
Knowledge in Action Efficacy Study Over Two Years

FEBRUARY 22, 2021

USC DORNSIFE CENTER FOR ECONOMIC AND SOCIAL RESEARCH

Anna Rosefsky Saavedra, *Principal Investigator*

Ying Liu

Shira Korn Haderlein

GIBSON CONSULTING GROUP

Amie Rapaport, *Co-Principal Investigator*

Marshall Garland

Danial Hoepfner

PENN STATE UNIVERSITY

Kari Lock Morgan, *Co-Principal Investigator*

Alyssa Hu

PREFACE

Knowledge in Action is a project-based learning approach to Advanced Placement (AP). Developers designed KIA intervention—comprised of curriculum, instructional materials, and robust professional development supports for teachers—to bolster students’ deeper understanding of content and skills by encouraging their active engagement through projects. With funding from the George Lucas Educational Foundation between 2008-15, University of Washington professors, in collaboration with local teachers, designed the KIA curriculum to cover an entire subject-specific AP curriculum framework through five project units taught over the course of an academic year. Lucas Education Research, the research division of the George Lucas Educational Foundation, asked the University of Southern California’s Dornsife Center for Economic and Social Research to conduct a randomized controlled trial efficacy evaluation of the Knowledge in Action intervention in 2016-17 with a follow-up study of RCT schools in 2017-18, funding this work with grants from March 2015 through March 2021. This report describes the study motivation, context, research methods, results, and implications.

ACKNOWLEDGEMENTS

The authors* extend our deep gratitude to the students, teachers, school leaders, and district staff who voluntarily participated in Knowledge in Action data collection. We are grateful to Lucas Education Research staff and leadership, AnneMarie Baines, Britte Cheng, and Nathan Warner, and for their comprehensive support. We thank Jill Carle, Sarah Jean Johnson, Janet Li, and Elizabeth Marwah for their many research contributions, and PBLWorks staff and coaches for their participation in data collection activities. We recognize the College Board, including Trevor Packer, Jeff Olson, Daryl Morris, Sherby Jean-Leger, and many other colleagues, for their support through sharing de-identified student assessment data. We are thankful to Beth Tipton and anonymous reviewers for two rounds of feedback to our methodology, results, and reporting. We thank, for their advice at varying points throughout the course of the study, Howard Everson, Jennifer Hamilton, Jane Lo, Richard Murnane, Susan Nolen, Walter Parker, Juan Saavedra, and Sheila Valencia. Also, we are grateful to Charlotte Holms and Karlen Lusborough for their support to district staff, Azucena Carabez and Tricia Goebel for serving as student survey and College and Work Readiness Assessment test proctors, and Mike Branom for writing support. We thank the staff of Council for Aid to Education for facilitating assessment administration and providing assessment data and analytic support. Finally, we credit John Engberg, Laura Hamilton, Tiffany Tsai, and Jennifer Tamargo for their initial contributions to the study. Any flaws are solely the authors' responsibility.

We performed computations for this research on the Pennsylvania State University's Institute for Computational and Data Science's Roar supercomputer.

We gratefully acknowledge financial support for this research from the George Lucas Educational Foundation. Findings and conclusions are those of the authors' and do not necessarily reflect Foundation views.

TABLE OF CONTENTS

PREFACE	2
ACKNOWLEDGEMENTS	3
EXECUTIVE SUMMARY	5
STUDY OVERVIEW	7
STUDY MOTIVATION	8
THE KNOWLEDGE IN ACTION INTERVENTION	10
BUSINESS-AS-USUAL SUPPORT FOR CONTROL TEACHERS	12
RESEARCH QUESTIONS	12
RANDOMIZED CONTROLLED TRIAL RESEARCH DESIGN	13
STUDY SCHOOL DISTRICTS	14
STUDENT OUTCOME MEASURES	15
RESEARCH QUESTION 1: SAMPLE, ANALYTIC METHODS, AND RESULTS	16
RESEARCH QUESTION 2: SAMPLE, ANALYTIC METHODS, AND RESULTS	23
RESEARCH QUESTION 3: SAMPLE, ANALYTIC METHODS, AND RESULTS	26
RESEARCH QUESTION 4: SAMPLE, ANALYTIC METHODS, AND RESULTS	31
STUDY-WIDE LIMITATIONS	33
IMPLICATIONS AND CONCLUSION	34
REFERENCES	36

EXECUTIVE SUMMARY

Background. Knowledge in Action (KIA) is a project-based learning (PBL) approach to Advanced Placement (AP). Developers designed the KIA intervention—comprised of curriculum, instructional materials, and robust professional development supports for teachers—to bolster students’ deeper understanding of content and skills by encouraging their active engagement through projects. With funding from the George Lucas Educational Foundation between 2008-15, University of Washington professors, in collaboration with local teachers, designed the KIA curriculum to cover a subject specific AP curriculum framework over the course of an academic year.

The Knowledge in Action Intervention. KIA curriculum and instructional materials are available with open access through the Sprocket portal, hosted by Lucas Education Research, for AP U.S. Government, AP Environmental Science, and AP Physics; this study addressed the first two. The KIA intervention supports are course-specific, designed to align to the College Board’s course curriculum frameworks for AP U.S. Government (APGOV) and AP Environmental Science (APES). However, the same design principles apply to both courses, so both versions of the KIA intervention include similar resources. AP exams administered and scored by the College Board serve as ideal measures of student outcomes. Developers designed both courses with the goals of developing students’ civic, political, and/or environmental awareness and engagement. Ongoing job-embedded professional learning provided by PBLWorks included a four-day summer institute, four full days during the year, and on-demand virtual coaching support.

Research Design. After one year, we evaluated the efficacy of the KIA intervention using a randomized controlled trial (RCT) with school-level randomization. Our research design featured a staggered roll-out, such that Year One (2016-17) impact analyses compared student outcomes of teachers with one versus zero years of KIA experience, while Year Two (2017-18) impact analyses compared student outcomes of teachers with two years of KIA versus one. Since the staggered rollout did not allow for a pure Year Two control group, we used two methods to estimate differences in AP outcomes between students of teachers with two and zero years of KIA experience. Our study of teachers’ KIA classroom implementation accompanied the impact analyses.

Sample. Teachers and their students were from five large school districts across the country. A higher proportion of the student sample, compared to typical AP exam-takers, was from low-income households. Four of five participating districts serve majority Black and Hispanic students.

Results. Results have notable implications for practitioners and policymakers. Under optimal conditions of teacher support, the Year One pattern of results suggests a positive KIA impact on students’ propensity to earn qualifying scores on AP examinations and underlying continuous AP scores. Earning qualifying AP scores can translate into college credit, and relates to enrolling and persisting in college. At the end of the first year, we observed positive results both within each course and pooled across courses. The pattern also was positive within respective groups of students from lower- and higher-income households; in districts serving a majority of students from lower-income households; in districts serving a majority of students from higher-income households; and within each of the five participating districts. Thus, one subgroup of students did not drive overall results.

Strengthening causal claims is the RCT design, as well as the statistical significance, magnitude, and robustness of estimated effect sizes across multiple covariate-adjusted sensitivity analyses. Weakening causal claims are high school-level attrition post-randomization, and differences between unadjusted results compared to those statistically adjusted to address observed baseline differences.

Results contribute to a body of evidence, currently narrow, on whether teachers' proficiency implementing inquiry-based pedagogy, including PBL, changes after their first year. Year Two results were less conclusive due to a lack of significance in second- versus first-year estimates, and substantial limitations to estimates of two years of KIA experience relative to none. Though student AP exam performance may continue to benefit from a teacher's second year of experience with KIA, AP performance gains occurred primarily in teacher's first KIA year. The only outcome with a different second-year trend was students' propensity to take the AP exam, for which we observed nearly all the effect in a teacher's second KIA year.

Investigation of implementation in Year One revealed teachers felt KIA was more engaging for students, offering the opportunity for them to develop real-world skills. Though KIA treatment teachers found the change considerable, with challenges in pacing and groupwork facilitation, the majority recommended the approach, citing benefits for themselves and students. Treatment students voiced benefits related to civic engagement, group work, engagement with learning, and examination preparation. For students of KIA teachers to outperform control students is even more practically meaningful given both teachers and students perceived benefits beyond examination performance.

Implications. Particularly in a high-stakes AP setting, shifting from primarily transmission to PBL instruction is a substantial change for teachers, suggesting the need for high levels of professional development and coaching support. During teachers' first year of KIA, professional learning supports included four days in the summer, four days during the school year, and on-demand coaching. In a teacher's second year, KIA supports were optional and did not include access to on-demand coaching; few second-year teachers participated. That the impact on student AP exam performance occurred primarily in teachers' first KIA year aligns with the ongoing, job-embedded professional learning occurring during that time. The lack of observed erosion of impact on student outcomes in teachers' second year of KIA suggests costs to shift to PBL do not require annual professional development expenses for teachers.

Related to scaling KIA beyond the KIA RCT study are treatment teachers' self-reported perception that KIA aligned to the AP curriculum framework and examinations, and students' feelings of learning more deeply and being prepared for the AP examinations. Also critical to scaling are KIA teachers' positive perceptions of the approach across courses and their recommendations of KIA to others.

In conclusion, the results of the Knowledge in Action Efficacy Study support teacher-driven adoption of the KIA approach in both APGOV and APES courses, among districts with open-enrollment AP policies that support project-based learning, and for students from both lower- and higher-income households.

STUDY OVERVIEW

Knowledge in Action (KIA) is a project-based learning (PBL) approach to Advanced Placement (AP). Developers designed the KIA intervention—comprised of curriculum, instructional materials, and robust professional development supports for teachers—to bolster students’ deeper understanding of content and skills by encouraging their active engagement through projects. With funding from the George Lucas Educational Foundation between 2008-15, University of Washington professors, in collaboration with local teachers, designed the KIA curriculum to cover an entire subject-specific AP curriculum framework through five project units taught over the course of an academic year. KIA curriculum and instruction materials are available through the online Sprocket portal hosted by Lucas Education Research. Ongoing job-embedded professional learning, provided by PBLWorks during the study, included a four-day summer institute, four full days during the year, and on-demand virtual coaching support throughout the year.

KIA curriculum and instructional materials are available with open access for AP U.S. Government, AP Environmental Science, and AP Physics; this study addressed the first two. The KIA intervention supports are course-specific, designed to align to the College Board’s course curriculum frameworks for AP U.S. Government (APGOV) and AP Environmental Science (APES). However, the same design principles apply to both courses, so both versions of the KIA intervention include similar resources. Serving as ideal measures of student outcomes for both courses are well-defined, well-known, end-of-year tests with strong psychometric properties: AP exams administered and scored by the College Board. In addition, designed both courses with the goals of developing students’ civic, political, and/or environmental awareness and engagement, which are areas of needed focus (e.g., Duncan & Ryan, 2021; Levine & Kawashima-Ginsberg, K. 2017; Valant & Newark, 2017).

After one year, we evaluated the efficacy of the KIA intervention using a randomized controlled trial (RCT) with school-level randomization and student-level outcomes. The study design featured a staggered roll-out such that teachers in control schools could participate in KIA in the second year. In the second year, we harnessed the RCT design to compare AP outcomes between students of teachers with one year of KIA versus two. Also, to estimate differences in AP outcomes between students of teachers with two years of KIA experience versus those with none, we used a novel method of leveraging the sample of teachers who volunteered in the RCT. We complemented this two-year estimate with a propensity-score matched analysis comparing students of teachers with two years of KIA to students of teachers in the same districts who did not volunteer for the RCT. Our study of teachers’ classroom implementation of the KIA approach accompanied the impact analyses.

Sample students, in five predominantly urban school districts across the country, were composed of a higher proportion of students from low-income households than the typical AP exam-taking community. This report summarizes our research approaches, and impact and implementation results across the Year One (2016-17) and Year Two (2017-18) school years, concluding with discussion of the implications of findings.

STUDY MOTIVATION

The traditional “transmission” model of instruction, in which teachers transmit knowledge to students through lectures and assigned readings, may be suboptimal for supporting students’ ability to think and communicate in sophisticated ways, demonstrate creativity and innovation, and transfer their skills, knowledge, and attitudes to new contexts (Gardner, 1999; Perkins, Jay, & Tishman, 1993). In contrast, through project-based learning (PBL), teachers primarily play a facilitator role while students actively engage in teacher- and student-posed learning challenges, working alone and in groups on complex tasks organized around central questions leading to a final product (Hmelo-Silver, 2004; Thomas, 2000).

A large body of observational studies has revealed positive associations between student exposure to PBL instruction and outcomes including: transfer of knowledge and skills (Barron & Darling-Hammond, 2008; Dochy et al., 2003; Gijbels et al., 2005); student engagement with learning (Boaler, 1997; Cognition and Technology Group at Vanderbilt, 1992; Maxwell, Bellisimo & Mergendoller, 2001; Strobel & Van Barneveld, 2009; Wieseman & Cadwell, 2005); standardized test performance (Geier et al., 2008; Schneider, Krajcik, Marx & Soloway, 2002); civic engagement (Saavedra, 2014; Youniss, 2012); metacognition (Kolodner et al., 2003); long-term retention (Strobel & Van Barneveld, 2009); problem-solving (Drake & Long, 2009); disciplinary thinking (Hernandez-Ramos, P., & De La Paz, S. 2009); and collaboration (Wieseman & Cadwell, 2005).

In a meta-analysis of 82 observational studies comparing PBL to other instructional approaches, Walker and Leary (2009) found PBL was most strongly associated with student gains in strategic thinking and designing solutions to complex challenges. A synthesis of eight meta-studies of observational studies from the past four decades showed transmission instructional approaches lagged behind PBL in the areas of students’ long-term retention and skill development, as well as satisfaction among both teachers and students. Transmission approaches were, on average, more effective for short-term retention, as measured through standardized exams (Strobel and van Barneveld, 2009). However, causal conclusions are difficult to draw from observational studies because teachers who choose to teach using a PBL approach may differ from those who do not—and those differences, rather than the PBL approach, may drive observed associations between students’ exposure to PBL and their outcomes.

More limited is causal evidence of PBL instruction’s impact on student outcomes (Condliffe et al., 2017). Using an RCT design, Finkelstein et al. (2010) demonstrated that a PBL economics course increased high school students’ performance on measures of economic literacy and problem-solving, with effect sizes of approximately 0.3 standard deviations. In an RCT evaluation of the effects of a middle school PBL-based science program on end-of-unit earth and physical science tests, Harris et al. (2015) estimated impacts of 0.22-0.25 standard deviations. Evaluations also detected effects for social studies and informational reading at the elementary school level using an RCT design (Duke et al., 2020). In a related RCT study, Jackson and Makarin (2018) demonstrated inquiry-based instructional materials improved middle school students’ state standardized mathematics scores by 0.06 standard deviations. When those instructional materials were paired with online professional development support, the improvement was 0.09 standard deviations.

No evaluations of PBL interventions harnessing an RCT design have documented program effects for more than one year. Multi-year studies of curricular interventions have demonstrated that teachers may not be able to fully implement new curriculum during the first year (e.g., Fullan and Miles 1992; Fullan 1993), with impacts on student outcomes lower in the first year than in subsequent years (e.g., Balfanz et al. 2006; Freer-Alvarez 2016; Rimm-Kaufman et al. 2007).

Building on earlier studies, we estimated the impact of the Knowledge in Action approach upon students' AP scores—outcomes with concrete relevance to students' post-secondary enrollment and success—after one and two years of KIA. The study also documents teachers' experiences using the KIA approach and students' experiences in those AP classes.

AP courses are intended to serve as rigorous preparation for college. When students earn qualifying AP exam scores—often a 3 or higher, though determined by individual post-secondary institutions—they earn credit accepted by many colleges as work towards graduation, which can help lower tuition costs. Earning qualifying AP exam scores also relates to other critical college outcomes, such as enrolling and persisting (Sadler, 2010; Smith, Hurwitz, & Avery, 2017).

The AP program has grown exponentially over the last two decades due to a concerted effort on the part of the College Board and school districts nationwide to expand AP course enrollment beyond already higher-performing and advantaged students—including relaxing prerequisites and encouraging more students to enroll (Finn & Scanlan, 2019; Sadler, 2010). Over the past 15 years, the number of 12th-graders who took at least one AP exam nearly doubled, from approximately 20 to 40% (College Board, 2020). Equity has been one of the College Board's core growth objectives, which they have realized with “impressive gains in access” (Kolluri, 2018, 1) among low-income (College Board, 2019) and Hispanic students (College Board, 2020). Over the past 15 years, the proportion of AP exam-takers from low-income families nearly tripled, from 11% in 2003 to 30% by 2018 (College Board, 2019).

AP is an especially challenging setting for shifting away from transmission instruction to a student-centered approach (e.g., Dole, Bloom, and Kowalske, 2016). AP teachers may be particularly inclined to follow the transmission model because of the prevailing AP culture emphasizing breadth of coverage over depth of learning, and exam preparation over any other outcome (Parker et al., 2013). Also, students have expressed concerns that anything other than transmission may not be the most effective means of preparation for exam success (Barron & Darling-Hammond, 2008; Parker et al. 2011, 2013).

Thus, the impetus for this study is to contribute to a growing body of evidence on the efficacy of PBL, this time in the AP setting and with our primary outcomes of interest the widely-known AP scores. To date, no studies have harnessed an RCT approach to determine the impact of a PBL intervention on classroom practice in an AP setting (Condliffe et al., 2017), nor examined impacts on outcomes as widely familiar—and policy-relevant—as AP examination scores. In addition, to our knowledge none have attempted to estimate PBL effects after the first year of an intervention's use within the context of an RCT study.

THE KNOWLEDGE IN ACTION INTERVENTION

KIA is a PBL approach to AP designed to support students' deeper learning of content and skills. The University of Washington researchers and local teachers who collaboratively developed the KIA curriculum envisioned a means to realize the potential yet under-realized impact AP courses could have on deeper learning for all students (Parker et al., 2011, 2013).

Six closely related concepts drawn from the learning sciences serve as KIA's theoretical foundation (Parker et al., 2013).

1. Accelerated coverage of material at a rapid pace is not strongly associated with learning depth.
2. Depth of transferable learning is preferable to breadth of temporary, non-transferable learning.
3. Critical to aligning the focus of instruction with deeper learning are assessments requiring students to demonstrate transferable deeper learning, rather than just knowledge and skills.
4. Courses should include transmission and participatory approaches, and their sequencing is critical.
5. Students' receptivity to a forced shift in the balance between transmission and inquiry approaches matters, particularly in the context of a high-stakes examination.
6. Students can develop knowledge through active approaches to learning.

KIA's "engagement first" principle—based on Schwartz and Bransford's (1998) research on instructional sequence—states that initiating learning about a topic through project work will prime students' interest and create a context for learning content through reading or lecture. KIA employs transmission instruction, but in a way intended to maximize its benefits. Other KIA design principles grounded in research on deeper learning include projects as central to the curriculum rather than supplementary; the curriculum looping "quasi-repetitively" over content and skills throughout the course of the year; teachers as co-designers, adapting the curriculum in response to their students' needs and their own; and a scalable approach.

AP U.S. Government and AP Environmental Science Courses

Though the intervention's curriculum and instruction course materials, as well as professional development (PD) supports, are course-specific, designed to respectively align to the College Board's APGOV and APES course curriculum frameworks, the same design principles apply to both courses with similar types of support resources. KIA's original developers and funders chose to focus on these AP courses for two reasons. First, well-defined, well-known, end-of-year examinations with strong psychometric properties—the AP exams administered and scored by the College Board—serve as ideal measures of student outcomes. Impact, as demonstrated through AP scores, is universally easy to interpret as practically meaningful. Second, the courses are designed to develop students' civic, political, and/or environmental awareness and engagement, areas of needed focus (e.g., Valant & Newark, 2017). Aligning with several best practices of civics education (Gould et al., 2011; Levine &

Kawashima-Ginsberg, K. 2017), KIA includes classroom instruction in civics and government; simulations of adult civic roles; student voice; “action civics;” and discussion of current events and controversial issues. Through KIA, students should learn how and why to engage civically rather than simply absorbing facts about citizenship, as is more typical in civics education (Hillygus & Holbein, 2020). Given the KIA APGOV and APES content focus, in combination with their pedagogical approaches, the courses provide opportunities for students to develop civic engagement skills.

Knowledge in Action Curriculum and Instructional Materials

The developers designed KIA to cover an entire subject-specific AP curriculum framework through project units taught over the course of an academic year. Curricula for both KIA APGOV and APES consist of five units designed to address the knowledge, concepts, and skills included in the College Board’s AP curriculum frameworks for those respective courses and examinations. Projects are the core structure within each unit and are strategically sequenced to build upon the course’s driving question. Example APGOV projects include student debates over historical and contemporary constitutional issues, mock presidential elections, and, for the culminating project, creating a political action plan intended to move an agenda item (e.g., immigration policy) through the political system (Parker & Lo, 2016). Example APES projects include students considering the ecological footprints of themselves and their families, environmental resources for farming, and, for their culminating project, assuming the role of a delegate to an international climate accord convention (Tierney et al., 2020).

Teachers access KIA’s curriculum and instruction materials through the Sprocket online portal hosted by Lucas Education Research. Sprocket offers teachers a number of tools: they can look at curriculum pages; download curriculum and/or instructional materials; upload materials to share with others; adapt existing materials to share with others; participate in an online forum discussion; request support; and organize their calendar. Though Sprocket is now open access, during the study, teachers only had access to the portal if they participated in the PD program.

Knowledge in Action Professional Development

The overarching objectives of KIA’s PD are to familiarize teachers with the KIA design principles, curriculum, and resources; support teachers’ planning for KIA curriculum and instruction; and develop teachers’ PBL instructional abilities. During the RCT, PBLWorks provided KIA’s PD and included a four-day summer institute, four full in-person days of during the year, virtual coaching, and on-demand support throughout the year.

Critical to the first and second objectives, KIA PD emphasizes to teachers the approach’s lack of curriculum “scripting.” Rather, successful KIA teaching and learning depends upon adapting the approach to specific classroom contexts. To inform the third objective of developing teachers’ PBL instructional practices, PBLWorks integrated their “Gold Standard” definitions of PBL design and practice into all aspects of the professional development program. The Gold Standard Design Elements are a comprehensive, research-based model for best practices in PBL instruction (Larmer, Mergendoller, & Boss, 2015).

The KIA PD model aligns with PD best practices by stipulating that support should be: curriculum- or content-specific (Correnti, 2007; Murray, Savin-Baden, 2000; Supovitz & Turner, 2000); provided on

an ongoing basis for substantial amounts of time (Yoon et al., 2007); aligned with standards (Geier et al., 2008); and featuring active learning and collective participation of teachers from the same communities (Garet et al., 2001). In addition, coaching can help teachers transfer the classroom skills and knowledge gained through PD, and develop norms of collaboration, experimentation, development, and collegiality (Joyce & Showers, 2002). KIA also seeks to build teacher community through its online curriculum portal and in-person PD convenings. Professional learning communities, essential for capacity-building within schools (Stoll et al., 2006), can be valuable supports for instructionally-focused reforms (Andrews & Louis, 2007; Little, 2002). We provide further details about the Knowledge in Action intervention in Appendix A.

BUSINESS-AS-USUAL SUPPORT FOR CONTROL TEACHERS

Participating districts continued to provide teachers randomized to the control condition—i.e., those offered the opportunity to participate in the KIA program the following year—business-as-usual professional development. Supports primarily included College Board-administered PD and school-level AP PD grants. Several districts provided additional support to AP teachers, including those with and without access to the KIA program, above and beyond the College Board’s annual 30-plus hour summer institutes. As examples: one district provided PD each August for all its teachers of AP, International Baccalaureate, honors, and dual enrollment courses; In another, teachers could attend monthly trainings on Saturdays through a university partnership program; another offered an AP Institute day each fall; one offered PD not specific to AP for all environmental science teachers; finally, another provided documents highlighting overlap between state standards and AP objectives. With the exception of the PBLWorks PD provided to treatment-group teachers as part of the study, no KIA districts offered PBL PD specific to AP courses. Further details about the business-as-usual support for control teachers can be found in Appendix A.

RESEARCH QUESTIONS

Our research questions address the impact of the offer to use the KIA intervention, or the intent-to-treat (ITT). Insight into the extent to which the offer to use the KIA intervention impacted students’ AP performance and other outcomes can inform whether other teachers, schools, and districts should adopt the program, and justify time and resource investments.¹

As the APGOV course addresses social studies and the APES is a science course, it is important to know whether effects were detected separately within each course group. In light of the growth and diversification of students taking the exam over the past 15 years, we also investigated the KIA effect separately within students from lower and higher income households, using student eligibility for free or reduced-price lunch as a proxy for family income. Our first research question asked:

¹ We also estimated the impact of participating in the KIA intervention, or treatment-on-the-treated (TOT). As compliance rates were high, ITT and TOT estimates were essentially equivalent.

RQ1: What is the impact after one year of the offer of the Knowledge in Action intervention—including curriculum, instructional materials, and professional development supports—on student AP examination-taking and performance, and other student outcomes?

- A.** After one year overall among the full sample
- B.** After one year within each course subsample
- C.** After one year within students of lower- or higher-income households

Since the effects of KIA after only one year of use may not have been representative of KIA’s potential after teachers have a year to familiarize themselves with the full sequence of KIA curriculum and instructional approaches, we added a second year of data collection in 2017-18. During this time, we followed teachers in schools randomized to the treatment condition into their second year of the KIA offer as well as teachers in control schools into their first year of KIA. Our second research question asked:

RQ2: Did the effect of the KIA offer on student AP examination-taking and performance differ after two years relative to after one?

In the 2017-18 year, all teachers, no matter their original assignment to treatment or control, had received the KIA offer and therefore we did not have a pure experimental comparison group with no KIA experience. Our third research question addressed whether the effect of the KIA intervention offer differed after two years relative to no access to the KIA intervention.

RQ3: What is the effect of the second year of the KIA offer on AP examination-taking and performance relative to no access?

Insight into the extent to which AP teachers who accessed the KIA intervention changed their pedagogy, as well as teachers’ and their students’ experiences with KIA, substantiates impact results while also informing decisions about program adoption. With qualitative and quantitative research methods, we thus addressed the fourth research question:

RQ4: How did teachers and students experience the KIA intervention?

RANDOMIZED CONTROLLED TRIAL RESEARCH DESIGN

We evaluated the efficacy of the KIA intervention using an RCT, randomizing schools to the conditions of treatment or control. Recruiting districts and teachers into the study depended upon factors such as approval from districts’ research review boards and teachers’ willingness to enroll. To encourage both, we, in collaboration with our funder Lucas Education Research (LER), intentionally designed the KIA evaluation as a one-year “delayed entry” RCT, in which consented teachers within randomized schools received their first offer to use the KIA intervention in either 2016-17 or the next year (i.e., a staggered rollout design).

Teachers volunteered in spring and summer 2016 to participate in the KIA RCT. One of LER’s guiding principles is that instructional shifts must be teacher-driven. For that reason—and due to the

intensity of the study’s curriculum and instruction, professional development, and research-related requirements—we recruited teachers directly. Though principals’ roles varied by district, for the most part they had minimal involvement in teachers’ decisions other than providing consent as per district requirements. Appendix B includes specific criteria for the inclusion of teachers and schools. We did not recruit students to take part in the study, nor did we advertise or announce to parents or school staff any changes in APES or APGOV curriculum.

Randomization of schools within the participating districts took place prior to the 2016 Summer Institute held locally in each school district. We used the re-randomization approach proposed in Morgan and Rubin (2012) to randomly assign 74 schools, with their 86 volunteering/consented teachers, to the treatment or control condition.² Participating treatment teachers were in schools randomized to receive the KIA intervention offer starting in 2016-17; control starting in 2017-18.

STUDY SCHOOL DISTRICTS

We partnered with five primarily urban districts, distributed geographically throughout the nation and all serving more than 50,000 students. As we show in Table 1, four of the five are in large cities. Four districts serve majority Black and Hispanic students. In three districts, approximately three-quarters or more of enrolled students live in lower-income households, as measured by eligibility for free or reduced-price meals. Each district sought to develop teachers’ capacity for PBL instruction, and required open-access AP course enrollment, such that students did not have to meet any course or prior academic achievement benchmarks if they wanted to enroll.

TABLE 1: Approximate characteristics of KIA RCT participating districts.

	DISTRICT A	DISTRICT B	DISTRICT D	DISTRICT E	DISTRICT E
Locale	City, large	City, large	City, large	Suburb, large	City, large
Enrollment > 50,000	Yes	Yes	Yes	Yes	Yes
Proportion Black and Hispanic students	76%	62%	83%	37%	82%
Proportion students qualifying for free or reduced-price meals*	72%	35%	76%	29%	83%
Participating course(s)*	APGOV & APES	APGOV & APES	APGOV & APES	APGOV	APES
Count (proportion) of randomized school sample	11 (15%)	6 (8%)	12 (16%)	12 (16%)	33 (45%)

Sources: National Council for Education Statistics Common Core of Data; district administrative data, district websites. Year of statistic provided (2014-2019) varies by source.*

² We originally stated in our summer 2017 RCT preregistration with the American Economic Association that we randomized 76 schools; however, we later determined two schools had been deterministically randomized. We did not include those two schools, or teachers and students within, in any analyses. In spring 2019, we also preregistered with the Society for Research on Educational Effectiveness.

Three districts participated in the study with both APGOV and APES teachers and students, one with only APGOV, and one with only APES. District D participated only with the APGOV course, and District D students were from higher-income households than the other districts. Districts A, C, and D each composed approximately 15% of the school sample, while District B schools represented 8%, and District E 45%. We further describe the districts, including their motivation to participate in the study and district AP policies, in Appendix C.

STUDENT OUTCOME MEASURES

Student outcome measures included AP performance outcomes from the May 2017 and May 2018 APGOV or APES exam, depending on the course in which they were enrolled and in which year. We examined KIA impacts on students' AP performance in four ways:

- » Exam-taking outcome: Whether students took the APES or APGOV exam.
- » Qualifying score outcome (full sample): This outcome encompasses all students of sample APES or APGOV teachers, no matter whether they took the exam. Students who took the AP exam and scored a 3, 4, or 5 “earned a qualifying score;” students who either did not take the exam or took it but failed to obtain at least a score of 3 did not earn a qualifying score.
- » Qualifying score outcome (exam-takers only): This outcome includes only sample teachers' students who took the APGOV or APES exam. Students who took the AP exam and scored a 3, 4, or 5 “earned a qualifying score;” students who took the exam but scored a 1 or 2 did not earn a qualifying score. Students who did not take the exam are excluded from this analysis.
- » Continuous scale score outcomes: These are the scale scores underlying students' ordinal scores, including a total score, multiple choice sub score, and free-response sub score. Like the qualifying score outcome among the exam-taker sample, this outcome is conditional on taking the examination.

The fourth measure was available in four districts, excepting District D, that provided the necessary permissions for the College Board to share de-identified student records with the research team. District D composed approximately 70% of the APGOV exam-taking sample.

In the first year, we administered an end-of-year student survey measuring students' inter- and intra-personal skills, as well as their inclination for civic engagement. We also administered the College and Work Readiness Assessment (CWRA+), which measures students' skills in critical thinking and written communication.³ In Appendix D, we provide further details on all study data sources; in Appendix E, we describe transformation of student achievement variables.

³ This summary document focuses exclusively on AP outcomes because of high student-level attrition and lack of baseline equivalence on other student outcomes. We provide other student outcome results in the accompanying appendices.

RESEARCH QUESTION 1: SAMPLE, ANALYTIC METHODS, AND RESULTS

RQ1: What is the impact after one year of the offer of the Knowledge in Action intervention—including curriculum, instructional materials, and professional development supports—on student AP examination-taking and performance, and other student outcomes?

We first addressed the study's primary research question about the impact of the offer of the Knowledge in Action intervention on student outcomes after one year by comparing Year One student outcomes of teachers in treatment schools with student outcomes of teachers in control schools. We explored the answer to this question overall, and also within each course, and within groups of students from lower- and higher-income households.

Research Question 1 Sample

We included complier and non-complier teachers, and their students, in our ITT causal analyses of the impact of KIA on student outcomes.

Without consideration for missing student outcomes, six schools, all treatment, attrited from the randomized sample before the end of September 2016. Following What Works Clearinghouse (WWC) version 4.1, this school-level attrition from randomization to the Year One sample exceeded acceptable thresholds at 8% overall, with a 16-percentage point differential between treatment and control conditions.

School attrition from the experimental sample is an important consideration relevant to districts' school-level implementation of the KIA intervention. A key driver of attrition was the lack of schools offering the APGOV or APES course every year, sometimes determining whether to offer the course based on student enrollment as of a September cut-off date, other times offering APES and AP Biology every other year, or APGOV and AP U.S. History every other year. The latter applied to teachers who taught in schools offering APGOV or APES in 2015-16 and consented to enroll in the study with the hopes that their school would again offer the course in 2016-17, and to second-year attrition described later in this report. Another driver of school-level attrition was principals not assigning consented teachers to teach APGOV or APES across consecutive years, even if the school offered the course across consecutive years. Other drivers included teachers switching schools within districts or leaving the district altogether. All school-level attrition from the randomized sample was due to the consenting teacher either not teaching APGOV or APES in 2016-17 or no longer teaching in their randomized school even if they continued to teach APGOV or APES.

School-level attrition also affected interpretation of impact results by potentially changing the composition of the schools, teachers, and students within the randomized treatment group.

After losing six schools post-randomization by the end of September of Year One (i.e., 2016-17), 74 teachers across 68 schools—and their 3,645 students—participated in the study by teaching APGOV or APES during the school year in their randomized school (intent-to-treat sample). Among the 74 teachers' students, 43% were from lower-income households and 47% of the first-year KIA student sample was Black or Hispanic.

With one school each attriting from Districts A, C, and D, and three from District E, the post-randomization school-level sample loss did not change the district composition of the school sample in a meaningful way (Appendix F). We provide details about Year One school and teacher characteristics in Appendix G, showing that changes in observed sample composition over time were quite minimal from randomization to Year One. Located in Appendices H and I are respective tables describing teachers and their students, overall and by treatment status and course, at randomization and Year One.

Research Question 1 analytic methods

We followed standard protocol for RCT analysis, assessing baseline equivalence, estimating ITT effects, and addressing sensitivity to modeling choices. Our primary analytic method for properly accounting for nested data was Hierarchical Linear Modeling (HLM). With school-level randomization and student-level outcomes, we fit two-level HLM models, grouping students within schools, with district-fixed effects to account for blocked randomization by school within districts.

Regression-adjusted models⁴ included student-, teacher-, and school-level covariates to account for remaining imbalance between treatment and control groups after randomization, and to improve statistical precision. Drawing from education literature, we chose for consideration in our impact models 22 substantively important covariates across students (e.g., prior achievement, race/ethnicity), teachers (e.g., course, baseline students' AP performance, years teaching APGOV/APES), and schools (e.g., proportion of school eligible for FRLP), as well as all student-level covariates averaged at the school level (Altonji and Mansfield, 2018). Our primary approach to covariate selection was to include in impact models all covariates with absolute baseline standardized mean differences of greater than 0.05 as well as those determined through automated covariate selection to improve model fit.

We used multiple imputation for missing covariate data, multiply imputing 20 data sets and combining results following Rubin's rules (Rubin, 1987).

Though student-level attrition did not exceed WWC thresholds for any AP performance outcomes, what did was the combination of overall and differential school-level attrition. Thus, we conducted baseline equivalence analysis to investigate the extent to which the groups differed after attrition. To meet WWC baseline equivalence standards, baseline differences between treatment and control groups on respective relevant student-level covariates must be less than an absolute value of 0.25 standard deviations. In addition, we included any relevant student-level covariates with absolute effect sizes greater than 0.05 in the impact model. Though baseline equivalence analysis of teacher- and school-level covariates is not required per WWC, the most substantial forms of attrition were at these levels, so we included those covariates as part of our baseline equivalence analysis. Therefore, baseline equivalence analysis was a necessary step to selecting which covariates to include in impact models, and informing interpretation of impact estimates.

We conducted several Year One robustness checks. To check model type, we fit our two-level HLM model as a three-level model, grouping students within schools within districts, and as an ordinary least squares model (i.e., not accounting for nesting with random effects). In addition to fitting

unadjusted models without covariates, we also fit our primary two-level HLM model with several combinations of covariates. In addition, as a check to the homogeneity of overall results by district, we estimated the magnitude of treatment effects separately by district.⁵

We also addressed the possibility that non-random sorting of students into KIA treatment or control classrooms could have biased results.

We conducted pre-registered exploratory subgroup analyses within courses and student household income groups. In addition to defining household income subgroups at the student level, as a robustness check we separately estimated effects within the two districts with lower proportions of students from lower-income households (Districts B and D) and within the three districts with higher proportions of students from lower-income households (Districts A, C, and E). We did not design the RCT sample size with enough power to detect significant effects within subgroups should they truly exist. Appendix J includes further details about these analytic steps.

Research Question 1 results

YEAR ONE OVERALL BASELINE EQUIVALENCE

For the qualifying score (exam-takers only) outcome (n=2,963), student-level attrition was 19%, with a differential of 4 percentage points, meeting the WWC “cautious boundary” threshold. Within the subgroup of exam-takers in schools with at least one continuous score outcome (n=1,599), overall attrition was 29% with a 6-percentage point differential, meeting the WWC optimistic threshold. We detail student-level analytic sample attrition across years and outcomes in Appendix K.

To inform our overall impact analyses, we analyzed baseline equivalence for three AP outcome analytic samples: the full sample of 3,645 students, all with an exam-taking outcome and a qualifying score outcome; the sample of students with a qualifying score outcome (exam-taking sample); and the sample of students with continuous score outcomes. For the first two samples, for all covariates baseline equivalence standardized mean differences fell within the 0.25 threshold defined by the WWC standards. For the third sample, including students who took the AP exam in four of five districts with which we examined the KIA impact on continuous AP outcomes (i.e., total score, multiple choice sub-score, and free-response sub-score), baseline imbalance does not exceed the WWC threshold for any student-level covariates. The covariates for which the WWC requires baseline balance, student-level prior achievement and socio-economic status, meet that threshold. Baseline imbalance for this sample exceeded the WWC threshold on two school-level covariates, both with values higher for treatment teachers: proportion of a teachers’ students who took at least one AP examination in May 2016, and proportion of a teachers’ students who were Asian-American.

No p-values associated with standardized mean differences at baseline fell below the 0.05 significance threshold. Despite school-level attrition exceeding WWC thresholds, these p-values indicated balance between treatment and control groups as expected in an RCT. To the extent we could calculate given missing data, post-attrition baseline equivalence on key covariates describing the teachers’

⁵ Though we are unable to report district-specific effects due to agreements with each participating district’s research review board, we conducted these analyses to determine whether results were concentrated in one or more district or district type (e.g., proportion of district students from lower-income households).

baseline students' AP performance did not get worse from the randomization sample to the Year One sample (Appendix G).

We included, per WWC requirements, all covariates with baseline equivalence standardized mean differences of greater than 0.05 in absolute value in the corresponding impact models. Because treatment groups were not perfectly balanced, either due to random chance or attrition, covariate inclusion in our impact models “adjusted” these baseline imbalances. For example, the prior students of teachers assigned to the treatment group had lower average AP outcomes than the prior students of teachers assigned to the control group—so without covariate adjustment, we would expect a naïve comparison of 2016-17 treatment versus control AP outcomes to be biased against the treatment group. We adjusted our impact estimates accordingly by including in the impact model imbalanced baseline covariates, such as baseline students' 2015-16 AP averages. However, a consequence is our estimated effect sizes depend on modeling the relationship between covariates and outcomes. (For details, see baseline equivalence analysis tables and figures informing Research Question One overall analyses [i.e., non-subgroup] in Appendix L.)

YEAR ONE SUBGROUP BASELINE EQUIVALENCE

We also looked at baseline standardized mean differences for AP performance outcomes within course and student household-income subgroups. With the sample sizes smaller for subgroups, we expected less balance here than for the overall sample. Focusing on the qualifying score outcome (full sample), with a few exceptions described below, we found balance within WWC thresholds on most student-level covariates for APGOV and APES subgroups, and for lower- and higher-household income students. Among the imbalanced covariates, APGOV treatment students were more likely to have taken at least one AP exam the prior year (i.e., in May 2016, $SMD=0.658$) and scored higher on their eighth-grade English Language Arts (ELA) tests—though not on their tenth-grade national ELA (i.e., PSAT, SAT, or ACT) exam. Within APES, the treatment group was composed of a higher proportion of White ($SMD=0.255$) and Asian ($SMD=0.304$) students. Within the lower-income group, White students composed a higher proportion of the treatment group ($SMD=0.279$) and treatment students scored higher on their eighth-grade math and ELA tests, though not on their 10th-grade national math or ELA exams. All student-level covariates were balanced within the higher-income group.

Within subgroups, most though not all teacher- and school-level covariate standardized mean differences were within the 0.25 threshold. Arguably the most important was the teacher-level covariate describing the proportion of teachers' baseline APGOV or APES students who earned qualifying scores on the exam out of all of their baseline APGOV or APES students. There were sizable standardized mean differences on this covariate in APGOV ($SMD=-0.434$) yet minimal in APES ($SMD=-0.031$), and sizable in the higher-household income student group ($SMD=-0.391$) yet minimal in the lower-household income group ($SMD=0.028$).

Due to smaller sample sizes, we could not have expected complete baseline equivalence within subgroups. Subgroup baseline equivalence results suggest the within-group impact results are more robust for lower-household income and APES students, and less robust for higher-income and APGOV students. (See all baseline equivalence analysis tables and figures for Research Question One subgroups in Appendix M).

YEAR ONE IMPACT ESTIMATES

OVERALL RESULTS

In covariate-adjusted models (Table 2), students of KIA teachers significantly outperformed students of control teachers on their May 2017 APES and APGOV exams, as measured by several indicators of AP performance. This held true among all 3,645 students of the sample teachers, as well as among the 2,963 students who took the exam.

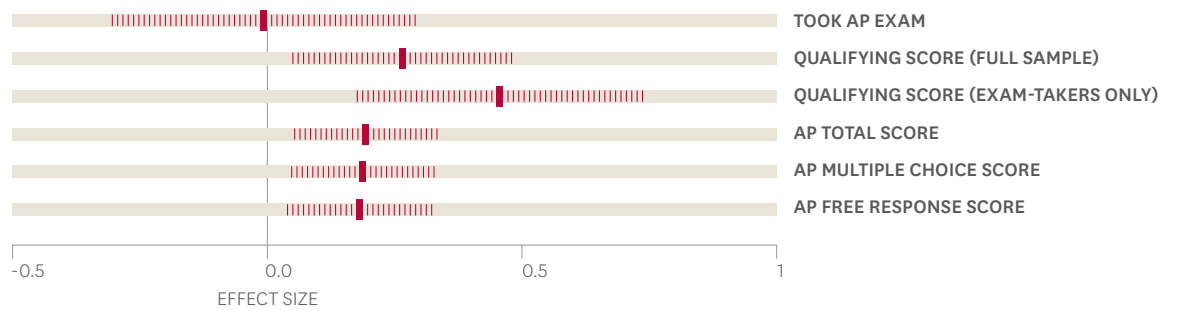
TABLE 2: Covariate-adjusted estimates of the overall Year One AP KIA impact

	EFFECT SIZE (SE)	95% CI	p-value	n
Took AP exam	-0.009 (0.15)	(-0.305, 0.286)	0.950	3645
Qualifying score (full sample)	0.264 (0.11)*	(0.049, 0.478)	0.016	3645
Qualifying score (exam-takers only)	0.457 (0.14)**	(0.177, 0.737)	0.002	2963
AP Total Score	0.192 (0.07)**	(0.054, 0.331)	0.006	1599
AP Multiple Choice Score	0.188 (0.07)**	(0.048, 0.328)	0.009	1599
AP Free Response Score	0.181 (0.07)*	(0.039, 0.323)	0.012	1599

Notes: Table columns show standardized effect sizes, standard errors, confidence intervals, p-values, and sample sizes. Asterisks denote statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We present the results from Table 2 visually in Figure 1.

FIGURE 1: Covariate-adjusted estimates of the overall Year One AP KIA impact



Notes: Figure shows standardized effect sizes and 95% confidence intervals.

We found no overall exam-taking effect (ES=-0.009, $p=0.95$), indicating no discernable overall relationship between KIA and whether students took the exam.

At the end of Year One, for all students in the Year One sample ($n=3,645$) we estimated the percentage of students who would earn a qualifying score with KIA is 4 percentage points higher than the percentage that would earn these scores without KIA (ES=0.264, $p=0.016$); dependent on the particular students in our sample, we estimated 31.1% of the students in our sample who took the AP exam would have earned a qualifying score without KIA, as compared to 35.1% with the KIA intervention.

Among students who took the exam (n=2,963), the equivalent difference was 7.6 percentage points (ES=0.457, p=0.002); again, dependent on the students, we predicted that about 37.2% of those taking the AP exam would have passed without KIA, as compared to 44.8% with the KIA intervention.

In the four districts with continuous AP score outcomes, the estimated effect sizes are significant for total scores (ES=0.192, p=0.009), as well as the multiple-choice (ES=0.188, p=0.009) and free-response (ES=0.181, p=0.012) subsection scores.

In the first column of Table 3, we show covariate adjustment was necessary to reported treatment effects with unadjusted models estimating effect sizes close to zero. Once we conditioned upon including in the impact model all covariates with absolute baseline equivalence effect sizes greater than 0.05, as required per WWC, the direction of estimated effect sizes was robust to method of selecting which covariates to include. Robustness checks also showed results were consistent across model type (i.e., 2-level HLM with district fixed effects, 3-level HLM, ordinary least squares), see Appendix N.

TABLE 3. Robustness of overall Year One KIA impact estimates to covariates.

	NO COVARIATES	PRIMARY MODEL	ALL COVARIATES WITH BE > 0.05	ALL COVARIATES
Took AP exam	0.125 (0.18)	-0.009 (0.15)	-0.007 (0.15)	0.056 (0.16)
Qualifying score (full sample)	-0.005 (0.21)	0.264 (0.11)*	0.16 (0.13)	0.351 (0.13)**
Qualifying score (exam-takers only)	-0.043 (0.21)	0.457 (0.14)**	0.145 (0.13)	0.353 (0.14)*
AP Total Score	0.035 (0.17)	0.192 (0.07)**	0.196 (0.07)**	0.188 (0.08)*
AP Multiple Choice Score	0.049 (0.17)	0.188 (0.07)**	0.196 (0.07)**	0.17 (0.08)*
AP Free Response Score	0.016 (0.16)	0.181 (0.07)*	0.185 (0.07)**	0.191 (0.07)*

Notes: Table columns show standardized effect sizes and standard errors. Asterisks denote statistical significance: *p<0.05, ** p<0.01, ***p<0.001.

In addition to the robustness tests, we also investigated the possibility of systemic sorting of students into KIA classrooms in a way that could have threatened the internal validity of our estimates. As we show in Appendices O and P, we did not find evidence of this threat.

Additionally, we estimated treatment effects separately within each district, finding estimated effect sizes on the qualifying score (full sample) outcome were substantively consistent, positive, and of meaningful magnitude.

SUBGROUP RESULTS

By course, 32 APGOV teachers across 29 schools taught 1,693 students (46%) and 42 APES teachers across 42 schools taught 1,952 students (54%); teachers of both courses taught in three schools.

Whereas across courses, the estimated exam-taking effect was zero, within courses we detected exam-taking effects. Though the exam-taking effects were not significant for either course, we did not power the study to detect significant effects within subgroups. Within the APGOV subsample, the effect was negative such that in covariate-adjusted models, KIA students took the exam at lower rates than control (ES=-0.4, p=0.09). In contrast, KIA students within the APES subsample took the exam at higher rates than control (ES=0.225, p=0.22).

Due to differences within courses concerning the characteristics of students who did and did not take the exam, with exam-takers demonstrating higher prior academic performance (Appendix Q), the KIA exam-taking effect affects interpretation of estimated impacts on the qualifying score outcome (exam-takers only) and continuous AP outcomes. The KIA exam-taking effect does not affect interpretation of the qualifying score outcome among the full subgroup samples, as we estimated this outcome for all students regardless of whether they took the AP examination. As reported above, across courses on the qualifying score (full sample) outcome, KIA students outperformed control students (ES=0.264, p=0.016) in covariate-adjusted models. When we disaggregated this estimate by course, the APGOV treatment subgroup outperformed control students, although the difference was not significant (ES=0.227, p=0.13). Within the APES subgroup, KIA students also outperformed control (ES=0.304), and this latter effect size was significant at the 0.05 level (p=0.048).

Among the full sample of 2016-17 students, 43% were from low-income households, as were 38% of exam-taking students. KIA students significantly outperformed control when we estimated effects only within the group of lower-household income students, and when we estimated effects only within the group of higher-household income students.

KIA effect sizes generally were comparable between students from lower-income households and those from higher-income households. We did not detect exam-taking effects within either student group: lower-income (ES=-0.016, p=0.92); higher-income (ES=-0.002, p=0.99). Among lower-income students, KIA students significantly outperformed control on the qualifying score outcome for exam-takers (ES=0.386, p=0.028), total score (ES=0.206, p=0.006), multiple-choice score (ES=0.191, p=0.013), and free-response score (ES=0.208, p=0.009). Among higher-income students, KIA students significantly outperformed control students on the AP qualifying score outcome in both the full sample and among those who took the exam (respectively, ES=0.319, p=0.007; ES=0.496, p=0.002), as well as the outcomes for total score (ES=0.173, p=0.03) and multiple-choice score (ES=0.183, p=0.024).

Robustness checks also showed positive KIA effects within both the lower-household income Districts A, C, and E and the higher-household income Districts B and D, and with effect sizes of comparable magnitude within district groups and, overall, across districts. (Figures and tables describing all Research Question One subgroup impact results can be found in Appendix R.)

OTHER YEAR ONE STUDENT OUTCOMES

In addition to examining the impact of KIA on AP outcomes, after the AP examination period we also administered the online College and Work Readiness Assessment (CWRA+), designed to measure students' skills in critical thinking, program solving, and written communication; and an end-of-year paper survey to students measuring their inter- and intra-personal skills, as well as their inclination for civic engagement. Due to challenges administering the 90-minute online CWRA+ during the final two weeks of school, when computers also were being used for state standardized tests, the sample of students with valid CWRA+ outcomes suffered from student-level attrition exceeding WWC thresholds in addition to the aforementioned school-level attrition. The CWRA+ and survey outcome samples lacked equivalence on measures of students' prior achievement as well as other covariates. Given these limitations, we do not highlight the results from either measure in this report, though we do share details on the accompanying technical appendices. None of the estimates for the survey or CWRA+ outcomes were significant. (Data collection, samples, and results for these outcomes are summarized in Appendices S [CWRA+] and T [survey measures].)

Research Question 1 limitations

School-level attrition exceeded WWC thresholds for the sample used to address Research Question 1, raising concerns about the extent to which attrition changed the composition, both observed and unobserved, of treatment schools relative to control schools. By performing baseline equivalence analyses, we assessed the extent to which such differential attrition could have impacted results. Baseline equivalence analysis demonstrated acceptable balance for observed covariates—partly mitigating this concern, although not completely. The differential cluster-level attrition may have changed the composition of treatment schools relative to control schools—and teachers—in ways we cannot measure or statistically adjust, and that also relate to student performance. In addition, the presented results depend on including covariates in the corresponding impact models, and modeling the relationship between covariates and outcomes.

RESEARCH QUESTION 2: SAMPLE, ANALYTIC METHODS, AND RESULTS

RQ2: Did the effect of the KIA offer on student AP examination-taking and performance differ after two years relative to after one?

Our second research question asked whether the effect of the KIA offer on student AP examination-taking and performance differed after two years relative to after one. We compared student outcomes of teachers in their second year of the KIA offer (i.e., those in schools randomized to the treatment group in Year One) with student outcomes in their first year of the KIA offer (i.e., those in schools randomized to the control group in Year One). In treatment teachers' second KIA year, they continued to have full access to curriculum and instructional materials through Sprocket. For treatment teachers in their second KIA year, professional development supports were optional, and on-demand coaching was not offered; few second-year teachers participated. In 2017-18, control teachers in their first year of the KIA offer had access to the full KIA intervention, including all the same professional development supports received by treatment teachers in 2016-17.

Research Question 2 sample

Schools and teachers from the Year One sample were included in the Year Two sample if teachers taught APGOV or APES in both 2016-17 and 2017-18 in their randomized school. Of the 74 teachers across 68 schools in the Year One sample, 53 treatment and control teachers across 50 schools persisted into the Year Two sample.

Therefore, after losing six schools post-randomization by the end of September of Year One, an additional 18 schools—10 treatment, 8 control—left the sample by September of Year Two. A school left the sample if it did not include a volunteering teacher of APGOV or APES teaching the course in 2017-18. From Year One to the Year Two, school-level attrition was 27% overall with a 11-percentage point differential between treatment and control schools, showing higher attrition in the treatment group. From randomization to the Year Two sample, school-level attrition was 32% overall with a 22-percentage point differential. Both levels of attrition exceeded WWC thresholds.

School-level attrition was not uniform across districts; rather, it was concentrated in Districts A and E, with 14 of 18 schools dropping from those two districts (seven from each; see Appendix F). According to districts' historical records of when schools offered APGOV or APES, 33% of District A schools offered APGOV or APES every year while 17% of District E schools offered APES every year.⁶ Thus, the observed high overall level of school attrition from Year One to Year Two in these two districts was unavoidable. Because both districts served high proportions of students from low-income households—71% in District A, participating in the study with APGOV and APES, and 83% in District E, which participated with APES only—the composition of schools serving primarily students from lower-income households decreased across the five districts between Years One and Two (Appendices F and G). As a result, a lower proportion of Year Two schools were Title 1 (approximately 65%) relative to Year One schools (approximately 72%).

Among the 53 teachers' 2017-18 students (n=2,946) included in the Research Question 2 sample, across treatment conditions 47% were eligible for free or reduced-price lunch (46% among exam-takers) and 45% were Black or Hispanic (43% among exam-takers).

Research Question 2 analytic methods overview

We used the same analytic methods to address Research Question 2 used to address our first research question (Appendix J). Due to effects of attrition on the sample's composition by district-level socio-economic status—as well as concerns about power—we do not report sub-group analyses in our Year Two analyses (i.e., research questions 2 and 3). Our Year Two robustness checks included addressing sensitivity to covariates and investigating the possibility that non-random sorting of students into KIA treatment or control classrooms could have biased results.

⁶ Of the five districts, three (Districts A, C, and E) provided historical course offerings in the three academic years prior to KIA implementation (2013-14 to 2015-16), whereas District B provided information for two years prior (2014-15 and 2015-16) and District D provided for one prior year (2015-16). In District A, 33% of schools offered APGOV or APES every year. The corresponding percentages for Districts B, C, D, and E were 45%, 18%, 81%, and 17%.

Research Question 2 results

BASELINE EQUIVALENCE

Attrition on the qualifying score outcome among exam-takers only (n=2,311) was within WWC thresholds. However, student-level attrition on the continuous score outcome samples exceeded the WWC thresholds, so we do not report results for this outcome (Appendix K).

For the analysis of Research Question 2 in Year Two, no student-level covariates exceeded WWC baseline equivalence thresholds, and no p-values were smaller than 0.05 on any standardized mean differences at the student, teacher, or school level. However, two teacher-level variables exceeded the WWC threshold for the qualifying score outcome in both the full sample and that of exam-takers only. Notably, baseline (2015-2016) students of treatment teachers in their second year of KIA had a lower propensity to earn qualifying scores on the APES/APGOV exam than did the baseline students of the control teachers in their first year of the KIA offer. This difference implies that, in the absence of any treatment, we would expect fewer qualifying scores among students of teachers in their second year of KIA compared to what we would expect of first-year KIA teachers' students. In addition, teachers in their second year of the KIA offer had fewer years of experience teaching APGOV or APES relative to teachers in their first year of the KIA offer, and this difference also exceeded the WWC threshold. (In Appendix U, we further discuss Research Question 2 baseline equivalence.)

IMPACT ESTIMATES

When fitting covariate-adjusted models, estimated effect sizes are all positive but non-significant (Table 4). The positive nature of the effect sizes suggests students of second-year KIA teachers may take the exam at higher rates and may perform better on the exam than do students of first-year KIA teachers. The lack of significance means these results may have been due simply to chance. We cannot determine with certainty that students of teachers in their second year of the KIA offer performed better than students of teachers in their first KIA-offer year, though we did not observe evidence of erosion of effects.

TABLE 4. Covariate-adjusted estimates of the differences in AP performance between students of teachers with two years of the KIA offer relative to one year of the KIA offer

	2 years vs. 1 year of the KIA offer			
	EFFECT SIZE (SE)	95% CI	p-value	n
Took AP exam	0.222 (0.18)	(-0.125, 0.568)	0.33	2946
Qualifying score (full sample)	0.212 (0.13)	(-0.05, 0.473)	0.34	2946
Qualifying score (exam-takers only)	0.087 (0.14)	(-0.196, 0.371)	0.57	2311
AP Total Score	0.163 (0.11)	(-0.057, 0.383)	0.13	1424
AP Multiple Choice Score	0.151 (0.11)	(-0.067, 0.37)	0.17	1424
AP Free Response Score	0.126 (0.11)	(-0.085, 0.337)	0.20	1424

Notes: Table columns show standardized effect sizes, standard errors, confidence intervals, p-values, and sample sizes.

Like for Year One, in Year Two we did not find evidence of systematic sorting of students into treatment or control classrooms (Appendices P and O).

Research Question 2 limitations

As was the case for Research Question 1, limitations to Research Question 2 results included school-level attrition and model dependence, with results contingent upon covariate adjustment. An additional limitation to the results of Research Question 2 was lack of baseline equivalence on key measured teacher-level characteristics, such that in the year prior to the KIA offer (2015-16), students of teachers in their second year of the KIA offer had notably lower APGOV/APES performance outcomes compared to students of teachers in their first KIA offer year. Adjustment for this baseline difference was crucial to our estimated impacts, with most of the effect size difference between unadjusted and adjusted results explained by the teachers' baseline students' May 2016 APGOV/APES exam performance.⁷

RESEARCH QUESTION 3: SAMPLE, ANALYTIC METHODS, AND RESULTS

RQ3: What is the effect of the second year of the KIA offer on AP examination-taking and performance relative to no access?

In the 2017-18 school year, all teachers, no matter their original assignment to treatment or control, had received the KIA offer. Therefore, we did not have a pure experimental comparison group with no KIA experience. We pursued two approaches, each with its own defined sample of teachers, to estimate differences in student outcomes after two years of the KIA intervention offer relative to zero years. Approach 1 relied on comparisons within the sample of teachers who volunteered to participate in the KIA RCT, while Approach 2 compared students of teachers who volunteered to participate in the RCT against those whose teachers did not. Limitations are more pronounced for results of this third research question relative to the first two.⁸

APPROACH 1: ESTIMATING THE TWO-YEAR EFFECT OF KIA WITHIN THE RCT SAMPLE OF VOLUNTEERING TEACHERS

First, we used a novel approach to estimate the difference between two and zero years of the KIA offer, using the sample of teachers who volunteered to participate in the RCT and continued to teach APGOV or APES in 2017-18. To estimate the effect, we broke down the estimate into two components:

⁷ To explore the consequence of the baseline difference, we fit models with the teachers' baseline students' May 2016 APGOV/APES exam performance and treatment status as the sole predictors of outcomes.

⁸ Two potential alternatives to Approach One may have been to compare treatment teachers' students in Year Two (2017-2018) to their baseline (2015-2016) pre-KIA students, or to directly compare treatment teachers' Year Two students to control teachers' Year One students. However, comparisons over time, including before-after comparisons, are well known to be inferior as causal estimands. Any natural changes over time—in students, teaching, schools, context, or outcome measures themselves—become completely confounded with the estimated treatment effect. For this reason, we restricted our analyses to approaches directly comparing treatment and control students within the same year.

the effect of one year of the KIA offer compared to none, and then any additional effect from the first year of KIA to the second.

APPROACH 1 SAMPLE

Our Research Question 3 Approach 1 sample was comprised of the 53 teachers across 50 schools who volunteered to participate in the RCT, taught APGOV or APES during the 2016-17 and 2017-18 school years, and kept teaching the course in their originally randomized school. The teachers are the same as those examined in Research Question 2—but in this approach, we included the students of the 53 teachers across both their 2016-2017 (n=3,100) and 2017-2018 (n=2,946) cohorts.

APPROACH 1 ANALYTIC METHODS OVERVIEW

We used most of the same analytic methods to address Research Question 3 as we used to address our first two research questions (Appendix J), with an important exception. To estimate the impact of two years of KIA as compared to no KIA, we first estimated the effect of one year relative to zero years of the KIA offer (like Research Question 1, but limited to students of teachers who persisted into Year Two). Then we estimated the effect of two years of the KIA offer relative to one year (Research Question 2). Finally, we added the two estimated coefficients and translated them into effect sizes as to estimate the difference between two and zero years of the KIA offer.

APPROACH 1 RESULTS

The 2017-18 student-level attrition (Appendix K) and baseline equivalence (Appendix U) results for the outcome samples used to address Research Question 2 also apply to the Research Question 3 2017-18 student sample (n=2,946). Baseline equivalence patterns for the 2016-17 student sample (n=3,100) informing Research Question 3 followed a similar pattern. Teacher-level covariates describing first-year KIA teachers' 2015-16 (baseline) students' propensity to earn a qualifying score on the APES/APGOV examination and years of experience teaching APGOV/APES did not exceed WWC thresholds for the full sample outcome (they did for the exam-takers only sample), though we observed the same direction of differences as Year Two, with standardized mean differences close to the WWC threshold.

By adding the covariate-adjusted estimates from Year 1 and Year 2 models, we observed students of teachers with two years of the KIA offer were significantly more likely to earn a qualifying score among the full sample and among only exam-takers, relative to students of teachers without the offer (Table 5).

TABLE 5. Covariate-adjusted estimates of differences in AP performance outcomes between students of teachers with one and zero years of the KIA offer, one and two years of the KIA offer, and two and zero years of the KIA offer (n=53 teachers across 50 schools).

	1 VS. 0 YEARS OF KIA OFFER			2 VS. 1 YEARS OF KIA OFFER			2 VS. 0 YEARS OF KIA OFFER	
	EFFECT SIZE (SE)	p-value	n	EFFECT SIZE (SE)	p-value	n	EFFECT SIZE (SE)	p-value
Took AP exam	0.025	0.910	3100	0.222	0.33	2946	0.246	0.410
Qualifying score (full sample)	0.374*	0.017	3100	0.212	0.34	2946	0.586*	0.046
Qualifying score (exam-takers only)	0.440*	0.016	2537	0.087	0.57	2311	0.527*	0.047
AP Total Score	0.186	0.091	1318	0.163	0.13	1424	0.349	0.052
AP Multiple Choice Score	0.204	0.081	1318	0.151	0.17	1424	0.355	0.057
AP Free Response Score	0.154	0.130	1318	0.126 0.33	0.20	1424	0.279	0.084

* $p < .05$ We calculated all 2-year vs. 0-year p -values via randomization-based inference, harnessing the original randomization scheme. We also used randomization-based inference to calculate p -values on models with singularity across one or more of 20 imputed datasets—the qualifying score (full sample) and qualifying score (exam-takers only) impact models.

While not significant, the magnitude of the effect sizes describing the KIA exam-taking effect for two versus zero years of KIA was positive and relatively high in magnitude (ES=0.246, $p=0.410$), suggesting more students of teachers with two years of the KIA offer took the exam than students of teachers with no KIA offer. Notably, 90% of the two-year exam-taking effect stems from increases in exam-taking comparing second-year KIA teachers’ students to those of first-year KIA teachers (ES=0.222), relative to 10% coming from first-year teachers’ students relative to no KIA (ES=0.025). This implies that if the KIA two-year exam-taking effect is a function of KIA, as opposed to just chance, the increases in exam-taking happen predominantly in a teacher’s second year of their KIA offer.⁹

We estimated the proportion of KIA students who would earn a qualifying score on the exam with second-year KIA teachers—among all sample students not just those who actually took the exam—would be greater than the proportion of control students by approximately 10 percentage points¹⁰ (ES=0.586, $p=0.046$). Depending on the unique characteristics of the students in the sample, we predicted 34.1% of the full sample would have earned a qualifying score on the exam without KIA,

⁹ Research Question One results showed no exam-taking effect among 3,645 students of 74 teachers across 68 randomized schools (ES=-0.009, $p=0.95$).

¹⁰ For each of the 53 teachers’ 3,100 students in the Year One sample, based on their covariates and the model comparing one year of the KIA offer to zero years of the offer, we estimated their probability of earning a qualifying score on the AP exam with no KIA. We then obtained a predicted probability of earning a qualifying score with a second-year KIA teacher by scaling up each of the estimated control probabilities according to the effect size describing two-years of the KIA offer relative to no KIA offer.

compared to 44.1% for students of teachers in their second year of KIA experience. Of the total two-year qualifying score (full sample) effect, 64% of the effect size stems from the first year and 36% from the second year. On the qualifying score (exam-takers only) outcome, conditional upon exam-taking, the effect size (ES=0.527, p=0.047) translates to an increase of 9.7 percentage points; 41.6% of the exam-taking students in our sample would earn qualifying scores without KIA (conditional upon exam-taking), as compared to 51.3 % with two years of the KIA intervention. Of this total two-year effect for exam-takers only, 83% is from the first year and 17% from the second year—meaning that the vast majority of improvement in exam scores themselves comes with a teacher’s first year of the KIA offer. (See Approach 1 sensitivity analyses in Appendix V.)

ADVANTAGES OF AND LIMITATIONS TO APPROACH 1 ESTIMATES

The key advantages of our first approach are randomization and lack of self-selection bias. Each component of our Approach 1 estimates come from a comparison of groups originally randomly assigned, albeit with subsequent attrition. Moreover, all teachers included in this approach taught in randomized schools and volunteered to participate in the RCT study, thus sharing similar unobserved characteristics, like motivation to try a new instructional approach.

Approach 1 suffers from the same limitations as Research Question 2: School-level attrition exceeded acceptable thresholds, and the analytic outcome samples lacked baseline equivalence on key measured teacher-level covariates (i.e., measures of their baseline students’ APGOV/APES performance), such that all estimates required covariate adjustment and were model-dependent.

In addition, Approach 1 assumes the impact of KIA for teachers in their first year in 2016-17 was the same as the impact of KIA for teachers in their first year in 2017-18. This assumption was likely violated for various reasons; for example, because of year-to-year contextual changes, changes in the composition of student groups from year to year, and/or because the KIA program designers implemented changes designed to improve to the KIA curriculum and professional development supports over time. Another limitation was inflated variance from summing two different estimates, resulting in decreased statistical power and greater uncertainty around estimated effect sizes (Appendix W).

APPROACH 2: ESTIMATING THE TWO-YEAR EFFECT OF KIA BY COMPARING YEAR TWO RCT TEACHERS TO A MATCHED GROUP OF TEACHERS WITH NO KIA EXPOSURE

Our second approach to estimating the difference in student outcomes between two and zero years of the KIA offer used propensity-score matching to select a comparison group of teachers with no KIA exposure from within the five participating KIA districts and comparing their students’ May 2018 AP performance to that of treatment teachers with two years of the KIA offer.

APPROACH 2 SAMPLE AND METHODS OVERVIEW

Non-volunteering teachers were among the universe of eligible teachers who did not enroll in the KIA Efficacy Study, either because they chose not to, could not attend the 2016-17 Summer Institute (a prerequisite for enrolling), or did not know about it. They did not have any access to the KIA intervention. Like volunteering teachers, non-volunteering teachers had to have taught APGOV or APES in 2016-17 and 2017-18.

Pre-match, volunteering RCT treatment teachers differed markedly from non-volunteering teachers. Notably, 29% of volunteering treatment teachers' students earned a qualifying score on the baseline (i.e., May 2016) APGOV or APES exam, relative to 47% of non-volunteering teachers' baseline students. Other measures of teachers' baseline students' academic achievement followed this same pattern. Determining which non-KIA teachers were most similar to KIA teachers at baseline presented challenges requiring multiple decision points (e.g., whether or not to allow matches within experimental schools) with tradeoffs to every decision. Ultimately, we used two different matching algorithms to choose two different matched groups, then examined the quality of both sets of matches and compared student outcomes between students of teachers with two years of the KIA offer to each of the two matched groups.¹¹

APPROACH 2 RESULTS

Even after matching, each set of comparison groups suffered from persistent baseline imbalance on student-, teacher-, and/or school-level covariates (Appendices BB and CC). Subsequent impact models fit with the two sets of matches resulted in estimates in opposite directions—one negative, one positive (Appendices DD and EE).

The marked discrepancy in baseline balance on observed measures pre-match—suggesting similar discrepancy in unobserved measures that cannot be addressed through matching or statistical adjustments—persistent baseline imbalance in the matched set of teachers (after two matching attempts), and opposing results from the two sets of models deem results from Approach 2 uninterpretable.

ADVANTAGES AND LIMITATIONS TO APPROACH 2 ESTIMATES

The key advantage of Approach 2, relative to Approach 1, was that it did not require assuming a constant first-year treatment effect from Year One to Year Two. Also, without having to sum two estimates, we would have expected greater precision of estimates from Approach 2.

The major limitation of this approach was bias due to self-selection into study conditions. The AP performance of volunteering teachers' prior students was dramatically different from non-volunteering teachers' students prior to matching. While matching and subsequent covariate adjustment can improve upon observed differences, they cannot resolve unobserved differences that may exist between teachers who volunteered to participate in an innovative and intensive PBL intervention and those who did not. The measured and potential unmeasured differences between volunteering and non-volunteering teachers after matching and covariate adjustment injects doubt into either set of results from Approach 2.

In summary, Approach 2 results were inconclusive. Future research may show the distribution of effect sizes across a range of approximately 30,000 possible matches.

¹¹ Appendix X provides the details of the first matching algorithm, in which we did not permit matched non-experimental teachers to have taught in the same school as experimental teachers, In Appendix Y are tables describing teachers in their second year of the KIA offer compared to unmatched and matched teachers, as well as the teachers' students, and schools. In the second matching algorithm, we permitted matched non-experimental teachers to have taught in the same school as experimental teachers (Appendix Z). Appendix AA describes teachers in their second year of the KIA offer compared to unmatched and matched teachers, as well as the teachers' students, and schools.

RESEARCH QUESTION 4: SAMPLE, ANALYTIC METHODS, AND RESULTS

RQ4: How did teachers and students experience the KIA intervention?

Our fourth research question addressed how teachers and students experienced the KIA intervention, including their perceived challenges and benefits. To address this question, we collected implementation-related data in both years. In Year One, we collected data from students, teachers, instructional coaches, school leaders, and district staff. In Year Two, we collected data only from teachers. The Year Two samples were limited, and we do not include those results in this document. (In Appendices FF and GG, respectively, we provide details about the Year One and Year Two implementation analyses.)

Implementation sample

We limited our implementation analyses to teachers who “complied” with their schools’ assigned treatment status. This is a subset of the RCT teacher sample. We define compliance as having participated in the KIA professional development program for at least one day or logging into Sprocket at least once.¹²

In Year One, 31 of 35 (89%) treatment teachers taught APGOV or APES during the 2016-17 school year and complied with their assigned treatment status. The 31 teachers across 27 schools were treatment “compliers,” and all 39 control teachers across 37 control schools also were compliers. Of the 70 teachers, all 31 treatment teachers used materials accessed through Sprocket while none of the 39 control teachers accessed Sprocket. KIA Summer Institute attendance among complier treatment teachers ranged from 90 to 97% (one did not participate at all). KIA Professional Development Session attendance ranged from 90% for the first session to 68-74% attendance across the three subsequent sessions.¹³

For our Year One implementation analysis, we included at least one form of collected data from all complier teachers. Collected data included surveys, instruction logs, and interviews with students, teachers, and instructional coaches, with most response rates by teacher data type above 80% and balanced between treatment and control groups.

Implementation analysis methods

To address our Year One implementation analysis research questions, we analyzed first-year field data describing teachers’ implementation of the KIA program—including surveys, instruction logs, and interviews with students, teachers, and instructional coaches—using appropriate quantitative and qualitative methods and triangulated results across participant types.

¹² Appendix II provides details about teachers’ use of Sprocket in Year One. Appendix JJ gives details about teachers’ participation in, and provision of, the KIA professional development program in Year One. In Appendix KK we share a summary of Year One professional development observations.

¹³ In Year 2, for treatment teachers in their second year of KIA implementation, we carried over complier status from Year One, as this indicator reflected their participation in the intervention when it was offered. That is, a non-complier in 2016-17 could not become a complier in 2017-18, and vice versa. For control teachers, all “complied” with control status in Year One because none received the treatment. In Year Two, we created a complier flag based upon teachers’ participation in the KIA intervention in 2017-18. Of the 23 KIA second-year teachers in 2017-18, 20 were compliers (87%). Of the 30 first-year KIA teachers in 2017-18, 26 were compliers (87%).

Year One implementation results

Drawing from triangulation of Year One data collected from complying teachers, their students, and KIA instructional coaches, this study demonstrated that teachers who used KIA's curriculum and supports changed their APGOV and APES pedagogy. Across courses, and under conditions of optimal professional learning supports and compared to randomly assigned, business-as-usual control teachers, treatment teachers in their first year of KIA implementation placed greater emphasis on deeper learning objectives, more frequently used student-centered pedagogy and in ways their students felt were authentic, and less frequently lectured or relied upon explicit exam preparation. For the most part, teachers across courses sustained their use of the KIA approach throughout the year and felt the KIA approach aligned to the AP curriculum framework and examinations. Their students reported feeling prepared for their relevant AP examinations, such as one student who explained, "I feel like the exam was a lot easier than any test we've ever taken in this class." Another student elaborated:

"I take a couple of AP classes, and I think this is one of the easier ones for me to understand and grasp more because I am a hands-on learner. All the other ones, it's kind of like you sit in a class and you take notes, and then you don't understand those notes, and then you fail the test and so on. I think this class made us more involved in what we were learning, so it was easier to grasp."

The study also illustrated a duality in which teachers and students perceived KIA as beneficial and helpful, while also challenging. Both teachers and students found the transition difficult in the area of students' comfort with responsibility for driving their own learning—a new experience for many students. Students experienced "project fatigue," wanting to rest and listen to lectures between back-to-back project units. In a group interview, a student expressed desire for balance between projects and more traditional, direct transmission modes of instruction:

"Projects were more fun than sitting and watching someone, but I feel like there needs to be a better balance between just sitting and watching someone lecture and projects. I think it needs to be almost 50/50, because you'd be able to retain the information more. Like, the projects were fun—it was just too much, too many, and not enough time."

Teachers found it difficult to balance transmission and student-centered approaches, while also experiencing issues with pacing and facilitating group work. Additionally, they voiced concern over whether the curriculum sufficiently prepared students to earn a qualifying score on the AP examinations.

The AP context is particularly challenging because of the sheer amount of content covered in the course-specific AP curriculum frameworks and the looming end-of-year, high-stakes examination. The unique AP contextual demands—on top of the challenge of shifting from a transmission to PBL approach in any context—meant the observed, sustained shift in KIA teachers' AP practice in virtually all critical aspects of instruction was not assured. Yet despite the challenges and changes in

pedagogy, including less reported time spent on explicit AP examination-taking skills, students across the two courses still felt prepared for the examinations, and the majority of teachers recommended the approach, citing benefits for themselves and their students.

Limitations to Year One implementation analysis

All data from interviews, focus groups, instruction logs, and surveys was reported from the perspectives of individual principals, teachers, and students, and so was subject to the drawbacks of self-reported data. Teachers' self-reported responses about their KIA implementation may have been particularly subject to potential over-reporting bias, as teachers may have said they implemented more KIA practices than they actually did. For this reason, we cross-referenced, or "triangulated," data across sources to verify self-reported responses from one group against other groups' responses. Though we triangulated responses, highlighting any discrepancies, response bias was a concern.

STUDY-WIDE LIMITATIONS

This study has several limitations, some of which apply across all presented analyses, while others are specific to one analysis type. We presented analysis-specific limitations throughout the report, and here state study-wide limitations.

Limitations related to external validity apply to all study results. All results apply to teachers within the participating districts who chose to enroll in the KIA RCT. The KIA Efficacy Study was, by definition, a test of the efficacy of KIA under ideal conditions. The five participating districts were not representative of all districts offering AP courses. Rather, they are districts supporting a teaching and learning approach philosophically aligned with KIA; offering AP courses at a great enough number of individual high schools to warrant inclusion in the RCT; interested enough in KIA to agree to participate; and requiring open-access AP course enrollment.

The teachers and schools volunteering to participate in the KIA Efficacy Study were "early adopters," and may not have delivered the courses in a way representative of large-scale implementation. As our complementary analysis including non-volunteering teachers demonstrates, those who volunteered to enroll in the KIA RCT were not representative of teachers who did not volunteer. Non-volunteering teachers' baseline students' APGOV/APES exam performance exceeded those of volunteering RCT treatment teachers by 18 percentage points. For the second-year analyses, results are applicable to a select group: teachers who volunteered to enroll and persisted to teach APGOV or APES in 2016-17 and 2017-18 in their randomized school.

In addition, the inextricability of the PBL approach's effect from the effect of the professional development and coaching part of the intervention means that in this study it is impossible to disentangle the separate influences of professional development and PBL on students' AP performance outcomes. Also, due to the level of professional learning supports provided, "efficacy" study conditions for teachers in their first year of the KIA offer in 2016-17 and 2017-18 were optimal compared to conditions that would be realistic in an "effectiveness" study.

IMPLICATIONS AND CONCLUSION

Despite the limitations, the results contribute to the PBL research base in several ways, with notable implications for practitioners and policymakers. Under optimal conditions of teacher support, the Year One pattern of results suggests a positive KIA impact on students' propensity to earn qualifying scores on AP examinations and on underlying continuous AP scores. Based on our primary Year One covariate-adjusted models, we estimated the percentage of all students in our sample earning a qualifying score would be about 4 percentage points higher with KIA, and those earning a qualifying score among exam-takers would be about 8 percentage points higher. Earning qualifying scores on AP examinations can earn students college credit, and relates to enrolling and persisting in college (Finn & Scanlan, 2019; Sadler, 2010; Smith, Hurwitz & Avery, 2017). At the end of the first year, we observed the positive pattern of results within both courses as well as pooled across courses. The pattern also was positive within respective groups of students from lower- and higher-income households, in districts serving a majority of students from lower-income households, in districts serving a majority of students from higher-income households, and within each of the five participating districts.

Strengthening causal claims is the RCT design, as well as the statistical significance, magnitude, and robustness of estimated effect sizes across multiple covariate-adjusted sensitivity analyses. Weakening causal claims are high school-level attrition post-randomization, and differences between unadjusted results and those statistically adjusted to address observed baseline differences.

Results also contribute to a narrow body of evidence on whether teachers' proficiency implementing PBL curriculum and instructional approaches—and, more generally, complex curricula—changes after their first year. Year Two results were less conclusive due to a lack of significance in second-versus first-year estimates, and substantial limitations to estimates of two years of the KIA offer relative to none. Though the impact of a teacher's continued experience with KIA may benefit their students, we observed the majority of gains in AP exam performance within the first year of the KIA offer. Teachers did not require more than one year of KIA experience before their students' AP performance benefitted. The only outcome with a different second-year trend was students' propensity to take the AP examination. In contrast to student AP exam performance for which we observed the majority of the effect in Year One, we observed nearly all of the exam-taking effect in Year Two. Notably, we did not observe erosion of student AP performance benefits in a teacher's second year of the KIA offer after teachers discontinued participation in KIA professional development. We estimate that the percentage of students in our sample who would earn a qualifying score is about 10 percentage points higher for students taking the course from a second-year KIA teacher, as opposed to students learning without KIA. This estimate is limited to students of teachers who persisted into our Year Two sample and applies to both the full sample of these students and only exam-takers.

Investigation of Year One implementation revealed teachers felt KIA was more engaging for students, offering the opportunity for them to develop real-world skills. Though KIA treatment teachers found the change considerable, with challenges in pacing and groupwork facilitation, the majority recommended the approach, citing benefits for themselves and students. Treatment students voiced benefits related to civic engagement, group work, engagement with learning, and examination preparation. The take-away message—students of KIA teachers outperformed students of control

on AP exams—is even more practically meaningful given both teachers and students perceived other benefits of KIA beyond exam performance.

The KIA intervention was comprised of a combination of curriculum and instructional materials, and professional development supports. Particularly in a high-stakes AP setting, shifting from primarily transmission to PBL instruction is a substantial change for teachers, suggesting the need for ongoing, job-embedded professional learning and coaching support. In a teacher’s second year, KIA supports were optional and did not include access to on-demand coaching; few teachers participated. Impact on student AP performance occurring primarily in teachers’ first year of the KIA offer aligns with the intensive provision of professional development in the first year. Lack of observed erosion of the KIA impact on students’ AP performance in teachers’ second year of the KIA offer suggests the costs to shift to PBL do not require annual payment of professional development expenses.

Another study finding that can inform districts’ implementation of the KIA intervention is the high school-level attrition between randomization and Year One, and between the first and second years. Important for district and school leaders to consider as they plan for use of the KIA intervention is the expectation that schools may not offer APGOV or APES courses across consecutive years.

Another consideration for district implementation of KIA is the marked differences between teachers who volunteered to enroll versus those who did not enroll in the KIA RCT. The prior students of teachers who volunteered to enroll performed considerably worse on the May 2016 APGOV or APES examinations than students of teachers who did not.

Of particular importance to scaling the KIA approach beyond the RCT study are treatment teachers’ self-reported perception that KIA aligned to the AP curriculum framework and examinations, and students’ feelings of learning more deeply and being prepared for the AP examinations. Critical to scaling will be KIA teachers’ positive perceptions of the approach across courses and their recommendations of KIA to others.

In conclusion, the results of the Knowledge in Action Efficacy Study support teacher-driven adoption of the KIA approach in both APGOV and APES courses, among districts with open-enrollment AP policies that support project-based learning, and for students from both lower- and higher-income households.

REFERENCES

- Altonji, J. G. & Mansfield, R. K. (2018). Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects. *American Economic Review* 108 (10) 2902-46.
- Andrews, D., & Lewis, M. (2007.) Transforming Practice from Within: The Power of the Professional Learning Community. In L. Stoll & K.S. Louis (Eds.), *Professional Learning Communities: Divergence, Depth and Dilemmas*. Maidenhead, UK: Open University Press.
- Boaler, J. (1997). *Experiencing School Mathematics: Teaching Styles, Sex and Settings*. Buckingham, UK: Open University Press.
- Balfanz, R., Mac Iver, D. J. and Byrnes, V. (2006). The implementation and impact of evidence-based mathematics reforms in high-poverty middle schools: A multi-site, multi-year study. *Journal for Research in Mathematics Education*, 33-64.
- Barron, B., & Darling-Hammond, L. (2008). *Teaching for Meaningful Learning: A Review of Research on Inquiry-Based and Cooperative Learning*. Powerful Learning: What We Know About Teaching for Understanding. Marin County, CA: Edutopia.
- College Board (2020). Archived AP data. Retrieved from <https://research.collegeboard.org/programs/ap/data/archived/ap-2019>
- College Board (2019). Student Participation and Performance in Advanced Placement Rise in Tandem Retrieved from <https://newsroom.collegeboard.org/student-participation-and-performance-advanced-placement-rise-tandem>
- Cognition and Technology Group at Vanderbilt (1992). The Jasper Series as an Example of Anchored Instruction: Theory, Program Description, and Assessment Data. *Educational Psychologist*, 27(3), 291-315.
- Condliffe, B., Visher, M. G., Bangser, M. R., Drohojowska, S., & Saco, L. (2017). Project-Based Learning: A Literature Review. Working Paper. MDRC.
- Correnti, R. (2007). An Empirical Investigation of Professional Development Effects on Literacy Instruction Using Daily Logs. *Educational Evaluation and Policy Analysis*, 29(4), 262-295
- Dochy, F., Segers, M. R., Van den Bossche, P., & Gijbels, D. (2003). Effects of Problem-Based Learning: A Meta-Analysis. *Learning and Instruction*. 13. In Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (Eds). *The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions* *Child Development*, 82(1), 405-432.
- Dole, S., Bloom, L., & Kowalske, K. (2016). Transforming Pedagogy: Changing Perspectives from Teacher-Centered to Learner-Centered. *Interdisciplinary Journal of Problem-Based Learning*, 10(1)
- Drake, K. N., & Long, D. (2009). Rebecca's in the Dark: A Comparative Study of Problem-Based Learning and Direct Instruction/Experiential Learning in Two Fourth-Grade Classrooms (Abstract). *Journal of Elementary Science Education*, 21(1), p 1-16.
- Duke, N. K., Halvorsen, A.-L., Strachan, S. L., Kim, J., & Konstantopoulos, S. (2020). Putting PjBL to the Test: The Impact of Project-Based Learning on Second Graders' Social Studies and Literacy Learning and Motivation in Low-SES School Settings. *American Educational Research Journal*.
- Duncan, A. & Ryan, A. (2021). After the siege, we need civics education. *New York Daily News*. January 9, 2021 retrieved from <https://www.nydailynews.com/opinion/ny-oped-after-the-siege-civics-education-20210109-w3zfpwu66zhdhp5f2jy3sw67re-story.html>
- Finkelstein, N., Hanson, T., Huang, C. W., Hirschman, B., & Huang, M. (2010). Effects of Problem Based Economics on High School Economics Instruction (NCEE 2010-4110). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West.
- Finn, C. & Scanlan, A. (2019). *Learning in the Fast Lane: The Past, Present, and Future of Advanced Placement*. Princeton, NJ: Princeton University Press.

- Freer-Alvarez, T. A. (2016). *The Impact of a Project-Based Learning Comprehensive School Reform on Student Achievement in a Group of High-Population Bilingual Urban Campuses* (Doctoral dissertation, Texas A&M University).
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform* (Vol. 10). New York: Psychology Press.
- Fullan, M. G. and Miles, M. B. (1992). Getting reform right: What works and what doesn't. *Phi delta kappan*, 73(10), 745-752.
- Gardner, H. (1999). *Disciplined Mind: What All Students Should Understand*. New York, NY: Simon & Schuster.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Educational Research Journal*, 38(4), 915-945.
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized Test Outcomes for Students Engaged in Inquiry-Based Science Curricula in the Context of Urban Reform. *Journal of Research in Science Teaching*, 45(8), 922-939.
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of Problem-Based Learning: A Meta-Analysis from the Angle of Assessment. *Review of Educational Research*, 75(1), 27-61.
- Gould, J., Jamieson, K. H., Levine, P., McConnell, T., & Smith, D. B. (2011). *Guardian of Democracy: The Civic Mission of Schools*. Philadelphia, PA: Lenore Annenberg Institute for Civics of the Annenberg Public Policy Center and the Campaign for the Civic Mission of Schools.
- Harris, C. J., Penuel, W. R., D'Angelo, C. M., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., Krajcik, J. S. (2015). Impact of Project-Based Curriculum Materials on Student Learning in Science: Results of a Randomized Controlled Trial. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/tea.21263/abstract>.
- Hernandez-Ramos, P., & De La Paz, S. (2009). Learning History in Middle School by Designing Multimedia in a Project-Based Learning Experience (Abstract). *Journal of Research on Technology in Education*, 42(2), 151-173.
- Hillygus, S. & Holbein, J. (2020) *Making Young Voters: Converting Civic Attitudes into Civic Action*, Cambridge, UK: Cambridge University Press.
- Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, 16(3), 235-266.
- Jackson, K., & Makarin, A. (2018). Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment. *American Economic Journal. Economic Policy*, 10(3), 226-254.
- Joyce, B., & Showers, B. (2002). Student Achievement Through Staff Development. Retrieved Jan. 16, 2018 from <https://www.nationalcollege.org.uk/cm-mc-ssl-resource-joyceshowers.pdf>.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by Design into Practice. *Journal of the Learning Sciences*, 12(4), 495-547.
- Kolluri, S. (2018). Advanced Placement: The Dual Challenge of Equal Access and Effectiveness. *Review of Educational Research*, 88(5), 671-711.
- Larmer, J., Mergendoller, J. R., & Boss, S. (2015). *Setting the standard for project based learning: A proven approach to rigorous classroom instruction*. Alexandria, VA: ACSD.
- Levine, P. & Kawashima-Ginsberg, K. (2015). *Civic Education and Deeper Learning. Students at the Center: Deeper Learning Research Series*. Boston, MA: Jobs for the Future.
- Levine, P., & Kawashima-Ginsberg, K. (2017). *The republic is (still) at risk and civics is part of the solution*. Washington, D.C.: Democracy at a Crossroads National Summit.
- Little, J. W. (2002). Locating Learning in Teachers' Communities of Practice: Opening up Problems of Analysis in Records of Everyday Work. *Teaching and Teacher Education* 18(8), 917-946.
- Maxwell, N., Mergendoller, J. R., & Bellisimo, Y. (2005). Problem-Based Learning and High School Macroeconomics: A Comparative Study of Instructional Methods. *The Journal of Economic Education*, 36(4), 315-331.

- Morgan, K. L. and Rubin, D. B. (2012), "Rerandomization to Improve Covariate Balance in Experiments," *The Annals of Statistics*, 40, 1263-1282.
- Murray, I. & Savin-Baden. (2000). Staff Development in Problem-based Learning. *Teaching in Higher Education* (5) 107-126.
- Parker, W. C., Lo, J. C. (2016). Reinventing the High School Government Course: Rigor, Simulations, and Learning from Text. *Democracy and Education*, 24(1), Article 6.
- Parker, W., Lo, J., Yeo, A. J., Valencia, S., Nguyen, D., Abbott, R., Vye, M. (2013). Beyond Breadth-Speed-Test: Toward Deeper Knowing and Engagement in an Advanced Placement Course. *American Educational Research Journal*, 50(6), 1424-1459.
- Parker, W., Mosborg, S., Bransford, J., Vye, N., Wilkerson, J., & Abbott, R. (2011). Rethinking Advanced High School Coursework: Tackling the Depth/Breadth Tension in the AP U.S. Government and Politics course. *Journal of Curriculum Studies*, 43(4), 533-559.
- Perkins, D. N., Jay, E., & Tishman, S. (1993). New Conceptions of Thinking: From Ontology to Education. *Educational Psychologist*, 28(1), 2867-2885.
- Rimm-Kaufman, S. E., Fan, X., Chiu, Y. J. and You, W. (2007). The contribution of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, 45(4), 401-421.
- Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.
- Saavedra, A. (2012). From Dry to Dynamic Civic Education Curricula. In D. Campbell, F. Hess, & M. Levinson (Eds.), *Making Civics Count: Citizenship Education for a New Generation*. Cambridge, MA: Harvard Education Press.
- Sadler, P. M., Sonnert, G., Tai, R. H., & Klopfenstein, K. (2010). *AP: A Critical Examination of the Advanced Placement Program*. Cambridge, MA: Harvard Education Press.
- Schneider, R., Krajcik, J., Marx, R. W., & Soloway, E. (2002). Performance of Students in Project-Based Science Classrooms on a National Measure of Science Achievement. *Journal of Research in Science Teaching*. 39(5), 410-422.
- Schwartz, D., & Bransford, J. (1998). A Time for Telling. *Cognition and Instruction*. 16(4), 475-522.
- Smith, J., Hurwitz, M., & Avery, C. (2017). Giving college credit where it is due: Advanced Placement exam scores and college outcomes. *Journal of Labor Economics*. 35 (1)
- Strobel, J., & van Barneveld, A. (2009). When is PBL More Effective? A Meta-Synthesis of Meta-Analyses Comparing PBL to Conventional Classrooms (Abstract). *The Interdisciplinary Journal of Problem-Based Learning*, 3(1), 4
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., and Thomas, S. (2006). Professional Learning Communities: A Review of the Literature. *Journal of Educational Change*, 7(4), 221-258.
- Supovitz, J. A., & Turner, H. M. (2000). The Effects of Professional Development on Science Teaching Practices and Classroom Culture. *Journal of Research in Science Teaching*, 37(2), 963-980.
- Thomas, J. W. (2000). A Review of Research on Project-Based Learning, 1-46. Retrieved from http://bie.org/object/document/a_review_of_research_on_project_based_learning.
- Tierney, G., Goodell, A., Nolen, S.B., Lee, N., Whitfield, L. & Abbott, R.D. (2020). (Re)Designing for Engagement in a Project-based AP Environmental Science Course, *The Journal of Experimental Education*, 88:1, 72-102
- Tugend, A. (2017). Who Benefits from the Expansion of AP Classes? *New York Times Education Issue*, September 7, 2017. Retrieved from <https://www.nytimes.com/2017/09/07/magazine/who-benefits-from-the-expansion-of-ap-classes.html>.
- Valant, J., & Newark, D. A. (2017). My Kids, Your Kids, Our Kids: What Parents and the Public Want from Schools. *Teachers College Record*, 119(12), n12.

- Walker, A. & Leary, H. (2009). A Problem-Based Learning Meta Analysis: Differences across Problem Types, Implementation Types, Disciplines, and Assessment Levels (Abstract). *Interdisciplinary Journal of Problem-based Learning*, 3(1), 12-43.
- Wieseman, K.C. & Cadwell, D. (2005). Local History and Problem-Based Learning. *Social Studies and the Young Learner*. 18 (1) 11-14.
- Yoon, K. S., Duncan, T., Lee, S. W-Y., Scarloss, B., and Shapley, K. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. U.S. Department of Education Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf
- Youniss, J. (2012) How to Enrich Civic Education and Sustain Democracy. In D. E. Campbell, M. Levinson, & F. Hess (Eds.), *Making Civics Count*. Cambridge, MA: Harvard Education Press.