newness of the schools at the time of the evaluation. A more thorough evaluation has since been commissioned but, unfortunately, the results were not available in time to consider it in this paper.

Wisconsin: Although not required in the charter school law, Wisconsin had one comprehensive evaluation conducted by the Wisconsin Legislative Audit Bureau (1998). We did not include this in our analyses, because it was quite dated and many schools were excluded due to a lack of data.

Three other states (Idaho, Illinois, and Ohio) have had recent and comprehensive evaluations that were also excluded from our study since they either lack data on student achievement or they had not yet been released to the public.

In total, eight states with one or more evaluations are included in our analysis because they are relatively comprehensive, relatively recent, and include an examination of student achievement results in charter schools. It should also be pointed out that some studies were excluded from these 8 states when there were more up-to-date or more comprehensive evaluations available that superceded them. For example, in Michigan, we excluded the first Horn & Miron evaluation (1999) since it was superceded by the second study which used the same design and included one extra year of data. We also excluded 2 evaluations conducted by Public Sector Consultants (Khouri, Kleine,

White, & Cummings, 1999; Kleine, Scott, & White, 2000) since they used a similar design, but included fewer years of data and only covered half the charter schools in the state. On the other hand, we included five annual studies from Colorado since they all considered different cross-sectional sets of data.

2.3 States Without Program Evaluations

Other things being equal, one would expect to see evaluations in states with longstanding charter school laws and/or a large number of charter schools in operation. However, we found a number of states with such characteristics but no independent evaluations. This section describes some of those states.

The absence of an evaluation of charter schools in Florida was perhaps the greatest surprise, given the presence of a number of factors that would lead one to expect such an evaluation. First, the state has a large number of schools in operation (152 in 2000), and the charter school law dates back to 1996. Second, the state's department of education is expected to submit a report each year to the state board, Commissioner of Education, and the legislature. A review panel was appointed to examine charter school policies and to suggest possible modifications in the charter school law. We have not, however, been able to obtain documents from the Commission that suggest any analysis of student achievement results.

Delaware, Hawaii, Kansas, New Mexico, and South Carolina are states where we expected–and failed–to find an evaluation of charter schools, since these states had laws in place by 1996 or earlier and have 6 or more schools in operation.

Massachusetts, California, and Wisconsin are also examples of states with long standing charter school laws and a large number of charter schools. Unlike the states discussed above, these states have conducted accountability studies of charter schools, but have not produced any comprehensive evaluations that examine charter schools' impact on student achievement. Massachusetts is a particularly interesting example, since it has perhaps the most rigorous and comprehensive accountability system for individual charter schools, but has not yet produced nor commissioned a comprehensive evaluation of the charter school reform. In 1996 the state prepared a comprehensive

summary of the schools in operation at the time. However, most of the data appeared to be self-reported, and no comparisons to noncharter schools were made. Massachusetts is also one of the three states where an evaluation is clearly specified in the charter school law. However, it is unclear when any such evaluation will be commissioned. Similarly, California's Office of Legislative Analysis is expected to contract with an independent evaluator in the near future. In Wisconsin, by contrast, officials say that the state will simply continue to compile summaries of self-reported information from the schools.

3. Summarizing and Evaluating the Studies: The Methodology

We turn now to a systematic review of what is currently known about the impact of charter school attendance on student achievement. Before providing this summary, it is important to dispense with a number of methodological issues, including:

- How should we measure student achievement, and which studies should be included? (Section 3.1)
- How should we aggregate findings from various studies to come up with a bottom line summary judgment? Specifically, should we give more weight to some studies than to others? (Sections 3.2, 3.2, and 3.4)
- How should we address the fact that the charter school "treatment" varies considerably from state to state? (Section 3.5)

3.1 Measuring Achievement and Selecting Studies

Our summary will be restricted to studies that analyze standardized test results; this can include both criterion-referenced tests and norm-referenced tests. While we are well aware of the many problems with standardized tests, they meet important criteria for large-scale assessments of school quality. First, they are administered to most charter schools in most states since charter schools typically must participate in the state tests mandated for all public schools in the state. In other cases, charter school studies have focused on, or also included, results from commercially developed tests such as the Iowa Test of Basic Skills (ITBS), the Stanford Achievement Test (SAT), or the California Achievement Test (CAT). The second criteria stems from the fact that evaluations of charter school

impact require evaluators to gain some understanding of what charter school student achievement might have looked like in the absence of charter schools. The best practical way to do this is by observing the achievement scores of a similar "comparison group" of schools. Since standardized tests are administered to a wide range of noncharter schools, it is usually possible to use these test scores to construct such comparison groups. Third, any measure of student achievement should be both valid and comparable (reliable). Validity is the extent to which the test measures what it is supposed to measure. This is often the Achilles heel of standardized tests, since they often measure content that is not covered by all schools administering the exams. Thus, test scores might capture not only student learning in a school, but also the extent the school teaches material covered on that particular test. However, what standardized tests lack in validity, they often make up for in their ability to provide a comparable measure of student achievement across a large number of schools.

In addition to restricting our analysis to studies that analyze standardized test results, we focused mainly on relatively recent studies conducted of entire states or significant portions thereof. The reasons for this are twofold. First, including only large-scale studies reduces the likelihood of selection bias in the studies' findings. Quite simply, studies that focus on only a handful of "case" schools are unlikely to produce findings that can be generalized to the larger population of schools in a state. Second, since the legal frameworks under which charter schools operate are created largely by state statutes and administrative rules, conducting the analysis by state might yield some interesting insights about the impact of state policies on charter school performance.

In searching for studies we relied on a number of strategies. First, we read previous studies of charter schools in order to locate citations to other studies. Second, we reviewed Web sites for state education agencies, along with a number of Web sites that specialize in charter school and education reform issues.[4] Finally, where we suspected that the first two strategies had failed to unearth studies, we made extensive inquiries with professional colleagues and officials at state education agencies.

[4] These latter included the federally sponsored charter schools web site <www.uscharterschools.org>, the Center for Education Reform's Web site <www.edreform.com>, the Charter Friends National Network's Web site <wwww.charterfriendsnetwork.org>, and the Educational Resources Information Center (ERIC) <www.eric.ed.gov>.

A final issue related to selection of studies is "publication bias." Simply put, publication bias results whenever the sample of programs and policies being assessed are not a representative sample of all relevant policies and programs. For example, our analyses are based on evaluations from only eight states, yet there are 38 charter school laws. There are a number of states with charter school laws that have too few or no charter schools in operation; and some states have schools that are too young to evaluate, so excluding them is not going to bias our findings. There are, however, some states, such as such as Florida, Massachusetts, California, Minnesota, and Wisconsin, that have a large number of charter schools; yet these states are excluded due to an absence of evaluations that could be considered. If charter school states with high performing schools have not been included, then this analysis will underestimate the effectiveness of charter schools. Conversely, if schools in unstudied states are weak performers, then our analysis will overestimate the charter school effect. Unfortunately, existing data provide no reliable way to estimate the direction or severity of this bias.

3.2 Assessing Charter School Impact in the Studies

An additional methodological issue involves how to combine studies into a set of bottom-line summary judgments. We made two sets of judgments about each study examined. First, we gave each study an "impact rating" that summarizes the direction and magnitude of the observed impact of charter school attendance on student achievement. Second, we gave each study a set of quality ratings, with one including information about use of comparison groups and the other indicating the number of years from which data were analyzed. The remainder of this section outlines the system employed for assigning impact ratings to the studies. Subsequent sections detail the quality ratings and how the impact and quality ratings were combined to form a bottom-line assessment of charter schools' impact on student achievement.

"Impact," as we use the term, is assessed in relation to a comparison group of similar noncharter public schools. This is similar to the oft-discussed notion of "value added," in which educational evaluators and researchers seek to estimate the additional value (measured in terms of student achievement) that charter schools provide to their students. It is important to bear in mind that

impact is not necessarily synonymous with absolute achievement levels. Thus, a charter school with low absolute scores might still add a significant amount of value to its students if its change scores are larger (or growing at a faster rate) than noncharter public schools with similar students. Conversely, a charter school with high absolute scores might not be adding value to its students if its change scores are not as large (or are growing at a slower rate) than noncharter public schools with similar students.

An important complication in assigning impact ratings arises from the fact that studies include several distinct analyses of charter school achievement data. In most cases, these various analyses examine the same charter schools but employ different comparison groups, statistical controls, or estimation techniques. For these studies, we followed a two-step decision rule. First, where appropriate, we focused only on what we regarded as the methodologically strongest findings. Gronberg and Jansen's (2001) study of Texas charter schools, for instance, begins with simple descriptive analyses of charter school achievement trends, successively adds more sophisticated controls, and culminates in regression models with fixed effects for each student. Fixed effects allow the analyst to filter out otherwise unmeasured differences among all students and thus are a powerful way to account for student background factors. Thus, we include only the findings with the more sophisticated analyses, since they are more reliable than the purely descriptive findings. Where one analysis was not clearly superior (i.e., where there was a tie), we considered each of the best analyses as a separate "study"–each with its own impact and quality ratings–for the purposes of combining and weighting studies. Table 3 summarizes the findings we have used from each study.[5]

Ideally, a meta-analysis study such as this would attempt to combine information from the studies into an overall "effect size" (see, e.g., Light & Pillemer, 1984; Hedges & Olkin, 1985). Unfortunately, the wide variety of measures and methods employed across the studies makes it difficult, if not impossible, to derive an overall effect size. Instead, we have given each study an ordinal rating according to the scale shown in Table 1, with positive values indicating that charter schools serve to increase student achievement and negative values indicating that they decrease

[5] Readers who wish to make their own assessments of each study's findings are encouraged to consult the citations at the end of the paper.

student achievement.[6] We emphasize, however, that quantification is used primarily as a heuristic. Unlike the more traditional "literature review" approach, the quantitative rating system allowed us to systematically combine information about both the studies' findings and aspects of their design quality.

Table 1. Scale for Charter School Impact Ratings

Scale Value	*Description*
2	Positive overall impact
1	Mixed to slightly positive overall impact
0	Mixed impact
-1	Mixed to slightly negative overall impact
-2	Negative overall impact

3.3 Assessing the Quality of the Studies

Our method for assigning quality ratings to studies includes three types of information: (a) the strength of the sample of charter schools included; (b) the strength of the quasi-experimental design used to compare charter school students with comparable noncharter school students; and (c) the number of years included in the study. Before discussing this system, it is important to note that the poor quality of some studies is usually the result of data limitations and not the incompetence of researchers and evaluators. As discussed near the end of the paper, limitations in existing data place severe restrictions on evaluations of charter school impacts in most states.

The first factor included in our assessment of study quality was the scope of schools and students included. We refrain from using the term "sample" in this context, because none of the studies we examined set out to make inferences about the population of charter schools in a state based on a sample. Instead, given the relatively small number of schools, the studies sought to examine a census of all charter schools. Thus, we evaluated the selection of schools and students mostly on the basis of response rate, with higher response rates getting higher ratings.[7] Interestingly,

[6] Readers should bear in mind that our 5-point scale might understate the variation in impacts found across studies.

[7] Response rates were calculated by comparing the number of schools included in the analyses with the total number of charter schools with students in the tested grades. Thus, we did not count as "missing" schools that did not enroll students in tested grade levels. For the few studies that attempted to match student records across years, we based our evaluation largely on match rates.

there was relatively little variation among the studies in the quality of charter school samples; thus, we will not discuss it further. However, we did exclude a small number of studies offering analyses of just a handful of schools in any given state (e.g., SRI, 1997; WestEd, 1998).

A second factor, and one on which there was considerable variation among the studies, involves the selection and use of comparison groups. Estimating the impact of charter school attendance requires information or informed speculation about what student achievement would have looked like in the absence of charter schools. Since we cannot directly observe the counterfactual, evaluators usually do this by comparing charter school performance with that of a control or comparison group of similar students. Ideally, such groups are created through random assignment. For a variety of reasons, this is difficult to do with charter schools.[8] More often, evaluators and researchers seek out "naturally occurring" comparison groups of similar students.

Generally, we have given more weight to studies that examine *changes* in student achievement over time. The strongest of these studies track individual students or cohorts of students over time. Here each student (or cohort of students) serves as his or her (its) own comparison group. Some particularly strong studies in this category also include a comparison group of similar students in order to determine whether observed changes are truly unique to charter schools. Unfortunately, few states' testing regimes allow for truly longitudinal analyses. More common is the analysis of "consecutive cohorts" in which the evaluator compares, for example, this year's fourth graders with

[8] Inasmuch as many charter schools are oversubscribed and that most state laws require schools to select students at random from their waiting lists, randomized experiments of the sort used to evaluate voucher programs ought to be possible, in principle. However, a number of logistical challenges must first be overcome. As voucher research has shown, there is often considerable attrition in control groups constructed from waiting lists (see, e.g., Rouse, 1997; Witte, 1997). Moreover, our fieldwork in charter schools suggests that charter school waiting lists are often insufficient for the construction of a good randomized experiment. In many cases, such lists are out of date or contain a cumulation of names over a number of years. In the most extreme cases, these lists cannot be produced for review when requested and may exist only in the minds of school administrators. Moreover, it is nearly impossible to assess whether students on the lists had subsequently enrolled in other charter schools or had been exposed to other educational reforms. In order to be convincing, any such analysis would have to include an audit of the waiting lists. While this is certainly possible, it would likely be time-consuming and costly. Finally, even assuming that researchers could overcome these logistical challenges, such randomized experiments would be, like any such experiments, of limited external validity. Indeed, they would be generalizable only to those students whose parents/guardians had attempted to enroll their children.

previous year's fourth graders. The most important problem with consecutive cohort studies is that observed changes over time might be due to changes in the composition of the student population instead of, or in addition to, the school's educational practices. Some studies include statistical controls for differences in charter and noncharter school students, often in addition to the aforementioned time series controls.

Generally speaking, the strongest studies employ a combination of these strategies. Solmon, Paark, and Garcia's (2001) analysis of Arizona charter schools, for instance, tracks individual test score gains over three years *and* includes a comparison group of traditional public school students matched on prior test score, grade level, and a number of other factors.[9] In addition, the study employs a number of rigorous statistical controls, including fixed and random effects models that control for unmeasured student-level differences. The weakest studies, by contrast, simply make cross-sectional comparisons between charter and noncharter schools, often without controls for preexisting differences among charter and noncharter students. Table 2 summarizes the rating system used to evaluate the quality of the studies. Note that studies rated as "0" were excluded from the analysis.

Table 2. Criteria for Design Quality

Grade	Attributes of the Study
4	Analysis of individual-level change scores along with inclusion of strong statistical controls and/or blocking
3	Analysis of consecutive cohorts with comparison group and statistical controls
2	Analysis of consecutive cohorts with a comparison group but no statistical controls
1	Analysis of consecutive cohorts without a comparison group; or, cross-sectional analysis with comparison group
0	Cross-sectional analysis without a comparison group. *These studies were excluded from the analysis*

[9] See Nelson & Hollenbeck (2001) for detailed critical evaluation of this study.

An important caveat about the design quality ratings is that the scale implicitly assumes that movement from one level to the next is equal across the levels. For instance, it is assumed that the quality advantage of a "4" study over a "3" study is exactly the same as the advantage of a "3" over a "2" and a "2" over a "1." This assumption is somewhat arbitrary. Having issued this caution, we proceed with the analysis in the absence of a clearly superior alternative.

Finally, we weighted studies according to the number of years included. This implies that studies that employ cross-sectional designs are given less weight than studies with long time series.[10]

3.4 Combining Impact and Quality Ratings

In order to combine information on findings regarding both impact and study quality, we calculated an average impact rating from Table 1 weighted by the quality scale shown in Table 2 and the number of years included in the studies. In some analyses we added a weighting term for number of schools since, other things equal, it is reasonable to give more weight to studies of a large number of schools.[11] Formally, the weighted average impact rating is given by the following expression:

$$\overline{I_w} = \frac{\sum_{i=1}^{n} I * w}{\sum_{i=w}^{n} w}$$

where I_w is the weighted impact rating, I the unweighted impact rating for a given school i, and w the final weight. In order to test the robustness of the findings, we calculated the average impact rating using a number of different final weights, each of which was a different combination of the design quality and "number of years" weights discussed above.

[10] Thus, even though there are a relatively large number of studies on Colorado charter schools, these findings are downweighted because of the studies' relatively weak data and designs.

[11] Readers should recall that, given that there was little variation across studies in school response rate, we have not included sample quality as a weighting term.

3.5 Heterogeneity of the Charter "Treatment"

A final methodological issue concerns the heterogeneity of the charter school "treatment." As numerous studies and analyses have pointed out, there are considerable variations among state charter school laws (RPP International, 2000). Moreover, evaluation studies have noted that even charter schools operating under a single state law often implement that law in very different ways (see, e.g., Henig et al., 1998; Miron & Nelson, 2000). Thus, in some sense, the studies included in this analysis involve apples-to-oranges comparisons of a variety of charter school interventions. This variation, however, opens up important avenues of exploration. As Elazar (1966) noted, states may serve as the "laboratories of democracy," each providing evidence from its own "experiments" in policymaking. Thus, a finding that charter schools in states with certain types of laws perform better than schools in states with other types of laws might provide important formative information for program improvement. We provide some very preliminary state-by-state comparisons in Section 5.

4. What Do the Studies Tell Us About The Impact of Charter Schools on Student Achievement?

This section presents the findings of our meta-analysis of studies of student achievement in charter schools. We begin by describing the range of findings represented in the studies. Next, we present an analysis that combines the ratings of key findings and the quality ratings in the form of weighted indices. The section concludes with a discussion of effect sizes in the studies and of what can be learned from variations in student achievement across the states.

4.1 Summary of Findings Without Respect to Quality

Table 3 provides a brief summary of the studies selected for review and analysis. For each state, the table includes information used in the analysis, including grade levels covered, the number of years included in the study, a brief summary of key findings, the impact rating (according to the criteria shown in Table 1), and a design quality rating. As discussed above, we located only 15 studies that met the criteria outlined above, covering student achievement in the charter schools of only 7 states

Table 3. Summary of Studies

State	*Citation*	*Grade Levels*	*Years Covered*	*Summary of Findings*	*Impact Rating*	*Design Quality Rating*
AZ	Solmon, Paark, & Garcia (2001)	2-11	1996/97-1998/99	Positive in reading; slightly positive in math	Positive (2)	4
AZ	Mulholland (1999)	2-11	1996/97-1997/98	Mixed, with nonsignificant gains in some grades and subjects. No difference overall	Mixed (0)	3
CO	Colorado Department of Education (2001)	4 & 7	1998/99-1999/00	Change score analysis: equal numbers of schools showed increases and decreases in reading, with more decreases than increases on writing	Mixed (0)	1
			1999/00	Cross-sectional analysis: charter schools higher than demographically similar schools and state average on writing; slightly higher in reading.	Positive (2)	1
CO	Colorado Department of Education (2000)	3, 4, & 7	1998/99	Generally positive–approximately 2/3 of charter schools outscored demographically similar schools; approximately 3/4 outscored host district and state.	Positive (2)	1
CO	Colorado Department of Education (1999)	3 & 4	1997/98	Generally positive–approximately 2/3 of charter schools outscored host districts; most outscored demographically similar schools.	Positive (2)	1
CO	Colorado Department of Education (1998)	4	1996/97	Generally positive–charter school average proficiency rate was 73% in reading and 44% in writing, compared with statewide averages of 57% and 31%.	Positive (2)	1
CO	Colorado Department of Education (1997)	4	1995/96	Mixed to slightly positive; stronger in reading than math.	Mixed/ positive (1)	1
CT	Horn & Miron (2001)	4, 6, 8, & 10	1997/98-2000/01	Generally positive, with limited exceptions at the 4th and 10th grade levels	Mixed/ positive (1)	3

State	*Citation*	*Grade Levels*	*Years Covered*	*Summary of Findings*	*Impact Rating*	*Design Quality Rating*
DC	Henig, et al. (2001)		1999/00-2000/01	Consecutive cohorts: Charter schools less likely to have improved, more likely to have declined than host district schools	Negative (-2)	2
			2000/01	Cross-sectional analysis: More charter schools scored "below basic" than host district schools; differences hold up under statistical elaboration	Negative (-2)	1
GA	Georgia Department of Education (2000)	3, 5, high school grades	1997/98-1999/00	40% stable or increasing; 15% decreasing; 45% mixed (averaged across all grade levels and subject areas).	Mixed/ positive (1)	1
MI	Eberts & Hollenbeck (2001)	4 & 5	1996/97-1998/99	Negative charter schools score 3 to 7 percent lower than comparable noncharter public schools.	Negative (-2)	3 [12]
MI	Horn & Miron (2000)	4, 5, 7, 8, & 11	1995/96-1999/00	Mixed positive for 5th grade science; no difference for writing, math, reading, and 8th grade science; negative for 4th grade math and reading.	Mixed/ negative (-1)	2
MI	Bettinger (1999)	4 & 7	1996/97-1998/99	Generally negative, though charter schools show some success (nonrobust) in moving students out of the "low" category.	Negative (-2)	3
PA	Miron & Nelson (2000)	5, 6, 8, 9 & 11	1997/98-1998/99	Consecutive cohorts: Charter schools outgained host districts	Positive (2)	2
				Cross-sectional comparisons: Charter schools scored lower than demographically similar schools.	Negative (-2)	1
TX	Gronberg & Jansen (2001)	3-8	1996/97-1999/00	Negative for all schools (-.909 on TLI) and non-at-risk schools (-1.56 on TLI). Positive for at-risk schools (+.759 on TLI)	Mixed/ positive (1)	4

[12] While this paper uses individual-level student data, by and large it cannot trace individual changes over time. Thus, we have given this study a design quality rating of “3" rather than “4."

(Arizona, Colorado, Connecticut, Georgia, Michigan, Pennsylvania, and Texas) and the District of Columbia. As discussed in Section 3, we focused only on the methodologically strongest findings from each study. Where two or more sections of a particular study were tied for "best," we treated the results from each analysis as separate studies (e.g., District of Columbia and Pennsylvania). According to these criteria, studies produced 2 sets of findings, providing 18 studies for the analysis.

Close examination of the table reveals a wide range of findings. A few studies provide evidence of a substantial positive charter effect. Solmon, Paark, & Garcia's (2001) analysis of Stanford Achievement Test scores in Arizona, for instance, found that charter school attendance has a significant positive impact on SAT9 scores in reading and a mixed to positive impact in math. Another study of Arizona by Mulholland (1999) found modest but statistically indiscernible gains over two years.[13] At the other end of the spectrum, all three studies of achievement in Michigan charter schools are, for the most part, negative. Eberts & Hollenbeck (2001), for instance, found that Michigan charter schools scored between 3 and 7 percent lower than comparable host districts on state criterion reference examinations. Horn & Miron's (2000) earlier study was slightly less negative, finding that charter school trends were either indistinguishable from or lower than those of their host districts in all grades and areas except fifth grade science. Similarly, Bettinger (1999) reported generally negative findings; however, he found weak evidence[14] that charter schools had moved some students out of the "low" category on the state examination. Finally, findings from Texas by Gronberg & Jansen (2001) represent the middle of spectrum. These authors found that while charter schools overall scored lower on the state test than noncharter schools, schools classified by the state as "at-risk" outscored comparable noncharter public schools.

Employing the quantitative rating system enumerated in Table 1, we found that the unweighted average across all studies was 0.2. In substantive terms, this suggests that as a group the studies provide a mixed picture of the charter school effect. The range of the unweighted impact ratings is

[13] These latter findings are not incompatible with Solmon, Paark, & Garcia, since the latter also found a slight dip in scores during the first two years.

[14] The findings did not hold up under alternative specifications of the statistical model.

illustrated in Figure 3. The figure confirms that the impact ratings are widely dispersed, if very weakly positive overall. However, as discussed above, it is important to weigh study quality in summary statements about the studies. Thus, we turn now to a discussion of the quality of the studies examined and how this affects our bottom-line assessment of charter schools' impact on student achievement.

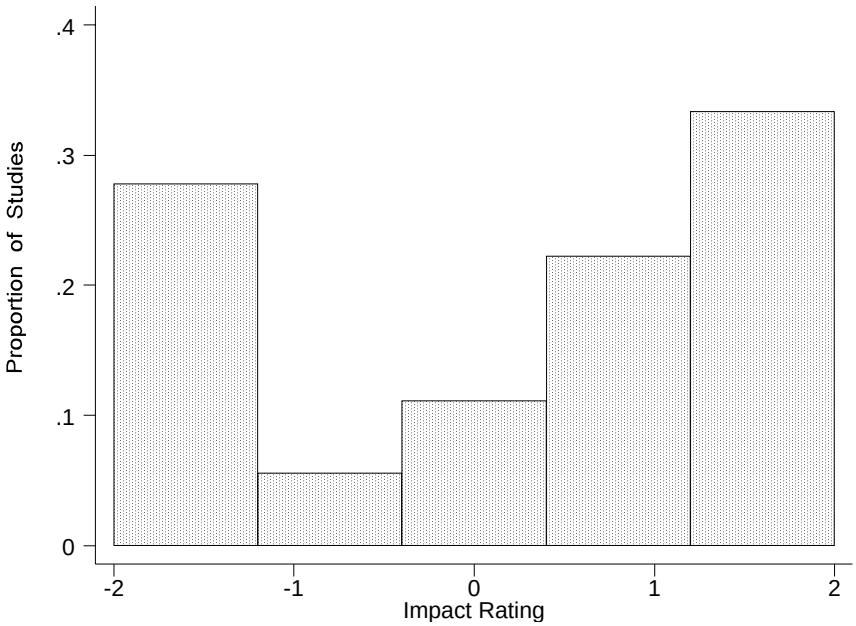

Figure 3. Summary of Impact Ratings Without Regard to Quality

4.2 Summary of Findings Including Quality

Examination of the studies revealed significant variations in design quality. Of the 18 studies coded, only 2 received the highest design quality rating. More than half, by contrast, received the lowest rating. Overall, the average design quality rating for all the studies was 1.9, which represents a study of only modest quality. Table 4 displays the number of studies receiving each design quality rating.

Table 4. Tabulation of Design Quality Ratings

Design Quality Rating	*Number of Studies*	*Percent of Studies*
4	2	12
3	4	23
2	2	12
1	9	53
Totals	18[†]	100

Note: A full discussion of the design quality ratings can be found in Section 3.
[†] As discussed in Section 3, a "study" includes separate analyses in the same report in cases where we were unable to identify a single best analysis.

Another indicator of study quality is the number of years of test data included. Once again, there is considerable variation among the studies; the cross-sectional studies, of course, include only 1 year of test data, while 1 study includes data from 5 consecutive school years. The average study included test data from 2.5 years.[15] It is important to bear in mind that the quality ratings employed are comparative and not absolute. Thus, even studies earning the highest rating may suffer from methodological limitations.[16] Once again, we emphasize that methodological weaknesses arise as much (if not more) from data deficiencies as from researcher incompetence.

As it turns out, introducing weights for quality changes the overall picture very little. Given that there are a number of ways to execute the weights, Table 5 presents results from analyses using a number of techniques. One analysis weights the impact ratings according to the number of schools covered in the study. Other things being equal, it is reasonable to give more weight to a study of a state with many charter schools (such as Arizona) than to a state (like Connecticut) with just a handful of charter schools. A second analysis weights the impact ratings according to the design quality rating system enumerated earlier in Table 2. A third analysis weights the impact ratings according to the number of years of test data included in the study, since it makes sense, other things being equal, to give more weight to studies with several years of test data. Finally, two analyses combine all three types of information, one by adding the weights, the other by multiplying them. The primary difference between the two is that the multiplicative weighting system implies that a weakness (or strength) on any one of the weighting criteria can dramatically affect the overall weight given to that study. For instance, under the multiplicative weighting system, the "credit" that a study gets for employing genuinely longitudinal analysis of individual student data with statistical controls would be severely discounted if that same study had only two years of data and/or included just a handful of charter schools.[17]

---

[15] Given that time-series studies received higher weights under the design quality index, it is not surprising that there was a positive relationship between the number of years covered in studies and their design quality ratings (Spearman's $\rho = 0.7$).

[16] See, for instance, Nelson & Hollenbeck (2001) for a critique of Solmon, Paark, & Garcia (2001), which received the highest rating in our design quality index.

[17] Thus, it is not surprising that there is more variation among studies on the multiplicative weights than the additive weights. The coefficient of variation for the multiplicative weights is 1.4, compared with 0.9 for the additive weights.

Table 5. Summary of Impact Ratings

Weights	*All Studies*	*Excluding Studies with Design Quality=1*
None	0.3	-0.1
Number of schools	0.3	0.1
Design quality	0.2	0.0
Number of years	0.0	-0.3
Schools + quality + years	0.3	0.1
Schools x quality x years	0.1	0.1
Number of studies[†]	18	9
Average of all methods	0.2	0.0

[†] As discussed in Section 3, a "study" includes separate analyses in the same report in cases where we were unable to identify a single best analysis.

Given that data constraints placed serious methodological limitations on many of the studies, we recalculated the average impact ratings (both weighted and unweighted) after excluding studies that received a design quality rating of 1 (see Table 5).. As explained in Section 3, these studies either analyze consecutive cohorts without a comparison group or provide a cross-sectional analysis with a comparison group (but no statistical controls). This action cuts the number of usable studies in half. Among the excluded studies are all of the Colorado studies.[18] Also excluded is the Georgia study which, although it charts changes over time, provides no comparison group against which to gauge these changes. Indeed, without a comparison group, it is impossible to know whether observed changes in charter school test scores are unique to charter schools. Finally, we excluded the cross-sectional analysis from Miron & Nelson's (2000) analysis of Pennsylvania charter schools.

Excluding the lowest quality studies depresses the average impact rating to a very small degree, but changes the overall picture very little. Where the impact rating (averaged over all weighting schemes) was 0.2, the same figure excluding the lowest quality studies was 0.0. One reason that dropping the lower quality studies has little effect on the overall impact ratings is that there is almost

[18] It is curious that Colorado has produced five studies, four of which provide only cross-sectional analyses. Only the most recent study provides any analysis of gains over time, and then only across two years.

no relationship between the impact ratings of the studies and their quality indicators. Figure 4 graphs impact ratings for each study against the design quality index, with larger circles indicating studies that included more years of data. The graph shows little or no evidence of a pattern.[19]

In summary, a variety of weighting schemes and decision rules for including and excluding studies all suggest that charter schools have a mixed impact on student achievement. Focusing only on the highest quality studies, a study of Arizona provides relatively strong evidence of a positive charter school impact on student achievement. Other relatively high-quality studies of Texas and Connecticut produce weaker, though still positive, findings. The positive findings, however, are counterbalanced by relatively high quality studies that provide evidence of negative charter school

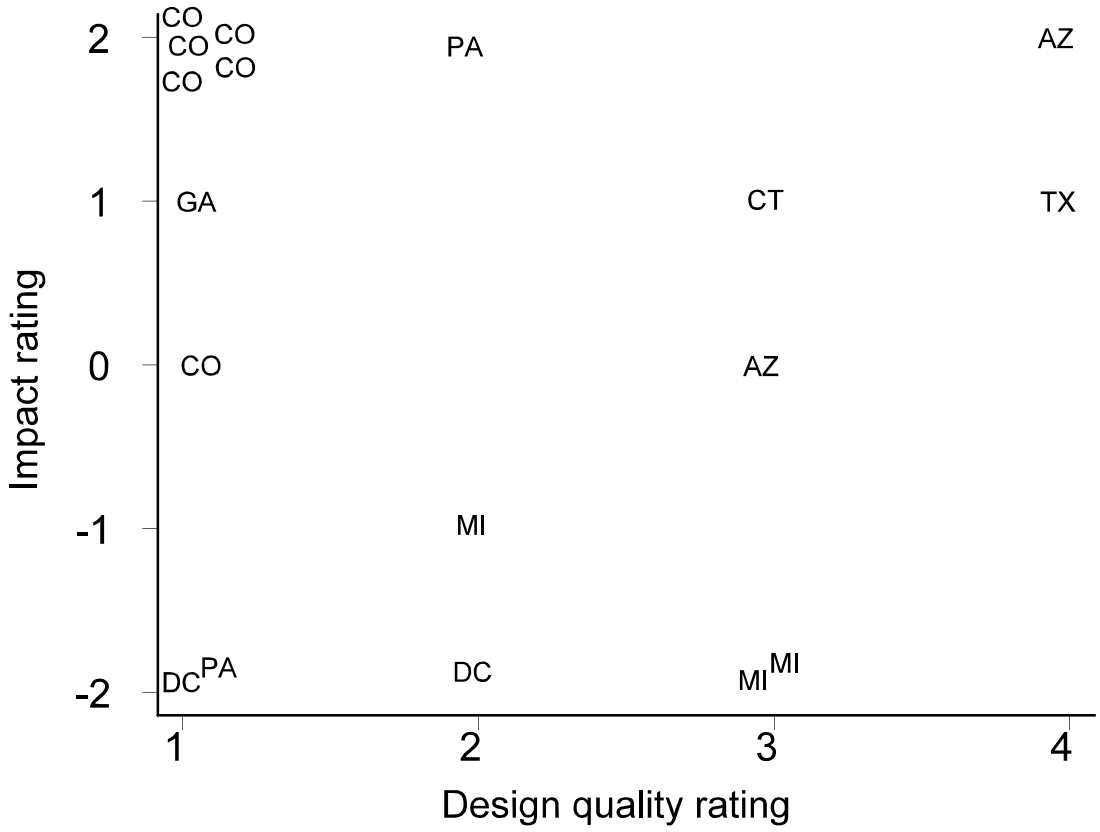

Figure 4. Impact Rating and Design Quality Rating for Charter School Evaluations

[19] We also checked the correlations between impact ratings and all of the weighting schemes discussed above. In no case did Spearman's ρ exceed 0.3.

impacts in Michigan and the District of Columbia. The lower-quality studies, as a group, provide a slightly more positive picture of charter schools' impact on student achievement, with a number of Colorado studies and a study of Georgia showing positive impacts and only two studies showing negative impacts. However, the addition of the lower-quality studies does relatively little to change the overall conclusion that evidence of charter schools' impact on student achievement is mixed.[20]

4.3 Effect Sizes

As discussed above, few of the studies provide clear discussions of effect sizes. It is for this reason that we resorted to the somewhat imprecise system of ordinal impact ratings presented in Table 1. Because some studies do, in fact, specify effect sizes, we briefly discuss that evidence here. The clearest statements of effect size came in the studies that utilized multivariate regression techniques. Most of these studies estimated several models, some stronger (i.e., including more controls for background differences among schools and students) than others. For the most part, we restrict our discussion to effect sizes derived from the strongest regression models.

The clearest discussion of effect size is presented in Solmon, Paark, & Garcia's (2000) study of Arizona charter schools. Comparing the typical size of the charter coefficients across their many regression specifications with the observed standard deviation in achievement scores, they conclude that the average effect of Arizona charter schools approximate 0.1 standard deviations. In order to place this value in context, they draw comparisons with the estimated effect sizes of other educational interventions, noting, for instance, that class size research has suggested an effect size of 0.3 associated with reducing class sizes from 35 to 20. The authors also contextualize their effect size estimates by comparing them with costs in a benefit-cost ratio. They note that since average per-pupil expenditures are significantly lower in charter schools, charter schools provide more education for less money.[21]

---

[20] Even the unweighted averages indicate that excluding the lower-quality studies has relatively little impact on the overall picture (see the first row of Table 5).

[21] Nelson & Hollenbeck (2001) argue that the appropriate way to address costs is in terms of marginal, not average costs.

Other regression-based studies provide information about effect sizes by reporting regression coefficients, but do not express them in terms of the standard deviation of the outcome variable. Bettinger's (1999) study of Michigan charter schools found that the two- and three-year growth in charter schools' pass rates on the state exam lag behind growth in comparable noncharter schools by a factor of 9 to 10 percent. He also found that increases in the percentage of charter school students in the "low" category on the state exam increased by some 6 to 7 percentage points over the same period. Similarly, Eberts & Hollenbeck's (2001) regressions of Michigan test scores suggest that charter school students score 2 to 7 percent lower on the state exam than comparable noncharter students. Finally, Gronberg & Jansen's (2001) study of Texas charter schools found that the impact of charter schools overall is approximately -0.9 points on the Texas Learning Index (TLI). However, when they partitioned the data into at-risk and non-at-risk schools, they found that while non-at-risk schools still have a negative impact of 1.56 points, at-risk schools have a positive impact of 0.8 points on the TLI.[22] Unfortunately, without further analysis it is not possible to determine whether the Arizona findings are more positive than the Michigan findings are negative.

5. Explaining Variations Across States

As noted in Section 3, there is substantial heterogeneity across states in the nature of the charter school treatment, different laws and educational administrative structures create different ways of actualizing the charter school concept. This heterogeneity across states creates an opportunity for researchers to identify the policy-related and other conditions under which charter schools can have a positive (or negative) impact on student achievement. In the spirit of states as the "laboratories of democracy," researchers might be able to identify "policy levers" in some states that might guide reforms across the states. We investigated a number of fairly straightforward explanations for variations in observed charter impact across the states, including "strength" of law and the frequency

[22] It is important to note that the authors show that at-risk students choosing to attend charter schools score higher on the TLI than those choosing to remain in traditional public schools, thus providing some evidence of cream-skimming from the ranks of at-risk students.

with which charter schools have been closed. Neither variable appears to explain variations in charter schools' impact on student achievement. As for strength of law, one finds low, middling, and high performers (e.g., Michigan, Texas, and Arizona) among the ranks of states receiving a grade of "A" in the Center for Education Reform's rating system.[23] The same holds for the frequency of closures, where both high and low performing states (e.g., Arizona and the District of Columbia) have closed a relatively high proportion of their schools. (Table 6 presents data on these background factors). Clearly, there are other variables at work, and we encourage other scholars to explore these factors.[24] These might include forms of oversight other than school closures. In Michigan, for instance, there is little evidence that university authorizers (who hold the lion's share of charters in the state) provide effective oversight of the schools they authorize.

Table 6. Summary of Findings, Study Quality, and Explanatory Factors by State

State	*Year of Law*	*Average Impact Rating*	*Average Design Quality Rating*	*CER Rating*	*% Schools Closed*	*# Schools Closed*
AZ	1994	2.0	3.5	A	4.7	21
DC	1996	-2.0	1.5	A	5.0	2
MI	1993	-1.7	2.7	A	3.1	6
TX	1995	1.0	4.0	A	5.9	10
CO	1993	1.5	1.0	B	2.0	2
PA	1997	-2.0	1.0	B	1.5	1
CT	1996	1.0	3.0	C	5.9	1
GA	1993	1.0	1.0	C	0.0	0

[23] In the Center for Education Reform's (CER) rating system, the laws most supportive of charter schools receive an A, while less supportive laws receive Bs, Cs, Ds and Fs.

[24] Demographic factors are already factored into these analyses, since most studies we considered have at least rudimentary controls for background factors and estimate the charter impact compared with some sort of comparison group of similar students. To test for sensitivity to variations in the quality of the studies, we examined the relationships only among studies that received a 2 or higher on our design quality index. Once again, there was little or no apparent relationship between these factors and the strength and direction of charter schools' impact on student achievement.

6. Why We Know So Little About Student Achievement in Charter Schools

This paper set out to answer two questions about charter schools' impact on student achievement. First, what do we know about it? Second, why do we know so little? Indeed, perhaps more striking than the substantive findings of the studies is how few studies there are and how few states with charter school laws these studies cover. While it is impossible to know with any certainty, in this section we briefly speculate about some of the reasons for the paucity of good empirical evidence.

One set of reasons resolves around limitations in the availability of achievement data. First, some states (e.g., New Hampshire) have only a few charter schools, limiting the utility of statewide evaluations of charter school achievement. Second, even in states with large numbers of charter schools many of the schools have been open for only a year or two, leaving many schools with little or no student achievement data to report. Third, many states test students at only a few grade levels. Indeed, in a state that does not test students until the fifth grade, there will be no broadly comparable student achievement from charter schools offering only the primary grades (K-3). Fourth, student confidentiality policies often mean that state testing systems provide little or no data from small schools (which often test fewer than 10 students in any given subject and grade). Finally, a number of states have recently changed testing regimes (e.g., Illinois), thus severely limiting evaluators' ability to make reliable temporal comparisons in student achievement.

Data limitations notwithstanding, in some states charter schools–for a variety of reasons–apparently collect student achievement data but do not report it to a centralized source. This leaves evaluators with the difficult task of collecting data–often based on different tests–from schools scattered across large geographical areas. In other states, by contrast, charter schools report extensive accountability data (including student achievement), although these reports go to local authorization agencies, not the state education agency. Once again, this dramatically increases data collection costs for would-be evaluators. We have sought in many of our own statewide evaluations to collect norm-referenced test data directly from schools. However, a combination of reluctance on the part of charter school officials and the sheer scale of the task has yielded little usable data.

Finally, political factors may discourage state officials and others from commissioning, sponsoring, and funding statewide evaluations of charter school achievement. As Levin (2001) has

noted of cost-effectiveness analysis, high-profile, statewide evaluations of charter schools' impact on student achievement may appear to policymakers as a threat to their discretion, as unexpected findings might diminish the credibility of popular policy approaches. Officials, in short, might decide that commissioning a high-profile statewide evaluation is not worth the risk. Similarly, many core policy questions related to charter schools are imbued with fundamental moral questions, such as the intrinsic value of choice, the moral weight one gives to the often competing goals of equity and efficiency, and so on (Miron & Nelson, in press). Since such moral judgments are often not grounded in empirical knowledge, evaluation studies about charter schools and student achievement may seem beside the point to policymakers and stakeholders attempting to grapple with these issues.

7. Conclusions

This paper sought to summarize what is currently known about charter schools' impact on student achievement. The question is of paramount importance given that at the heart of the charter school concept lies a bargain whereby the schools received enhanced autonomy in exchange for greater accountability for student outcomes. As we have stated, student achievement is not the only relevant student outcome, but certainly it is among the most important outcomes of charter schools.

In spite of the topic's importance to the debate over charter schools and school reform, it is striking how little we currently know about charter schools' impact on student achievement. As discussed above, we found useable independent evaluations of achievement impacts in only 8 of the 38 states with charter school laws. Leaving aside the recently enacted Indiana law, as well as the states with laws but no currently operating charter schools, this amounts to less than a quarter of the states with charter school laws. If we restrict our view to relatively high quality studies, there are such studies for only 5 states. Thus, it cannot be emphasized enough that any conclusions drawn from the body of existing research on charter schools must be treated with an appropriate degree of skepticism. The reasons for the paucity of evidence—as discussed in the previous section— include limitations in state testing regimes, charter schools' failure to report data, state agencies' failure to collect such data centrally, and political motivations.

Given these caveats, we find that the existing body of research on charter schools' impact on student achievement reveals a mixed picture, with studies from some states suggesting positive impacts, studies from other states suggesting negative impacts, and some providing evidence of both positive and negative impacts. Of the methodologically strongest studies, the strongest positive findings come from Arizona, while the strongest negative findings come from Michigan. Another methodologically strong study from Texas found evidence that while charter schools have an overall negative impact on student achievement, schools classified as at-risk have a positive impact. The addition of relatively lower-quality studies does little to change this mixed view of charter schools' impact on student achievement.

Given the available data, we cannot yet identify any clear explanations for variations in charter school performance among states. Exploring such explanations is important because the ways in which the charter school concept is operationalized and implemented varies substantially across the states. Identifying the factors behind such variations might yield important insights about what types of conditions are most likely to support successful charter schools.

In summary, our knowledge of charter schools' impact on student achievement is still in its infancy, in spite of the fact that the movement is now ten years old. Relatively few researchers have adequately tackled the achievement question. In most cases this is due to limitations in available data as well as a lack of requests by the policymakers that legislated these reforms to evaluate them. Indeed, state policymakers need to consider ways to improve the evaluability of charter schools through improvements in data collection and testing practices.[25] However, there remain unexploited opportunities to learn more about the topic as more and more states make testing and other data widely available. Evaluators and researchers should seek out these new opportunities and reanalyze data from existing studies in order to learn more about charter schools' impact on student achievement. Without such efforts by both policymakers and researchers, the charter concept's promises about accountability may eventually ring hollow.

[25] We provided some preliminary recommendations for improving state data collection and testing practices in Horn & Miron (2000) and Miron & Nelson (2000).

References

Antonucci, R.V. (1997). *1997 test results from Massachusetts charter schools*. Available: http://www.doe.mass.edu/cs.www/reports/1997/toc.html.

Barr, J.E., Caillouett, P.B., & Welsh, S. (1999). *Evaluation summary: Charter schools in Louisiana, how are they performing?* Baton Rouge, LA: Board of Elementary and Secondary Education.

Bettinger, E.P. (1999). *The effect of charter schools on charter students and public Schools.* New York: National Center for the Study of Privatization in Education, Occasional Paper No. 4.

Britsch, B., & Geiger, E., (2000). *Alaska charter schools program evaluation report.* Portland, OR: Northwest Regional Educational Laboratory.

Center for Applied Research and Educational Improvement. (1998). *Minnesota charter schools evaluation final report.* Minneapolis, MN: Center for Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota.

Colorado Department of Education. (1997). *The Colorado charter schools evaluation.* Denver, CO: University of Colorado-Denver.

Colorado Department of Education. (1998). *1997 Colorado charter schools evaluation study: The characteristics, status, and student achievement data of Colorado charter schools.* Denver Author.

Colorado Department of Education. (1999). *1998 Colorado charter schools evaluation study: The characteristics, status, and student achievement data of Colorado charter schools*. Denver: Author.

Colorado Department of Education. (2000). *1998-1999 Colorado charter schools evaluation study: The characteristics, status, and performance record of Colorado charter schools.* Denver: Author.

Colorado Department of Education. (2001). *The state of charter schools in Colorado: 1999-2000: The characteristics, status, and performance record of Colorado charter schools.* Denver: Author.

Delaware Department of Education. (2001). *Delaware charter schools: Fourth annual state report.* Dover, DE: author.

Eberts, R.W., & Hollenbeck, K.M. (2001). *An examination of student achievement in Michigan charter schools.* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Elazar, D. (1966). *American federalism: A view from the states.* New York: Crowell.

Geiger, E., Schwendiman, J., Britsch, B., Hornibrook-Hehr D. & Melvin, J. (2000). *Idaho charter schools: Program evaluation report year one*. Portland, OR: Northwest Regional Educational Laboratory.

Georgia Department of Education. (2000). *Annual charter school report: 1999-2000 school year*. Atlanta, GA: Author.

Gifford, M., Phillips, K., & Ogle, M. (2000). *Five year charter school study: An overview.* Phoenix, AZ: The Goldwater Institute's Center for Market-Based Education.

Goldwater Institute. (1996). *Arizona's charter schools: A survey of teachers.* Phoenix, AZ: author.

Gronberg, T., & Jansen, D.W. (2001) *Navigating newly chartered waters: An analysis of Texas charter school performance*. Austin, TX: Texas Public Policy Foundation.

Hedges, L., & Olkin, I. (1985). *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press.

Henig, J.R., Moser, M., Holyoke, T.T., & Lacireno-Paquet, N. (1998). *Making a choice, making a difference? An evaluation of charter schools in the District of Columbia.* Washington, DC: The Center for Washington Area Studies, The George Washington University.

Henig, J.R., Holyoke, T.T., Lacireno-Paquet, N., & Moser, M. (2001). *Growing pains: An evaluation of charter schools in the District of Columbia; 1999-2000.* Washington, DC: The Center for Washington Area Studies, The George Washington University.

Horn, J., & Miron, G. (1998) *First annual report of the evaluation of the charter schools and the charter school initiative in the state of Connecticut.* Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Horn, J., & Miron, G. (1999). *Evaluation of Michigan public school academy initiative.* Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Horn, J., & Miron, G. (2000). *An evaluation of the Michigan charter school initiative: Performance, accountability and impact.* Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Horn, J., & Miron, G. (2001) *Interim report of the evaluation of the charter schools and the charter school initiative in the state of Connecticut.* Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Kane, P. R. (1998). *New Jersey charter schools: The first year, 1997-1998.* New York: Teachers College, Columbia University.

Kansas State Department of Education. (2000). *Kansas charter schools accountability report for school year 1998-1999.* Kansas City, KS: author.

Khouri, N., Kleine, R., White, R., & Cummings, L. (1999). *Michigan's charter school initiative: From theory to practice.* Lansing, MI: Public Sector Consultants.

Kleine, R., Scott, C., & White, R. (2000). *Issues in Michigan's public school academy initiative phase II*. Lansing, MI: Public Sector Consultants.

Legislative Audit Bureau. (1999). *An evaluation charter school program.* Madison, WI: Author.

Levin, H. (2001, summer). Waiting for Godot: cost-effectiveness analysis in education. *New Directions for Evaluation,* 90, 55-68.

Light, R., & Pillemer, D. (1984). *Summing up: the science of reviewing research.* Cambridge, MA: Harvard University Press.

Massachusetts Department of Education.(1996). *The Massachusetts charter school initiative: 1996 report.* Boston: Charter School Office, Massachusetts Department of Education.

Massachusetts Department of Education. (1997). *Test Results from Massachusetts charter schools: A preliminary study 1997*. Boston: Charter School Office, Massachusetts Department of Education.

Massachusetts Department of Education. (1998) *The Massachusetts charter school initiative: 1998 report.* Boston, MA: Charter School Office, Massachusetts Department of Education.

Miron, G. & Nelson, C. (2000). *Autonomy in exchange for accountability: An initial study of Pennsylvania charter schools*. Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Miron, G. & Nelson, C. (in press). *What's public about charter schools: Lessons learned about school choice and accountability*. Thousand Oaks, CA: Corwin Press.

Mulholland, L. (1999). *Arizona charter school progress evaluation.* Phoenix: Morrison Institute for Public Policy, Arizona State University.

Nelson, C., & Hollenbeck, K. (2001). *Does charter school attendance improve test scores?: Comments and reactions*. Kalamazoo, MI: W.E. Upjohn Institute Working Paper.

North Carolina Department of Public Instruction. (1998). *Selected characteristics of charter schools, programs, teachers and students*. Raleigh: Division of Accountability Services, Evaluation Section, North Carolina Department of Public Instruction.

Rouse, C.E. (1997). *Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program*. Washington, DC: National Bureau of Economic Research Working Paper # 5964.

RPP International. (2000). *The state of charter schools.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Solmon, L., Paark, K., & Garcia, D. (2001). *Does charter school attendance improve test scores?* Arizona: Goldwater Institute 's Center for Market-Based Education.

SRI International. (1997). *Evaluation of charter school effectiveness.* Available: http://www.lao.ca.gov/sri_charter_schools_1297-part2.html.

Wells, A.S., Ariles, L., Carnochan, S., Cooper, C.W., Grutzik, C., Holme, J., Lopez, A., & Scott, J. (1998). *Beyond the rhetoric of charter school reform: A study of ten California school districts*. Los Angeles: UCLA Charter School Study.

WestEd. (1998). *The findings and implications of increased flexibility and accountability: An evaluation of charter schools in Los Angeles Unified School District*. Los Angeles, CA: WestEd and University of Southern California.

Wisconsin Legislative Audit Bureau. (1998). *An evaluation: Charter school program.* Madison, WI: Joint Legislative Audit Committee.

Witte, J. (1997). *Achievement effects of the Milwaukee Voucher Program.* Paper delivered at the annual meeting of the American Economics Association. New Orleans, LA.