Chapter 4

Bringing the Full Picture into Focus: A Consideration of the Internal and External Validity of Charter School Effects

Douglas Lee Lauen and Kyle Abbott

In the United States, students enrolled in primary and secondary school attend a wide variety of schools operating in different settings and under various institutional arrangements. Of those enrolled in public or private schools, about 71% attend a neighborhood-based assigned traditional public school. The rest attend alternatives that serve students from many neighborhoods: 15% attend a public magnet school, about 5% attend a public charter school, and about 10% a private autonomous school (Snyder, De Brey, & Dillow, 2019). In this chapter, we focus on public charter schools, which are publicly funded semiautonomous schools designed as alternatives to traditional public zoned schools attended by most students. Charter schools challenge the notion that the state should control the practices and curriculum students experience. They serve a small fraction of students in most areas of the U.S., but this portion is growing rapidly, and, in some big cities, charter schools serve more than one third of students. These schools began as laboratories of educational innovation in the early 1990s. They are freed from many important financial, hiring, and curriculum regulations state and local authorities impose on schools. In exchange, they have a *charter*, a performance agreement, that a chartering agency, or authorizer, monitors and can rescind if violated. Depending on the state, these authorizers are state and district boards of education, higher education institutions, or nonprofits. Some states have only one charter school authorizer; others have many. Charter schools, therefore, are an example of an alternative institutional approach to state oversight of education. The sociological setting of interest in this chapter is therefore the institutional arrangements that surround education—the norms, practices, laws, and regulations that govern how schools operate (DiMaggio

D. L. Lauen (✉) · K. Abbott

Department of Public Policy, University of North Carolina at Chapel Hill,
Chapel Hill, NC, USA

e-mail: dlauen@unc.edu; kyleta@live.unc.edu

© The Author(s) 2021, corrected publication 2022

T. Freytag et al. (eds.), *Space, Place and Educational Settings*, Knowledge and
Space 16, https://doi.org/10.1007/978-3-030-78597-0_4

& Powell, 1983; Meyer & Rowan, 1977; Meyer & Scott, 1992; Scott, 1998; Weick, 1976).

Some observers view charter schools as the central mechanism for educational reform because they have the potential to free students from low-performing traditional public schools and provide them with a different, and potentially better, education at no cost to the family (Budde, 1996; Chubb & Moe, 1990; Friedman, 1962; Hanushek, 2006). Others view charter schools as a neoliberal ploy to drain public schools of funding and public support, turning a public institution (the “common” school that educates everyone in the same way) into a fragmented collection of schools serving narrow interests in private markets in which schools compete for students (Fiske & Ladd, 2000; Lubienski, 2003; Ravitch, 2010).

Where one stands on this debate depends at least in part on how one feels about markets providing public goods. Clearly, education has positive externalities (spill-over effects that create benefits not only for the student who receives the education, but for society as a whole). For example, teaching a student to read helps that student progress in school and get a decent job later in their life. Teaching everyone to read creates an educated populace that can create jobs, foster innovation, and promote healthy behaviors, as well as many other social goods. If schooling’s only goal were to promote reading skills, we might not be too concerned if these skills were imparted by different types of schools. The key problem is that policymakers and parents want schools to teach many academic, social, and cultural skills; impart specific kinds of knowledge and not others; and promote cultural socialization. Moreover, many stakeholders want all students to attend the same types of schools with the same curriculum, so they receive the same educational opportunities and gain appreciation for cultural difference, thereby maintaining support for public schools, community ties, and the common welfare. Accordingly, critics of charter schools question whether liberalizing educational markets will promote the socialization mission of public schools and are troubled when taxpayer dollars are spent on a sector, such as charter schools, that is less accountable to elected officials.

In this chapter, we focus on the evidence of whether an institutional experiment in the delivery of education to students—charter schools—has succeeded in developing student academic skills in reading and math. We concentrate less on the theory of change and more on empirical evidence across a wide range of quantitative studies about test score effects. After 30 years of research in the effects of charter schools, there is now ample evidence on test score achievement effects to draw conclusions. This focus necessarily leaves aside the question of whether charter schools fray the social fabric, harm other types of schools, or promote harmful social values, as well as other important questions for which there is much less empirical research available to summarize.

We argue that charter school research provides an interesting case study for the interplay between internal validity and external validity. Most studies are strong in either internal validity or external validity, but not both. Therefore, policy conclusions are difficult to draw from any one study. But, when considering the studies in this review from over three decades of empirical research, we can draw a few conclusions: (1) charter school studies with the strongest internal validity (a

randomized controlled trial, RCT) are rare, involve small and unrepresentative samples, and produce quite positive effects on test-score achievement; (2) studies with good external validity are quite common and produce small and mixed results, but are viewed as having lower internal validity than RCTs; and 3) as the number of schools in the study sample increases, the charter school effect size (relative to the effect of traditional public schools) shrinks toward zero. This chapter is relevant to the theme of this volume in a second sense: Empirical findings can vary a great deal across both geographic settings and studies with smaller and larger collections of schools. Therefore, considering variation across settings in empirical research is essential to understanding effectiveness.

We begin this chapter with a brief introduction to the nature and extent of charter schools in the U.S., with some reference to similar types of schools in other countries. We then outline a theoretical framework that defines estimation error as a function of both treatment selection error (a threat to internal validity) and sample selection error (a threat to external validity). We argue that policy conclusions should be based on a critical assessment of the research base that comprises both internal and external validity. A critical summary of the empirical research on the effect of charter school enrollment on test-score achievement follows. In this review, we carefully consider sample size (the number of schools in the analysis), setting (urban centers, statewide, nationwide), and research design in drawing conclusions from the research. Following this summary, we conclude with a discussion of this review's implications for research and policy, with a focus on what types of evidence are needed and at what level (local, state, or federal).

Charter Schools in the U.S.

Over the last 30 years, charter schools have become increasingly popular attempts by state and local governments to insert market forces into education delivery, with the goals of improving outcomes for students and containing costs for units of government. Charter schools are public schools that are exempt from residential assignment and many other regulations that govern traditional public schools. In this section, we describe charter schools in detail, including how they came to be, how widespread they have become and where they are located, and we place them in the context of education internationally.

In the United States, education is decentralized with a limited role for the federal government. State and local governments have traditionally maintained nearly all responsibility for primary and secondary education, with local school districts playing the largest role. School districts are organized as their own local governments and corporate entities, with governing boards that are elected separately from other local units. School district boundaries do not necessarily follow from city or county borders. In total, the U.S. has 13,598 school districts. The state of Texas has the most at 1081 and the state of Hawaii has the least, with just a single statewide school district (Snyder et al., 2019). These local school boards have varying authority by

state, but most are able to raise revenue via property and sales taxes, issue bonds (with approval by voters), acquire and hold real property, set standards and curricula, hire staff, and supplement staff salaries.

One way to measure the federal government's role is to look at their education expenditures relative to state and local governments. In fiscal year 2016, the federal government provided \$56 billion in revenue to public schools, only 8.3% of the total. Most federal funding goes to economically disadvantaged schools and special education programs. In contrast, state governments provided \$318.5 billion and local governments provided \$303.8 billion (National Center for Education Statistics, 2018). As the federal government has a very limited role, we focus this section on the state and local government role in charter school oversight.

A charter school is a public school that operates under a contract or charter with the state or an agency empowered by the state to authorize the charter. This could be a university, local school district, or state board of education. Charter schools are generally exempt from most state regulations but are held to similar performance standards. As public schools, they cannot charge tuition or set admissions requirements. If more students want to enroll than seats are available, schools must admit students via lottery. Unlike traditional public schools (TPS), which have catchment zones, they do not serve students in particular neighborhoods or zones. Any student within a state may apply to enroll in a charter school. Nearly without exception, charter schools are schools of choice, where families must take affirmative steps to enroll, whereas the neighborhood TPS is the default schooling option (Epple, Romano, & Zimmer, 2016).

As of 2018, more than three million students in 44 states enrolled in over 7000 charter schools. Enrollment has increased by more than one million students since 2011. Although this is significant growth, charters enroll less than 5% of all students nationwide. Six states do not have authorizing statutes and therefore no charter schools operating within them (National Alliance for Public Charter Schools, n.d.). States vary widely in the proportion of students that enroll in charter schools. At 12%, Arizona has the most students enrolled in charter schools as a share of all public school students. In terms of total enrollment, California has the most students enrolled in charter schools, as they have 20% of all charter students nationwide (Epple et al., 2016).

Although there are many charter schools in rural and suburban areas, they are disproportionately located in densely populated areas. Only about one in four TPS's lie in urban areas, whereas more than half of charters schools are in cities. Of the 30 locales with the highest percentage of students enrolled in charters, all are majority urban or based in urban areas. Students in charter schools are more likely to be nonwhite and lower income than students in TPS's (Epple et al., 2016).

U.S. charter schools can be compared to similar school types in international contexts. Academy schools in the United Kingdom are similar. They are publicly funded and cannot charge tuition, yet they are operated by mostly nongovernmental entities such as nonprofit trusts, universities, other schools, or faith groups. They are also exempt from many state regulations, but must follow the same rules on admissions, special needs, and exclusions as state schools. They do not have to follow the

national curriculum, but still must perform to standards as measured by the same standardized tests. They are periodically inspected by the Office for Standards in Education, Children's Services, and Skills. Academies are much more common than charter schools. As of 2015, 61.4% of public secondary schools in the U.K. were academies. In the U.S., only 6.5% of secondary schools are charters (Government Digital Service, 2016; National Center for Education Statistics, 2018).

Other European governments have implemented similar arrangements. German private schools may only operate if they do not cause segregation by income. If the school abides by this rule, they receive public funding and maintain some degree of autonomy, similarly to U.S. charter schools. The Netherlands has a system of universal school choice, where families can choose public or private schools at no cost to them (Government of the Netherlands, 2017). Public schools operate similarly to charters in the U.S., where if schools are oversubscribed, the local government must provide a seat elsewhere.

Theoretical Framework

We posit that research to inform policy should have high validity—specifically, that policy decisions about charter schools should be based in large part on evidence from large-scale quantitative studies with high internal and external validity. In other words, researchers should be confident in assessments of policy effects from the past and reasonably sure that if they designed new policy, they would could predict positive results for students, on average. The key idea is that researchers should be able to draw a causal inference from their findings; in short, they should be confident that their observations are not the result of anecdotal evidence, biased reporting, confounding, selection bias, or random chance, but instead reflect some reliable and repeatable relationship from cause to effect. Consider a simple example: Suppose you want to get rid of a headache and are considering taking an aspirin. If you take the aspirin (the cause), you want to know if your headache is likely to go away (the effect). Similarly, researchers want to know whether educational interventions will reliably improve student outcomes for students on average.

Our definition of causality is informed by the extensive literature on causal inference (Holland, 1986; Imbens & Rubin, 2015; Morgan & Winship, 2015; Pearl, 2000). These authors refer to all outcomes as *potential outcomes*, some of which researchers can observe and some of which they cannot because they were never realized. For example, suppose one is considering a charter school's effect for student i on outcome Y and that one can observe whether the student attended the charter school (coded as 0 if she did not and 1 if she did). In theory, student i has two potential outcomes: $Y_i(1)$, if she attended the charter school, and $Y_i(0)$, if she did not. Suppose she in fact attended a charter school—then $Y_i(1)$ is observable as Y and $Y_i(0)$ is unobserved and is called her *counterfactual* outcome. Conversely, if she in fact did not attend a charter school, $Y_i(0)$ is observable as Y and $Y_i(1)$ is her unobserved counterfactual. The treatment effect for student i is $TE_i = Y_i(1) - Y_i(0)$, or the

difference in individual potential outcomes. In short, to determine the causal effect of attending a charter school for student i , one must compare her outcomes from simultaneous enrollment in a charter and in a non-charter school, which is obviously impossible in practice, but is useful to consider as a theoretical ideal.

Although the individual treatment effect for person i is not observable, under some conditions it is possible to calculate average treatment effects across many persons. Averaging over either the sample or the population produces either the sample average treatment effect (SATE) or the population average treatment effect (PATE), where I_i is an indicator for being in the sample, n is the total sample size, and N is total population size (Imai, King, & Stuart, 2008):

$$SATE = \frac{1}{n} \sum_{i \in \{I_i=1\}}^n TE_i$$

$$PATE = \frac{1}{N} \sum_{i=1}^N TE_i$$

The credibility of either SATE or PATE rests on assumptions for making causal inferences, such as randomization balancing units on pre-treatment observable and unobservable confounds in either the sample or a population. For example, suppose there are 100 charter schools in a state. A researcher conducts a study of ten of these, each of which has more student applicants than seats (i.e., they are oversubscribed) and are willing to allocate seats through a random lottery. We will call this the *lottery study*. For this study, we will assume that the ten schools volunteered to be part of the study. The estimate from this study would be the SATE, because not all charter schools participated. Although the individual treatment effect for a student who attended this school is unknown, one can usually safely presume that the average counterfactual outcome for students who were randomized to attend one of the charter schools is the average outcome for students who were randomly denied admission to a charter school. In other words, because the charter schools randomly accepted students, one can construct a reasonable counterfactual for the students who attended.

Now, consider a nonrandomized study. A researcher conducts a study of all 100 charter schools in the state but does not use the lottery information as part of his research design. One can call this the *observational study* and assume that schools did not need to consent to be part of the study and were not necessarily oversubscribed (i.e., some had more seats than applicants). Assuming the population of interest is the state (rather than a national estimate), the estimate from the observational study is the PATE.

If potential outcomes are generated through an additive model (no interactions between the component parts of the equation below), one can decompose total estimation error, Δ as (Imai et al., 2008):

$$\Delta = \Delta_S + \Delta_T = \Delta_{SX} + \Delta_{SU} + \Delta_{TX} + \Delta_{TU}$$

This decomposition makes clear that error can arise from sample selection error (Δ_S) and from treatment selection error (Δ_T). Both types of error can be further decomposed into observable (X) and unobservable (U) errors (ibid). An ideal design is one in which units are first randomly drawn from a well-defined population and then individually randomized to treatment. Assuming no loss to follow up, this design would have no sample selection error and no treatment selection error, in expectation. This type of design is costly and rare unless sites can be forced or strongly incentivized to participate in the study. It is perhaps not surprising that perhaps the only two U.S.-based studies of this type in social research are large federally mandated studies of federal programs, Upward Bound and Head Start (U.S. Department of Education, 2009; U.S. Department of Health and Human Services, 2010).

Returning to the charter school example, the lottery study would likely have high internal validity because $E(\Delta_T) = 0$. Unless there is selective attrition, the lottery should remove treatment selection error as a threat to validity because randomization balances units on pretreatment observable (Xs) and unobservable (Us) confounds. For this reason, the lottery study may be a good estimate of the SATE. On the other hand, because the researcher relies on volunteer schools that were all oversubscribed, one might expect that $E(\Delta_S) \neq 0$. In other words, perhaps the more effective charter schools were oversubscribed, could hold a lottery, and were willing to agree to evaluation. These facts could lead to misleading estimates, where $SATE \neq PATE$.

Conversely, the observational study may have lower internal validity because $E(\Delta_T) \neq 0$ due to omitted confounders in the analysis. For example, suppose the study does not adequately capture parenting skill, student motivation, or family socioeconomic status. The omission of these factors could result in biased estimates of the charter-school effects if these factors cause families to choose charter schools and result in observed student outcomes. On the other hand, the observational study may have $E(\Delta_S) = 0$ because the study includes all charter schools in the population of interest (i.e., the entire state). Therefore, even though the observational study may include all schools in the state, it may produce a biased estimate of the PATE due to treatment selection error.

In summary, studies about school effects face a tradeoff between internal validity (treatment selection error) and external validity (sample selection error). Studies generally fall into two categories: (1) small convenience samples of nonrepresentative schools that hold lotteries and consent to be part of a study, or (2) large district-, state-, or multi-state population wide samples of nonlottery estimates. We argue in this chapter that policy conclusions should be made on the totality of the evidence from many different studies and that both internal and external validity are important for this purpose.

Evidence About Effectiveness

The literature on charter school effectiveness is vast and summarizing it all is well beyond the scope of this section. However, we can summarize what we consider a representative sample of the literature. We categorize studies along two dimensions, internal validity and external validity. Generally, high-quality studies have one or the other, depending on their specific methodology. We will examine each type of methodology by describing a sample of studies within each category in chronological order. Below, you will find a graph categorizing each study by internal and external validity. The categorization scale is simply ordinal: A study rated a five is not 20% better than a study rated a four along any dimension. Each is judged relative to the others in the sample, so we are not saying that any study is the best or worst ever done on the subject. We judge internal validity based on the study design's ability to reduce treatment selection error. We evaluate the external validity of each study by the number and type of settings where it takes place. For example, a single school lottery study has very low external validity, whereas a near population level study (e.g., CREDO, 2013) has high external validity. We define single locale studies as those conducted in only one city. Multiple local studies are either done in an entire state or across multiple cities or states.

Single Locale Lottery Based Studies

Lottery-based studies in a single locale have high internal validity and low external validity. They also tend to produce the largest and most consistently positive effects. This may be because schools that conduct lotteries are the highest performing with the best reputations, and high-performing schools are generally more willing to subject themselves to evaluation. Of the five single-locale, lottery-based studies (see Table 4.1), none found negative effects of attending a charter school on student achievement. However, because of their low external validity, it is very difficult to generalize these results to charter schools or students in the broader population.

Hoxby and Murarka (2009) conducted a lottery study of virtually all charter schools in New York City (42 schools) and found a positive and significant impact of 0.09 standard deviations per year of attendance in math and 0.04 standard deviations per year in reading on standardized tests for grades 3–8. The largest contributor to these gains was the extended school year. Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011) found positive and statistically significant results of eight Boston charter middle and high schools using lottery data. The most consistently positive effects were in oversubscribed, *No Excuses* charter schools. *No Excuses* charters emphasize frequent testing, increased instructional time, strict discipline relative to TPS's, and a constant focus on math and reading achievement.

Table 4.1 List of studies in review

Study	Location	Number of charter schools	Number of students	Method	Result	Effect size: Mathematics	Effect size: ELA/Reading
Fortson, Gleason, Kopa and Verbitsky-Savitz (2015)	Multiple	N/A	22,152	Random assignment by lottery, within study comparison	No statistically significant impact except OLS specification (+)	.06 (OLS), 0 otherwise	.06 (OLS), 0 otherwise
Gleason, Clark, Tuttle and Dwoyer (2010)	Multiple	36	2904	Random assignment by lottery	No statistically significant impact	Null	Null
Abdulkadiroğlu, Angrist, Narita and Pathak (2017)	Denver, CO	31	10,203	Random assignment with propensity score	Positive effects in math, reading, and writing	0.42	0.17
Hoxby and Murarka (2009)	New York City	42	37,242	Random assignment by lottery	Positive effects in reading and math	0.09	0.04
Abdulkadiroğlu et al. (2011)	Boston, MA	8	3370	Random assignment by lottery, supplemented with observational analysis	Positive effects in reading and math	0.38	0.20
Angrist, Cohodes, Dynarski, Pathak and Walters (2016)	Boston, MA	6	3205	Random assignment by lottery	Positive effects in reading and math	0.59	0.41
Curto and Fryer Jr. (2014)	Washington, DC	1	5487	Random assignment by lottery	Positive effects in reading and math	0.20	0.20
Credo (2009)	Multiple	2403	1,733,758	Matching	Negative effects in math and reading	-0.03	-0.01
Credo (2013)	Multiple	5068	20,172,202	Matching	Negative effects in math, positive in reading	-0.01	0.01

(continued)

Table 4.1 (continued)

Study	Location	Number of charter schools	Number of students	Method	Result	Effect size: Mathematics	Effect size: ELA/Reading
Davis and Raymond (2012)	Multiple	2242	618,522	Matching, fixed effects	Negative effects in math and reading	-.074 (FE), -.063 (Matching)	-.034 (FE), -.024 (Matching)
Zimmer, Gill, Booker, Lavertu and Witte (2012)	Multiple	617	N/A (only reported as student-year observations)	Student fixed effects	Insignificant or negative effects with the exception of math in Denver and Milwaukee	-.18 to .17	-.08 to null
Baude, Casey, Hanushek, Phelan and Rivkin (2020)	Texas	366	Approximately 11,000,000	Matching, fixed effects	Positive effects in reading and math	0.07	0.03
Bifulco and Ladd (2006)	North Carolina	34	495,943	Student fixed effects	Negative effects in math and reading	-0.16	-0.10
Ni and Rorrer (2012)	Utah	65	467,216	Student fixed effects	Negative effects in math, language, and reading	-0.09	-0.13
Zimmer and Buddin (2006)	Los Angeles and San Diego, CA	62	3,035,150	Student fixed effects	Mostly negative or null. Only one positive effect across 8 outcomes	-.09 (SD) -1.12 (LA)	-.08 (SD) .10 (LA)
Witte, Weimer, Shober and Schlomer (2007)	Milwaukee	22	??	Student fixed effects	Positive effects in math, no significant results in language	0.12	0.09
Imberman (2011)	Large, urban district in the southwest	32	Approximately 100,000 (only student-year n reported)	Student fixed effects	Positive effects in math, no significant results in reading	0.07	Null

Note: Source: Design by authors

Curto and Fryer Jr. (2014) use lottery data from one charter school, SEED Washington, which combines *No Excuses* instruction with a 5-day-per-week boarding program. They find 0.2 standard deviation increases in math and reading per year of attendance, with effects largely driven by female attendees. This study has the weakest external validity in our sample because SEED schools are unique in their treatment (*No Excuses* plus boarding), the sample sizes are small because the authors only used a single school's lottery, and all lottery applicants are black.

Angrist, Cohodes, Dynarski, Pathak, and Walters (2016) use lottery data from six Boston charter and pilot schools to build on the work of Abdulkadiroğlu et al. (2011). They find that attendance at one of Boston's charter high schools increases scores in the tenth-grade Massachusetts standardized test by 0.4 standard deviations in English and almost 0.6 standard deviations in math. This study matches a pattern we see in the literature about age of the charter sector. Outcomes tend to improve as high-performing schools improve and low-performing schools lose students or are shut down by an authorizing body.

Abdulkadiroğlu, Angrist, Narita, and Pathak (2017) use a design with randomized elements in Denver, Colorado public schools (DPS) from 2011–2012 to 2013–2014. This study is unique because the Denver school district has central assignment, where families request a seat at any public school in the district and are not bound by residential assignment. Parents rank up to five schools of any type, then students are assigned to a school based on eligibility for free or reduced lunch status, whether a sibling attends that school, and other factors. This helps correct for selection into charters, as each family must state school preferences. The authors used these lists to construct propensity scores to summarize the assignment mechanism, similarly to how researchers employ a stratified randomized design to estimate a causal effect conditional on one or more blocking variables. The lists allow them to match on each student's stated school preferences, not just observed demographic characteristics and test scores. Their matching estimates of pooled 4th–10th graders show large, positive effects on the state standardized test. Students offered admission into charters score 0.4 standard deviations higher in math, 0.2 higher in reading, and 0.3 higher in writing. These gains were largely driven by students in *No excuses* charter schools operated by large charter management organizations (CMO's). This study has high internal validity because of its unique matching design (includes observable demographics, a pre-test, and preferences of all DPS students). It has low external validity for a matching study because it is only in one locale.

Thus far, all the studies mentioned have been in a single locale, which limits external validity. If we expand that scope and maintain the lottery structure, we can improve external validity while maintaining internal validity. This is very difficult to do in practice, so there are fewer studies that have been able to achieve that result. The few studies we discuss in the following section do so and report much smaller or null effects.

Multi-Locale Lottery Studies

Gleason, Clark, Tuttle, and Dwyer (2010) conducted a multi-state lottery study that included 36 charter middle schools across 15 states. In total, they had 2330++applicants in their sample. All schools were recruited and voluntarily enrolled in the study. On average, they found no significant results in test scores after 2 years of charter school attendance. When broken out by subgroup, they did find that lottery winners with higher incomes (defined by free or reduced-price lunch eligibility) performed significantly worse on standardized tests than high-income lottery losers. For low-income students, the reverse was true. Low-income lottery winners scored 0.17 standard deviations higher than lottery losers. There was no significant result in reading. They found the same when comparing schools with higher or lower proportions of low-income students.

Fortson, Gleason, Kopa, and Verbitsky-Savitz (2015) built on Gleason et al. (2010) by comparing lottery results to nonexperimental designs including ordinary least squares (OLS) regression, matching, and student fixed effects regression. In other words, they conducted a within-study comparison by benchmarking nonexperimental estimates against experimental estimates (Cook, Shadish, & Wong, 2008). They collected data from some of the sites' studies in Gleason et al. and restricted the nonexperimental sample to students who attended the same TPS's at baseline. In other words, treatment and control students attended the same school before treatment but treated students attended charter schools. Control students remained in the same TPS's. Internal validity was strong in both studies. Their experimental (RCT) estimates matched Gleason's with no statistically significant differences. With regard to nonexperimental designs, Fortson et al. (2015) found OLS was biased upward relative to experimental estimates, suggesting their models had unobserved confounding variables and weaker internal validity. With their exact matching, propensity score matching, and student fixed effects designs, they found no statistically significant results, suggesting all but OLS were unbiased estimates.

Single Locale Fixed Effects Studies

Many of the first and most-cited studies of charter schools fall under the fixed effects umbrella. A student fixed effects analysis requires longitudinal data with repeated outcome observations on the same students over time, at least some of which must switch into or out of a charter school. This technique estimates the within-student deviation in test scores (relative to the student's average test score) over time from switching into or out of a charter school (Allison, 2009). This approach has some advantages over lottery studies. Fixed effects designs tend to have higher external validity than single locale lottery studies, as they do not rely on a sample of schools that are oversubscribed and may have to agree to be evaluated. With fixed effects, researchers examine all students who start in a TPS and subsequently enter a charter

school, because they can identify a treatment effect given that the treatment (switching from a TPS to a charter) is time varying. Fixed effects also allow researchers to control for all time-invariant factors that may affect individual students' success or failure at a given school. However, at least in theory if not in practice, fixed effects designs tend to have lower internal validity due the absence of a randomized design.

Witte, Weimer, Shober, and Schlomer (2007) examined the effects of 22 charter schools in Milwaukee, Wisconsin using student fixed effects. The authors only used data from the Milwaukee school district because the rest of the state's charter schools are geared toward at-risk students and have less of an emphasis on academic achievement. Results indicate that students who switch into charters scored, on average, 0.09 standard deviations higher on the Terra Nova test than students who remained in TPS's. Gains were especially prevalent in math, where students in charters scored 0.12 standard deviations higher on average than students in TPS's. The authors suggested that this could be due to a high concentration of charters that focus on science and mathematics curricula more than a typical TPS.

Imberman (2011) used 9 years of data from an anonymous urban district in the American southwest to examine the effect of charter attendance on cognitive and noncognitive skill formation. His outcomes of interest were math, reading, and language scores on the Stanford Achievement Test. Using student fixed effects, he found that students who switch into a charter school score 0.07 standard deviations higher in math on average than students who stay in TPS's. He found no significant results on reading and language test scores.

Multi-Locale Fixed Effects Studies

Bifulco and Ladd (2006) used student fixed effects to examine the effect of charter attendance on end of grade test scores for five cohorts of 4th–8th graders in the entire state of North Carolina from the 1995–1996 to 2001–2002 academic years. In both reading and math, they find significant negative effects of charter school attendance. Students who switch into a charter school score, on average, 0.10 standard deviations lower in reading and 0.16 standard deviations lower in math. Ni and Rorrer (2012) used student fixed effects to measure outcomes for charter school switchers in Utah. They used longitudinal data of every public school student in the state from 2004–2009, finding that elementary students who switch into charters score 0.10 standard deviations lower in math and language arts than students who stay in TPS's. Effect sizes are smaller as students age, with no significant effects in grades 7–11.

Zimmer, Gill, Booker, Lavertu, and Witte (2012) used fixed effects to estimate the effect of switching into a charter school across seven locations: Chicago, Denver, Milwaukee, Philadelphia, San Diego, Ohio, and Texas. Years of data vary by location but span the academic years 1994–1995 to 2006–2007. Because standardized tests varied by location, the authors standardized each student test score. Overall, effects are null or negative for math and reading. The only exceptions are a 0.17

standard deviation increase in math for students who switch to charters in Denver and a 0.05 standard deviation increase in math for charter entrants in Milwaukee. The Milwaukee results replicate Witte et al. (2007)'s findings.

Multi-Locale Propensity Score Matching Studies

The final major approach researchers take in evaluating charter schools is propensity score matching (PSM). Pioneered by Rosenbaum and Rubin (1983), PSM is a way to estimate program impacts on a matched sample of treatment and comparison units who are matched on a summary statistic: the probability each unit took up treatment, based on observable background variables. For PSM to produce unbiased impacts, the matching model must include a complete set of confounds to construct a matching model capable of matching truly exchangeable treatment and control units. The stronger the set of confounds, the less likely unobservable characteristics will bias the model. This method can have higher external validity but will typically have lower internal validity than lottery/RCT designs. Unlike fixed effect studies, PSM designs do not require baseline test scores. However, evidence presented above (Fortson et al., 2015) suggests that baseline test scores are essential for bias reduction and strong internal validity. As both types of designs require the same covariates (crucially at least two consecutive test scores), we rate them the same on internal validity.

Researchers can tailor this method further by choosing to match with or without replacement (i.e., whether any control unit can be matched onto more than one treated unit) or using different numbers of control cases to compare to each treated unit. *Nearest neighbor* matching refers to a single control unit with the propensity score closest to the matched treatment unit. Researchers can also use a decision rule called a caliper, which establishes a clear boundary (10% of a standard deviation is common) within which one can compare control and treatment units. There are a variety of propensity score methods, such as coarsened exact matching (Iacus, King, & Porro, 2012), full matching (Hansen, 2004), and inverse probability of treatment weighting (Cole & Hernán, 2008), but in the charter school literature, variants of nearest neighbor matching are used most often.

Baude, Casey, Hanushek, Phelan, and Rivkin (2020) used matching in a state-wide study of the state of Texas. They tracked the growth of charter schools from 2001–2011 and compared student outcomes across these same years to gauge how the sector performed. They constructed the comparison group based on students of the same grade, school, and demographic group as those that exited the school and entered a charter. Then they estimate a value-added model conditional on factors not used in the matching process: prior behavioral infractions, family factors, and school fixed effects. They found that students in *No Excuses* schools performed 0.12 standard deviations higher in math on standardized tests, although there was no difference in reading.

Stanford University's Center for Research on Education Outcomes (CREDO, 2009) nationwide evaluation of charter schools is an example of matching with a

baseline test score, which strengthens internal validity. Because of its large sample (over 70% of charter students nationwide across 16 states), this study also has very strong external validity. CREDO's model matches on grade, gender, race, poverty status, English language-learner status, special education status, and prior score on state achievement tests. Their approach is unique in that they only use potential matches from schools that have schools transfer into charters. All student records are then pooled within schools and a virtual control student is built for each student that attended a charter. This method resembles a hybrid between PSM and synthetic control, a method for constructing control units (see Abadie, Diamond, & Hainmueller, 2010 for more detail) CREDO's study found different effects by state. Five states had positive effects, seven had negative effects, and four were null. When pooled together, charter school students showed 0.01 standard deviations lower scores in reading and 0.03 lower standard deviations in math.

Davis and Raymond (2012) built on CREDO's prior work by using similar multistate data and comparing CREDO's matching estimates to the more commonly used fixed effects impacts. They found very similar results, which validated CREDO's results against designs many scholars found more legitimate. Their fixed effects estimates were of the same sign and similar in magnitude to results using matching. Pooled results showed students in charters scored 0.06 standard deviations lower in math and 0.02 lower in reading than matched control cases. They found negative effects in most states using both methods. Davis and Raymond (2012) and Fortson et al. (2015) both showed that impact estimates from matching designs that include a pre-test are virtually identical to fixed effects designs that also require a pre-test.

In 2013, CREDO updated the estimates from the original sample of 16 states. Students in charters in these states scored 0.01 standard deviations higher than TPS students. Effects in math improved but were still negative: -0.01 standard deviations in 2013 vs. -0.03 in 2009. This report also included estimates from 11 additional states. In total, their sample included 95% of charter school students in the U.S. A pooled analysis of all 27 states estimated gains of 0.01 standard deviations in reading and no significant impact in math. Although the nationwide estimates for charter schools were not impressive, effects varied significantly across states. They found positive effects in reading for 16 states, negative in eight, and null in three. In math, they reported positive effects in 12 states, negative effects in 13, and null effects in two. This study has the highest external validity of any in our review, because it included nearly all students in charters across the United States.

Overall Assessment

In Table 4.1 and Fig. 4.1, we classify studies on two dimensions: internal validity on the horizontal axis and external validity on the vertical axis. We score studies on internal validity with only two scores: a 3 for fixed effects and matching studies, because we view them as equivalent as all the matching studies incorporate

Fig. 4.1 External and internal validity. Source: Design by authors

pre-tests, and a 5 for lottery-based analyses. We score studies on external validity with five scores: 1 for single school studies, 2 for single type studies (*No Excuses* charters) with more than one school, 3 for studies based in one or two urban areas, 4 for whole state studies, and 5 for multiple state studies. For example, to take two extremes, the CREDO (2013) study covers nearly all charter school students in the country, but uses a nonexperimental matching design, so it is scored as 5 in external validity and 3 on internal validity; the Curto and Fryer Jr. (2014) lottery study, on the other hand, is of only one very unique charter boarding school, so we score it a 1 on external validity and a 5 on internal validity. In our view, the highest quality studies are those that use randomized designs across multiple states, scoring 5s on both internal and external validity. We note, however, that even these studies had their limitations given that they focused on middle schools, which reveals about elementary and high school charter impacts.

Several patterns emerge from this review. Lottery studies tend to have the largest positive effects, which is perhaps not surprising if oversubscribed schools have higher average quality than non-oversubscribed schools. Though systematic evidence on this point is challenging to obtain, authors of a national evaluation of charter middle schools discussed above found that only about one in four charter middle schools are oversubscribed (Gleason et al., 2010, p. 6), which means that if

this statistic holds across elementary and high schools as well, the vast majority of charter school could not hold lotteries. Therefore, impacts from nonlottery schools are essential if the most important question to answer is not whether some charters can be effective but whether charter schools are more or less effective than TPS overall. Fixed effects studies tend to show negative results. These studies are older, so they could be reflecting that charter schools were ineffective in the past. More recently, researchers have employed various matching approaches in larger, national samples and found not negative, but mostly null effects in math and small positive effects in reading. In general, the larger and more nationally representative the sample, the smaller the effects. We also see that as the sector ages, it tends to improve, as shown in national studies (CREDO, 2009, 2013). Initially, school quality was quite variable, but has grown more consistent and improved over time.

Conclusion

In this chapter, we have discussed charter schools in the United States in terms of the institutional environment they occupy, the methods researchers use to study them, and the body of available evidence as they approach 30 years in existence. This chapter is relevant to the theme of this volume in two respects: (1) Charter schools are a challenge to the standard institutional arrangement of state-run and controlled schooling in the sense that they are an alternative and deregulated setting in which education is delivered to students; (2) empirical findings can and do vary across geographic settings, across studies with different research designs, and across studies with smaller and larger collections of schools. Therefore, understanding this variation in research findings about ostensibly the same intervention is essential to understanding its effectiveness.

We see a consistent tradeoff between internal and external validity in each method. This raises the question of whether internal or external validity is more important for drawing conclusions about effectiveness of charter schools. We do not find randomized controlled trials in one locale particularly convincing for one reason: Charter schools likely vary a great deal in their instructional approaches, so there is no reason to expect that their effects would be uniform if implemented elsewhere. If charter schools were like educational aspirin manufactured with the same ingredients to solve the same educational problem, we would argue that there is consistency of the intervention and be more convinced by RCT evidence from one place. This does not accurately describe the reality of charter schools in the United States, however. Due to the wide diversity of charter school types, external validity becomes an important consideration for drawing policy conclusions about relative effectiveness. Therefore, we argue that both internal validity and external validity should be weighed in these assessments.

Although they often have lower internal validity, authors of cross-state studies point out a key empirical puzzle that deserves future research. Charter-school effects vary a great deal by state. It should be noted that education in the U.S. has always

been decentralized. Charter schools are perhaps a clear expression of this desire for local control. Today, the federal government spends less than 10% of all education dollars in the country. States have always taken the lead in this policy area. Almost all states delegate much of the operational aspects of education to special local governments called school districts. These districts manage the day-to-day aspects of running a school and in many cases can raise their own revenue through property and sales taxes. States generally have a more policy-relevant role in setting standards and mechanisms by which schools are financed.

States have developed quite different approaches to authorizing and regulating charter schools. Some types of schools consistently raise test scores, such as *No Excuses* institutions. It is possible that the prevalence of these schools varies by state. State charter laws vary among many dimensions, including who can authorize a charter, what accountability measures are in place, and the types and thresholds of sanctions states can place on schools that fail to perform. What types of incentives do operators respond to when improving the school? Do nonprofit operators respond differently than for-profits? Answering these types of questions would allow us to further explore the mechanisms charters use to improve student outcomes.

In general, charter school studies report larger effects for disadvantaged groups, which could reflect the effectiveness of the charter schools or the quality of the nearby traditional public schools. Effect sizes are larger for nonwhite, poor, and urban students than their white, nonpoor students, and rural students. CREDO's 2009 and 2013 studies allow comparison of these groups across states. In their 2013 study, they found that all learning gains are driven by the lower half of the achievement distribution. All effects in the top half are null or negative. They also find gains concentrated in poor students, among all races. English Language Learners (ELL's) also show significant gains from attending a charter school. However, more advantaged groups in charter schools (white, nonpoor, non-ELL) show declines in learning relative to their peers in TPS's (CREDO, 2009, 2013). Therefore, future researchers could unpack the reasons for these larger effects for these important subgroups and why charter school do not seem to have positive effects for high-achieving and more affluent students.

Finally, without evidence from so-called within-study comparisons of RCT and non-RCT impacts from the same treatment group (Cook et al., 2008), one would not know that these types of designs produce very similar estimates of effectiveness for test-score outcomes with controlled baseline test scores. This is critical because it strengthens the argument that studies with strong external validity such as CREDO's may have functional equivalent internal validity as well. Another approach is to attempt to generalize the impacts from charter schools that have produced RCT impacts to charter schools that have not. Authors of more recent show that conducting this type of analysis requires one to assess sample selection and heterogeneity of the treatment effect across the same factors that predict sample selection (Stuart, Cole, Bradshaw, & Leaf, 2011). If the same factors that predict sample selection (e.g., family poverty) also predict causal heterogeneity in the charter school effect, one can use these factors to adjust the charter school effect for the target population.

In future, researchers could attempt to use the techniques applied to within-study comparisons and generalization to better understand charter school effects.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105, 493–505. <https://doi.org/10.1198/jasa.2009.ap08746>
- Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics*, 126, 699–748. <https://doi.org/10.1093/qje/qjr017>
- Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., & Pathak, P. A. (2017). Research design meets market design: Using centralized assignment for impact evaluation. *Econometrica*, 85, 1373–1432.
- Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks: Sage.
- Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, 34, 275–318. <https://doi.org/10.1086/683665>
- Baude, P. L., Casey, M., Hanushek, E. A., Phelan, G. R., & Rivkin, S. G. (2020). The evolution of charter school quality. *Economica*, 87(345), 158–189. <https://doi.org/10.1111/ecca.12299>
- Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy*, 1, 50–90. <https://doi.org/10.1162/edfp.2006.1.1.50>
- Budde, R. (1996). The evolution of the charter concept. *Phi Delta Kappan*, 78(1), 72–73.
- Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Washington, D.C.: The Brookings Institution.
- Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168, 656–664. <https://doi.org/10.1093/aje/kwn164>
- Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27, 724–750. <https://doi.org/10.1002/pam.20375>
- CREDO. (2009). *Multiple choice: Charter school performance in 16 states*. Stanford, CA: Stanford University, Center for Research on Education Outcomes (CREDO). Retrieved from https://credo.stanford.edu/sites/g/files/sbiybj6481/f/multiple_choice_credo.pdf
- CREDO. (2013). *National charter school study 2013*. Stanford, CA: Stanford University, Center for Research on Education Outcomes (CREDO). Retrieved from https://credo.stanford.edu/sites/g/files/sbiybj6481/f/ncss_2013_final_draft.pdf
- Curto, V. E., & Fryer, R. G., Jr. (2014). The potential of urban boarding schools for the poor: Evidence from SEED. *Journal of Labor Economics*, 32, 65–93. <https://doi.org/10.1086/671798>
- Davis, D. H., & Raymond, M. E. (2012). Choices for studying choice: Assessing charter school effectiveness using two quasi-experimental methods. *Economics of Education Review*, 31, 225–236. <https://doi.org/10.1016/j.econedurev.2011.08.012>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.
- Epple, D., Romano, R., & Zimmer, R. (2016). Charter schools: A survey of research on their characteristics and effectiveness. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 139–208). Vol. 5. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-444-63459-7.00003-8>
- Fiske, E. B., & Ladd, H. F. (2000). *When schools compete: A cautionary tale*. Washington, D.C.: The Brookings Institution.

- Friedman, M. (1962). *Capitalism and freedom*. Chicago, IL: The University of Chicago Press.
- Fortson, K., Gleason, P., Kopa, E., & Verbitsky-Savitz, N. (2015). Horseshoes, hand grenades, and treatment effects? Reassessing whether nonexperimental estimators are biased. *Economics of Education Review*, 44, 100–113. <https://doi.org/10.1016/j.econedurev.2014.11.001>
- Gleason, P., Clark, M., Tuttle, C. C., & Dwyer, E. (2010). The evaluation of charter school impacts: Final report (NCEE 2010-4029). U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Retrieved from <https://files.eric.ed.gov/full-text/ED510573.pdf>
- Government Digital Service. (2016). Types of school. Retrieved from <https://www.gov.uk/types-of-school/academies>
- Government of the Netherlands. (2017). Public-authority and private schools. Retrieved from <https://www.government.nl/topics/freedom-of-education/public-authority-and-private-schools>
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99, 609–618. <https://doi.org/10.1198/016214504000000647>
- Hanushek, E. A. (2006). *Choice, charters, and public school competition*. Cleveland: Federal Reserve Bank of Cleveland.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960. <https://doi.org/10.1080/01621459.1986.10478354>
- Hoxby, C. M., & Murarka, S. (2009). Charter schools in New York City: Who enrolls and how they affect their students' achievement (NBER Working Paper No. 14852). Cambridge: National Bureau of Economic Research. <https://doi.org/10.3386/w14852>
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20, 1–24. <https://doi.org/10.1093/pan/mpr013>
- Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 481–502. <https://doi.org/10.1111/j.1467-985X.2007.00527.x>
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. New York, NY: Cambridge University Press.
- Imberman, S. A. (2011). Achievement and behavior in charter schools: Drawing a more complete picture. *Review of Economics and Statistics*, 93, 416–435. https://doi.org/10.1162/REST_a_00077
- Lubienski, C. (2003). Innovation in education markets: Theory and evidence on the impact of competition and choice in charter schools. *American Educational Research Journal*, 40, 395–443. <https://doi.org/10.3102/00028312040002395>
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83, 340–363. <https://doi.org/10.1086/226550>
- Meyer, J. W., & Scott, W. R. (1992). *Organizational environments: Ritual and rationality*. Newbury Park: Sage.
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). New York, NY: Cambridge University Press.
- National Alliance for Public Charter Schools. (n.d.). About charter schools. Retrieved from <https://www.publiccharters.org/about-charter-schools>
- National Center for Education Statistics. (NCES). (2018). Revenues for public elementary and secondary schools, by source of funds. Retrieved from https://nces.ed.gov/programs/digest/d18/tables/dt18_235.10.asp?current=yes
- Ni, Y., & Rorrer, A. K. (2012). Twice considered: Charter schools and student achievement in Utah. *Economics of Education Review*, 31, 835–849. <https://doi.org/10.1016/j.econedurev.2012.06.003>
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/S0266466603004109>
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55. <https://doi.org/10.1093/biomet/70.1.41>

- Scott, W. R. (1998). *Organizations: Rational, natural, and open systems*. Englewood Cliffs: Prentice Hall.
- Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 369–386. <https://doi.org/10.1111/j.1467-985X.2010.00673.x>
- Snyder, T. D., De Brey, C., & Dillow, S. A. (2019). Digest of Education Statistics 2018 (NCES 2020-009). Retrieved from <https://nces.ed.gov/pubs2020/2020009.pdf>
- U.S. Department of Education. (2009). The impacts of regular upward bound on postsecondary outcomes seven to nine years after scheduled high school graduation. Retrieved from <https://files.eric.ed.gov/fulltext/ED505850.pdf>
- U.S. Department of Health and Human Services. (2010). Head start impact study: Final report. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/hs_impact_study_final.pdf
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19. <https://doi.org/10.2307/2391875>
- Witte, J., Weimer, D., Shober, A., & Schlomer, P. (2007). The performance of charter schools in Wisconsin. *Journal of Policy Analysis and Management*, 26, 557–573. <https://doi.org/10.1002/pam.20265>
- Zimmer, R., & Buddin, R. (2006). Charter school performance in two large urban districts. *Journal of Urban Economics*, 60(2), 307–326.
- Zimmer, R., Gill, B., Booker, K., Lavertu, S., & Witte, J. (2012). Examining charter student achievement effects across seven states. *Economics of Education Review*, 31, 213–224. <https://doi.org/10.1016/j.econedurev.2011.05.005>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.