

A Re-Analysis of the Effects of KIPP and the Harlem Promise Academies

Stuart S. Yeh

University of Minnesota

Background/Context: Existing evaluations of the Knowledge is Power Program (KIPP) and the Harlem Children's Zone (HCZ) charter school programs suggest that these programs may potentially be very effective in closing the academic achievement gap.

Purpose/Objective: The purpose of the current study is to investigate the possibility of an internal contradiction in the assumptions underlying the KIPP and HCZ models that would prevent the models from being scaled up with the same positive results.

Research Design: The method involves: (1) review of the results of the key impact studies; (2) review of empirical evidence of high teacher attrition; (3) review of statements by the founders of the HCZ and KIPP models articulating the core assumption; (4) a narrative explanation of the internal contradiction in these models; and (5) a formal analysis demonstrating the consequences of this contradiction.

Findings/Results: The analysis suggests that gains from the KIPP and HCZ models would fall to zero once these programs are implemented in every school across the nation. The analysis suggests why existing impact results are likely due to artifacts stemming from high teacher attrition and hoarding of a disproportionate share of the nation's limited pool of highly dedicated teachers, rather than gains that could be sustained when the programs are scaled up and implemented nationwide.

Conclusions/Recommendations: The article recommends that national implementation of the KIPP and HCZ models should be delayed until research has been performed that would substantiate the assumption that there is a sufficient supply of highly dedicated teachers to fill the large number of vacant teaching positions that arise as a consequence of these models.

The No Child Left Behind Act has created enormous pressure to find effective ways to raise student achievement and to close the achievement

gap between economically disadvantaged students and their more privileged peers. Many districts have struggled to raise achievement, and national trends are discouraging. Recently, however, dramatic results from evaluations regarding the Knowledge is Power Program (KIPP) and the Harlem Children's Zone (HCZ) charter school programs have raised hope that these types of reform-oriented charter schools may be especially effective in raising student achievement. A rigorous evaluation concluded that, "the effects of attending an HCZ middle school are enough to close the black-white achievement gap in mathematics. The effects in elementary school are large enough to close the racial achievement gap in both mathematics and English language arts" (Dobbie & Fryer, 2011, p. 158). Similarly, the best available evaluation of KIPP concluded that, "Estimated impacts are frequently large enough to substantially reduce race- and income-based achievement gaps within three years of entering KIPP" (Tuttle, Teh, Nichols-Barrer, Gill, & Gleason, 2010, p. xi). Significantly, the federal government's Race to the Top initiative emphasizes expanded implementation of these and similar types of charter schools to address the achievement gap.

The Knowledge is Power Program (KIPP), begun in 1994, comprises a national network of charter schools that aim to equip students, who are drawn primarily from low-income and minority families, with the knowledge, skills, and character traits needed to succeed in top quality high schools and colleges. KIPP emphasizes high expectations and standards for students, and seeks to achieve its goals by recruiting highly dedicated teachers who are willing to work long hours.

The Harlem Children's Zone (HCZ) is a non-profit organization that funds and operates a neighborhood-based system of education and social services in a 97-block area in central Harlem, New York. HCZ combines reform-minded charter schools with a web of community services created to support children from birth through college. HCZ charter schools are similar to KIPP schools in emphasizing high expectations and standards for students and reliance on the recruitment of highly dedicated teachers to improve student achievement.

Thus, a core assumption of both charter school approaches is that raising student achievement requires highly dedicated teachers who are willing to work long hours. However, both approaches exhibit high teacher attrition. To scale up these approaches, a secondary assumption is required: there must be a sufficient supply of highly dedicated teachers to fill the large number of vacant teaching positions that arise as a consequence of these models. This assumption appears to be incorrect since principals of "no excuses" charter schools such as KIPP and HCZ report

that they have to “scour the country” for suitable teachers (Curto, Fryer, & Howard, 2010, p. 24).

The analysis presented in this article indicates that the core assumption, in combination with empirical evidence regarding high teacher attrition and evidence that principals have to scour the country for suitable teachers, implies an internal contradiction: if this assumption is correct, it is *logically impossible* to replicate the impact results on a national level, given the high level of teacher attrition. While other researchers have suggested that the HCZ and KIPP models may be *difficult* to scale up (Curto, et al., 2010; Henig, 2008; Woodworth, David, Guha, Wang, & Lopez-Torkos, 2008), the contribution of the analysis presented here is that it demonstrates the existence of an internal contradiction that would effectively *prevent* the models from being scaled up.

The distinction between a *difficult* and a *logically impossible* task is important. If it is logically impossible to scale up KIPP and HCZ, the task should not be attempted, because it would divert scarce resources from improvement strategies that are feasible and worthy. If it is logically feasible to scale up KIPP and HCZ, then the effort may be worthwhile. However, the analysis reported here suggests reasons why scaling appears to be impractical. Unless and until KIPP/HCZ supporters can demonstrate that the analysis is incorrect, the appropriate policy conclusion is that scale up should be delayed until research has been performed that would substantiate the secondary assumption that there is a sufficient supply of highly dedicated teachers to fill the large number of vacant teaching positions that arise as a consequence of these models.

While the validity of the current analysis depends on the models’ core assumption, the assumption is derived from statements by the founders of the HCZ and KIPP models, as described below in Section I. If it can be shown that any intervention based on this assumption, in combination with empirical evidence of high teacher attrition, cannot be scaled up in a way that maintains the putative impacts of the HCZ/KIPP models, then serious questions arise about the external validity of the models. Either the assumption is correct, and the models cannot be scaled up nationally, or the assumption is incorrect and the models are based on an incorrect assumption.

The analysis proceeds in two parts. Section I of this article reviews the results of the key impact studies and empirical evidence of high teacher attrition, draws upon statements by the founders of the HCZ and KIPP models articulating the core assumption, and presents a narrative explanation of the internal contradiction in these models. This analysis suggests why the impact results are likely due to artifacts stemming from

high teacher attrition and hoarding of a disproportionate share of the nation's limited pool of highly dedicated teachers, rather than gains that could be sustained when the programs are scaled up and implemented nationwide. Section II of this article formalizes the analysis. Under the assumption that any gains depend on the proportion of highly dedicated teachers, the gains would fall to zero once these programs are implemented in every school across the nation. The contribution of the formal analysis presented in Section II is that it demonstrates that neither HCZ nor KIPP can be scaled up nationally while maintaining their putative impacts on student achievement, given the founders' core assumption, empirical evidence regarding high teacher attrition, and evidence that the characteristics of HCZ and KIPP are insufficient to generate the waiting list of highly dedicated teachers that would be necessary for scale up—the analysis suggests that no other assumption or evidence is required to reach this conclusion. To refute the analysis, it would be necessary to show that additional assumptions or empirical evidence are required. Section III discusses this result.

The purpose of the analysis presented in this article is to investigate the possibility of an internal contradiction in the assumptions underlying the KIPP and HCZ models, given the available evidence of high teacher attrition. This possibility suggests the need for extreme caution in extrapolating the results of the impact studies of KIPP and HCZ. The purpose of the analysis is not to resolve empirical questions about the level of teacher attrition, or whether teachers who stay are truly more dedicated or more effective than teachers who leave, or the reasons why teachers leave the KIPP and HCZ schools.

SECTION I. A THREAT TO VALIDITY

Researchers who evaluate the effectiveness of various interventions are well aware that sample attrition can lead to false conclusions that an intervention is effective. Attrition occurs when, for example, a greater number of less motivated students drop out of a treatment group, compared to a control group. What is not widely recognized is that a similar threat to the validity of a research study may occur when an intervention causes less motivated teachers to drop out, thereby raising the measured performance of the remaining group of teachers and the putative effectiveness of any intervention based on that performance (see Campbell & Stanley, 1963). If student achievement depends on having motivated teachers, and if an intervention causes less motivated teachers to drop out, then it necessarily follows that the measured performance of the

remaining group of teachers will be higher, resulting in the false conclusion that the intervention caused student achievement to improve. The conclusion is false because the intervention did not cause student achievement to improve; instead, the improvement is an artifact of teacher attrition.

At first glance, it may seem desirable if a treatment causes less motivated teachers to drop out. If it is always possible to substitute highly motivated teachers for less motivated counterparts, then this phenomenon could be considered a desirable feature of the intervention, and the measured impact of the intervention is a valid measure of program impact. However, a problem arises if the intervention relies entirely on teacher attrition to improve student performance and there are insufficient teacher candidates who meet the selection criterion and are available to fill the teaching slots that become available as a consequence of the treatment's high attrition rate. This may not be a problem when the intervention has only been implemented in a few scattered schools across the nation, because these schools are free to pull highly motivated teachers from neighboring schools and districts, as well as distant schools from across the nation. These schools may also pull a disproportionate number of the limited supply of highly dedicated college graduates seeking to become teachers. Significantly, principals of "no excuses" charter schools such as KIPP and HCZ report that they have to "scour the country" for suitable teachers, which suggests that the model of teacher recruiting employed by these charter schools may not be scalable (Curto et al., 2010, p. 24).

This siphoning effect may be hidden in the large flows of teachers from school to school and district to district across the nation every year. Typically, researchers do not pay attention to this effect and do not observe or measure it. They focus on controlling bias in the selection of students, rather than bias in the retention of highly dedicated teachers. When researchers focus on attrition, they focus on students, not the attrition of less dedicated teachers.¹

Recently, several research studies have evaluated the impact on student achievement of the Knowledge is Power Program (KIPP) and the Harlem Children's Zone (HCZ). These studies sought to control for sample bias, found positive impacts on student achievement, and concluded that the interventions are promising (Angrist, Dynarski, Kane, Pathak, & Walters, 2010; Dobbie & Fryer, 2011; Tuttle, et al., 2010; Woodworth, et al., 2008). However, these studies did not examine the possibility that the results may be artifacts of high teacher attrition.

RESEARCH ON HCZ

Dobbie and Fryer (2011) provide the first and perhaps the most highly regarded empirical test of the causal impact of HCZ on student achievement. Using lottery-based randomization and two-stage least squares (2SLS), they found that Promise Academy middle-school lottery winners gained 0.229 SD per year in math and 0.047 SD per year in English language arts, compared to students who did not win admission. Promise Academy elementary school lottery winners gained 0.191 SD per year in math and 0.095 SD per year in English language arts. Significantly, while HCZ includes a variety of social services in addition to the charter school component, a Brookings Institution analysis found no evidence that HCZ influences student achievement through those social services, suggesting that the impact of HCZ on student achievement results primarily from the influence of the HCZ charter schools (Whitehurst & Croft, 2010).

The Promise Academy schools emphasize the recruitment and retention of high quality teachers in order to raise student achievement, and use a test-score value-added measure to incentivize and evaluate current teachers (Dobbie & Fryer, 2011). The view that teacher quality and teacher motivation is paramount is clear from an interview with Geoffrey Canada, the founder of the HCZ Promise Academies, on the CBS news program 60 Minutes. Canada stated that he would “fire the teachers” if they do not raise student achievement to a level where students are college-ready (Cooper, 2009). In the documentary, “Waiting for Superman,” he stated “I want to be able to get rid of teachers that we know aren’t able to teach kids” (Collins, 2010, p. A35). Elsewhere, he has stated that “Finding great teachers is the ‘secret sauce’ of great schools and, in particular, great charter schools” (Weber, 2010, p. 194). This indicates that the theory of action underlying the Academies is that teachers are the main factor influencing student achievement and low student achievement is due to poor teachers. If instead Canada believed that the most important factor was the length of the school day and the school year, or the level of resources, or any factor other than the effort and quality of the teachers, it would make more sense to simply address those other factors. It would not make sense to threaten to fire teachers if the most important factors are beyond their control.

At present, it is not possible to determine whether the impact of the Academies on student achievement is due primarily to the recruitment of highly dedicated teachers. This type of analysis would require a large sample of HCZ schools and a statistical analysis of all of the factors influencing student achievement. The limited number of HCZ schools makes this analysis impossible. However, it is possible to test the *implications* of

Canada's assumption that finding great teachers is the "secret sauce" of great charter schools. The analysis in Section II, below, suggests that this assumption contains an internal contradiction—if Canada is correct, then the impacts identified by Dobbie and Fryer (2011) cannot be replicated when the intervention is scaled up. While Canada's analysis may be faulty, and it is possible that the impacts are due to factors other than highly dedicated teachers, serious questions arise about the viability of the intervention if it can be demonstrated that the founder's most important assumption implies that impacts fall toward zero when the intervention is scaled up nationally. Either the assumption is correct, and the HCZ model cannot be scaled up nationally, or the assumption is incorrect and the model is based on an incorrect assumption.

Canada's assumption, however, is fully consistent with the available evidence. Highly dedicated teachers are needed because the HCZ Promise Academy approach requires that teachers work long hours. HCZ Promise Academy students who are behind grade level attend school for approximately twice as many hours as a traditional public school student in New York City (Dobbie & Fryer, 2011). As a consequence, HCZ teachers work several hundred more hours than regular New York City teachers. These extra hours are accumulated through a longer work day, a lengthened school year and, for many teachers, a summer school session that is added to the school year and runs through the first week of August (Tough, 2008). Perhaps as a consequence, both schools have had high teacher turnover as they search for highly dedicated teachers willing to work long hours: 48% of Promise Academy teachers did not return for the 2005-2006 school year, 32% left before 2006-2007, and 14% left before 2007-2008, suggesting very high 3-year attrition rates that almost certainly exceed 48% (Dobbie & Fryer, 2011). The decline in the rate of teacher turnover in each subsequent year is consistent with the hypothesis that turnover primarily serves to rid the school of less -dedicated teachers. As the proportion of highly dedicated teachers increases each year, the proportion of less -dedicated teachers declines, resulting in a decrease in turnover during each subsequent year. Thus, the cumulative 3-year teacher attrition rate is a more valid indicator of the degree of teacher replacement that is occurring than any single-year attrition figure.

Qualitative evidence suggests that highly dedicated teachers are needed for a second reason, in addition to long hours. The Promise Academies require HCZ teachers to undertake strenuous efforts to raise student achievement. According to Terri Grey, the former principal at the Promise Academy middle school, the HCZ approach requires not only long hours but an emphasis on test preparation, leading to teacher

burnout and attrition (Tough, 2008). Opposed to the heavy emphasis on test preparation, she was fired (Tough, 2008). As recounted by Paul Tough in his book about HCZ:

Test prep was under way by the third week in September. There were morning test-prep sessions, a test-prep block during the school day, test prep in the afterschool program, and test prep on Saturdays ... As the year went on, the time dedicated to test prep only grew, and the time dedicated to everything else was forced to shrink further. (Tough, 2008, p. 165)

A month later superintendent Doreen Land quit and more than a dozen teachers followed Grey and Land out the door (Tough, 2008, p. 172). However, despite all of the test prep, the HCZ board of directors feared that the eighth-grade cohort of students would perform poorly on standardized tests. In order to protect the reputation of the Harlem Children's Zone, the cohort was disbanded and all of the students in the cohort were reassigned to other schools (Tough, 2008, p. 251). This depressed teacher morale. Chris Finn, the dean of students, reported that the experience was exhausting and engendered a feeling of failure (Tough, 2008, p. 252). In sum, the evidence suggests that HCZ relies heavily on highly dedicated teachers who are willing to work long hours, endure conditions that cause less-dedicated teachers to quit, and undertake strenuous efforts to raise student achievement. Teachers who are unable to maintain the exhausting pace either quit or are fired.

RESEARCH ON KIPP

Studies of KIPP suggest a parallel narrative: promising impact results as a consequence of highly dedicated teachers working long hours under exhausting conditions, resulting in high teacher attrition. While early evaluations of KIPP were limited to a small number of schools or included only weak controls for selection bias, Tuttle et al. (2010) used a matched comparison group design with achievement data for a nationwide sample of 22 KIPP schools and found that after 3 years KIPP students gained an average of 0.14 SD per year in math and 0.08 SD per year in reading. The researchers did not examine teacher attrition. However, an SRI study investigated teacher attrition at five San Francisco Bay Area KIPP schools and found that among the 84 teachers who taught in the five schools in 2006-2007, nearly half (49%) left the classroom before the start of the 2007-2008 school year (Woodworth et al., 2008). In the spring

of 2007, the median tenure of teachers at the Bay Area KIPP schools was only 2 years (Woodworth et al., 2008).

Thus, evaluations of both HCZ and KIPP suggest positive impacts on student achievement, but at the cost of teacher attrition that ranges up to 49% each year as a consequence of the “heroic” efforts that are required of teachers (Curto, et al., 2010, p. 24). While the SRI report points out that high teacher turnover is not uncommon in urban schools serving poor and minority students, studies suggest that the annual turnover rate in these schools is substantially lower than attrition at the HCZ and KIPP schools—about twenty percent (Woodworth, et al., 2008). The SRI study is useful because it explores the reasons for KIPP teacher attrition in detail. While SRI studied the KIPP schools, the explanation may also apply to the HCZ schools as well because both approaches rely on highly dedicated teachers working long hours. The SRI study suggests that the long work hours and personal sacrifices demanded by this type of approach cause high teacher attrition. Excerpts from the interviews conducted by SRI with KIPP teachers and school leaders included the following:

“I can’t do this job very much longer. It is too much. I don’t see any solution. . . . No one has really presented any way to solve that problem.” (p. 35)

“The big question is the sustainability question. . . . We are really tired.” (p. 34)

“You’re taking on the place of the family, giving up your own family. I respect and admire that in others, and I don’t know that I can do that again.” (p. 35)

“That’s the biggest KIPP challenge: How do you keep teachers coming back here year after year? A lot of the workload I have I put on myself. . . . When do I stop worrying about them and take care of me? It’s hard to find that balance. That’s going to be the most challenging thing, retaining teachers and keeping them rested and healthy.” (p.34)

“Turnover is so high that teachers are constantly coming in and reinventing the wheel.” (p. 34)

According to the SRI report, veteran KIPP teachers in every school, including the founders, expressed similar sentiments, and regret that

they need to choose between teaching at KIPP and finding balance in their lives (Woodworth et al., 2008). Bay Area KIPP teachers spent a median 65 hours per week on all school-related activities (a range of 60 to 67 hours), whereas urban middle school teachers worked an average of 52 hours per week nationally (Woodworth et al., 2008).

At KIPP, teachers carry cell phones and are expected to be available 24 hours a day to respond to any concerns that students may have (Carter, 1999). Nine-and-a-half hour days, class on Saturday, and school during the summer “are all non-negotiable” (Carter, 1999, p. 17). According to one of the co-founders, the whole KIPP framework is built around maximizing teaching time and teacher accountability (Carter, 1999). KIPP teachers are personally held accountable for student progress and are “contractually obligated to see that their students succeed. They know they *have* to teach until the kids get it” (Carter, 1999, p. 19, italics in original). KIPP teachers sign commitments to do “whatever it takes” to get students to learn (Carter, 1999, p. 19). KIPP teachers regularly visit students in their homes to teach parents the importance of checking homework, reading with their children, and fostering aspirations to attend college (Carter, 1999). Both co-founders believe that it is impossible to scale up and replicate the KIPP model on a national scale, given the current pool of teachers (Carter, 1999). “What we do isn’t easy. First, we need to find a way to make this level of commitment the standard. Then we need to make it attractive, livable, and affordable for teachers” (Carter, 1999, p. 20).

SECTION II. A FORMAL ANALYSIS

A formal analysis may be conducted regarding the implications of Canada’s assumption that the impact of charter schools such as HCZ and KIPP depends on the recruitment of highly dedicated teachers. This analysis indicates that the gains of HCZ and KIPP would fall to zero once these programs are implemented in every school across the nation.

Underlying both KIPP and HCZ is the core assumption that the recruitment and retention of a highly dedicated teaching staff is central to student achievement.

Let $f[x]$ describe this theory of student achievement, where $f[x]$ is a linear, monotonically increasing function that depends on the proportion (a) of the teaching staff (S) that is composed of highly dedicated teachers: $f[x] = f[aS]$, $f'[x] > 0$

If $f[x] = f[aS]$ is assumed to be linear, then $f[aS] = k aS$, for some $k > 0$. If the units of achievement are chosen such that $k = 1$, then student achievement = $f[aS] = aS$.

Let b = the proportion of highly dedicated teachers before implementation of KIPP or HCZ

Let c = the increase in the proportion of highly dedicated teachers after implementation of KIPP or HCZ (pulled from non-KIPP/HCZ schools or college graduates who would otherwise be hired by non-KIPP/HCZ schools),

where:

$$a = b + c$$

and

$$f[(b + c)S] > f[bS]$$

Then the initial level of achievement is

$$bS$$

before implementation of KIPP or HCZ. Achievement increases to

$$(b + c)S$$

after implementation of KIPP or HCZ.

Thus, student learning increases by an amount equal to

$$(b + c)S - bS = cS$$

as a consequence of implementing KIPP or HCZ.^a

If, however, every school implements KIPP or HCZ, the value of (c) must (by definition) equal zero ($c = 0$). The proportion of the teaching staff that is highly dedicated must equal (b), in other words, $a = b$, and the increase in student achievement as a consequence of implementing KIPP or HCZ must fall to zero:

$$(b + c)S - bS = bS - bS = 0$$

The key to understanding this result is that once every school is a KIPP or HCZ school, it is no longer possible for any school to pull teachers from non-KIPP/HCZ schools: recruitment necessarily pulls from the supply of highly dedicated teachers to other KIPP/HCZ schools, reducing the performance of the other schools. In essence, recruitment becomes

a zero-sum game in which any single KIPP/HCZ school can only recruit additional highly dedicated teachers if some other KIPP/HCZ school loses highly dedicated teachers.²

Take, for example, three schools, each with a number of highly dedicated teachers equal to (bS) . School 1 introduces KIPP or HCZ and attracts cS highly dedicated teachers, pulled equally from schools 2 and 3. The boost to school 1 is equal to cS , consistent with the published evaluation results. School 2 loses $cS/2$, as does school 3, resulting in a total loss of achievement at Schools 2 and 3 equal to $cS/2 + cS/2 = cS$. The aggregate gain in achievement across the three schools is therefore zero.

SECTION III. DISCUSSION

The results of the formal analysis suggest that it is worth considering the full implications of high teacher attrition. While the KIPP co-founders concede that it may be difficult to scale up and replicate the KIPP model on a national scale, and other researchers have suggested the same conclusion, no previous analysis has suggested that it would be impossible to scale up the KIPP/HCZ models, given their core assumption and evidence of high teacher attrition.

To explain why the putative impacts of KIPP and HCZ necessarily fall to zero when the programs are implemented nationally, under Canada's assumption that finding great teachers is the "secret sauce" of great charter schools, it is useful to consider an analogy. If KIPP and HCZ rejected 49% of the least motivated students every year and only accepted the most motivated students, the measured impact of the KIPP and HCZ approaches would be artificially boosted, simply because of this creaming effect. A similar effect would occur if KIPP and HCZ screen out the least motivated 49% of teacher recruits by setting grueling work hours and conditions. This creaming effect would leave only the most motivated, dedicated teachers, artificially raising the measured performance of KIPP and HCZ in any impact evaluation. Unfortunately, it is impossible to detect this bias in any published evaluation of KIPP or HCZ because those evaluations simply state the gains achieved by the corps of teachers who remain with KIPP or HCZ after substantial creaming has already occurred; the results do not tell us what would happen if the KIPP/HCZ approaches were to be used with an average group of teachers. However, it is clear that the impact results would be biased in any case where the least motivated teachers were regularly screened out.

The resulting bias in existing KIPP/HCZ impact evaluation results has been hidden because KIPP and HCZ schools can pull highly motivated teachers from non-KIPP/HCZ schools (or from the supply of those teachers

to non-KIPP/HCZ schools) in order to fill slots that open up due to attrition. However, the bias would become clear if every school across the nation adopted KIPP or HCZ. Under that condition, KIPP/HCZ schools could only pull teachers from other KIPP/HCZ schools, or from the supply of new college graduates who would ordinarily go to those other schools. The effectiveness of every KIPP/HCZ school would be reduced as every school lost dedicated teachers recruited by other KIPP/HCZ schools. Without the ability to stock their teaching staffs with highly dedicated teachers, each KIPP/HCZ school would be forced to recruit from the less-dedicated corps of teachers that was rejected when KIPP and HCZ were only implemented in a few schools across the nation. The performance of each KIPP and HCZ school would necessarily decline to a normal level—the level that prevails when highly dedicated teachers are evenly distributed across all schools instead of being concentrated in a few KIPP/HCZ schools. Evaluations of the KIPP and HCZ schools would reflect this lower level of performance. The magnitude of the reduction in KIPP and HCZ school performance would provide information about the magnitude of bias in current estimates of KIPP/HCZ impact. The results of the analysis in Section II imply that the reduction would equal the difference in performance between current KIPP/HCZ schools and non-KIPP/HCZ schools, under Canada’s assumption that the main reason for the outstanding performance of current KIPP/HCZ schools is their recruitment of a staff of unusually dedicated teachers. In sum, the analysis in Section II implies that if the unusually dedicated staff goes away, the unusually high performance goes away.³

A question that arises is whether the national implementation of the HCZ and KIPP charter schools might attract a much larger pool of highly dedicated individuals to the teaching profession, thereby filling the empty teaching slots that are created by high teacher attrition. In fact, a central claim of market-oriented reformers is that the use of merit pay and the practice of retaining teachers on the basis of student learning gains might attract a more talented pool of individuals to the teaching profession. Hanushek (2009), for example, makes this argument. However, he concedes that, “we do not know how teacher quality responds to different levels of salaries” (Hanushek, 2009, p. 12). In other words, his claim that merit pay might elicit a stronger pool of teachers is not based on empirical data. Similarly, it is possible that KIPP and HCZ may, in the long run, induce a larger flow of highly dedicated individuals into the teaching profession—individuals who are attracted by the KIPP/HCZ emphasis on good teaching. In addition, KIPP and HCZ might inspire and transform teachers who are currently not highly dedicated, such that they become teachers who are highly dedicated. Either

effect could serve to address the problem of high teacher attrition. At present, however, there is no evidence that either of these effects is significant. In the absence of evidence, it would not be appropriate to assume that KIPP or HCZ would elicit a substantially greater flow of talented individuals into the teaching profession. On the contrary, reports from principals of “no excuses” charter schools such as KIPP and HCZ that they have to “scour the country” for suitable teachers suggests that, to date, neither KIPP nor HCZ has inspired an adequate flow of such individuals to take up the profession of teaching (Curto, Fryer, & Howard, 2010, p. 24). This is consistent with the view of the KIPP co-founders that it is impossible to scale up and replicate the KIPP model on a national scale, given the current pool of teachers (Carter, 1999). Whatever KIPP and HCZ have accomplished, they have not inspired a vast pool of talented individuals to switch to the teaching profession. Given that the principals of the KIPP and HCZ schools have to scour the country for suitable teachers, the only way that the schools can acquire the necessary teachers is by hoarding.

The validity of the current analysis is independent of explanations about why some teachers fail to thrive in KIPP/HCZ schools. It does not matter whether failure is a consequence of their lack of dedication or commitment or their ineffectiveness relative to their peers who remain in the KIPP/HCZ schools. For the purpose of the current analysis, it is sufficient to demonstrate three conditions: (1) The KIPP/HCZ models are based on the assumption that high performance is a consequence of highly dedicated teachers; (2) There is high attrition among KIPP/HCZ teachers; and (3) There is a shortage of highly dedicated teachers who can replace teachers who leave. All of the conclusions of the analysis follow from these three conditions. No other assumption is necessary. It is conceivable, for example, that teachers who leave are just as talented as teachers who stay, but simply lack the necessary endurance and fortitude. The end result is that there is a shortage of the type of teacher that is required by the KIPP/HCZ models. If there is a shortage, then—by definition—the models cannot be scaled up nationally. They can only succeed in scattered examples here and there because they can pull a disproportionate share of dedicated teachers. If an attempt is made to scale up the models nationally and the option of pulling teachers from other schools is eliminated, it becomes logically impossible to maintain the gains that were achieved in the small-scale research studies.

It is not helpful to compare the current gain scores of the students taught by teachers who stay, versus teachers who leave. Lower gain scores for teachers who leave would support the interpretation presented throughout the article—that creaming of highly effective teachers occurs

because less dedicated, lower-performing teachers quit. However, equal gain scores would not be inconsistent with the thesis that creaming occurs. Highly effective teachers may be attracted to the KIPP and HCZ schools in disproportionate numbers (rather than being culled from a larger set of teachers, some of whom are less dedicated and less effective). When the highly effective teachers quit, they scatter across the rest of the nation's schools. Researchers who compare the performance of KIPP/HCZ schools to non-KIPP/HCZ schools would continue to find that the KIPP/HCZ schools outperform the non-KIPP/HCZ schools—until these models are scaled up nationwide.

Finally, the validity of the analysis presented here is independent of evidence that there are plenty of entrants to the teaching profession, but poor organizational conditions, including high pressure from accountability systems, cause many teachers to leave the profession (Ingersoll, 2001). Whatever the reasons that teachers leave the profession, the existence of KIPP and HCZ schools has clearly not reversed this trend. If the organizational changes implemented by KIPP and HCZ were sufficient to reverse this trend, then we should observe long waiting lists of teachers seeking to transfer to KIPP and HCZ schools. This is clearly not the situation if charter schools such as KIPP and HCZ report that they have to “scour the country” for suitable teachers (Curto et al., 2010, p. 24). Thus, there is no reason to think that scaling up the KIPP and HCZ models would reverse the enormous teacher attrition that occurs across the nation.

The basic problem with the KIPP and HCZ models is that if they are implemented in every school, there are no longer any non-KIPP/HCZ schools where teachers may be pulled. Recruiting schools can only pull teachers from the limited national supply of highly dedicated teachers that is available to all KIPP/HCZ schools. If one school garners a disproportionate number of those teachers, other KIPP/HCZ schools must lose those teachers. Recruitment becomes a zero-sum game. Since it would no longer be possible for all KIPP/HCZ schools to maintain a teaching force of the most highly dedicated individuals from across the nation, performance would inevitably decline. KIPP/HCZ schools would lose the essential character that made them successful when implemented in a few scattered schools across the nation, simply because there is no waiting, unemployed army of individuals seeking to become teachers whose level of performance exceeds the 49th percentile. This is true whether teachers are pulled directly from other KIPP/HCZ schools, or indirectly, out of the supply of brand new teacher candidates.

In essence, existing evaluations of the KIPP/HCZ approach lack external validity. To address the issue of external validity, it would be necessary

to implement KIPP/HCZ in every school within a defined geographical area and prohibit the schools from recruiting teachers from outside of that area. To the extent that the schools recruit nationally, they would be drawing down the small corps of highly dedicated teachers that would otherwise be available to schools outside of the area.

The KIPP/HCZ approach can only be scaled up if working conditions are changed so the life of a KIPP/HCZ teacher is less exhausting and more attractive. However, relaxing the bar would permit less dedicated teachers to remain. The bar would have to be relaxed substantially if the KIPP/HCZ approach were to be extended to all schools. It is likely that the bar would have to be lowered to the current level for non-KIPP/HCZ schools. If every school is a KIPP or HCZ school and teachers have no option but to remain at a KIPP or HCZ school or leave the profession, a bar that is any higher would drive teachers out of the profession and exacerbate the current shortage of teachers, increasing class sizes or leaving large numbers of students without teachers.

The policy implication that has been drawn from published impact evaluations of KIPP and HCZ is that they represent promising approaches for raising student achievement, if only the issue of teacher burnout can be addressed (see, for example, Curto et al., 2010). However, this issue cannot be treated as a minor issue. The high burnout rate is not only a fundamental flaw in the approach, but suggests that positive interpretations of KIPP's and HCZ's effects on student achievement are based on flawed analyses. Reported gains in student achievement are most likely artifacts due to the attrition of up to 49% of the teaching force every year, leaving the most dedicated teachers and reflecting brief spurts of 65-hour workweeks that cannot be sustained over time.

The KIPP and HCZ approaches work by recruiting highly dedicated teachers, creating a work environment that only the top half of all teachers can survive, and constantly recruiting additional teachers across the nation in a wide-ranging attempt to fill the empty teaching slots. The consequence, however, is that this process inevitably pulls the best teachers from across the country, leaving an insufficient number of those teachers to implement the same approach in every school nationwide.

The implication is that the KIPP/HCZ model can only be scaled up if the number of teacher applicants is twice the size of the current teaching force, permitting KIPP and HCZ schools to reject the bottom half of those applicants. However, there is currently a teacher shortage, and the shortage is projected to become increasingly severe over the foreseeable future (Gordon, Kane, & Staiger, 2006). The vast army of unemployed, highly qualified, and highly dedicated teachers that is required to implement KIPP and HCZ on a nationwide basis simply does not exist.

This issue has largely been overlooked by previous researchers. At first glance, the promising results of impact analyses have held out hope that the KIPP/HCZ recipe has solved the problem of low student achievement, if only this approach is scaled up and implemented nationwide. Researchers have claimed that this approach can close the achievement gap between poor minority students and their more advantaged peers (Dobbie & Fryer, 2011). However, those research studies overlooked KIPP's and HCZ's high teacher attrition and overlooked the pulling of highly dedicated teachers from non-KIPP and non-HCZ schools in order to fill the large number of empty teaching slots created each year when teachers are unable to endure the exhausting KIPP and HCZ workdays. Thus, the results of the research studies lack external validity—the type of validity that is required if KIPP and HCZ are to be successfully scaled up and implemented nationwide. The problem is not a research issue that can be addressed in future studies—it is inherent in the KIPP/HCZ approach. The only way the issue can be addressed is by making KIPP/HCZ less grueling for teachers and relaxing the bar for hiring and retaining teachers. That bar must be relaxed to the point where teachers do not leave the KIPP/HCZ schools in disproportionate numbers. However, if the bar is relaxed, those schools cease to maintain the essential character of KIPP and HCZ—the character established by recruiting the most-dedicated teachers in the nation.

Policymakers may wonder how it is possible that KIPP and HCZ can draw the most dedicated teachers, produce impressive results as measured by carefully controlled research studies, but not be scalable. To draw an analogy, an automobile that is not working can be pushed by a team of very strong, highly dedicated athletes. In a few cases, these athletes may even sustain an impressive speed for a short period of time. However, this type of athlete is rare, and it is impractical to recruit a sufficient number of these athletes to push all of the stalled cars nationwide.

The task of raising student achievement is not simple. However, it may be useful to diagnose the reasons for low student achievement, much as it is useful to diagnose the reasons when a car is stalled, before attempting a solution. A proper diagnosis might suggest that attention should be focused in a different direction, and might be more efficient, just as it would be efficient to fix a faulty electrical system in order to start a stalled car, instead of recruiting a team of super athletes to push the car. A number of studies comparing the cost effectiveness of various approaches for raising student achievement are now available, and suggest the nature of the direction that may be fruitful (Yeh, 2007, 2008, 2009a, 2009b, 2010a, 2010b; Yeh & Ritter, 2009).

Acknowledgments

The author wishes to thank Joseph Ritter for extremely helpful comments on an earlier draft of the manuscript.

Notes

1. When researchers examine teacher attrition, they typically compare sample statistics regarding observable characteristics such as gender, teaching experience, and credentials. If both groups of teachers are comparable on these observable measures, then researchers tend to conclude that differential attrition was not a problem. However, the problem that I have identified is different. It is invisible to researchers because it is typically unmeasured and unobserved. It is the difference in the level of dedication between teachers who stay, versus teachers who quit. To measure and observe this, it would be necessary for researchers to design and administer a measure of “dedication” to both groups of teachers and then to use the results to investigate the possibility of differential attrition. This was not done in either of the studies of HCZ and KIPP which supporters are relying upon in stating that these charter school approaches are highly effective (see Dobbie & Fryer, 2011; Tuttle et al., 2010).

2. If $f[x]$ is nonlinear, it would be possible to obtain an increase in aggregate student achievement if teachers are systematically redistributed in a purposeful manner. For example, if $f[x]$ is concave with respect to each school, aggregate gains in student achievement could be obtained by shifting highly dedicated teachers from a high- b school to a low- b school. When $f[x]$ is concave, the gain obtained by the low- b school is larger than the loss to the high- b school. The opposite would be true if $f[x]$ is convex. Thus, it would be possible to obtain an increase in aggregate achievement even if $c = 0$ in the aggregate. However, any gains from global implementation would have to come via systematic redistribution, and the magnitude of those gains would be limited by the available pool of highly dedicated teachers. There is no obvious reason to expect this type of systematic redistribution of teachers.

3. It is possible that performance might remain high even if the unusually dedicated staff goes away. However, that result would undermine the basic assumption underlying the KIPP/HCZ models, which emphasizes the importance of recruiting dedicated teachers. It is unlikely that the high performance of the KIPP/HCZ schools is due to the extended school day. Previous research indicates that lengthening the school day by 60 minutes improves student achievement by only 0.03 SD in math and 0.07 SD in reading (Levin, Glass, & Meister, 1987). This only accounts for a small portion of the effects attributed to the KIPP and HCZ schools. However, the performance of the KIPP/HCZ schools might be attributable to the use of an extended school year. Meta-analytic results of summer school programs estimate a median effect size of 0.19 SD (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). The results of the Tennessee class size experiment (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001; Nye, Hedges, & Konstantopoulos, 1999, 2001) suggest that a portion of the effects of KIPP and HCZ may be due to small class sizes, yet evidence from natural experiments suggests that the Tennessee results are likely due to Hawthorne effects (Hoxby, 2000; Levin, 2001; Wößmann, 2007). Thus, the small class sizes that are characteristic of the KIPP and HCZ schools are unlikely to explain the putative effects of KIPP and HCZ on student achievement. Finally, it is unlikely that the high performance of the KIPP/HCZ schools is due to the enhanced expenditure per pupil that is associated with KIPP and HCZ schools. Based on meta-analytic results, Greenwald, Hedges, and Laine

(1996) estimated that a 10% increase in expenditure per pupil increases student achievement in math and reading by only 0.083 SD per year. In contrast, Hanushek (1992) found that student achievement may be improved by an entire grade level if high-performing teachers are substituted for low-performing teachers, suggesting that the most likely source of the high performance of the KIPP and HCZ schools is the recruitment of highly dedicated, high-performing teachers.

References

- Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., & Walters, C.R. (2010). *Who benefits from KIPP?* (NBER Working Paper No. 15740). Cambridge, MA: National Bureau of Economic Research.
- Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research on teaching*. In N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.
- Carter, S.C. (1999). *No excuses: Seven principals of low-income schools who set the standard for high achievement*. Washington, DC: Heritage Foundation.
- Collins, G. (2010, September 30). Waiting for somebody. *The New York Times*, p. A35.
- Cooper, A. (2009, December 6). *Harlem's education experiment gone right*. Retrieved September 27, 2010, from http://www.cbsnews.com/stories/2009/12/04/60minutes/main5889558_page2.shtml?tag=contentMain;contentBody
- Cooper, H., Charlton, K., Valentine, J.C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. [Serial No. 260]. *Monographs of the Society for Research in Child Development*, 65(1).
- Curto, V.E., Fryer, R.G., Jr., & Howard, M.L. (2010). It may not take a village: Increasing achievement among the poor. Unpublished manuscript.
- Dobbie, W., & Fryer, R.G., Jr. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3), 158–187.
- Finn, J.D., Gerber, S.B., Achilles, C.M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2), 145–183.
- Gordon, R., Kane, T.J., & Staiger, D.O. (2006). *Identifying effective teachers using performance on the job (Discussion paper 2006-01)*. Washington, DC: The Brookings Institution.
- Greenwald, R., Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361–396.
- Hanushek, E.A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1), 84–117.
- Hanushek, E.A. (2009). *Teacher deselection*. Retrieved June 9, 2011, from <http://edpro.stanford.edu/hanushek/admin/pages/files/uploads/Hanushek%202009%20CNTP%20ch%208.pdf>
- Henig, J.R. (2008). *What do we know about the outcomes of KIPP schools?* Tempe, AZ: Arizona State University.
- Hoxby, C.M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4), 1239–1285.
- Ingersoll, R.M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Education Research Journal*, 38(3), 499–534.
- Levin, H.M., Glass, G.V., & Meister, G. (1987). A cost-effectiveness analysis of computer-assisted instruction. *Evaluation Review*, 11(1), 50–72.
- Levin, J. (2001). For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics*, 26, 221–246.

- Nye, B., Hedges, L.V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2), 127–142.
- Nye, B., Hedges, L.V., & Konstantopoulos, S. (2001). Are effects of small classes cumulative? Evidence from a Tennessee experiment. *Journal of Educational Research*, 94(6), 336–345.
- Tough, P. (2008). *Whatever it takes: Geoffrey Canada's quest to change Harlem and America*. Boston, MA: Houghton Mifflin Harcourt.
- Tuttle, C.C., Teh, B., Nichols-Barrer, I., Gill, B.P., & Gleason, P. (2010). *Student characteristics and achievement in 22 KIPP middle schools*. Washington, DC: Mathematica Policy Research.
- Weber, K. (Ed.). (2010). *Waiting for 'superman': How we can save America's failing public schools*. New York: Public Affairs.
- Whitehurst, G.J., & Croft, M. (2010). *The Harlem Children's Zone, promise neighborhoods, and the broader, bolder approach to education*. Washington, DC: Brookings Institution.
- Woodworth, K.R., David, J.L., Guha, R., Wang, H., & Lopez-Torkos, A. (2008). *San Francisco Bay Area KIPP schools: A study of early implementation and achievement: Final report*. Menlo Park, CA: SRI International.
- Wößmann, L. (2007). *International evidence on expenditures and class size: A review*. In T. Loveless & F. Hess (Eds.), *Brookings papers on education policy: 2006/2007* (pp. 245–272). Washington, DC: Brookings Institution Press.
- Yeh, S.S. (2007). The cost-effectiveness of five policies for improving student achievement. *American Journal of Evaluation*, 28(4), 416–436.
- Yeh, S.S. (2008). The cost-effectiveness of comprehensive school reform and rapid assessment. *Education Policy Analysis Archives*, 16(13). Retrieved from <http://epaa.asu.edu/epaa/v16n13/>
- Yeh, S.S. (2009a). Class size reduction or rapid formative assessment? A comparison of cost-effectiveness. *Educational Research Review*, 4, 7–15.
- Yeh, S.S. (2009b). The cost-effectiveness of raising teacher quality. *Educational Research Review*, 4(3), 220–232.
- Yeh, S.S. (2010a). The cost-effectiveness of 22 approaches for raising student achievement. *Journal of Education Finance*, 36(1), 38–75.
- Yeh, S.S. (2010b). The cost-effectiveness of NBPTS teacher certification. *Evaluation Review*, 34(3), 220–241.
- Yeh, S.S., & Ritter, J. (2009). The cost-effectiveness of replacing the bottom quartile of novice teachers through value-added teacher assessment. *Journal of Education Finance*, 34(4), 426–451.

STUART S. YEH is Associate Professor and Coordinator of the University of Minnesota Evaluation Studies Program. He recently published *The Cost-Effectiveness of 22 Approaches for Raising Student Achievement* (Information Age Publishing), which integrates impact and cost information for a range of approaches, including voucher programs, charter schools, increased accountability, and improvements in teacher quality through value-added teacher assessment, NBPTS teacher certification, and a strategy that combines higher teacher salaries with a minimum SAT test score requirement.