EdWorkingPaper No. 23-755

A Kindergarten Lottery Evaluation of Core Knowledge Charter Schools: Should Building General Knowledge Have a Central Role in Educational and Social Science Research and Policy?

David Grissmer
University of Virginia

Thomas White
University of Virginia

Richard Buddin
University of Virginia

Mark Berends
University of Notre Dame

Daniel Willingham
University of Virginia

Jamie DeCoster
University of Virginia

Chelsea Duran
University of Virginia

Chris Hulleman
University of Virginia

William Murrah
Auburn University

Tanya Evans
University of Virginia

The Core Knowledge curriculum is a K-8 curriculum focused on building students General Knowledge about the world they live in that is hypothesized to increase reading comprehension and Reading/English-LA achievement. This study utilizes an experimental design to evaluate the long term effects of attending Charter schools teaching the Core Knowledge curriculum. Fourteen oversubscribed kindergarten lotteries for enrollment in nine Core Knowledge Charter schools using the curriculum had 2310 students applying from parents in predominately middle/high income school districts. State achievement data was collected at 3rd-6th grade in Reading/English-LA and Mathematics and at 5th Grade in Science. A new methodology addresses two previously undiscovered sources of bias inherent in kindergarten lotteries that include middle/high income families. The unbiased confirmatory Reading-English-LA results show statistically significant ITT (0.241***) and TOT (0.473***) effects for 3rd-6th grade achievement with statistically significant ITT and TOT effects at each grade. Exploratory analyses also showed significant ITT (0.15*) and TOT (0.300*) unbiased effects at 5th grade in Science. A CK-Charter school in a low income school district also had statistically significant, moderate to large unbiased ITT and TOT effects in English Language Arts (ITT= 0.944**; TOT = 1.299**), Mathematics (ITT= 0.735*; TOT = 0.997*) and positive, but insignificant Science effects (ITT= 0.468; TOT = 0.622) that eliminated achievement gaps in all subjects.

VERSION: April 2023

Suggested citation: Grissmer, David, Thomas White, Richard Buddin, Mark Berends, Daniel Willingham, Jamie DeCoster, Chelsea Duran, Chris Hulleman, William Murrah, and Tanya Evans. (2023). A Kindergarten Lottery Evaluation of Core Knowledge Charter Schools: Should Building General Knowledge Have a Central Role in Educational and Social Science Research and Policy?. (EdWorkingPaper: 23-755). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/nsbq-hb21

A Kindergarten Lottery Evaluation of Core Knowledge Charter Schools: Should Building General Knowledge Have a Central Role in Educational and Social Science Research and Policy?

David Grissmer[1], Thomas White[1], Richard Buddin[1], Mark Berends[2], Daniel Willingham[1], Jamie DeCoster[1], Chelsea Duran[1], Chris Hulleman[1], William Murrah[3], Tanya Evans[1]

[1] University of Virginia

[2]University of Notre Dame

[3]Auburn University

This draft has not been submitted for publication nor been peer reviewed. Results and interpretations may change during the review process. This draft paper should not be copied or circulated without the permission of the authors. Research supported by the Institute for Education Research (R305E090003), National Science Foundation (1252463), Arnold Foundation and Smith-Richardson Foundation. Views expressed of those of the authors alone, and do not necessarily represent positions of the sponsors. We are grateful to our staff in Colorado for their critical contributions to this project including Jennifer Larson and Shannon Altenhofen and the Colorado Department of Education for providing achievement scores. Errors of fact or interpretation are the sole responsibility of the authors. Communication should be directed to dwg7u@virginia.edu or willingham@virginia.edu.

A Kindergarten Lottery Evaluation of Core Knowledge Charter Schools: Should Building "Cumulative Knowledge" Have a Central Role in Education Research and Policy?

ABSTRACT

The Core Knowledge curriculum is a K-8 curriculum focused on building students General Knowledge about the world they live in that is hypothesized to increase reading comprehension and Reading/English-LA achievement. This study utilizes an experimental design to evaluate the long term effects of attending Charter schools teaching the Core Knowledge curriculum. Fourteen oversubscribed kindergarten lotteries for enrollment in nine Core Knowledge Charter schools using the curriculum had 2310 students applying from parents in predominately middle/high income school districts. State achievement data was collected at 3rd- 6th grade in Reading/English-LA and Mathematics and at 5th Grade in Science. A new methodology addresses two previously undiscovered sources of bias inherent in kindergarten lotteries that include middle/high income families. The unbiased confirmatory Reading-English-LA results show statistically significant ITT (0.241***) and TOT (0.473***) effects for 3rd-6th grade achievement with statistically significant ITT and TOT effects at each grade. Exploratory analyses also showed significant ITT (0.15*) and TOT (0.300*) unbiased effects at 5th grade in Science. A CK-Charter school in a low income school district also had statistically significant, moderate to large unbiased ITT and TOT effects in English Language Arts (ITT= 0.944**; TOT = 1.299**), Mathematics (ITT= 0.735*; TOT = 0.997*) and positive, but insignificant Science effects (ITT= 0.468; TOT = 0.622) that eliminated achievement gaps in all subjects.

Keywords: General Knowledge, Core Knowledge curriculum, kindergarten lottery, Charter schools

INTRODUCTION

Over the last 40 years, there have been several education reform efforts at the national and state level to improve achievement in both Mathematics and Reading/English-LA from K-12 (A Nation at Risk, 1983; National Education Goals Panel, 1998; No Child Left Behind, 2001, Common Core State Standards, 2013). A puzzling pattern of achievement gains has emerged over this 40 year period. At 4th and 8th grade, students have made very small gains in Reading/English-LA, while much larger gains occurred in Mathematics achievement, while at 12th grade, the math gains have also been small and no gains occurred for English-LA (Shakeel, M. D. & Petersen, P., E., 2022).

Figure 1 shows the contrasting long term achievement gains in Mathematics and Reading/English-LA at 4th and 8th grade using two different National Assessment of Educational Progress (NAEP) data sources. Gains in Mathematics at both 4th and 8th grade have been 3-4 times larger than Reading/English-LA gains. For instance, the median student taking Long Term NAEP in 2012 compared to the median student in 1978- 34 years earlier- would be scoring about 7 percentile points higher in Reading/English-LA, but 25 percentile points higher in Mathematics. Adding to the puzzling contrast between achievement gains in Mathematics and Reading/English-LA is the amount of weekly K-8 classroom time spent on teaching Mathematics (4.9. to 5.7 hours) is far less than Reading/English-LA (10.5 to 11.6 hours) and has changed little in decades (Morton, B., & Dalton, B., 2007; Hoyer, K., M., & Sparks, D., 2017; Perie et al, 1997).

Two questions that arise for researchers and policymakers from this data are:

- Why has so little progress been made in increasing English-LA achievement at 4th, 8th and 12th grade in the last 30-40 years?
- What kinds of new educational and social reforms and policies are needed to improve future Reading/English-LA achievement?

Future national, state and local K-12 school reforms cannot be successful unless research can address these two questions.

Several national and select panels have been convened over 40 years to specifically address the lack of progress in Reading/English-LA achievement (Anderson et al, 1985; National Research Council, 1998; National Reading Panel, 2000; National Research Council, 2000; RRSG, 2002; Pearson & Cervetti, 2015). A consensus emerged in these reports that early

Reading/English-LA instruction needed to proceed by developing skills in five areas: phonics, phonemic awareness, vocabulary, reading comprehension and fluency. However, while there is currently little disagreement on whether and how to teach phonics, phonemic awareness and vocabulary, there is much uncertainty and intense debate on the best approaches to improving a student's comprehension and fluency of what they read. Comprehending what is heard and read is a complex cognitive process that lies at the heart of learning, yet is still not well understood.

In the last 30 years, two approaches to improving reading comprehension have been articulated. The first approach assumes that students primarily need a learned set of "procedural skills" that enable comprehension of what is read. These "procedural skills" are learned by students during their Reading/English-LA instruction that presumably allow them to comprehend what they read. This "procedural skills" based approach to reading comprehension has been a major focus of the national and select panels recommending educational policy and classroom practice during the period of time allocated to Reading/English-LA instruction for forty years.

A second approach to improving reading comprehension assumes that building a stronger base of previously stored General Knowledge allows better comprehension (Hirsh et al, 1988, Hirsch, 2003, Hirsch, 2006; Willingham, 2006, Hirsch, 2011; Willingham and Lovette, 2014; Willingham, 2017, Willingham and Riener, 2019; Cabell and Hwang, 2020; Hwang & Duke, 2020). However, the approaches and methods to building a stronger base of General Knowledge are much more complex, less understood and more difficult to measure than implementing the "procedural skills" approach.

For instance, building a stronger base of General Knowledge can involve changing out-of-school environments from birth and during early schooling and also can involve a multi-year, multi-grade effort during schooling that starts at kindergarten and requires more exposure to subjects that build General Knowledge- typically more time spent on Science and Social Sciences. While the reading and select panels have acknowledged the need to build a stronger base of General Knowledge, none of the panels suggested the more comprehensive curriculum changes required to accomplish this.

One of the reasons that long term Mathematics gains may be much greater than Reading/English-LA is that Mathematics requires a much more limited and well defined amount of background knowledge that can be provided in earlier mathematics instruction, and except for word problems, does not require a wide ranging amount of General Knowledge. Comprehending the wide-ranging texts that are used in Reading/English-LA often requires a breadth and depth of General Knowledge about a wide range of domains of knowledge. Unlike Mathematics, the General Knowledge required to comprehend these diverse texts can encompass virtually any of the domains of knowledge that typical students experience in school or in their out of school environment. And, unlike Mathematics, it seems likely that a significant part of their capacity for comprehending Reading/English-LA texts relies on General Knowledge acquired out of school and before students enter school in their family and community environment. Thus, the origin of the often repeated ideas that Mathematics is more school dependent, whereas Reading/English-LA is more widely dependent on out-of- school environments.

Significant long term improvements in Reading/English-LA achievement similar to gains in Mathematics achievement would have required a much more comprehensive strategy than simply improving the quality of instruction during the time spent on Reading/English teaching improved "procedural skills". Over the last 30-40 years, there has been little change in the time spent teaching different subjects The time spent on different subjects in elementary grades in public schools shows the largest time spent on Reading/English-LA (10.5 to 11.6 hours) weekly followed by Mathematics (4.9 to 5.7 hours) with much less time on Science (2.8 hours) and Social Science (2.9 hours) and this pattern shows little variation across public schools with different student populations and has changed little over the last 30 years (Morton, B., & Dalton, B., 2007; Hoyer, K., M., & Sparks, D., 2017; Perie et al, 1997).

This data would suggest the absence of any long term widespread curriculum strategy to increase Reading/English-LA achievement nationally by placing increased emphasis on subjects that build General Knowledge. The small gains (7 percentile points) that have occurred in 4^{th} and 8^{th} grade Reading/English-LA over the last 30-40 years may be accounted for by improving the early skills involved in phonics, phonemic awareness and vocabulary and improving the "procedural skills" approach to improving comprehension. However, there has been no large scale, long term experimental evidence that provides support for the current reading curriculum using the "procedural skills" approach. A few experimentally designed interventions have been implemented for periods of 1-3 years with typically significant small or null effects on standard measures of reading comprehension (Conor et al, 2013, Kim, et al., 2021; Kim et al., 2023). However, no longer term follow-up measures were collected.

Recent Reading Panels convened to address the lack of progress in improving Reading/English-LA achievement have increasingly pointed to the need to address the more complex issue of building General Knowledge through "knowledge-rich" curriculum , but have not suggested the more basic structural curriculum reforms across grades and subjects that might be needed (Pearson et al, 2020). Wexler, 2019 provides a compelling case for such knowledge-rich curriculum and is a leading advocate (Wexler, 2018; Wexler, 2022). Several states have recently developed a variety of approaches to address the "Science of Reading" problem (Schwartz, 2022; Schwartz, 2021a; Schwartz, 2021b). However, the conceptual basis for the building of General Knowledge as well as a curriculum designed to build General Knowledge pre-dates current interest by three decades.

E. D. Hirsch—the leading long term proponent of the knowledge-based approach—led a research and consensus-building effort in the late 1980's to develop a comprehensive curriculum directed at building students "cumulative General Knowledge" of the world they live in. This "knowledge-based" curriculum (The Core Knowledge curriculum) included all subjects (language arts, history and geography, mathematics, science, music, and visual arts) and all K-8 grades (Core Knowledge Foundation, 2010). Since its inception, the Core Knowledge foundation has produced teacher instructional manuals for all subjects and K-8 grades and offered professional development support for teachers, and the curriculum has been implemented in hundreds of schools nationally (Core Knowledge Foundation, 2010). Despite its presence for over 25 years in hundreds of schools nationwide, there have been no experimental evaluations of schools teaching the Core Knowledge curriculum.

This study utilizes an experimental approach through a kindergarten-based lottery to assess the long term effects (after 4-7 years of intervention from K to 3rd-6th grade) on Reading/English-LA, Science and Mathematics achievement in Charter schools teaching the Core Knowledge curriculum (CK-Charter). This evaluation is also the first to assess effects of a Reading/English-LA intervention using a kindergarten lottery that includes predominately middle/high income students. A new methodology addresses two sources of bias inherent in kindergarten lotteries that include middle/high income families.

Fourteen oversubscribed kindergarten lotteries for enrollment in nine CK-Charter schools using the K-8 Core Knowledge curriculum had 2360 students applying from parents in predominantly middle/high income school districts. State achievement data was collected at 3rd -

6th grade in English Proficiency and Mathematics and at 5th Grade in Science, and "intent to treat" (ITT) and "treatment of the treated" (TOT) effects estimated.

The confirmatory Reading-English-LA results show statistically significant ITT (0.241***) and TOT (0.473***) effects for combined 3rd-6th grade achievement with statistically significant ITT and TOT effects at each 3rd-6th grade. Exploratory analyses also showed significant ITT (0.15*) and TOT (0.300*) effects at 5th grade in Science and positive, but insignificant effects in 3rd-6th grade Mathematics. A CK-Charter school in a low income school district also had statistically significant moderate to large ITT and TOT effects in English Language Arts (ITT= 1.089**; TOT = 1.737**), Mathematics(ITT= 0.807*; TOT = 1.271*) and Science (ITT= 0.667*; TOT = 1.032*). These effects were large enough to close achievement gaps for disadvantaged students by 3rd-6th grade in all subjects measured.

The policy-relevant TOT effect size of 16 percentile points for all students equals the 40 year difference in gains between Mathematics and Reading/English-LA. The size of these gains could also close the international gap in Reading/English-LA for U.S. students. U.S. students placed 15th among 50 countries taking the 2016 PIRLS 4th grade Reading/English test, but national student gains similar to gains in this intervention would place the U.S. among the top five countries (PIRLS, 2016; Mullis et al, 2017).

The characteristics of this intervention as well as the results are atypical in the intervention research literature involving RCTs to improve Reading/English-LA achievement. Interventions in Reading/English-LA typically report null to small size, short term effects when standardized outcome measures of Reading/English-LA are used with declining effects whenever longer term outcomes are measured (Kraft, 2020; Bailey, 2017). The atypical moderate size, statistically significant TOT effects in this intervention for students from all family income groups together with the much larger, statistically significant TOT effects for a lower income CK-Charter school may be due to several aspects of the intervention that are also atypical.

These include an intervention that was implemented from K-6 providing at least 4 years and up to 7 years of dosage to students. Perhaps more importantly, the intervention changed not only the instruction in Reading/English-LA, but changed the emphasis and coordination of the instruction across all subjects and grades to increase a student's General Knowledge of the world (Hirsch, 2011; Hirsch 2019). Teachers in the Core Knowledge schools utilized completely different instructional material across all subjects and grades and implemented different classroom methods such as read-alouds (Core Knowledge Foundation, 2010). From this perspective, the intervention changed instruction and teacher preparation across all subjects and grades for up to 7 years. No Reading/English-LA intervention has been designed to have such a potentially comprehensive and long term effect on student achievement. Nagy, 2005 suggests that only long term and comprehensive interventions can be expected to improve long term measures of reading comprehension.

Students exposed to a curriculum that is aligned across grades and subjects such as Core Knowledge may bring greater efficiency from building on knowledge learned in previous grades and avoiding unnecessary content repetition (Engel et al., 2013). This kind of integration and focus across all grades and subjects is not typically prioritized in schools or school districts or states where, typically, two subjects (Mathematics and Reading/English-LA) are given the highest priority and little integration occurs across all subjects.

The level of General Knowledge is highly positively correlated with the more traditional measures of SES, parental education and income. However, unlike these measures, General Knowledge is malleable and can be increased with interventions and can be targeted to students with lower levels. Perhaps the most intriguing result from this study is that the K-6 intervention implemented in a single low income school had effects that eliminated achievement gaps at 3rd-6th grade in Reading/English-LA, Science and Mathematics. If these results replicate, early interventions that build General Knowledge may be a new direction for eliminating achievement gaps across all subjects.

There has been substantial **non-experimental** evidence linking gains in measures of General Knowledge to later achievement in Reading/English-LA, Science and Mathematics (Claessens et, 2009, Duncan et al, 2007; Duncan et al, 2020, Grissmer et al, 2010; Morgan et al, 2016). The experimental results in this study are very similar in magnitude and in their pattern across subjects with these non-experimental results. This evidence suggests that gains in General Knowledge would have a larger effect on future achievement than similar gains in the more widely studied non-cognitive skills including executive function, visuo-spatial/fine motor and socio-emotional skills (Grissmer & Eiseman, 2008: Grissmer, et al, 2010).

Together the experimental and non-experimental evidence linking the level of early General Knowledge to later achievement in Reading/English-LA, Science and Mathematics suggest that increasing General Knowledge should be an important area for future research. New research initiatives need to assess experimental replication and explore causative mechanisms through collection of mixed methods data from classroom observations, surveys of teachers, parents and students.(Grissmer, 2017). If future research supports these findings and the underlying theory, there are significant long term implications for the direction of educational and social science research and policy.

Federal data collections are designed to monitor and better understand the most important economic, educational, and social trends in society. Educational policies over decades have considered increasing achievement in Mathematics and Reading/English-LA to be the primary building blocks for later achievement and positive educational outcomes, and are among the primary early predictors for a range of later outcomes in life. Achievement in Mathematics and Reading/English-LA are also the primary current measures used to evaluate and compare the effectiveness and quality of schools and teachers. For these reasons, early achievement levels in Mathematics and Reading/English-LA are tracked at national, state and local testing levels.

This paper has suggested that the primary measures currently collected to monitor education including measures of Mathematics and Reading/English-LA (as well as other subjects) do not capture an important aspect of learning, namely, the level of General Knowledge. Well-designed measures of General Knowledge should be considered as an important addition to our routinely collected national measures for students in elementary grades. However, designing nationally collected measures of General Knowledge will pose a substantial challenge for researchers and policymakers, not unlike the challenge of measuring the Gross National Product as an economic indicator.

However, measures of General Knowledge will carry an additional challenge. Characterizing and measuring the General Knowledge that young students have in lower elementary grades will need not only scientific validity, but also political viability. Adults will offer different judgments as to which General Knowledge matters most for students, and that will undoubtably trigger debates and a variety of viewpoints. But characterizing the General Knowledge that is needed to better understand the books actually read by children and the textbooks used in future education seems essential to educational efficiency and meeting long term educational goals.

To some extent, common ground may be established by taking an empirical approach that seeks to link the gains in early General Knowledge (however defined) to later rising achievement. Whatever the method to make the decision, it's important to remember that the decision cannot be avoided. Making no active decision about curriculum—that is, doing nothing—is still a choice, but it's a choice that means students may not have the option of a knowledge-rich, logically sequenced curriculum and the nation may miss the opportunity to have a better educated future work force.

The implications of raising long term Reading/English-LA achievement by an average of 16 percentile points for a sample of students across all income levels and closing achievement gaps for low income students carries implications for possible wider and longer term effects. The achievement gains in this study, if mediated through increased reading and verbal comprehension due to General Knowledge, would suggest that future achievement gains might also occur in later schooling across all subjects that require increasing reading and verbal comprehension- virtually all subjects to different degrees. Students that carry higher levels of verbal and reading comprehension into later schooling might be expected to have higher achievement across a range of subjects, have higher educational attainment, increases in high school completion, college entrance and years of education as well as higher wages and labor force productivity.

A significant weakness in previous research and theory may be embedded in the theory and language involving human capital that places the emphasis on "skill building" as the primary developmental cognitive process involved in learning (Bailey et al, 2017). This theory characterizes development as the process of building increasingly complex skills summarized in the phrase "skill begets skill" (Heckman, 2006). The results of this study would suggest that there are two separate but complementary, cognitive processes involved in development and learning: "skill building" and "knowledge accumulation". Perhaps the phrases that better capture cognitive development would be- "skill begets skill; knowledge begets knowledge; and almost certainly- skill *x* knowledge begets skill *x* knowledge".

However, the results from a single intervention and evaluation can never provide sufficient evidence for achieving a new longer term research or policy consensus among researchers or policymakers or for understanding of the causative mechanisms. Rather the implications of these results should primarily initiate a large new research and policy agenda directed to replication as well as research that identifies and leads to better understanding of the causative mechanisms and theory involved in the Core Knowledge curriculum. In the long term, it is stronger theories and increased understanding of the causative mechanisms that predict the results of new experimental evidence that moves science forward.

Background

This study takes advantage of two large-scale interventions in U.S. education directed at improving student achievement: charter schools and the Core Knowledge K-8 curricular design. Charter schools enrolled 3.4 million students in school year 2019-2020- over 5% of students in public schools nationwide (Digest of Educational Statistics, 2021). In 2019, there were approximately 1700 pre-K-8th public schools- about 2.5% of public schools- using the Core Knowledge K-8 curriculum (Core Knowledge Foundation, personal communication).

This study incorporates data from 14 kindergarten lotteries from 9 oversubscribed Charter schools teaching the Core Knowledge curriculum (CK-Charter) to provide the first experimental evidence for the long term achievement effects (3rd to 6th grade) of the Core Knowledge curriculum. To interpret the results of this study, it is important to understand the rationale and previous research findings underlying both the Core Knowledge curriculum and Charter schools.

<u>Charter Schools - Rationale</u>

Charter schools are publicly funded but granted more autonomy than traditional public schools in shaping major decisions about how to educate students. There are four rationales for granting such flexibility. First, public schools were constrained by a political and bureaucratic process more influenced by local, state, and federal requirements (Chubb & Moe, 1990). Granting charters greater flexibility in choosing policies and curriculum was expected to lead to a more diverse set of schools that could be evaluated for improving student outcomes. The charter school movement has focused on improving three aspects of schools: autonomy, innovation, and accountability. Advocates for charter schools argue that these aspects of reform will produce organizational innovations which in turn will lead to better student outcomes (Chubb & Moe, 1990; Walberg & Bast, 2003).

The second rationale for charter schools is to offer parents a wider variety of schools, enabling them to select one that fits the developmental needs of their children (Betts, 2005). Practices and conditions related to autonomy, innovation, and accountability are expected to differ across schools (and school types) in response to parental and community preferences, further promoting student achievement (Walberg, 2011). The third rationale is that an evolving competitive process between public and charter schools would lead to improvements in both charter and public schools and in student achievement (Betts, 2005; Hoxby, 2001, 2003).

The fourth rationale for charter schools is to promote their potential as incubators of innovation, experimenting with different school policies and organization, curriculum, and pedagogical practices (Nathan, 1999), which could be more rigorously evaluated using the random lottery selection process often mandated for over-subscribed charter schools (Hoxby & Murarka, 2008). Such evaluations could identify specific interventions that raise achievement, and these interventions could then be adapted to improve both charter and public schools. This study is made possible by the flexibility given to Charter schools who implemented the Core Knowledge curriculum and allows the experimental measurement of the effects of Charter Schools teaching the Core Knowledge curriculum.

Charter Schools - Empirical Evidence for Impacts

The advent and rapid growth of charter schools has been accompanied by a voluminous research literature to assess whether charter schools are achieving the envisioned outcomes. There have been several reviews or meta-analyses of this literature that have been done at different points in time as well as large scale studies that inform our literature review, as do the original studies covered by these reviews (Buddin & Zimmer, 2003; Bifulco & Ladd, 2006; Sass, 2006; Zimmer, et al., 2011; Baude, Casey, Hanushek, & Rivkin, 2013; CREDO, 2009, 2011, 2013; Clark, Gleason, Tuttle, 2015; Berends, 2015; Cheng et al., 2017; Epple et al., 2016; Chabrier et al., 2016; Gamoran & Fernandez, 2018; Austin & Berends, 2018, 2020; Betts & Tang, 2019; Ferrare, 2020; Zimmer et al., 2020; Cohodes and Parham, 2021; Zimmer, Buddin, Smith, and Duffy, 2020).

Overall, these studies reach no consensus on the key question of whether charter schools produce higher achievement than public schools, but rather show a mix of positive and negative differences. This evidence suggests that simply freeing public schools from their perceived bureaucratic constraints through charter schools does not reliably produce higher achievement. However, Berends (2015, 2020, in press) suggests that asking whether typical charter schools can outperform public schools may be the wrong question. There is a large amount of variability among both charter schools and non-charter public schools in both educational policies, pedagogical practices, and student/family characteristics. Instead, the question is whether there are conditions under which charter schools consistently outperform non-charter public schools, and whether the conditions causing these differences reside primarily in innovative charter school practices or the characteristics of non-charter public schools.

The research has suggested that Charter schools have distinct advantages over non-charter public school counterfactuals located in inner city schools, but not for suburban schools. In Massachusetts, charter schools in inner city, urban districts have consistent, small-to-moderate positive effects on reading and math, while charter schools in suburban school districts show no effects (Angrist, Pathak, & Walters, 2013). A series of CREDO city analyses show fairly consistent results of small to very large statistically significant achievement gains in math and reading in inner city charter schools over non-charter inner city public schools, but no consistent evidence for achievement gains outside of inner cities (Center for Research on Education Outcomes, 2013). Gleason, Clark, Tuttle & Dwoyer's (2010) lottery-based evaluation of charter school impacts included 36 middle schools across 15 states located in both inner city and suburban areas (see also Clark et al., 2015). Gleason et al. found no overall significant impacts on math or reading test scores. However, consistent with the findings of Angrist et al. (2013), schools in large urban areas had statistically significant positive effects on math (+0.16 SD), whereas schools outside urban regions had statistically significant negative effects (- 0.14 SD). Cohodes and Parham, 2021 suggest that charter schools may outperform public schools in certain urban school districts, but there is no consistent evidence for charter schools outperforming public schools outside of the inner city.

A few studies have measured what happens inside charter schools outside the inner city. Their results shed light on why particular kinds of charter schools might be more effective (Hess & Loveless, 2005; Zimmer & Buddin, 2007; Berends, Watral, Teasley, & Nicotera, 2008; Betts et al., 2006; Berends et al., 2010, 2019; Berends, 2015, 2020; Zimmer et al., 2020). For example, research from Texas charter schools using quasi-experimental methods suggests that charter schools in operation longer tend to show stronger effects, and the overall effectiveness of charter schools seems to improve over time from the combination of more years of operation and other schools' exits from the market (Baude, Casey, Hanushek, & Rivkin, 2014).

Charter schools outside the inner city have often been designed as "test-beds" for identifying policies and curriculum that, if successful, could be implemented more broadly in charter and public schools (Betts & Tang, 2019). A movement to implement and financially support groups of charter schools (Educational Management Organizations (EMO); Charter Management Organization (CMO) with similar policies is spreading within and across states (Farrell, Wohlstetter, & Smith, 2013; Berends, in press).

Dynarski et al. (2018) used a lottery-based design to examine the effects of a large for-profit set of schools (EMO-National Heritage Academy) on student achievement. Using lotteries from 44 schools in Michigan between 2003 and 2014, they estimated the impacts on achievement for the EMO National Heritage Academy. They found that attending a National Heritage Academy charter school resulted in a 0.04 SD gain in mathematics; effects on other outcomes like reading, attendance, grade progression, disciplinary actions, and special education placement were not significant. Heterogeneous effects also occurred among non-profit Charter Management Organizations that estimated effects for 22 middle schools (Furgeson et al. (2012)). Results showed more positive than negative effects with substantial variation in each direction.

Overall, this literature supports three main conclusions:

Charter schools in urban school districts show consistent, statistically significant gains in Reading/English-LA and Mathematics over their inner-city traditional public counterfactuals with ITT effect sizes that range from small to moderate.

Typical charter schools in non-urban school districts show no consistent, significant effects for higher achievement than traditional public schools in Reading/English-LA and Mathematics at the elementary, middle, or high school level. Almost all studies involving large samples of similar traditional public and charter schools have shown a large variance in results, from significant negative to positive results.

Evaluation of instructional designs for charter schools outside of inner cities is an important avenue for future research to extend our knowledge base (Berends & Dallavis, 2020). However, this research requires starting or finding a large enough sample of charter schools with similar instructional design to provide the needed lottery data.

In this context, the current study takes advantage of the large number of charter schools teaching the Core Knowledge curriculum in Colorado. CK-Charter schools were started in the 1990s and their popularity spread such that by 2012 there were more than 50 CK-Charter schools in Colorado. Many of these schools were oversubscribed mandating kindergarten lotteries for admission. This study utilizes data from 14 oversubscribed school lotteries in 9 CK-Charter schools to generate experimental evidence about the long term achievement effects (3^{rd} - 6^{th} grade) of Charter schools using the Core Knowledge K-8 instructional design and curriculum.

Core Knowledge Charter Schools

CK-Charter schools were first implemented in Colorado in the early 1990s and have grown steadily to over 50 of these schools in Colorado. We identified schools with a history of excess enrollment demand and associated use of kindergarten lotteries for admission. We offered $1000 for each year of participation in the study. We prospectively tracked lotteries for one or two entry cohorts in these schools. Some of these lotteries were eliminated because they made offers to all applicants, leaving the 14 CK-Charter lotteries in 9 schools with excess demand.

All schools in the study had been in operation for a minimum of four years and up to 14 years when the study began. We tracked each lottery and lottery outcome prospectively at each school at least two months before the lottery until one month after kindergarten entry to provide internal validity of the study. Each school developed a spreadsheet with applicant information that also tracked email and phone communication with lottery applicants. The information on each student included first and last name, lottery number, and the timing of any offer and acceptance/rejection of offers. Eleven of the lotteries included data on birth date—about 75% of applicants. Most lotteries included gender- and using software that assigns gender from first names- gender was identified for applicants in lotteries not having gender information, resulting in gender missing on only 3% of applicants with androgynous names.

Core Knowledge Curriculum

The Core Knowledge curriculum takes an unconventional approach to improving achievement. It uses a comprehensive approach to specifying curriculum in every subject (Language Arts, Mathematics, Science, History, Geography, Visual Arts and Music) at each K-8 grade (Core Knowledge, 2010). The curriculum is designed to build students' cumulative General Knowledge and their range and depth of vocabulary to boost their capacity to comprehend the world they live in (Hirsch, 1988, 2003, 2006, 2011). General knowledge comprises content about people, objects in the world, facts and meanings of words, and associations among these entities. The curriculum is controversial because it suggests that student's Reading/English-LA and Science achievement in later grades may depend not only on the quality of direct instruction in these subjects but on the content and quality of curriculum outside of these subjects.

The Core Knowledge curriculum has early and continuing emphases on: (a) following a planned sequence of specific topics that integrates knowledge from the seven subject areas across K-8 grades to systematically build their knowledge and comprehension of the world; (b) exposing children to broad, information-rich curricula across subjects to build oral vocabulary; and (c) using read-alouds to build oral vocabulary, knowledge, and listening skills (Core Knowledge, 2010). Throughout this study Core Knowledge had a K-8 integrated set of teacher manuals for the specific topics in each year, as well as associated student reading materials (Core Knowledge, 2010), and it offered access to aligned lesson plans from a variety of authors. The Core Knowledge foundation also provided professional development and support opportunities for teachers and principals across the nation.

Research on General Knowledge in Learning

Measures of General Knowledge have not typically been included in major longitudinal data that are used to study factors that influence achievement from early grades. However, the 1998 Early Childhood Longitudinal Survey (ECLS-K) that tracked students from kindergarten to 5th and 8th grade collected family and student characteristics as well as a range of early cognitive measures including early reading and math skills along with executive function (attention), visuo-spatial/fine motor skills, socio-emotional skills and General Knowledge prior to kindergarten entry. Duncan et al, 2007 utilized this data in combination with international longitudinal data sets having measures of executive function to estimate significant effects from measures of executive function on 5th grade Mathematics and Reading/English-LA achievement. Grissmer et al, 2010 and Murrah, 2010 utilized the ECLS-K to predict 5th and 8th grade Mathematics, Reading/English-LA and Science achievement that added variables for visuo-spatial/motor skills and the level of General Knowledge. Figure 2 shows the relative significance of the predictors for 5th grade Math, English and Science.

Fifth grade math is most strongly predicted by early domain specific Mathematics skills, followed by level of General Knowledge, and reduced, but still significant impacts from executive function and visuo-spatial/fine motor skills. Fifth grade Reading/English-LA and Science is most strongly predicted by the level of General Knowledge followed by early Mathematics, executive function and visuo-spatial/fine motor skills. The variable that would have the largest cumulative impact across later subjects is the early level of General Knowledge. These non-experimental results suggest that early General Knowledge acquisition may be as critical to long term achievement as the more well-studied early skills in Mathematics and Reading/English-LA, executive function and visuo-spatial/motor skills. Despite this empirical evidence for the role of General Knowledge in predicting later achievement, a consensus in educational research on its importance, its measurement or a theory specifying its role in cognitive development has been slow in developing.

Early General Knowledge can emerge informally through everyday interactions and experiences, conversations with peers, teachers, parents and other adults, as well as through more prescriptive activities such as travel, reading books and visits to museums (National Research Council, 2009). By the time children enter school they may have considerable prior General Knowledge gained from early childhood environments (National Research Council, 2007). Once in school, students with high prior knowledge tend to learn course content better than those with low prior knowledge (Steinkamp & Maehr, 1983; Tobias, 1994).

The link between General Knowledge and later achievement may operate through improvements in reading comprehension as well as improvements in motivation to learn. The systematic building of General Knowledge has been long identified as an important component in building skills needed for reading and verbal comprehension (Hirsch, 2003, 2006; RRSG, 2002; National Research Council, 2000, 2007, 2009; Pearson, P.D, et al, 2020).

For instance, according to many contemporary theories of comprehension, the reader may construct a "situation model" integrating relevant prior knowledge with a "textbase" or mental representation of the meanings of the words and sentences in the text (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; Zwaan & Radvansky, 1998). The situation model enables the reader to fill in gaps and unstated ideas in the text, disambiguate the meaning of words and sentences, integrate information across sentences, and make inferences. Readers who lack prior General Knowledge often fail to fill in conceptual gaps within texts and fail to make inferences that go beyond information that is explicitly stated in the text (e.g., Beck et al, 1991; Voss & Silfies, 1996), even when they receive training in comprehension strategies (McNamara, 2004).

The effects of prior General Knowledge on reading comprehension are also revealed in studies of adults and children who are (a) high or low on measures of knowledge or expertise in a domain and (b) high or low on measures of reading skill or general cognitive ability (Adams, Bell, & Perfetti, 1995; Recht & Leslie, 1988; Schneider, Korkel, & Weinert, 1989; Walker, 1987). These studies show that prior knowledge can compensate for low reading skill or cognitive ability. For example, 3rd, 5th, and 7th graders with high levels of knowledge about a topic but low performance score better on comprehension posttests than high-performance novices, and third-grade experts outperform fifth-grade novices (Schneider et al., 1989).

Reading/English-LA, Mathematics and Science at later grades may demand increasing student contextual understanding of their physical and social world. The National Council of Teachers of Mathematics advocates for students to understand the practical, real-world applications of mathematics in order to best learn mathematical concepts (National Council of Teachers of Mathematics, 2000). Also, word problems in mathematics often demand comprehension of the world. Further, having extensive prior knowledge may aid in developing higher order cognitive skills, such as mathematical and scientific reasoning, since such processes are thought to require a combination of content General Knowledge and process skills (Zimmerman, 2000).

Improved verbal and reading comprehension may also increase student motivation. Expectancy value theory posits that students have increased motivation when they see the value of what is learned for their everyday lives (Wigfield & Eccles, 2000). Increased General Knowledge can impact student motivation by enabling students to make connections between their personal lives and academic subject knowledge. Such connections are the foundation of inquiry-based education, and serve as a foundation for both interest and knowledge development (Renninger & Hidi, 2011; National Research Council, 2000). Research demonstrates that students' perceived utility value of academic course content is closely related to achievement, engagement, interest, and to perceived utility in their lives (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009). For instance, when students write essays linking science learned in school to their lives, students showed increased interest in taking more science courses and overall grades in their science class.

A final source of empirical evidence on the effect of General Knowledge on achievement comes from the psychological literature- the Cattell-Horn-Carroll (CHC) theory- that identifies the differences in cognitive characteristics between higher and lower scoring students. One of the most important characteristics identified that is hypothesized to explain these differences is the level of General Knowledge defined as "the student's knowledge about the world that they live in" (Evans et al, 2002; Zaboski et al, 2018).

The Core Knowledge theory of action suggests that students from middle/high income families present a stringent test for the Core Knowledge curriculum for three reasons. First, these advantaged children typically have family environments that have provided wide exposure to vocabulary and sources of General Knowledge (Lareau, 2011). Second, these middle/high income parents who lose the lottery have access to a wide range of high quality public, charter, private and home schooling alternatives (Murnane and Riordan, 2018; Lareau & Goyette, 2014). Third, previous research on charter schools has found no consistent impacts for charter schools in middle/high income school districts. Thus, it may be difficult for middle/high income students in Core Knowledge charter schools to show achievement gains above those in control schools. The range of school and family characteristics in our study from very high income suburban school districts to a school in a low income district enables a test of whether Core Knowledge charter schools are effective for a wide range of family income levels.

Research on Effects of the Core Knowledge Curriculum

Existing research on the effects of Core Knowledge is limited in several ways. There are no experimental studies. The few quasi-experimental studies have been conducted under conditions of partial or early implementation, and the sample sizes have been relatively small with a wide variance in results. Datnow et al. (2003) compared the standardized test performance of fourth grade students in two urban Core Knowledge schools and two matched comparison schools and found no difference in reading achievement. Taylor compared scores on the Iowa Tests of Basic Skills for more than 300 matched pairs of Core Knowledge and non-Core Knowledge students in Grades 3-5 in an urban school district with statistically significant effects across several subtests including reading comprehension (d = +.17) (Richards, O. H. (2001).

Datnow, Borman, and Stringfield (2000) examined achievement outcomes for two student cohorts in four Core Knowledge schools in four states. Each Core Knowledge school had a within-district, demographically matched non-Core Knowledge comparison school. Results on norm-referenced test of reading achievement showed no effects. However, using classroom observations, they found evidence for wide variance in implementation, and adjustments showed Core Knowledge had positive effects on norm-references reading. Borman et al. (2003) reported a mean effect size of +0.03 for six studies of Core Knowledge, combining results for reading, math, and other subjects. However, the results also appeared to be sensitive to fidelity and years of implementation.

Our study has many unique aspects that extend the current research on both the Core Knowledge curriculum, Charter schools and research on educational interventions. It is the first study to provide experimental evidence about the Core Knowledge curriculum. Second, schools were located predominantly in middle/high income school districts, whereas most intervention research using lotteries focused on urban, inner city schools (Betts & Tang, 2019). Third, almost all educational intervention research has focused on interventions lasting for shorter time periods than 4-7 years. Fourth, interventions using lotteries have focused more on lotteries at later elementary, middle, or high school and measured impacts over shorter time periods. This focus on later grades may underestimate the impact of charters since research suggests that early interventions may be more effective than later interventions (Heckman, et al, 2006; Cunha and Heckman, 2007). Fifth, the Reading/English-LA achievement test (Partnership for Assessment of Readiness for College and Career) has a combined achievement score that includes both a reading and writing component with achievement scores available from 3rd-6th grade.

DATA

School Sample and Characteristics

Nine CK-Charter schools participated in the study. Five of the schools participated in consecutive lotteries for kindergarten entrance in school years 2009-2010 and 2010-1011. Four schools participated in only one lottery to provide a total of 14 lotteries for analysis. The nine schools were located in six school districts that stretched from the Denver area to northern districts including Loveland and Ft. Collins.

Table 1 shows the characteristics of the schools and their school districts. Six of the schools are in the Denver-Aurora-Lakewood SMSA that includes 10 counties around Denver. Four of these schools are in very high income suburbs of Denver in the Douglas county school district. Two schools were northwest of Denver in middle-income Jefferson County and Littleton school districts, and one school is in Aurora- a low income, urban school district east of Denver. Three schools are outside of the Denver SMSA in northern Colorado near Loveland and Ft Collins in the Thompson and Pourde school districts. The study schools vary in size having 2016 K-8 enrollments ranged from 466 to 2359.

Table 1 also shows the school district median family income (2010) and percentage of families with children below the poverty level. Three schools in the Douglas school district have a median family income of $114,233, while five schools are in districts with median income from $92,137 to $75,105. The median family income in these districts is above the Colorado median income ($74,000) and also the national median income ($64,000). One school in the lower income Adams-Arapahoe school district near Denver has a median income of $51,424, with 28.8% of families under the poverty level. This school participated in only one lottery, leaving 13 of our 14 lotteries in middle to high income school districts.

Lottery Samples and Characteristics

The lotteries at each school occurred in the December to February window prior to kindergarten enrollment in September. Each lottery was monitored to ensure adherence to randomization and until enrollment in kindergarten could be verified. Matching all lottery applicants using first and last names across 14 lotteries revealed that some students applied to more than one of the 14 lotteries. Almost all of those applying to more than one lottery applied to two or more schools for kindergarten entry in the same year, while some made applications in two consecutive years for kindergarten entry. The former group lived in catchment areas where more than one of our schools allowed multiple applications. The latter were in schools participating in consecutive lotteries who applied and lost the first lottery, and delayed kindergarten entry and reapplied in a lottery in the next year. We treat these two groups differently in the analysis and use the term "multiple applier" only to the former group, while referring to the latter group as "delayed entrants or red-shirts".

We estimated the effects of the intervention for single appliers and for the larger sample including all lottery applications from single and multiple appliers ("all-applications"). As shown in Table 2, there were 2853 applications across all 14 lotteries submitted by 2310 students. The student sample of single appliers is 1831 or 79.3% of all students, while 479 or 20.7% of students submitted more than one application. Of the 2853 applications, 1831 or 64.2% were from single appliers, while 35.8% were from multiple appliers.

Table 3 summarizes the lottery outcomes for the all-applications and single-applier samples. For single appliers, 37.6% were lottery winners. Across all-applications, 35.4% were winners, and 41.3% of all students won at least one lottery. Table 4 shows the lottery characteristics and lottery outcomes for the "all-applications" sample for each of the 14 lotteries. Overall, 35.4 % of the "all-applications" were lottery winners, and 47.0 % of applications resulted in enrollment leaving a 53.0 % rate of non-compliance (not enrolled) by lottery winners. The non-compliance for lottery losers who entered the lottery kindergarten after losing the lottery was very small, and was mainly due to being admitted under a sibling provision.

The lottery characteristics show wide variation across lotteries. The total lottery applications varied from 62 in lottery 5 to 397 in lottery 4, while the percentages winning the lottery varied from 13.5 % in lottery 13 to 76.8 % in lottery 11. Moreover, the percentage of winning applications that resulted in an enrollment varied from 29.7% in lottery 3 to 70.2% in lottery 2. The number of applications that resulted in an enrollment varied widely from 16 in lottery 5 & 14 to 71 in lottery 4. The diversity in the characteristics of the lotteries likely strengthens the generalizability of the results.

Table 5 presents the data on the two available co-variates (gender and age). Statistically significant differences between lottery winners and losers would imply potential bias in the randomization process. Two of the 14 lotteries (lottery 2 & 3) showed marginally significant differences in the gender distribution, while no lotteries showed any significant differences in the age distribution. This evidence suggests adherence to the randomization process.

Fidelity Data

To determine the extent to which the study schools were implementing the Core Knowledge curriculum, we reviewed school websites, surveyed and interviewed teachers and principals, and observed classrooms in kindergarten through third grade. Key findings were as follows:

- 100% of the teachers reported that Core Knowledge (CK) will “definitely” be a major element in their school’s curriculum in the next several years.
- 100% of teachers reported having a copy of the CK teacher handbook in their grade.
- Depending on the subject, between 86% and 95% of the teachers reported that they planned to teach “all or almost all” of the topics in the Core Knowledge Sequence.
- 91% of the teachers participated in a professional development workshop conducted by the Core Knowledge Foundation.
- Principal interviews suggested that the teachers were enthusiastic about teaching the CK curriculum.
- Based on the 15-minute classroom observations by an expert on Core Knowledge implementation was rated as medium to very high, and the percentage of the curriculum being implemented was between 80-100%.

The evidence suggests that teachers were adhering to the curriculum as developed by the Core Knowledge Foundation, and that students were being exposed to the curriculum’s knowledge and concepts.

Achievement Data

Achievement data in Reading/English-LA and Mathematics from 3^{rd}, 4^{th}, 5^{th} and 6^{th} grades comes from Colorado state-wide PARRC (Partnership for Assessment of Readiness for College and Careers) tests. The PARRC tests were a multi-state effort that built new tests to measure how well students are learning the Common Core standards. The Common Core Standards and associated PARCC tests are designed to test complex "critical thinking" skills. The PARRC English-LA test contains a reading/literacy component as well as a writing component. Scores are combined into one overall score. The Science test is only given at 5^{th} grade and is designed by CTB/McGraw-Hill.

The PARRC Reading/English-LA and Mathematics tests were implemented in Colorado in the 2015-16 school year. Six of our lotteries were in 3^{rd} grade before the change to PARRRC tests, and took an older Colorado Transitional Assessment Program (TCAP) given from 2012-2014 in Reading, Math and Writing that was developed by the Colorado Department of Education and CTB/McGraw-Hill. We have eliminated this 3^{rd} grade data for these six lotteries due to the absence of comparable 3^{rd} grade achievement scores, but have comparable PARRC tests at 4^{th}, 5^{th} and 6^{th} grade for these lotteries.

Research Questions

The confirmatory research question is:

- whether CK-Charter schools teaching the K-8 Core Knowledge curriculum have a long term positive and significant effect on 3^{rd} ,4^{th}, 5^{th} and 6^{th} grade English-LA achievement and combined 3^{rd}-6^{th} grade achievement.

The exploratory questions include:

- whether CK-Charter schools teaching the K-8 Core Knowledge curriculum have a long term positive and significant effect on 5^{th} grade Science achievement and on 3^{rd}, 4^{th}, 5^{th} and 6^{th} grade Mathematics achievement and combined 3^{rd}-6^{th} grade Mathematics achievement.
- whether each of the estimated effects are positive and significant for females and males, and are there significant differences between genders?
- whether the ITT and TOT 3^{rd}-6^{th} grade effects show significance differences between a CK-Charter school in a low income school district and the remaining lotteries in middle/high income school districts

The exploratory research questions on Math and Science effects flow from both the underlying theory of the Core Knowledge curriculum and empirical work with the ECLS-K (Grissmer et al, 2010, Murrah, 2010). The exploratory questions concerning gender differences and differences by school district income level reflect the common finding of such differences in educational outcomes.

Attrition Data

Each lottery applicant's first and last name was used to search Colorado enrollment files to identify the Colorado identification code (CIC) for each student. This CIC code was then used to extract their Colorado state achievement scores for 3rd to 6th grade in Reading/English-LA and Mathematics and 5th grade Science. This matching process identified 7 categories of lottery applicants that were either missing achievement data or had achievement data from a different year than "On-Track" applicants and are all considered part of study attrition. These categories of attrition and the associated assumptions include the following:

- **"No Achievement Data"**-Applicants with no valid 3rd – 6th grade achievement data).
    - **"Attended Private/Home school"**- No enrollment data in K-6th grade
    - **"Moved Out of State/Transfer to Private/Home School"**- Partial or complete enrollment records from K-2rd but no enrollment data from 3rd-6th grade
- **"Not Tested Due to LEP/IEP Exclusion"-** applicants with a valid enrollment record in the 3rd-6th grade but having no achievement data at one of more grades).
- **"Off-Track"-** applicants with enrollment records, but with achievement data from ***a year earlier or later*** than their cohort
    - **"Skippers"-** applicants who skipped a grade between K-3rd grade.
    - **"Retained in Grade"**-applicants who started kindergarten with their cohort, but were retained in grade before 3rd grade
    - **"Delayed Entrants or Red-shirts**"-applicants who started kindergarten a year after their cohort.

Table 6 shows the overall attrition statistics by gender and lottery status for the all-applications" sample. The overall attrition rate for all 10349 applications from 3rd to 6th grade is 35.5%. Lottery winners (32.5%) have lower attrition than lottery losers (37.1%) and a statistically significant level of differential attrition (-4.6***). There are strong gender differences in attrition. Females have lower overall attrition rates (31.3%) than males (36.4%); and females have no significant level of differential attrition (-0.4). In contrast, male differential attrition is highly statistically significant (-7.2***). Significant attrition differences commonly occur in RCTs by income or SES groups, but attrition differences between males and females is an unusual result.

The What Works Clearinghouse (WWC) provides guidelines for analyzing levels of overall and differential attrition to assess the potential risk for bias in results (What Works Clearing House, Standards Manual, 2020). These results suggest that any estimates using the full sample will have moderate bias risk. The male sample will have very high risk for bias, but estimates using the female sample have low bias risk.

Analytic Strategy for Estimating, Presenting and Interpreting Results

This intervention was registered in the Open Science registration in 2017-18 as part of receiving funding from the Arnold Foundation. The evaluation methodology registered at that time did not take account of the possibility of having high and significant levels of differential attrition and the associated bias. That is, the proposed methodology assumed no sources of bias linked to the estimates and proposed the standard RCT methodology for estimation of results. The framework makes researcher bias difficult by ensuring that the primary reported results follow a predetermined methodology using the entire sample. However, this framework does not work well when there is first time, unexpected sources of bias. The bias was unexpected since there were no previous kindergarten lottery evaluations including middle/high income parents. This was the first RCT to encounter this form of bias.

However, the new source of bias primarily affects the male sample. The male sample shows high and significant levels of differential attrition and the associated bias threat. However, the female sample shows no significant level of differential attrition and presents no bias threat. Thus we report all results by gender to highlight the stability of the female results across estimates and the rapid increase in male effects as the sources of differential attrition are eliminated.

Discussion of Analytical Issues

Addressing High and Significant Levels of Differential Attrition in the Male and Full Sample

This study is the first educational lottery analysis to report high levels of differential attrition with statistically significant levels for males and no significant effects for females. To identify potential bias on both the male and full samples linked to these attrition levels, we first examined the causes of this differential attrition. Our approach was twofold. First, we identified the specific parental decisions made by middle/high income parents that could produce high levels of differential attrition. Second, we assessed whether the differential attrition was primarily confined to a few lotteries or particular subgroups that could be eliminated from the estimation.

The presence of high levels of differential attrition appears to be linked to the parental decision-making process in middle/high income school districts. This self-selected group of middle/high income parents who applied to CK-Charter lotteries have a complex decision process. They are choosing where and when to start children in kindergarten and whether to retain or advance students in early grades. There is a wide choice of schooling alternatives in middle/high income school districts in Colorado, especially in suburban areas. These include high-quality public non-charter schools, a variety of charter school types, sectarian and nonsectarian private schools, and home schooling. Colorado is also an open enrollment state, so parents can apply to any public school regardless of location.

Applicants commonly apply to multiple schools. Entering a lottery to a CK-Charter school suggests that parents prefer this alternative to at least one, and perhaps all of their other schooling options pursued. In addition to pursuing private schools or home schooling, these higher-income parents have the option to delay kindergarten entry by a year (red-shirt) and also chose to retain or advance students in a later grade.

In making these choices, parents are increasing attrition rates in our study. Colorado does not test students in private schools or home schools, and no achievement data is available. Students attending Colorado schools and taking achievement tests—but are red-shirted at kindergarten or retained or advanced in a grade from K-3rd—will not have comparable achievement data to "On-Track" students. Students who are delayed or retained or advanced in a grade take different achievement tests in different years and also experience a different intervention than the "On-Track" main sample, which enters kindergarten immediately after the lottery and proceeds from K-6th grade without interruption. Therefore, students who are delayed, retained, or advanced are also considered part of attrition. If these choices are different for lottery winners and losers, differential attrition and associated bias risk can result.

This bias risk appears to be unique to lotteries that include middle/high income parents and only for lotteries at kindergarten entry and only for male applicants. Lower-income parents do not have the option of delayed entry or private/home schooling due to the higher financial and non-financial costs of choosing private or home schooling and/or red-shirting students.

Middle/high income parents in lotteries at higher grades also do not have the option of delayed entry and are less likely to switch to private/home schooling at later grades. Finally, only the male sample shows significant levels of differential attrition.

Analyzing our enrollment and achievement files allows the differential attrition data to be displayed by three sources of attrition- “Off-Track,” “Not Tested,” and “Private/Home/Out of State”. “Off-Track” students have attended Colorado public schools but are in a different grade than their on-track peers and have no comparable achievement data. “Not Tested” students are enrolled in the appropriate grade, but not tested. These students could have been absent the day of the test due to parents that opted out of testing, or were excluded from testing for IEP or LEP reasons. “Private/Home/Out of State” students have no Colorado enrollment records or enrollment records that terminate before 3rd grade. These students are assumed to have enrolled in a private/home school at kindergarten or transferred at later grades or moved out of state.

Table 7 provides the attrition rates for lottery winners and lottery losers by gender for the three major sources of attrition. Females have no statistically significant levels of differential attrition for the Off-Track (1.8), Not Tested (-0.3), or Private/Home/Out of State sample (-1.7) or the entire female sample (-0.1). Therefore the estimated effects for females have no significant bias risk linked to differential attrition.

However, males have statistically significant differences in differential attrition for the Off-Track (-2.3*) and Private/Home/Out of State samples (-5.1**) and Total Sample (-7.2**), but not the “Not- Tested” sample (0.2). The percentage of male students not tested was 7.7%; 8.5% for females. This accounts for about 24% of all attrition. The absence of significant differential attrition for both genders for the “Not Tested” group implies that student absences and policies used to exclude students from testing on the day of the test do not correlate with lottery outcome. The results imply that CK-Charter schools and the public schools attended by students in our sample follow similar procedures in excluding students from testing. In addition, levels of student absences and student characteristics are similar across these schools.

Two other sources account for the differential attrition for males. Parents may decide to delay kindergarten entry or retain or advance a student in grade before 3rd grade. This places the student in the “Off-Track” group. Parents may also decide to use private/home school, either from kindergarten or later, or they might move out of state before 3rd grade.

These parental decisions have been studied in the literature. Bassock and Reardon, 2013, using a national sample, report that delayed entry (red-shirt) rates were about 3-5% of enrollments and rates were much higher rates for higher SES parents, twice as high for males than females, and much higher for younger students. The lower rate of red-shirting for females may indicate that males are behind females their age in kindergarten readiness measures. This result suggests that parents perceive similarly aged males as more problematic than females for school entry decisions (Chatterji, M. 2006; Ready et al, 2005; DiPrete, T. A., & Jennings, J. L. (2013). However, Bassock and Reardon, 2013, also suggest that parental assessment of readiness is more focused on males who are perceived to have marginal levels of readiness for kindergarten entry.

Huang, 2014 suggests that early grade retention is higher for males and strongly linked to being among the youngest at kindergarten entrance. This result suggests middle/higher income parents use grade retention as another strategy (in addition to red-shirting) to give a perceived long-term advantage to a student. The parental decision to advance a child in grade occurs less often than deciding to retain a child in grade. The evidence suggests that males and females have similar rates of grade advancement, which occurs mainly for older students.

Our Colorado enrollment data shows that red-shirting occurred in our sample of middle/high income parents with 5.9 % of males and 2.7% of females entering kindergarten a year later, while no red-shirting occurred in the lottery in the low-income school district. Red-shirts were almost all younger students with birthdays before or close to school cut-off dates. Male red-shirts were predominantly 5.4 years of age or younger, while female red-shirts were 5.2 years of age or younger. Grouping red-shirt students and those retained in early grades shows that 8.6% of males and 4% of females in our sample red-shirted or stayed in grade. Red shirts were predominately younger students.

Our data also shows that 3.2% of males and 1.8% of females skipped grades. These students are almost entirely older. However, males and females are equally likely to be lottery winners or losers, and pose little threat for bias.

The second source of potential bias from attrition includes parents choosing private/home schooling or moving out of state. It seems unlikely that a kindergarten lottery outcome would be correlated with later decisions whether to move out of state. So the differential attrition in this category likely arises from attendance at private/home schools. Some parents who apply in our lotteries may see a CK-Charter school as an alternative to private schools or home schools or as a back-up in case the student is not admitted to private school.

Buddin, 2012, analyzed national data from 2000-2008 and concluded that charter schools in suburban areas drew 68% of their students from public schools, 18% from nonsectarian private schools and 16% from religiously affiliated private schools. This suggests that the lottery applicants in our study could include a mix of parents with varying preferences for regular public schools or private sectarian or nonsectarian schools. In some cases, the CK lottery application could provide a back-up for not being admitted to a preferred public or private school. In other cases, the lottery applicant could prefer a CK-charter school as a first choice, but would enroll in one of the alternate schools if the lottery is lost.

Parents may have a higher propensity to enroll males than females in private/home schools if they lose the lottery. This higher propensity is much stronger in highly populated, suburban areas because of the expanded number of private schooling options,- some of which are tailored toward males.(Long, M.C., & Conger, D., 2013). These private schooling options may include "irregular" private schools that serve students with special needs as well as religious and nonsectarian schools.

Parents whose decisions involve private schools are also more likely to have higher incomes that can fund private school tuition. The average inflation-adjusted tuition in private elementary schools was $12,000 in 2011 with a much higher rate in private nonsectarian schools—$23,000, and the percentage of higher income families choosing a private school in a national sample was approximately 15-20% (Murnane and Reardon, 2018, Murnane, et al 2021). Choosing a private school at kindergarten entry could incur six years of tuition until middle school with approximate costs of $72,000 to $140,000. Winning the lottery may avoid substantial private school costs for some students in this lottery, and the sibling preference allows all children in the family to also attend without entering a lottery.

Not all CK-Charter schools are necessarily substitutes for private schools. CK-Charter schools with outstanding reputations in higher-income school districts might be expected to attract higher-income parents whose alternative would be a private school. This motivation might explain the very large number of applications combined with very low acceptance rates at lotteries 6, 8, 12.13 and 14 (see Table 4). Anecdotally, these lotteries are at CK-Charter schools with well-established reputations in high-income districts and can be seen as viable substitutes for private schools. These higher-income parents may be more likely to send their children to CK-Charter schools if they win the lottery, but to private nonsectarian schools if they lose.

We use two types of sensitivity analyses to address these potential sources of bias and to provide transparency in the analyses of results. The first sensitivity analysis **eliminates 4 of the 14 lotteries** with the highest differential attrition; the second sensitivity analysis also eliminates all young students with early birthdays. The first sensitivity analysis eliminated four of the fourteen lotteries that had the highest level of differential attrition and associated risk for bias. These lotteries were in higher-income school districts with a CK-Charter school having a strong, long term reputation and having among the largest number of applications in our 14 lotteries. Our hypothesis suggests that parents in these districts have the resources to send their children to private schools, but some of these parents view their CK-Charter option as a free, close substitute for a nonsectarian school. The CK-Charter option could save from $70,000 to over $125,000 in tuition over 6 years of attendance. The second sensitivity analysis eliminates all young students in the age window for delayed entry and/or grade retention that includes male students younger than 5.4 years and female students younger than 5.2 years.

A final source of potential bias stems from lotteries with small samples of achievement data from either winners or losers. Small samples can result in highly non-representative achievement levels for lottery winners or losers, which causes bias. Small samples are more likely in two cases: for lottery winners since only about one in three applicants win the lottery and in schools with smaller kindergarten enrollment. Small samples are also more likely when estimating results by gender. We eliminate lotteries in each analysis when the number of lottery winners or losers falls below six achievement scores.

Estimating Single and Multiple Appliers with Achievement Data Across Three Grades

Some schools are located close enough together that parents could apply to more than one study school (multiple appliers). We provide estimates for "single appliers" and also for "all applications" that includes both single appliers and all multiple applications. Estimates using all-applications has significantly more statistical power to detect smaller effects by increasing the sample of students by about 25% and the number of lottery applications by about 56%. The single applier sample can be estimated using the standard methods (Bloom, 1984). A more complex estimation method is needed for the all-application sample that uses multi-way clustering with pooling across grades and applications using a Huber-White adjustment for student-level variance (Wooldridge, 2002).

Non-Compliance

The non-compliance rate for lottery winners is 53.0%. This figure reflects the wide range of schooling choices in our middle/high income school districts. This result required estimating both ITT and TOT effects because TOT effects account for non-compliance. These effects are critical to interpreting policy impacts and in comparing results to other studies. The TOT effects have nearly identical levels of statistical significance as the ITT effects. TOT effects are in the range of 1.8 to 2.2 times larger than ITT effects in this study.

Estimation Methodology

The presence of multiple appliers requires a different estimation methodology than for single appliers. About 20% of students were in more than one lottery, and almost all applied to two lotteries. Our analysis for single appliers uses the standard RCT methodology for estimating ITT and TOT effects (Bloom, 1984: Raudenbush et al, 2012). The analysis including those applying to more than one lottery uses the multi-path clustering approach (Cullen et al, 2006) with each application (as opposed to each student) as the unit of observation and is referred to as the "all-application" sample.

We developed statistical models to identify how access to a CK charter school affected student achievement. For single appliers, the ITT model is estimated using Ordinary Least Squares (OLS) and reflects the impact of being offered a position in one of the CK-Charter schools. This so-called "intent to treat (ITT)" model is

$$A_{ig}=\beta W_{ij}+\gamma X_i+\sum_{ij}\delta_j D_{ij}+\varepsilon_{ijg} \quad (1)$$

where A_{ig} is an achievement score for student i in the gth grade, W is an indicator variable for the ith student winning the jth lottery, X is a set of student characteristics, D is an indicator for the student's application to the jth lottery, and e is a stochastic error term.

Multiple applier students will have multiple records for each achievement score and for each charter school application (Cullen et al. 2006) in Eq. 1, so student-level residuals are likely to be correlated with one another at both the grade- and application-levels. For example, a student's residual in 3rd grade math was unlikely to be independent of their 4th and 5th grade residual. Clustering methods were used to adjust for possible correlations of student residuals across different grades and applications. The adjustment was based on a Huber-White sandwich estimator of student- or applicant-level variance (Wooldridge, 2002).

If winners were required to attend a charter or all winners choose the charter alternative, then β would reflect the relative achievement benefit (or decrement) from attending a charter school. Applicants are not required to attend the charter, however, and many "winning" parents choose another alternative as their circumstances change or they acquire more information on schooling options. As a result, β reflects the average benefit from receiving a charter offer, where the average is comprised of some students in a charter and some in an alternative school.

We extended the model to estimate how actual charter enrollment affected student achievement, i.e., the "treatment on treated (TOT)" estimates. TOT is a two-equation instrumental variable model where charter enrollment (E) is a function of W, X, and D:

$$E_{ij}=\theta W_{ij}+\vartheta X_i+\sum_{ij}\tau j D_{ij}+\varphi_{ij} \quad (2)$$

The second equation estimates achievement (A) as a function of imputed enrollment from Eq. 2, as well as X, and D.

$$A_{ig}=\varphi E_{ij}+\pi X_i+\sum_{ij}\alpha_j D_{ij}+\omega_{ijg} \quad (3)$$

The TOT effect is φ, the effect of expected charter enrollment on achievement.

The ITT and TOT models were estimated separately by subject in grades 3 through 6 for English-LA and Mathematics as well as for a Colorado 5th grade science test. Several schools had lotteries in both cohorts, and these lotteries were treated separately for each cohort. The models also included demographic controls for grade, gender, race/ethnicity, and free/reduce lunch eligibility. Race/ethnicity and free/reduce lunch eligibility is only available for students who had at least one year of enrollment in Colorado schools. Achievement scores are standardized by grade and cohort.

Our TOT estimates use CK charter enrollment in kindergarten in the year of application as the enrollment variable. Additional models were estimated where enrollment was characterized as charter enrollment in any year or total years enrolled in a charter. These results were similar to those reported below.

RESULTS

We provide detailed results for the 'All applications" sample and both genders. The All-Application sample has the largest sample and highest power to detect effects. Both gender results are also included due to the higher potential for bias in the male sample. We also present a comparison of effects from single appliers with the "all applications" sample. Results showed slightly larger effects for the single applier sample.

Differential Attrition Characteristics of the Samples

Table 8 summarizes the sample sizes and the overall and differential attrition for the three estimated samples that include all applications. The three samples are the full sample (Full), the sample that eliminates the four lotteries with the highest differential attrition (Lotteries left out-L-LO), and the sample that, in addition, eliminates young students (L-LO & YS-LO). Table 8 shows that the high level of differential attrition for males in the full sample (-7.0***) is reduced in the L-LO sample (-3.4^{+}) and eliminated in (L-LO&YS-LO) when young students are also left out (-0.0). The L-LO & YS-LO sample also eliminates differential attrition (-0.01) in the full sample. Thus while the Full sample and L-LO sample have significant bias risk, the L-LO&YS-LO sample has little risk for bias. Each estimate of effects below includes estimates for the Full sample, the L-LO sample and the L-LO&YS-LO sample to assess the extent and direction of potential bias.

These results suggest that the sources of differential attrition and the associated bias threat arise from two groups: (1) the four lotteries that have the largest differential attrition and, (2) all younger, primarily male students with early birthdays who parents often choose to delay entry or retain in grade. The former group are located predominately in high-income school districts that have CK-Charter schools with reputations that parents likely perceive as substitutes for private sectarian schools. In both situations, some parents who win the lottery may enroll in a CK-Charter school; but to hedge against losing, they may also apply to a private school or red-shirt or later retain a student if they lose the lottery. In the first case, parents will avoid the cost of private sectarian tuition. In the second case, parents who delay entry have another opportunity to win the lottery in the next year or can retain a child at a later grade. Eliminating differential attrition from the estimated sample reduces the bias threat, but the direction of the bias and the associated changes in estimated effects cannot be predicted in advance. Whether effects weaken or strengthen is an empirical question that we address by comparing results across the three sample estimates.

Effect Estimates

The confirmatory hypothesis is that CK-Charter schools will have long-term, positive significant effects on English-LA achievement. We estimated intent-to-treat ITT results (Table 9) using achievement data from all grades as well as comparable estimates for grades 4, 5 and 6.

- The estimates using combined achievement from all grades show statistically significant effects that increase from the FULL sample (0.143**) to the L-LO sample (0.179**) to the L-LO&YS-LO sample (0.241***).
- The effects by grade show similar trends: statistically significant or marginally significant effects for each grade with similar trends that increase from the FULL sample to the L-LO sample to the L-LO&YS-LO sample. For instance, 6th grade effects increase from the FULL sample (0.114^{+}) to the L-LO sample (0.128^{+}) to the L-LO&YS-LO sample (0.208*).
- The estimates for the L-LO&YS-LO sample with no differential attrition show statistically significant, small size ITT effects at all grades (0.241***) and at 4th grade (0.196**), 5th grade (0.281**) and 6th grade (0.208*).

These results show a counterintuitive trend. As the sample declines from 6652 in the FULL sample to 4949 in L-LO sample to 4027 in the L-LO&YS-LO sample, the effect size increases and has stronger significance. These results suggest parental decision-making in middle/high income school districts may introduce a downward bias in kindergarten lotteries from parents who red-shirted or retain a student in grade and from higher income parents who can afford private schools but see a particular type of charter school as a close substitute. Since males are behind females in school readiness, much of the parental concern and bias is focused on males

Table 10 shows results by gender. Female effects show statistically significant effects for all three samples with small increases from the Full sample (0.223**) to the L-LO sample (0.242**) to the L-LO&YS-LO sample (0.267**). Since differential attrition was at low and nonsignificant levels in all three samples for females, the similarity in the effect size and significance might be expected.

However, the male results show a different pattern. We find insignificant positive effects for the Full sample (0.063) and the L-LO sample (0.068), and marginally significant effects for the L-LO&YS-LO sample (0.207^{+}). The somewhat weaker size and significance of male effects may reflect that the higher differential attrition for males (Table 8) only disappears for the L-LO&YS-LO sample. The reduced size of the male sample (1514) compared to females (2556) may also weaken male effects relative to female effects.

Table 11 compares estimated 3rd-6th grade effects for single applier students to the effects estimated for all applications that include students applying to more than one school. The single applier effects are statistically significant for all samples and show similar increasing effects as lotteries and young students are eliminated. The single applier effects show similar or modestly higher estimates for the FULL sample (0.155* vs. 0.143**), the L-LO sample (0.199* vs. 0.179**) and the L-LO&YS-LO sample (0.306** vs. 0.241**).

Table 12 compares both ITT and TOT estimates for all grades (3rd-6th), and for 4th, 5th and 6th grades for the L-LO&YS-LO sample. The TOT effects are all approximately twice as large as the ITT effects due to the 52% compliance rate. The TOT effects are statistically significant and moderate in size for the estimates, including all grades (0.473***), 4th grade (0.383**), 5th grade (0.543***) and 6th grade (0.404*). There are no upward or downward trends by grade suggesting the effects may have stabilized by fourth grade.

Table 13 provides the exploratory ITT and TOT effects for Mathematics and Science with effects also by gender for the L-LO&YS-LO sample. The science effects measured at 5th grade show statistically significant ITT effects (0.154*) for the all-gender sample, with marginally significant effects (0.184^{+}) for females and positive, insignificant effects for males (0.083). The results for mathematics show insignificant, positive ITT effects for female (0.146) and male (0.003) results and for the all- gender (0.081) sample. The TOT effects for females in math (0.273) and science (0.339^{+}) are much larger than the corresponding effects for males in math (0.006) and science (0.175), but the differences are not statistically significant.

Table 14 compares ITT effects for a single lottery for a CK-Charter school in a low income school district to the remaining 13 lotteries in middle to high income school districts. The ITT effects for the CK-Charter in a low income school district are statistically significant and large to very large in English-LA (0.944**) and Mathematics (0.735*), and positive, but insignificant for Science (0.468). The lotteries in middle/high income school districts show statistically significant effects in English-LA (0.201**), and positive, insignificant effects for Mathematics

(0.041) and Science (0.125). The effect differences between the two types of school districts are statistically significant for English-LA and Mathematics, but not Science.

Table 15 shows the more policy-relevant TOT effects comparing results from the low income school district to the middle/high income schools districts. The data show very large, statistically significant effects for the CK-Charter in a low income school district in English-LA (1.299**), and Mathematics (0.997*), and positive, bur insignificant effects in Science (0.622). The corresponding TOT effects for the schools in middle/high income school districts (English-LA (0.445**), Mathematics (0.090) and Science (0.270)) show statistically significant differences compared to the low-income charter in English-LA and Mathematics, but not Science.

Overall, these effects are large enough to eliminate achievement gaps between advantaged and disadvantaged students in all three subjects. This long-term intervention that changed curriculum from K-6th grade shows no achievement gaps at 3rd-6th grade in English/Proficiency or Math or Science at 5th grade between low income students and middle/high income students.

Table 16 summarizes the 3rd – 6th grade TOT results by subject for the total sample and by gender with the estimated percentile gain across subjects and genders. These results show that the percentile gains estimated from the TOT effects for the all-applications sample were statistically significant for Reading/English-LA (16.1 percentile points), Science (10.2 percentile points) and positive, but insignificant for Mathematics (5.4 percentile points). Female results showed statistically significant TOT gains for Reading/English-LA (17.0 percentile points), marginally significant gains for Science (11.5 percentile points) and positive, but insignificant for Mathematics (9.3 percentile points). Male results showed marginally significant TOT gains for Reading/English-LA (15.0 percentile points), marginally significant gains for Science (6.0 percentile points) and positive, but insignificant for Mathematics (0.00 percentile points).

Discussion

Summary

There has been substantial **non-experimental** evidence linking gains in measures of General Knowledge to later achievement in Reading/English-LA, Science and Mathematics (Claessens et, 2009, Duncan et al, 2007; Duncan et al, 2020, Grissmer et al, 2010;). This evidence suggests that gains in General Knowledge would have a larger effect on future achievement than similar gains in the more widely studied non-cognitive skills including executive function, visuo-spatial/fine motor and socio-emotional skills. However, the lack of experimental evidence for interventions directed at changing levels of General Knowledge has left improving non-cognitive skills as one of the best viable options for improving achievement even though it has proven challenging to design interventions using non-cognitive skills that raise later achievement.

The present study provides the **first experimental evidence** that suggests that a curriculum (Core Knowledge Curriculum) directed toward building General Knowledge from kindergarten to 8th grade leads to long term achievement gains. These results open a new category of interventions that build General Knowledge with potential effects predicted to be larger than interventions that target non-cognitive skills.

The current results (see Tables 15 & 16) show:

- statistically significant, moderate size, long-term TOT achievement gains (0.473***) from 3rd-6th grade in Reading/English-LA for the entire sample of students and schools spanning low to high income characteristics.
- statistically significant, small-size TOT achievement gains (0.300*) in 5th Grade Science for the entire sample of students and schools spanning low to high income characteristics.
- a small positive, but insignificant, TOT gain (0.159) in Mathematics for a sample of students and schools spanning low to high income characteristics.
- Large to very large, statistically significant, achievement gains in Reading/English-LA (1.299**) and Mathematics (0.997*) and moderate, positive, but insignificant Science effects (0.622) for a school in a low income school district that eliminated achievement score gaps in 3rd-6th grade in Reading/English-LA and Mathematics.

The evidence would suggest that the level of General Knowledge may be a critical, largely unmeasured, cognitive characteristic that may help explain the factors underlying achievement for students from all income levels as well as accounting for current achievement score gaps between advantaged and disadvantaged students. Moreover, this study suggests that the level of General Knowledge is malleable, and an intervention that increases General Knowledge may increase achievement for students from all family income groups with much larger effects that eliminate achievement gaps for disadvantaged students. The much larger effects for a school in a low income school district may simply reflect the greater opportunity that students from higher income homes have to acquire General Knowledge outside of school. However, it remains surprising that the academic disadvantages associated with students from lower income families may be largely reflected in their lower level of General Knowledge, and that the level of General Knowledge is malleable, and that a school curriculum from K-6th could ameliorate the differences in General Knowledge and close achievement gaps.

The size of the long term TOT effects for Reading/English-LA (~ 16 percentile points) in this intervention is approximately equal to the difference in the achievement gains over the last 30-40 years between Reading/English-LA (7 percentile points) and Mathematics (25 percentile points (Shakeel, M. D. & Petersen, P., E., 2021). The size of these 16 percentile gains could also close the international gap in Reading/English-LA for U.S. students. U.S. students placed 15th among 50 countries in the 2016 PIRLS 4th grade Reading/English test, but national student gains similar to gains in this intervention would place the U.S. among the top five countries (PIRLS, 2016).

This study may be the first experimental intervention that shows statistically significant ITT and TOT effects that improve long term achievement for students from all income groups. Conolly et al, (2018) has identified 1017 RCT's evaluating educational interventions in the 1980-2016 time frame, and the WWC has maintained a data-base of published results for experimental and quasi-experimental evaluations. Chabrier et al, 2016 has also summarized the results of high school, middle school and elementary school/kindergarten lottery based RCT's utilizing state achievement as outcomes. Few interventions in this universe are directed to raise achievement for students in all income groups and/or implement interventions that last for four or more years and/or measure long term effects, and none that have all of these characteristics.

Kraft, 2020 and Hill et al, 2008 characterizes how the size of educational intervention effects should be viewed that differs from the traditional view that labels small effects (~.25 SD), medium effects (~.50 SD and large effects (.75 SD). This new characterization recognizes that the size of actual intervention measurements are commonly insignificant, and those showing significant effects almost always lie in the range less than .25 SD for ITT effects. Unfortunately, this literature does not provide comparisons of TOT effects- which are often unreported. Using TOT effects provides better predictions of the effects for a student who actually experienced the intervention. The TOT effects in this intervention (.47 SD) would likely be in the range of the largest effects measured in previous interventions.

The absence of previous interventions with similar characteristics leaves open the question of whether the effects arise partly from the longevity of the intervention as opposed to the specific causative mechanisms arising from the Core Knowledge curriculum. The question becomes whether there are causative factors inherent in the intervention that might explain the effects.

Exploring Potential Causative Mechanisms

The Core Knowledge curriculum has many similarities and areas of agreement with more standard curriculum. For instance, the two curriculums do not differ on incorporating similar methods of building the early reading related skills associated with phonemic awareness and phonics. Both curriculum require teaching the subjects of Mathematics, Reading/English-LA, Geographical, History, Science and Mathematics. Core Knowledge incorporates a Mathematics Curriculum, but also allows schools to choose other Mathematics curriculum.

However, there are major differences in the Core Knowledge curriculum from the curriculum taught in typical public schools that might help account for the results. **The Core Knowledge curriculum is directed toward b**uilding accumulative knowledge which requires a reconceptualization of the teaching of all subjects and the time devoted to each, and unlike almost all previous reading comprehension interventions, is not simply an instructional change during the Reading/English proficiency part of the reading curriculum.

Darling-Hammond et al, 2015, Darling-Hammond et al, 2020, Osher et al, 2020 and Cantor, 2019 provide comprehensive syntheses of the research involving learning and development that includes classroom practices that have evidence for improved short and long term student learning. The Core Knowledge curriculum has several characteristics identified by these syntheses that might help account for the achievement gains, and should be the focus of future mixed methods research in the classroom. These include:

- Curriculum that takes advantage of and enhances a student's existing knowledge about the world they live in (Barron et al, 2015; Willingham, 2003)
- Building knowledge through more emphasis on History, the Social Sciences and the Arts (Elleman & Osmond, 2019)
- An early and sustained focus on developing background knowledge (Ellerman & Osland, 2019; Willingham, 2003; Willingham, 2006)

- Learning the unique structure, the particular modes of inquiry and different types of text analysis that are unique to each subject taught (Ellerman & Osmond, 2019; Goldman et al, 2016; Shanahan & Shanahan, 2008).
- Combining explicit instruction organized around a conceptual map or schema (Kim et al. 2021; Kim et al, 2023, Core Knowledge, 2010)
- Learning and taking advantage of each student's interest (Hidi et al, 2017)
- Greater efficiency from building on knowledge learned in previous grades and avoiding unnecessary content repetition (Engel et al., 2013).
- Reductions in cognitive load from a well-designed and integrated curriculum across all subjects and all K-8 grades (Paas & van Merrienboer, 2020; Engel et al., 2013; Willingham, 2006; Elleman & Osmond, 2019).

The kind of integration and focus across all grades and subjects that characterizes the Core Knowledge curriculum is not typically prioritized in previous literacy interventions or in school districts or states where, typically, two subjects (Mathematics and Reading/English-LA) are given the highest priority and little integration occurs across subjects and grades.

These experimental results also directly address the three decades debate about the causative mechanisms underlying reading and verbal comprehension (Hirsh et al, 1988, Hirsch, 2003, Hirsch, 2006; Willingham, 2006, Hirsch, 2011; Willingham, 2017, Wexler, 2019). This debate involved whether the causative mechanism involved in increasing reading comprehension is due mainly to increasing the level of previous General Knowledge or to the acquisition of "procedural skills" that enable comprehension. This debate could not be settled due to the absence of definitive, long term experimental research testing both hypotheses. The results from this study would provide the first experimental evidence suggesting that building General Knowledge leads to higher Reading/English-LA achievement. While there is no long term experimental evidence on the effects of improving "procedural skills", this approach has been the major focus to improving reading comprehension of Reading Panels and researchers for 30-40 years.

During this period, the major reading panels convened to recommend policies to improve reading comprehension focused primarily on improving the productivity of the time spent in Reading/English-LA instruction by teaching improved "procedural skills" to facilitate better comprehension. While acknowledging that the level of General Knowledge was an important factor in comprehension, the series of panel recommendations over this extended period made no recommendations for the kind of dramatic policy changes that would be needed to enable significant improvements in General Knowledge; i.e, shifting the curriculum to much more time on Science and Social Science vs. time on Mathematics and Reading/English-LA instruction.

Rather, the panels and researchers primarily confined their search for interventions to the quality and type of instruction provided during the classroom time devoted to Reading/English-LA instruction. Improving this instruction was almost entirely directed toward improving the "interpretative skills" of students. While the role of General Knowledge was generally acknowledged to be an important factor in comprehension, there was no strategy or intervention identified in the various reading panel reports that was directed toward dramatic changes to improve a student's General Knowledge. There has also been little change over 40 years on the time spent in different subjects in elementary grades (Morton, B., & Dalton, B., 2007; Hoyer, K., M., & Sparks, D., 2017; Perie et al, 1997). For instance, teacher reported weekly time spent at 3rd grade was 9.9 hours on Reading/English-LA, 5.8 hours on Mathematics, 2.9 hours on Science and 2.8 hours on Social Science.

This pattern of stable time usage over 30-40 years implies no dramatic changes in the time used to teach Reading/English-LA and rules out the type of curriculum changes required to implement an instructional strategy directed toward building reading comprehension through increasing General Knowledge. Rather, the large amount of classroom time devoted to Reading/English-LA is largely accounted for by teaching the early stages of reading comprehension (phonemic awareness, phonics, vocabulary). The achievement gains in Reading/English-LA of 7 percentile points over the 30-40 year period likely reflects the results of increasing the effectiveness on these early reading improvements, but there is little evidence that any dramatic changes occurred to increase reading comprehension.

The current results suggest that the "procedural skills" approach to teaching reading comprehension that has dominated reading comprehension instruction over the last 30 years in public schools is less effective than a "knowledge-based" approach that places cumulative General Knowledge as the main mechanism for increasing comprehension. Our conjecture is that the failure to significantly increase long-term English-LA achievement lies in the long term assumption that reading and verbal comprehension- the final phase of learning to read- is achieved by activities and subjects that attempt to teach students "procedural skills" rather than activities and subjects that enable them to increase their General Knowledge. Multiple models of reading comprehension (Cromley & Azevedo,2007;Talwar et al., 2018; O'Connor et al., 2017; Kim et al., 2017; Mol & Bus, 2011; Neuman et al., 2011; Reid et al, 2021: Shanahan et al., 2010) suggest that vocabulary and background knowledge (e.g., science and social studies topics) are essential to improved reading comprehension. If the current results prove replicable, the future theory underlying the "Science of Reading" will incorporate the acquisition of cumulative General Knowledge as a strong causative mechanism in later verbal and reading comprehension.

An intervention focused on increasing General Knowledge to improve reading comprehension appears to set off an unusual long term, compounding process whereby improved reading comprehension leads to increased knowledge, and increased knowledge leads to even better comprehension, leading to more increases in knowledge, etc. This compounding process would not only occur in Reading/English-LA linked to instruction, but across all subjects to the extent that they depend primarily on reading comprehension for learning. So achievement gains would likely spread across nearly all subjects. Moreover, these achievement gains in subjects would likely extend into future years as increased comprehension in one year leads to increased knowledge and comprehension in the next year, leading to even longer term gains. These cascading achievement effects across all subjects and over time would likely also increase years of educational attainment and future labor market success.

However, elevating General Knowledge to a more central place and higher priority in research and policy will require a significant conceptual shift from current impressions and understanding of the term "General Knowledge" as well as new research directions that aid in better understanding and articulating the role of General Knowledge in cognitive development. The term, "Building General Knowledge" does not readily trigger a conceptual map linking the intervention to higher achievement that occurs when other common interventions such as reducing class size, extending the school day, and raising teacher pay are considered.

Elucidating the possible causative links between increasing early General Knowledge and higher later achievement will be a necessary step in building this conceptual map. However, the first step is to provide a different conceptual understanding of the term "General Knowledge" and what it measures.

Understanding "General Knowledge"

The difficulty in conceptually linking General Knowledge to achievement may arise partly from the methodology of measuring General Knowledge that asks students a series of seemingly "simple" questions across a wide variety of topic domains. However, what makes a question simple and the answer remembered may be the presence of more in-depth knowledge and schema in a given domain (Willingham, 2003). That is, a student may be able to answer a "simple" question about baseball- how many outs in an inning- because he has experience and/or more in-depth knowledge from playing or watching or tracking baseball games. If a student has little knowledge of baseball, it is unlikely to either come across and/or remember any reference to baseball having three outs in each inning.

Measures of General Knowledge would then measure the number of domains in which a student has developed some in-depth/critical knowledge and understanding. Thus, measures of General Knowledge might then reflect both the breadth and depth of knowledge across all domains of a student's knowledge. Thus the term "Total Knowledge" may more accurately convey the meaning of "General Knowledge" and also provide a possible explanation of why it may be the single most important predictor of later achievement across all subjects and may also be the single best predictor of years of educational attainment, future wages and productivity.

Currently the measures that are used to characterize the overall status of educational progress and used to predict long term outcomes include early Mathematics and Reading/English-LA achievement. This emphasis leads to the identification and research on non-cognitive characteristics like executive function, socio-emotional and visuo-spatial skills as targets of interventions to improve achievement. However, this approach leaves out what may be the single most important variable linked to future achievement in all subjects and possibly later life outcomes - the students level of General Knowledge.

The level of a students General Knowledge is highly correlated with the more traditional measures of SES, parental education and income often used to predict future achievement and account for achievement score gaps. This correlation is likely linked to the greater opportunities in higher income and SES families to accumulate General Knowledge (Lareau, 2011). This level of General Knowledge may underlie the power of SES measures to predict a range of future outcomes. However, unlike such SES measures, General Knowledge is malleable and interventions that raise students General Knowledge may be the single most effective way to increase later achievement, close achievement gaps and raise other long term outcomes.

A recent study adds a new dimension to the role of General Knowledge in raising later achievement. Jirout et al, 2022 suggest that students early attitudes toward school and learning (curiosity, enjoyment of schooling) in pre-k and 1st grade is bi-directionally linked to their level of General Knowledge. Thus expanding early General Knowledge that links to students curiosity about the world they experience may play a pivotal role in shaping students attitudes, motivation and satisfaction with early schooling and possibly extend into later schooling.

It is hard to identify a more central and important cognitive learning capacity than continually being able to comprehend more and more of what is read. Such continually increasing comprehension appears to be built upon an increasing accumulation of General Knowledge. Success in this process not only affects all future learning, but also builds self-confidence, motivation and social connections. This capacity appears to lie at the heart of individual cognitive and social development.

Future Research Implications

Future research on the Core Knowledge curriculum needs a broad focus that:

- Identifies opportunities for replication by identifying oversubscribed charter schools teaching the Core Knowledge curriculum
- Uses historical empirical research to estimate whether there are significant differences in achievement in Core Knowledge schools vs. schools with similar characteristics that do not teach Core Knowledge.
- Designs newly implemented state or district level RCTs that randomly assign the Core Knowledge curriculum to schools using mixed methods data collections to discover causative mechanisms including classroom observations and surveys of parents, teachers and students.,
- Develops and refines specific causative hypotheses that might account for the results of this study and establish links to established theories and empirical research.
    - Identifies through case studies and classroom observations the differences in time spent on subjects and in student and teacher interactive behavior by subject and grade between schools teaching Core Knowledge and similar schools not teaching Core Knowledge
    - Integrates these results into the broader literature that includes Darling-Hammond et al, 2015, Darling-Hammond et al, 2020, Osher et al, 2020,

Cantor, 2019, Hanushek et al, 2019, Pellegrino, 2012 and Pellegrino, & Hilton, 2013; Schneider and McGrew, 2012).

- Addresses whether the Core Knowledge curriculum raises achievement for students from all family income levels with much larger effects for low income students.

These experimental results together with the extensive non-experimental evidence about General Knowledge seems sufficient to initiate a large scale research effort aimed at replication of results and better understanding of the causative mechanisms underlying the effects. The results from a single intervention and evaluation can never provide sufficient evidence for achieving a new longer term research or policy consensus among researchers or policymakers or for understanding of the causative mechanisms. However, unexpected experimental results often suggest that current theories underlying learning and educational interventions need to be examined to assess how to accommodate the new experimental results. In the long term, it is stronger theories and increased understanding of the causative mechanisms that predict the results of new experimental evidence that moves science forward.

Issues in Implementation

State and local policymakers will be the primary decision-makers involving implementation of Core Knowledge in Charter schools and/or in public schools. When making such decisions, a number of factors are involved besides the size of the expected impact and its uncertainty. RCTs that produce significant effects that can be replicated often pose a significant challenge to maintaining effects when implemented in the real world. Implementing new interventions from small scale RCTs often carries risk from decaying effects in the longer term, limited generalizability, lack of fidelity in transferring the intervention from the lab to widespread implementation in schools and uncertain costs.

One risk in implementing interventions that have shown significant effects in RCTs is that measurements of effects have only been made over short terms, and long term effects may decay (Guerrero-Rosada, et al, 2021; Kruger, 2011; Dick et al, 2019; Romero et al, 2021; Bailey et al, 2017: Bailey et al, 2020). However, this study measured the long- term effects of the Core Knowledge intervention implemented from K-6 through 3^{rd}-6^{th} grade, and results remained significant through 6^{th} grade.

Another implementation risk is that later replications would significantly reduce the size and significance of effects, as has occurred in other interventions (Ioannidis et al, 2017; Kirkham, et al, 2010; Dwan et al, 2013, Shah et al, 2020). While this risk is a possibility, not all RCTs have equal replication risk. Failure to replicate results is probably more likely when results are short term, sample sizes and/or effects are small, the statistical significance is in the marginal $p < .05$ range, and the intervention is research-based and not yet widely implemented making fidelity problematical. In this study, the replication risk is likely reduced by four factors: (1) the size of current TOT effects in the moderate size range, (2) their strong statistical significance ($p < .005$) due partly to a very large sample size together with their measurement in the long term over four grades, and (3) the widespread, long term implementation of the intervention in over 700 schools nation-wide and, (4) an associated professional development infrastructure exists to aid in implementation.

Finally, implementing new interventions from small scale RCTs often carries risk from limited generalizability and uncertain costs. This intervention included students from all income levels with different sizes of significant effects for low vs. middle/high income students that limits the risk from generalizability. Finally, the long-term cost of the intervention is low since the marginal costs involve mainly the professional development expenditures involving the Core Knowledge curriculum.

Policymakers should consider future implementations that combine the Core Knowledge curriculum intervention with other types of non-curriculum interventions that have experimental evidence for increasing achievement. Interventions that have measured both Mathematics and Reading outcomes have almost always shown larger effects in Mathematics then in Reading/English-LA (see, for instance, Chabrier et al, 2016). Thus, combining Core Knowledge with the non-curriculum intervention, "No Excuses", in KIPP schools would combine two interventions with experimental evidence that may better address eliminating achievement gaps in both Mathematics and Reading/English-LA.

Implications for Educational and Social Policy

Federal data collections are designed to monitor and better understand the most important economic, educational, and social trends in society. This paper has suggested that the primary measures currently collected to monitor education including measures of Mathematics and Reading/English-LA (as well as other subjects) do not capture an important aspect of learning- namely the level of General Knowledge. Well-designed measures of General Knowledge should be considered as an important addition to our routinely collected national measures for students in elementary grades. However, designing nationally collected measures of General Knowledge will pose a substantial challenge for researchers and policymakers, not unlike the challenge of measuring the Gross National Product as an economic indicator.

However, measures of General Knowledge will carry an additional challenge. Characterizing and measuring the General Knowledge that young students have in lower elementary grades will need not only scientific validity, but also political viability. There may be differences among groups of adults that vary by SES and/or cultural characteristics about what General Knowledge may be important for children to have to enable higher later achievement and other educational objectives. Attempting to define such General Knowledge will undoubtably trigger debates and a variety of viewpoints. But characterizing the General Knowledge that is needed to better understand the books actually read by children and the textbooks used in future education seems essential to educational efficiency and meeting long term goals. What seems essential, and less controversial, is to focus on linking the gains in early General Knowledge (however defined) to later achievement-

Finding a lack of progress over the last 40 years in raising long-term Reading/English-LA achievement is not an unusual pattern in fields of scientific research. Fallow periods are the rule rather than the exception, as are periods of very rapid progress, often triggered by unexpected experimental results, new experimental technologies, or new theories. The lack of more rapid progress in Reading/English-LA achievement cannot be assigned to poor quality research or failing to adhere to a scientific approach in research or the lack of research funding. Rather the primary problem in educational research is that the field of inquiry is one of the most difficult and challenging for the scientific method to address. Slow scientific progress is to be expected with the presence of a multitude of forces that can influence outcomes in a non-linear and interactive manner, together with the great difficulty in producing definitive experimentation with younger subjects.

Accumulating General Knowledge leading to better understanding of a student's experience in the world is certainly an unmeasured by-product of current curriculum and educational and social policies. However, the current results suggest that explicitly making building General Knowledge one of the primary objectives of early development and elementary education may lead to higher long-term achievement across all subjects. Currently, the primary measures used to characterize the performance of K-12 students are achievement scores in Mathematics and Reading/English-LA. These subjects and associated trend measures are used in policy to broadly characterize the quality and performance of students and of the education system leading to their prioritization in the curriculum.

The results of this study would suggest that this prioritization of the subjects of Mathematics and Reading/English-LA fails to adequately capture a critical measure of a student's cognitive development- namely their General Knowledge of the world. The level of General Knowledge may be the single best indicator of a wide range of future outcomes. The absence of this measure in educational and social science research and policy may have high opportunity costs from lower achievement for all students and larger achievement gaps between advantaged and disadvantaged students. Assuming future replication will support the current evidence, priority should be given to **making measures of General Knowledge a central objective of educational and social policy and future data collection.**

A significant interpretive weakness in previous research may be embedded in the theory and language involving the economic theory of human capital that places the emphasis on "skill building" as the primary developmental cognitive process involved in learning as described in Bailey et al, 2017. This theory characterizes development as the process of building increasingly complex skills summarized in the phrase "skill begets skill" (Heckman, 2006).

However, the results of this study would suggest that there are two separate, but complementary, cognitive processes involved in development and learning: skill building and General Knowledge accumulation. So, in addition to characterizing learning and development as a process of "skill begets skill", "knowledge begets knowledge" must be added as a critical process. And it would not be surprising to eventually find that "skill x knowledge begets skill x knowledge. Building skills and building knowledge both seem essential to learning, and are likely interactive and our research and policy frameworks should change to incorporate both cognitive processes.

Caveats

This RCT had no pre-test data. Ideally pre-test data can provide evidence for significant deviations from perfect randomization and also information to assess bias risk especially for smaller sample RCTs. However, the value of pre-tests depends on the quality and reliability of the measures and the attrition rate in the pre-tests. For kindergarten-based lotteries that have much later school achievement tests as outcome measures, any earlier pre-test would have to be administered outside schooling during a narrow window between the lottery application and the actual lottery after obtaining parent permission. Such measures would have less reliability and lower response rates than the state administered measures of achievement partly due to the younger age of students, more limited testing time and out-of-school administration. Pre-tests under these circumstances would appear to add little to the quality of the effect estimates. These limitations will always be present in experimental studies using kindergarten lotteries with later state collected achievement data as outcome measures.

The results in this study will generalize only to the types of CK-Charter schools that were selected for inclusion in the study. These CK-Charter schools had been in operation for over four years, spanned all income groups and were oversubscribed. Being oversubscribed is a limitation in all kindergarten lottery-based experimental studies. The characteristics of oversubscribed schools may differ from schools that were not over-subscribed. For instance, the over-subscribed schools may have characteristics better known by local parents that reflect their higher achievement in ways not linked to Core Knowledge. If the oversubscribed schools were similar to one another, the threat for bias and limited generalization could be significant. However, the oversubscribed lotteries in our sample have a wide diversity in characteristics including amount of oversubscription, size of school, location of school, sample size of winners and losers, percentage of lottery winners, size of queue and characteristics of the school districts and parents. Future experimental studies that rely on different methods for establishing randomization that do not exclude oversubscribed schools will be needed to assess whether effects differ between such schools.

The effects measured in the low income school district depend on a single lottery and a relatively small lottery sample size of 62. Although the results are highly significant, these results carry much more replication risk than do the main results from all lotteries. Also, the cause of the effect may be different than for the schools in middle/high income districts. For example, the public school alternatives in the low-income district may have weaker relative performance than the alternatives in middle/high income districts. For these reasons, it will be essential to replicate these effects in low income school districts.

Another important question is the extent to which these effects can be attributed to the Core Knowledge curriculum as opposed to being implemented in a charter school. However, the evidence reviewed in the study suggests that being a charter schools alone does not predict higher achievement outside of inner city charter schools, leaving Core Knowledge as a likely cause of the effects in suburban schools. But much more replication is needed utilizing the charter school methodology in this study.

Currently, the Core Knowledge curriculum is implemented widely in both charter and regular public schools suggesting the possible absence of substantial implementation issues that would differentially impact the effectiveness of Core Knowledge in regular and charter schools. However, the study provides no experimental evidence that the Core Knowledge curriculum in regular public schools would produce similar effects. There may be differences between charter and public schools that might lead to differences in impacts of the Core Knowledge curriculum. For instance, teachers may play a critical role in successfully implementing and taking advantage of the Core Knowledge curriculum – and any differences between charter and regular public schools in attracting, retaining and enabling high quality teachers could account for part of the measured effects.

An important consideration for policymakers deciding whether to implement the Core Knowledge curriculum is that the evidence shows achievement effects on 3rd-6th grade students only for students who have experienced Core Knowledge from kindergarten through 3rd-6th grade. Implementation in early grades may be critically important since early knowledge building appears to be a critical element. However, implementation at later grades or for shorter periods in early grades would be problematical. Finally, measurable effects may require experiencing Core Knowledge for several years, i.e., measuring no effects after one or two years may not test whether longer term effects exist.

Researcher bias is a potential threat whenever cited results are based on samples that eliminate a significant fraction of total observations. This study certainly carries that threat since some of the cited results are based on samples that have eliminated over one-third of the observations. Moreover the sources of the missing observations (attrition) cannot be accounted for by the often cited random non-response, but rather primarily occur due to parents making alternative schooling decisions for their children at kindergarten entry or between kindergarten entry and third grade. These decisions to delay entry, retain in-grade or attend private or non-sectarian schools prevented students from taking state administered achievement tests between 3rd-6th grade that were comparable across the entire sample. And these decisions caused both non-random attrition as well as attrition differences between lottery winners and losers, making using the full sample for estimation potentially biased.

However, the threat of researcher bias is not present for effect estimates for females, but only for estimates that include males. The absence of researcher bias for female estimates is due to the lower, insignificant level of female attrition and the absence of differential attrition in the female sample (see Table 6). This absence of the potential for researcher bias in female results is indicated by the similarity of effect estimates across the three estimated samples for females (see Table 8). On the other hand, the male estimates using the three samples have results that range from null effects in the first two samples to a statistically significant effect in the third sample (see Table 8).

There are two different interpretations of these different gender results. The first interpretation from the set of results with the full sample is that the intervention worked for females, but not for males. However, the interpretation from the results from the third sample would be that the results for the full sample of males is biased downward, but removing the bias through elimination of observations that cause high levels of differential attrition makes effects for both genders significant and similar in magnitude.

The theory underlying at least part of these gender differences in bias potential (articulated in the report) is that females have higher levels of parent perceived levels of school readiness than similarly aged males. This gender difference in school readiness leads to parent preferences to delay entry for some younger males. However, the decision to delay entry can be made more complicated for those winning the lottery- an opportunity that may not be present in the next year. Thus the data would suggest that some parents who win the lottery decide to start a child in a preferred school, but if that lottery is lost, parents decide to delay entry to the following year. A second mechanism that helps explain these gender differences is that parents of the marginally readiness males more often apply to private and non-sectarian schools as an additional alternative. However, these parents would have to be in higher income groups that could afford private/nonsectarian tuition.

Finally, this intervention was registered in the Open Science registration in 2017-18 as part of receiving funding from the Arnold Foundation. The evaluation methodology registered at that time did not take account of the possibility of having bias in the results, and utilized the standard RCT methodology for estimation. The bias was unexpected since it stems only in RCT's using kindergarten lotteries having middle/high income student samples. This was the first RCT to encounter this form of bias. The Open Science framework for reporting results works well for

RCTs that do not encounter new, unexpected forms of bias. Future RCTs using kindergarten lotteries that include middle/high income students can incorporate the methodologies used here to predict future effects using the Open Science framework.

The unexpected sources of bias in this study changed the method for reporting results. The Open Science Framework calls for reporting the primary results utilizing the methods specified in the Framework first, and then to present results using other specifications or samples as exploratory analysis. However, when unexpected bias is encountered, the presentation of results should change to reflect the presence of bias.

The results that are presented first as the best estimates arising from the study should be the unbiased results. In this study, the female results with the full sample show no differential attrition and are likely unbiased. The full male sample has significant differential attrition and likely significant associated bias making the sample using both genders also biased. However, identifying and eliminating the sources of bias by eliminating younger students and lotteries with the highest differential attrition leaves a sample with no differential attrition. Estimates from this "unbiased" sample are presented first in our analysis as the primary results. We also present estimates for the full sample of males and for the full sample including both genders as "biased" estimates.

References

A Nation at Risk : the Imperative for Educational Reform : a Report to the Nation and the Secretary of Education, United States Department of Education. Washington, D.C. :National Commission on Excellence in Education : [Superintendent of Documents, U.S. Government Printing Office distributor], 1983.

Abdulkadirnglu, A., Angrist, J.D., Dynarski, S.M., Kane, T.J., and Pathak. P.A., (2011). "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." Quarterly Journal of Economics 126(2): 669–748

Abdulkadirnglu, A., Angrist, J.D., Hull, P., Pathak, P. Charters Without Lotteries: Testing Takeovers in New Orleans and Boston, NBER Working Paper February 2014

Adams, B.C., Bell, L.C., & Perfetti, C.A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. Discourse Processes, 20, 307-323.

Afflerbach, P., Hurt, M., & Cho, B. Y. (2020). Reading comprehension strategy instruction. In Handbook of strategies and strategic processing (pp. 98-118). Routledge.

Anderson, R. C., Hiebert, E. H., Scott, J. A., Wilkinson, I. A. G., Becker, W., & Becker, W. C. (1985). Becoming a nation of readers: The report of the Commission on Reading. Education and Treatment of Children, 11(4), 389–396.

Angrist, Joshua D., Bettinger, Eric, Bloom, Erik, King, Elizabeth, and Kremer, Michael (2002). "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review*, 92(5), pp. 1535-58.

Angrist, J.D., Dynarski, S.M., Kane, T.M., Pathak, P.A., & Walters, C.R. (2012). Who benefits from KIPP? Journal of Policy Analysis and Management, 31, 837-860.

Angrist, J.D., Pathak, P.A., & Walters, C.R. (2013). Explaining charter school effectiveness. American Economic Journal: Applied Economics, 5, 1-27.

Austin, M., & Berends, M. (2018). School choice and learning opportunities. In B. Schneider (Ed.), *Handbook of the Sociology of Education in the 21st Century* (pp. 221-250). New York: Springer.

Bailey, D., Duncan, G. J., Odgers, C., Yu, W., (2017), Persistence and fadeout in the impacts of child and adolescent interventions, Journal of Research on Educational Effectiveness, J Res Educ Eff. 2017;10(1):7-39.

Bailey DH, Duncan GJ, Cunha F, Foorman BR, Yeager DS. (2020), Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest*. 2020;21(2):55-97.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

Barron, B., & Bell, P., (2015), "Learning environments in and out of school." In *Handbook of educational psychology*, pp. 337-350. Routledge, Bassock, D., and Reardon, S.F., (2013) Academic Redshirting in Kindergarten: Prevalence, Patterns, and Implications Educational Evaluation and Policy Analysis, 35:3, 283-297, https://doi.org/10.3102/0162373713482764

Baude, P., Casey, M., Hanushek, E., & Rivkin. 2014, "The Evolution of Charter School Quality," NBER Working Papers 20645, National Bureau of Economic Research, Inc.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman,J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*(3), 251-276.

Berends, M. (2015). Sociology and School Choice: What We Know After Two Decades of Charter Schools. Annual Review of Sociology, 41 (159-180). https://doi.org/10.1146/annurev-soc-073014-112340

Berends, M. (2020). Social perspectives on school choice. In *Handbook of research on school choice, 2nd edition* (pp. 32-45). New York: Routledge.

Berends, M. (In Press). The ever-evolving landscape of school choice in the United States. In *Handbook of Urban Education, 2nd Edition,* edited by H. Richard Milner IV & K. Lomotey. New York: Routledge.

Berends, M., & J. Dallavis. (2020). Charter schools after three decades: What are charter schools doing? Center for Reinventing Public Education, University of Washington-Bothel.

Berends, M., Goldring, E., Stein, M., & Cravens, X. (2010). Instructional conditions in charter schools and students' mathematics achievement gains. *American Journal of Education,* 116(3), 303-335.

Berends, M., Peñaloza, R. V., Cannata, M., & Goldring, E. B. (2019). Innovation in charter schools: An exploratory analysis. In M. Berends, R. J. Waddington, & J. Schoenig (Eds.), *School Choice at the Crossroads: Research Perspectives* (pp. 90-106). New York: Routledge.

Berends, M., Watral, C., Teasley, B., & Nicotera, A. (2008). Charter school effects on achievement: Where we are and where we're going. Charter school outcomes, 243-266.

Betts, J. R. (2005). The economic theory of school choice. In J. R. Betts & T. Loveless (Eds.), *Getting choice right: Ensuring equity and efficiency in education policy* (pp. 14–39). Washington, DC: Brookings Institution Press.

Betts, J., & Hill, P. (2006), Key Issues in Studying Charter Schools and Achievement. Review and Suggestions for National Guidelines, National Charter School Research Project Center on Reinventing Public Education, Seattle.

Betts, J. R., & Tang, Y. E. (2019). The effects of charter schools on student achievement. In M. Berends, R. J. Waddington, & J. Schoenig (Eds.), *School Choice at the Crossroads: Research Perspectives* (pp. 69-91). New York: Routledge.

Bifulco, R. & Ladd, H. (2006). The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina, Education Finance and Policy 1:1, 50-90.

Bloom, Howard S. (1984), "Accounting for no-shows in experimental evaluation designs." Evaluation Review 8.2 (1984): 225-246.

Bloom, H. S., & Michalopoulos, C. (2013). When is the story in the subgroups? Prevention Science, 14(2), 179-188.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2),125-230

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, 34, 319-337

Buddin, Richard, The Impact of Charter Schools on Public and Private School Enrollments (August 28, 2012). Cato Institute Policy Analysis No. 707, Available at SSRN: https://ssrn.com/abstract=2226597

Buddin, R. & Zimmer, R. (2003). Academic Outcomes. Charter school operations and performance: Evidence from California. 37-62.

Cabell, S. Q., & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly*, *55*, S99-S107.

Calkins, Lucy, (2022), Units, Tools, *and* Methods *for* Teaching Reading *and* Writing, Units of Study, https://www.unitsofstudy.com/

Calkins, Lucy, (2015) A guide to the reading workshop, primary grades. Heinemann, Portsmouth, NH.

Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2019). Malleability, plasticity, and individuality: How children learn and develop in context. Applied Developmental Science, 23(4), 307-337.

Chabrier, J., Cohodes, S. & Oreopoulos, P. (2016). "What Can We Learn from Charter School Lotteries," Journal of Economic Perspectives, vol 30(3), pages 57-84.

Cheng, Albert and Hitt, Collin and Kisida, Brian and Mills, Jonathan, (2015), No Excuses Charter Schools: A Meta-Analysis of the Experimental Evidence on Student Achievement , Journal of School Choice, 11:2, 209-238,

Center for Research on Education Outcomes (2013). National Charter School Study. Stanford University Stanford, CA. Retrieved June 18, 2020 at https://credo.stanford.edu/sites/g/files/sbiybj6481/f/ncss_2013

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology, 98*(3), 489–507. https://doi.org/10.1037/0022-0663.98.3.489

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Washington D.C.: Brookings Institution Press.

Clark Tuttle, Gill, B., Gleason, P., Knechtel, V., Nichols-Barrer, I., and Resch A., (2013). KIPP Middle Schools: Impacts on Achievement and Other Outcomes. Washington, DC: Mathematica Policy Research.

Clark Tuttle, Gleason, P. M., Tuttle, C. C., & Silverberg, M. K. (2015). Do charter schools improve student achievement? *Educational Evaluation and Policy Analysis*, 37(4), 419–

Clark Tuttle, C. C., Gleason, P., Knechtel, V., Nichols-Barrer, I., Booker, K., Chojnacki, G., Coen, T., & Goblel, L. (2015). Understanding the effect of KIPP as it scales: Volume I, impacts on achievement and other outcomes. Washington, DC: Mathematica Policy Research.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. Economics of Education Review, 28(4), 415-427.

Cohodes, S. R., & Parham, K. S. (2021). Charter schools' effectiveness, mechanisms, and competitive influence. NBER, Working Paper 28477

Common Core State Standards Initiative, (2013), National Governors Association Center for Best Practices & Council of Chief State School Officers. Washington, DC:

Connolly, P., Keenan, C., & Urbanska, K., (2018) The trials of evidence-based practice in education: a systematic review of randomized controlled trials in education research 1980–2016, Educational Research, 60:3, 276-91, DOI: 10.1080/00131881.2018.1493353

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. Psychological Science, 24(8), 1408-1419.

Core Knowledge Foundation, (2010), The Core Knowledge Sequence, ISBN 978-1-890517-25-0, Charlottesville, VA 22902

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. Journal of educational psychology, 99(2), 311.

Cullen, J. B., Jacob, B. A. and Levitt, S. (2006). "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica,* 74(5): 1191-1230.

Cunha, F., Heckman, J. J. (2007). The technology of skill formation. American Economic Review, 97(2), 31–47. Doi:10.3386/w12840

Darling-Hammond, Linda, Brigid Barron, P. David Pearson, Alan H. Schoenfeld, Elizabeth K. Stage, Timothy D. Zimmerman, Gina N. Cervetti, and Jennifer L. Tilson. (2015) Powerful learning: What we know about teaching for understanding. John Wiley & Sons

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron & Osher, D., (2020) Implications for educational practice of the science of learning and development, Applied Developmental Science, 24:2, 97-140, DOI: 10.1080/10888691.2018.1537791

Datnow, A., Borman, G. D., Stringfield, S., Overman, L. T., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. Educational Evaluation and Policy Analysis, 25(2), 143-170.

Datnow, A., Borman, G. D., & Stringfield, S. (2000). School reform through a highly specified curriculum: Implementation and effects of the Core Knowledge sequence. *The Elementary School Journal* 101(2), 167-191.

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(2), 225−236.

Dempster, D.M., (2015), The Determinants of Home Schooling, Home School Researcher, 31(4)

Dick, A.S., Garcia, N.L., Pruden, S.M. *et al.(2019),* No evidence for a bilingual executive function advantage in the ABCD study. *Nat Hum Behav* **3,** 692–701

Digest of Education Statistics, (2019), Table 206.10, National Center for Educations Statistics, Department of Education, Wash, D.C.

DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American Schools.* New York: Russell Sage Foundation.

DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*.

Duke, N.K., Ward, A.E. & Pearson, P. D., The Science of Reading Comprehension Instruction, The Reading Teacher, Volume 74, Issue 6, pp 663-672

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Japel, C. (2007). School readiness and later achievement. Developmental Psychology, 43, 1428–1446. Doi:10.1037/0012-1649.43.6.1428.supp

Duncan RJ, Duncan GJ, Stanley L, Aguilar E, Halfon N. (2020) The Kindergarten Early Development Instrument Predicts Third Grade Academic Proficiency. Early Child Res Q. 2020 4th Quarter;53:287-300.

Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., & Reporting Bias Group. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS one*, *8*(7), e66844.

Dynarski, S., Hubbard, D., Jacob, B., & Robles, S. (2018). Estimating the effects of a large for-profit charter school operator**.** (NBER Working Paper 24428). Cambridge, MA: National Bureau of Economic Research.

Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. Policy Insights from the Behavioral and Brain Sciences, 6(1), 3-11.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. Educational Evaluation and Policy Analysis, 35(2), 157-178.

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).

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School psychology review*, *31*(2), 246-262.

Farrell, C., Wohlstetter, P., Smith, J., 2013). Choices and Challenges Charter School Performance in Perspective. ISBN-13: 978-1-61250-541-1

Ferrare, J. J. (2020). Charter school outcomes. In M. Berends, A. Primus, & M. G. Springer (Eds.), *Handbook of research on school choice, 2nd edition* (pp. 160-173). New York, NY: Routledge.

Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., ... & Lake, R. (2012). Charter-school management organizations: Diverse strategies and diverse student impacts. Mathematica Policy Research, Inc.

Gamoran, A., & Fernandez, C. (2018). Do charter schools strengthen education in high poverty and urban districts? In (I. Rotberg & J. Glazer, eds.), *Choosing charters: Better schools or more segregation?* (pp. 133-152). New York: Teachers College Press.

Gleason, P. (2019). Searching for policies and practices that make charter schools successful: What can research tell us? In M. Berends, R. J. Waddington, & J. Schoenig (Eds.), Gleason, P., Clark, M., Tuttle, C. C., & Dwoyer, E. (2010). *The evaluation of charter school impacts: Final report.* (NCEE 2010-4029) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldman, S.R., Britt, M.A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C.D., Shanahan, C. and Project READi, 2016. Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. Educational Psychologist, 51(2), pp.219-246.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological review, 101, 371- 395.

Grissmer, D.W., and Eiseman, E. (2008). Can Gaps in the Quality of Early Environment and Non-Cognitive Skills Help Explain Persisting Black-White Achievement Gaps? In Waldfogel, J. and Magnuson, K. (Eds.), *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap.* (pp. 139-180). New York: Russell Sage Foundation.

Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. Developmental Psychology, 46, 1008-1017. Doi: 10.1037/a0020104.

Grissmer, D., 2017, A Guide to Incorporating Multiple Methods in Random Controlled Trials to Assess Intervention Effects, 2nd edition, American Psychological Association, Wash D.C.

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., Maier,M., Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study, Early Childhood Research Quarterly,Volume 54, 2021, Pages 1-12,

Hanushek, E., A., Peterson, P., E., Talpey, L., M., Woessmann, W., The Achievement Gap Fails To Close, Education Next, Summer, 2019Vol 19, No.1

Heckman, J., (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children, Science, 312, pp 1900-1902

Hess, F. M., & Loveless, T. (2005). How school choice affects student achievement. Getting choice right: Ensuring equity and efficiency in education policy, 85-100.

Hidi, S. E., Renninger, K. A., & Northoff, G. (2017). The development of interest and self-related processing. In F. Guay, H. W. Marsh, D. M. McInerney, & R. G. Craven (Eds.), International advances in self research. Vol. 6: SELF – Driving positive psychology and well-being (pp. 51–70). Information Age Press.

Hill, C.J., Bloom, H.S., Black, A.R., Lipsey, M.W. (2008), Empirical Benchmarks for Interpreting Effect Sizes in Research, Child Development Perspectives, Vol 2, Issue 3, pp 172-177

Hirsch, E.D. (1988), Cultural Literacy, What Every American Needs to Know,

Hirsch, E. D. (2003). Reading comprehension requires knowledge—of words and the world. American Educator, 27, 10-13.

Hirsch, E.D. (2006, Spring). Building knowledge: The case for bringing content into the language arts block and for a knowledge-rich curriculum core for all children. American Educator.

Hirsch, E.D. (2011, Winter), Beyond Comprehension: We Have yet to Adopt a

Common Core Curriculum that Builds Knowledge Grade by Grade—But We Need to, *American Educator*, v34 n4 p30-36

Hirsch, E. D. (2019). Why knowledge matters: Rescuing our children from failed educational theories. Cambridge, MA: Harvard Education Press.

Hoxby, C. M. (2001). Rising tide. Education Next, 1(4), 68–74.

Hoxby, C. M. (2003). School choice and school productivity (or, could school choice be a tide that lifts all boats?). In C. M. Hoxby (Ed.), *The economics of school choice* (pp. 287–341). Chicago: University of Chicago Press.

Hoxby, C. M., & Murarka, S. (2008). Methods of assessing achievement of students in charter schools. In M. Berends, M. G. Springer, & H. J. Walberg (Eds.), *Charter school outcomes* (pp. 7–37). Mahweh: Taylor & Francis Group.

Hoxby, C. M., Murarka, S., & Kang, J. (2009). *How New York City's charter schools affect achievement.* Cambridge, MA: New York City Charter Schools Evaluation Project.

Hoxby, C. M. and Rockoff, J. (2005). "The Impact of Charter Schools on Student Achievement." Harvard Institute of Economic Research Working Paper Series.

Hoyer, K. M., & Sparks, D. (2017). Instructional Time for Third-and Eighth-Graders in Public and Private Schools: School Year 2011-12. Stats in Brief. NCES 2017-076. *National Center for Education Statistics*.

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. Journal of Educational Psychology, 102(4), 880-895. Doi: 10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. Science, 326(5958), 1410-1412.

Huang, F. L. (2014) Further understanding factors associated with grade retention: Birthday effects and socioemotional skills, Journal of Applied Developmental Psychology, Volume 35, Issue 2, Pages 79-93.

Ioannidis, J. P., Caplan, A. L., & Dal-Ré, R. (2017). Outcome reporting bias in clinical trials: why monitoring matters. *Bmj*, *356*.

Jirout, J. J., Ruzek, E., Vitiello, V. E., Whittaker, J., & Pianta, R. C. (2022). The association between and development of school enjoyment and general knowledge. Child Development.

Kaefer, T., Neuman, S.B., and A. M. Pinkham, (2015) "Pre -existing Background Knowledge Influences Socioeconomic Differences in Preschoolers' Word Learning and Comprehension," Reading Psychology 36, no. 3, 203–231.

Kaefer, T. (2020). When did you learn it? How background knowledge impacts attention and comprehension in read-aloud activities. Reading Research Quarterly, 55, S173-S183.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635-650.

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., ... & McIntyre, J. (2023). A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension. Journal of Educational Psychology, 115(1), 73.

Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2021). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. Journal of Educational Psychology, 113(1), 3.

Kim, Y. S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). Scientific Studies of Reading, 21(4), 310-333.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge, UK: Cambridge University Press.

Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., & Williamson, P. R. (2010). The impact of outcome reporting bias in 61 randomized controlled trials on a cohort of systematic reviews. *Bmj*, *340*.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. Educational Researcher, 49(4), 241-253.

Kruger GHJ. Executive Functioning and Positive Psychological Characteristics: A Replication and Extension, (2011), *Psychological Reports*. 108(2):477-486.

Lareau, A. (2011). Unequal childhoods: Race, class, and family life, 2nd edition: A decade later. Berkeley, CA: University of California Press.

Lareau, A., & Goyette, K. (2014). (Eds.). Choosing homes, choosing schools. New York: The Russell Sage Foundation.

Lee, C. D. (2017). Integrating research on how people learn and learning across settings as a window of opportunity to address inequality in educational processes and outcomes. Review of research in education, 41(1), 88-111.

Long, M.C., & Conger, D. (2013) Gender Sorting Across K-12 Schools in the United States, American Journal of Education, Volume 119, number 3.

McNamara, D. S. (2004). SERT: Self-explanation reading training. Discourse Processes, 38, 1-30.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. Psychological bulletin, 137(2), 267.

Morgan, P., L., Farcas, G, Hillemeier, M. M., Maczuga, S., (2016) Science Achievement Gaps Begin Very Early, Persist, and Are Largely Explained by Modifiable Factors, Education Researcher, Vol 45, Issue 1 pp 18-35

Morton, B. A., & Dalton, B. (2007). Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 through 4. Stats in Brief. NCES 2007-305. *National Center for Education Statistics*.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Boston College, Boston, Mass.

Murnane, R. J., & Reardon, S. F. (2018). Long-term trends in private school enrollments by family income. *AERA Open*, *4*(1), 2332858417751355.

Murnane, R.J., Reardon, S.F., Mbekeani, P.P., Lamb, A.,(2021), Who Goes to Private Schools, Education Next, 21(2), 2021

Murrah, W. M. (2010). Comparing self-regulatory and early academic skills as predictors of later math, reading, and science elementary school achievement. Charlottesville, VA: University of Virginia. Unpublished doctoral dissertation.

Nagy, W. "‘Why Vocabulary Instruction Needs to Be Long-Term and Comprehensive'. In Hiebert, EH, Kamil, ML (Eds.), Teaching and Learning Vocabulary; Bringing Research To Practice. Mahwah, New Jersey, London: Laurence Erlbaum Associates." (2005).

Nathan, Joe. (1999). Charter schools: Creating hope and opportunity for American education. San Francisco: Jossey-Bass Publishers.

National Council of Teachers of Mathematics (Ed.). (2000). Principles and standards for school mathematics (Vol. 1). National Council of Teachers of Mathematics, Reston, VA.

National Education Goals Panel. (1998). The National Education Goals report: Building a nation of learners, 1998. Washington, DC: U.S. Government Printing Office.

National Reading Panel (U.S.) & National Institute of Child Health and Human Development (U.S.). (2000). Report of the National Reading Panel: Teaching children to read : an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

National Research Council. (2000). How people learn: Brain, mind, experience, and school: Expanded edition. National Academies Press.

National Research Council. 1998. *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Academies Press. https://doi.org/10.17226/6023.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Commitee on Science Learning, Kindergarten through Eighth Grade. Richard A. Duschl, Heidi A. Schweingruber, and Andrew W. Shouse, Editors. Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

National Research Council. (2009). Mathematics learning in early childhood: Paths toward excellence and equity: Committee on Early Childhood and Mathematics, The National Academies Press.

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. Reading Research Quarterly, 46(3), 249-272.

No Child Left Behind Act (2001), H. Rept. 107-63, Part 1- U.S. Government Publishing Office, Washington, D.C.

Nye B, Hedges LV, Konstantopoulos S. The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment. *Educational Evaluation and Policy Analysis*. 1999;21(2):127-142.

O’Connor, R. E., Beach, K. D., Sanchez, V., Bocian, K. M., Roberts, S., & Chan, O. (2017). Building better bridges: Teaching adolescents who are poor readers in eighth grade to comprehend history text. Learning Disability Quarterly, 40(3), 174-186.

Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development1. Applied Developmental Science, 24(1), 6-36.

Pearson, P. D, Palincsar, A.S., Biancarosa, G., Berman, A.I.,(2020), Reaping the rewards of the reading for understanding initiative, National Academy of Education, Wash D.C.

Pearson, P. D and Cervetti, G., (2015) Fifty Years of Reading Comprehension Theory and Practice in Research-Based Practices for Teaching Common Core Literacy P. David Pearson & Elfrieda H. Hiebert, eds, Teachers College, Columbia University

Pellegrino, J. W. (2012). From cognitive principles to instructional practices: The devil is often in the details. Journal of Applied Research in Memory and Cognition, 1, 260–262.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, DC: National Academies Press

Perie, M. (1997). Job satisfaction among America's teachers: Effects of workplace conditions, background characteristics and teacher compensation (Vol. 97, No. 471). US Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

PIRLS 2016 International Results in Reading. Copyright © 2017 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Raudenbush, S. W., Reardon, S. F., & Nomi, T. (2012). Statistical analysis for multisite trials using instrumental variables with random coefficients, *Journal of research on Educational Effectiveness*, *5*(3), 303-332.

Ray, Brian D., (2017), A systematic review of the empirical research on selected aspects of homeschooling as a school choice, Journal of School Choice, VOL. 11, NO. 4, 604–621

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference?. *The elementary school journal*, *106*(1), 21-38.

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. Journal of Educational Psychology, 80(1), 16.

Reid S., Snow, P., Serry, T., & Hammond, L., (2021) The Role of Background Knowledge in Reading Comprehension: A Critical Review, Reading Psychology, 42:3, 214-240, DOI: 10.1080/02702711.2021.1888348

Richards, O. H. (2001). Cultural literacy, afrocentrism, and education for citizenship. University of Pennsylvania.

Renninger, K. A., & Hidi, S. (2011). Revisiting the Conceptualization, Measurement, and Generation of Interest. Educational Psychologist, 46(3), 168-184. Retrieved from http://dx.doi.org/10.1080/00461520.2011.587723

Romeo, R. R., Leonard, J. A., Scherer, E., Robinson, S., Takada, M., Mackey, A.P., West, M.R., & Gabrieli, J.E., (2021) Replication and Extension of Family-Based Training Program to Improve Cognitive Abilities in Young Children, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2021.1931999

RRSG (RAND Reading Study Group). (2002). Reading for understanding: Toward a research and development program in reading comprehension. Santa Monica, CA: RAND Corporation.

Sass, Tim R. (2006). Charter Schools and Student Achievement in Florida. Education Finance and Policy, 1 (1), 91-122.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high-and low-aptitude children. Journal of educational psychology, 81(3), 306.

Schneider, J, and McGrew, K., "The Cattell-Horn-Carroll (CHC) model of intelligence v2. 2: a visual tour and summary." Institute for Applied Psychometrics (IAP) 1 (2012): 03-13.

Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology, 109*(8), 1120–1140. https://doi.org/10.1037/edu0000193

Schwartz, S., (2022), Which States have Passed “Science of Reading” Laws, Education Week, July 20, 2022.

Schwartz, S., (2021a), Popular Literacy Materials Get ‘Science of Reading’ Overhaul. But Will Teaching Change?, Education Week, October 13, 2021

Schwartz, S., (2021b), New Curriculum Review Gives Failing Marks to Two Popular Reading Programs, Education Week, November 9, 2021

Shah, K., Egan, G., Huan, L. N., Kirkham, J., Reid, E., & Tejani, A. M. (2020). Outcome reporting bias in Cochrane systematic reviews: a cross-sectional analysis. *BMJ open*, *10*(3), e032497.

Shakeel, M., & Peterson, P. E. (2022). A Half Century of Progress in US Student Achievement: Agency and Flynn Effects, Ethnic and SES Differences. Educational Psychology Review, 1-88.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. Harvard educational review, 78(1), 40-59.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). Improving Reading Comprehension in Kindergarten through 3rd Grade: IES Practice Guide. NCEE 2010-4038. What Works Clearinghouse.

Steinkamp, M. W., & Maehr, M. L. (1983). Affect, ability, and science achievement: A quantitative synthesis of correlational research. Review of Educational Research, 53(3), 369-396.

Talwar, A., L. Tighe, E., & Greenberg, D. (2018). Augmenting the simple view of reading for struggling adult readers: A unique role for background knowledge. Scientific Studies of Reading, 22(5), 351-366.

Tobias, S. (1994). Interest, prior knowledge, and learning. Review of Educational Research, 64(1), 37-54.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. Educational psychology review, 26(2), 265-283.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, "Results from the 2019 Mathematics and Reading Assessments," www.nationsreportcard.gov/mathematics/support files/2019_infographic.pdf

U.S. Department of Education, National Center for Education Statistics, *Characteristics of Private Schools in the United States: Results From the 2013-14 Private School Universe Survey (NCES 2013-316).*

Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. Cognition and Instruction, 14, 45-68.

Walberg, H. J. (2011). *Tests, testing, and genuine school reform.* Stanford, CA: Hoover Institution Press.

Walberg, H. J., & Bast, J. (2003). *Education and capitalism: How overcoming our fear of markets and economics can improve America's schools*. Stanford,| CA: Hoover Institution Press.

Walker, 1987, Relative Importance of Domain Knowledge and Overall Aptitude on Acquisition of Domain-Related Information, Cognition and Instruction, 4, 25-42.

Wexler, N., (2018) Why American Students Haven't Gotten Better at Reading in 20 Years, The Atlantic, April, 2018

Wexler, N. (2019). The knowledge gap: the hidden cause of America's broken education system--and how to fix it. New York: Avery, Penguin Random House LLC.

Wexler, N., 2022, What-works-in-reading-comprehension-and-what-doesn't, Forbes, 04/09/2022

What Works Clearing House. (2020), Procedures and Standards Handbook, Version 3.0, Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Wash, D.C.

Whitebread, D., Neale, D., Jensen, H., Liu, C., Solis, S. L., Hopkins, E., & Zosh, J. (2017). The role of play in childen's development: a review of the evidence. LEGO Fonden.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. Contemporary educational psychology, 25(1), 68-81. doi:10.1006/ceps.1999.1015

Willingham, D. T. (2017). The reading mind: A cognitive approach to understanding how the mind reads. John Wiley & Sons.

Willingham, D. T. (2015). Raising kids who read: What parents and teachers can do. John Wiley & Sons.

Willingham, D. T., & Lovette, G. (2014). Can reading comprehension be taught. Teachers College Record, 116, 1-3.

Willingham, D. T. (2006). How knowledge helps: It speeds and strengthens reading comprehension, learning-and thinking. American Educator, 30(1), 30.

Willingham, D. T. (2003). Students remember what they think about. American Educator, 27(2), 37-41.

Willingham, D. T., & Riener, C. (2019). Cognition: The thinking animal. Cambridge University Press.

Wooldridge Jeffrey, M. (2002). Econometric analysis of cross section and panel data.

Zaboski BA., Kranzler JH, Gage NA.(2018) Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-horn-Carroll theory. J Sch Psychol. Dec71 pp 42-56.

Zimmerman, C. (2000). The development of scientific reasoning skills. Developmental Review, 20(1), 99-149. Retrieved from http://www.sciencedirect.com/science/article/pii/S0273229799904975

Zimmer, R., & Buddin, R. (2007). Getting inside the black box: Examining how the operation of charter schools affects performance. Peabody Journal of Education, 82(2-3), 231-273.

Zimmer, R., Buddin, R., Jones, J., Liu, N. (2011). What Types of School Capital Projects are Voters Willing to Support? Public Budgeting and Finance, 31(1), 37-55.

Zimmer, R., Buddin, R., Smith, S. A., & Duffy, D. (2020). Nearly three decades into the charter school movement: What has research told us about charter schools? The Routledge Handbook of the Economics of Education

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123, 162-185.

Appendix A Figures

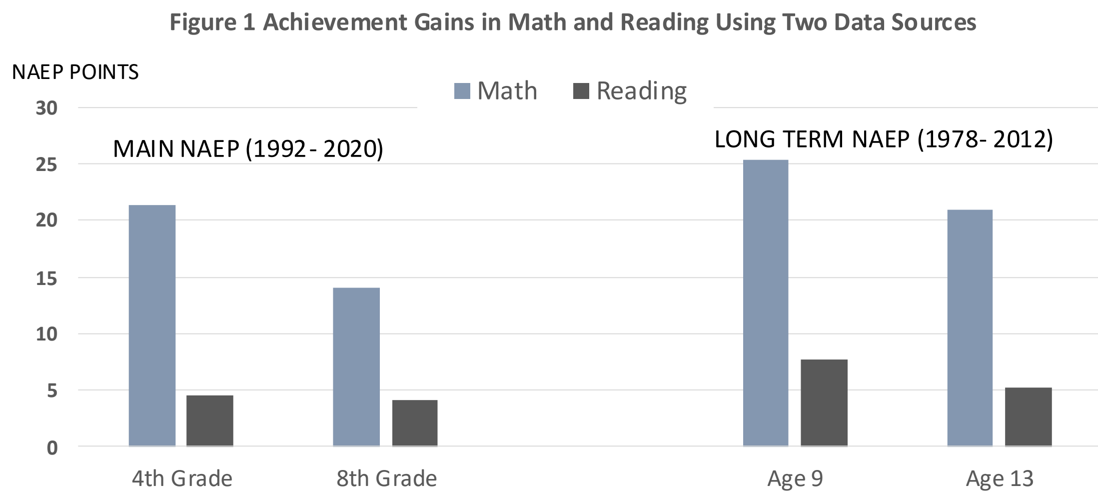
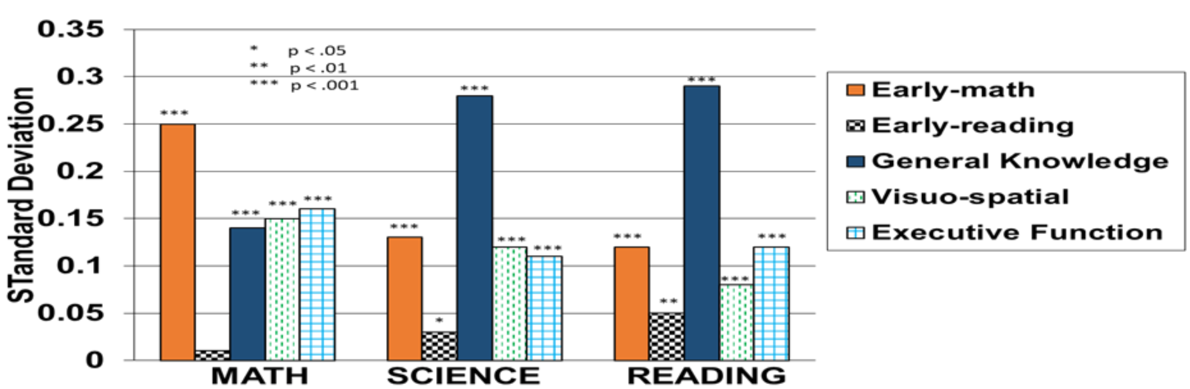


Figure 2 Estimated Effects of Early Skills at Kindergarten Entrance on 8th Grade Achievement



Appendix B Tables

Table 1 Characteristics of School Districts for Nine Participating Schools

Nearest City/Town	Number of lotteries	School District	Median Family Income[1]	% families with children under poverty level [1]	School Size (K-8)[2]
Aurora	**1**	**Adams-Arapahoe**	**51,424**	**28.8**	**544**
Loveland	**2**	**Thompson**	**75,105**	**10.5**	**481**
Ft Collins	**1**	**Poudre**	**77,491**	**11**	**798**
Ft Collins	**1**	**Poudre**	**77,491**	**11**	**571**
Arvada	**1**	**Jefferson**	**85,793**	**9.7**	**466**
Littleton	**2**	**Littleton**	**92,137**	**6.7**	**772**
Castle Rock	**2**	**Douglas**	**114,223**	**3.6**	**767**
Highlands Ranch	**2**	**Douglas**	**114,223**	**3.6**	**600**
Castle Pines	**2**	**Douglas**	**114,223**	**3.6**	**2359**

[1] Estimated using the School District Data from the American Community Survey (2014) five year estimates (2010-2014). Median Income in 2010 inflation adjusted $
[2] Estimated enrollments for 2016-2017 School Year

Table 2 Sample Sizes for Lottery Applications and Students

		Single Appliers	More Than One Lottery	Total		% Single
Applications		**1831**	**1022**	**2853**		**64.2%**
Students		**1831**	**479**	**2310**		**79.3**

Table 3 Lottery Outcomes

		Winners	Losers	% Winners	Winners	Losers	% Winners
		Single Appliers			All-Applications		
Applications		**688**	**1143**	**37.6%**	**1011**	**1842**	**35.4%**
Students		**688**	**1143**	**37.6%**	**954**[1]	**1356**	**41.3%**

[1] Student Won at Least One Lottery

Table 4 Lottery Applications and Outcomes for the All-Applications Sample by Lottery

		Winning Applications	Losing Applications	Total Applications		% Winning Applications	Winner's Accepting	% Accepting
Lottery 1		57	43	100		57.0%	35	61.4%
Lottery 2		47	116	163		28.8%	33	70.2%
Lottery 3		202	105	307		65.8%	60	29.7%
Lottery 4		233	164	397		58.7%	71	30.5%
Lottery 5		26	36	62		41.9%	16	61.5%
Lottery 6		73	240	313		23.3%	33	45.2%
Lottery 7		34	42	76		44.7%	20	58.8%
Lottery 8		44	160	204		21.6%	24	54.6%
Lottery 9		73	150	223		32.7%	30	41.1%
Lottery 10		36	73	109		33.0%	21	58.3%
Lottery 11		73	22	95		76.8%	24	32.9%
Lottery 12		39	228	267		14.6%	20	51.3%
Lottery 13		46	296	342		13.5%	23	50.0%
Lottery 14		28	167	195		14.4%	16	57.1%
Total		1011	1842	2853		35.4%	475	47.0%

Table 5 Randomization Tests for Gender and Age

		% Female Winners	% Female Losers	Z-Value		Average Age-Losers	Average Age-Winners	Z-Value
Lottery 1		55.81	54.39	-0.14		5.64	5.57	-0.198
Lottery 2		59.13	73.91	1.86		5.57	5.55	-0.063
Lottery 3		60.00	48.47	-1.93		5.61	5.64	0.086
Lottery 4		52.20	53.91	0.33		5.55	5.60	0.162
Lottery 5		51.43	42.31	-0.71		5.71	5.66	-0.151
Lottery 6		52.47	47.95	-0.67		5.55	5.58	0.078
Lottery 7		43.59	41.18	-0.21		na	na	na
Lottery 8		54.84	44.19	-1.24		5.58	5.56	-0.070
Lottery 9		53.38	49.32	-0.57		5.56	5.52	-0.094
Lottery 10		50.00	54.29	0.42		na	na	na
Lottery 11		30.00	40.28	0.87		5.54	5.60	0.189
Lottery 12		53.18	52.63	-0.06		5.66	5.52	-0.503
Lottery 13		50.69	53.33	0.33		5.59	5.55	-0.121
Lottery 14		42.59	29.63	-1.35		5.58	5.37	-0.606
Total		52.02	50.15	-0.943		5.58	5.59	0.012

Table 6 Attrition Statistics for 3rd -6th Grade English-LA Achievement for All-Applications

		Total Sample	Lottery Winners	Lottery Losers	Differential Attrition
		FEMALES			
Potential Achievement Scores (3rd-6th Grade)		5172	1798	3374	
Missing Achievement		1621	559	1062	
Overall Attrition		31.3%	31.1%	31.5%	-0.4
		M ALES			
Potential Achievement Scores (3rd-6th Grade)		4915	1779	3136	
Missing Achievement		1791	567	1224	
Overall Attrition		36.4%	31.9%	39.0%	-7.2***
		TOTAL SAMPLE (including missing gender)			
Potential Achievement Scores (3rd-6th Grade)		10349	3632	6717	
Missing Achievement		3674	1181	2493	
Overall Attrition		35.5%	32.5%	37.1%	-4.6***

+ p < .10, * p < .05, ** p < .01

Table 7 Sources of Attrition by Lottery Status and Gender

		Off-track			Not Tested			Private/Home/ Out of State			Total	
		No-offer	Offer		No-offer	Offer		No-offer	Offer		No-offer	Offer
<u>Males</u>												
% of Attrition		11.9%	9.6%		4.9%	5,1%		20.6%	15.5%		37.4%	30.2%
Std. Error		0.68%	0.82%		0.45%	0.62%		0.85%	1.01%			
Differential Attrition			-2.3*			0.2			-5.1**			-7.2**
<u>Females</u>												
% of Attrition		4.7%	6.5%		6.6%	6.3%		18.6%	16.9%		29.9%	29.8%
Std. Error		0.68%	0.82%		0.45%	0.62%		0.85%	1.01%			
Differential Attrition			+1.8			-0.3			-1.7			-0.1

+ p < .10, * p < .05, ** p < .01, *** p < .005

Table 8 Overall and Differential Attrition for Samples with Different Bias Vulnerability

	All Ages		Young Ages Excluded[1]	
	All Lotteries (FULL)	Lotteries Excluded[2] (LLO)	All Lotteries (YSLO)	Lotteries Excluded[2] LLO_YSLO)
	ALL APPLICATIONS[3]			
Overall Attrition	34.1%	33.8%	32.8%	31.6%
Differential Attrition	-4.5***	-1.6	-4.8***	-0.01
WWC Criteria	Liberal	Conservative	Liberal	Conservative
Sample	7496	6336	6620	5568
	FEMALES			
Overall Attrition	29.9%	29.7%	28.9%	28.6%
Differential Attrition	-0.0	0.02	-0.02	-0.0
WWC Criteria	Conservative	Conservative	Conservative	Conservative
Sample	3744	3408	3541	3223
	MALES			
Overall Attrition	35.0%	34.6%	32.7%	31.8%
Differential Attrition	-7.0***	-3.4+	-5.0**	-0.0
WWC Criteria	Liberal	Liberal	Liberal	Conservative
Sample	3562	2980	2889	2324

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

Sample excludes males younger than 5.4 years and females younger than 5.2 years as o October 1

2 Sample excludes lotteries with average 3rd-6th achievement samples less than 7 and/or high and significant differential attrition

3 All Applications includes applications with no gender- about 2.5% of the total sample- and none of these applications have achievement data

Table 9 Comparing English Proficiency ITT Effects[1] for Specific Grades and Different Bias Vulnerability

	All Grades	4th Grade	5th Grade	6th Grade
	Full Sample			
Effect Size	0.143**	0.109+	0.169**	0.114+
Sd. Error	(0.053)	(0.059)	(0.062)	(0.064)
Sample	6652	1888	1810	1728
	Eliminate Highest Differential Attrition Lotteries[2]			
Effect Size	0.179**	0.167*	0.220**	0.128+
Sd. Error	(0.063)	(0.072)	(0.074)	(0.074)
Sample	4949	1375	1307	1244
	Eliminate Young Students[3] and Highest Differential Lotteries			
Effect Size	0.241***	0.196**	0.281**	0.208*
Sd. Error	(0.068)	(0.075)	(0.080)	(0.085)
Sample	4027	1162	1098	1037

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

1 Estimated using multiway clustering (Cameron et al, 2011)

2 Sample excludes lotteries in middle/high income school districts with highest differential attrition and lotteries that have very small achievement samples (six or fewer observations)

3 Sample excludes males younger than 5.4 years and females younger than 5.2 years as of

Table10 Comparing 3rd-6th Grade English Proficiency ITT Effects by Gender for Samples with Different Bias Vulnerability

			Full Sample		Eliminate Highest Differential Attrition Lotteries[1]		Eliminate Young Students[2] and Highest Differential Attrition Lotteries
Females	Effect Size		0.223**		0.242**		0.267**
	Sd. Error		(0.077)		(0.081)		(0.085)
	Sample		3541		2960		2556
Males	Effect Size		0.063		0.068		0.207+
	Sd. Error		(0.074)		(0.094)		(0.112)
	Sample		3111		2200		1514

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

1 Eliminates lotteries in high/middle income school districts that have the largest differential attrition levels

2 Eliminates young applicants- males less than 5.4 years and females less than 5.2 years

Table 11 Comparing 3rd-6th English Proficiency ITT Effects for Single Appliers and All Applications

		Full Sample		Eliminate Highest Differential Attrition Lotteries[1]		Eliminate Young Students and Highest Differential Attrition Lotteries[2]
		Single Appliers				
Effect Size[3]		0.155*		0.199*		0.306**
Sd. Error		(0.071)		(0.081)		(0.086)
Sample		4114		3067		2647
		All Applications				
Effect Size[4]		0.143**		0.179**		0.241**
Sd. Error		(0.053)		(0.063)		(0.068)
Sample		6652		4949		4027

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

1 Sample excludes lotteries in middle/high income school districts with highest differential attrition and lotteries that have very small achievement samples (six or fewer observations)

2 Sample excludes males younger than 5.4 years and females younger than 5.2 years as of October 1

3 Estimated using one-way clustering across grades

4 Estimated using multiway clustering across grades and multiple applications (Cameron et al, 2011)

Table 12 Comparing ITT and TOT English Proficiency Effects by Grade for Sample Eliminating Lotteries with High Differential Attrition[1] and Young Students[2]

		All Grades		4th Grade		5th Grade		6th Grade
		ITT Effects						
Effect Size		0.241***		0.196**		0.281**		0.208*
Sd. Error		(0.068)		(0.075)		(0.080)		(0.085)
Sample		4027		1162		1098		1037
		TOT Effects						
Effect Size		0.473***		0.383**		0.543***		0.404*
Sd. Error		(0.135)		(0.146)		(0.156)		(0.162)
Sample		4027		1162		1098		1037

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

1 Sample excludes lotteries in middle/high income school districts with highest differential attrition and lotteries that have very small achievement samples (six or fewer observations)

2 Sample excludes males younger than 5.4 years and females younger than 5.2 years as of October 1

Table 13 Mathematics and Science ITT and TOT Effects by Gender for Sample Eliminating Lotteries with High Differential Attrition[1] and Young Students[2]

			Both Genders		Females		Males
			ITT Results				
Mathematics(3rd-6th Grade)	Effect Size		0.081		0.146		0.003
	Sd. Error		(0.071)		(0.089)		(0.118)
	Sample		4023		2557		1508
Science (5th Grade)	Effect Size		0.154*		0.184+		0.083
	Sd. Error		(0.075)		(0.097)		(0.115)
	Sample		1113		690		467
			TOT Results				
Mathematics(3rd-6th Grade)	Effect Size		0.159		0.273		0.006
	Sd. Error		(0.139)		(0.168)		(0.250)
	Sample		4023		2557		1508
Science (5th Grade)	Effect Size		0.300*		0.339+		0.175
	Sd. Error		(0.147)		(0.178)		(0.239)
	Sample		1113		690		467

+ $p < .10$, * $p < .05$

1 Eliminates lotteries in high/middle income school districts that have the largest differential attrition levels

2 Eliminates young applicants- males less than 5.4 years and females less than 5.2 years

Table 14 Comparing ITT Effects by Subject for a CK-Charter School in a Low Income School District to CK-Charter Schools in Middle/High Income School Districts

			English Proficiency (3rd-6th grade)		Mathematics (3rd-6th grade)		Science 5th Grade
Low Income School District	Effect Size		0.944**		0.735*		0.468
	Sd. Error		(0.295)		(0.350)		(0.316)
	Sample		167		166		41
Middle/High Income School Districts	Effect Size		0.201**		0.041		0.125
	Sd. Error		(0.070)		(0.073)		(0.077)
	Sample		3860		3857		1072
Effect Difference			0.743**		0.694*		0.343

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

Table 15 Comparing TOT Effects by Subject for a CK-Charter School in a Low Income School District to CK-Charter Schools in Middle/High Income School Districts

			English Proficiency (3rd-6th grade)		Mathematics (3rd-6th grade)		Science 5th Grade
Low Income School District	Effect Size		1.299**		0.997*		0.622
	Sd. Error		(0.459)		(0.439)		(0.380)
	Sample		167		166		41
Middle/High Income School Districts	Effect Size		0.445**		0.090		0.270
	Sd. Error		(0,155)		(0.160)		(0.166)
	Sample		3860		3857		1072
Effect Difference			0.854**		0.907*		0.352

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$

Table 16 Summary of 3rd-6th Grade TOT Effects and Percentile Gains for Sample with Excluded Young Students and High Differential Attrition Lotteries

			English	Math	Science
ALL APPLICATIONS	Effect		0.473***	0.159	0.300*
	Std Error		(0.135)	(0.139)	(0.147)
	Sample		4027	4023	1113
	Percentile Gain		16.1***	5.41	10.2*
FEMALE	Effect		0.500***	0.273	0.339+
	Std Error		(0.162)	(0.168)	(0.178)
	Sample		2556	2557	690
	Percentile Gain		17.0***	9.3	11.5+
MALE	Effect		0.440+	0.006	0.175
	Std Error		(0.238)	(0.250)	(0.239)
	Sample		1514	1508	467
	Percentile Gain		15.0+	0.002	6.0+

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .005$