High School Student Experiences and Learning in Online Courses: Implications for Educational Equity and the Future of Learning

By

Jennifer Darling-Aduana

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Leadership and Policy Studies

August 7, 2020

Nashville, Tennessee

Approved:

Carolyn J. Heinrich, Ph.D.

Jason A. Grissom, Ph.D.

Gary T. Henry, Ph.D.

Mark Warschauer, Ph.D.

Dedication

I dedicate this dissertation to my family – to my grandparents who supported my graduate education with quiet pride, to my parents who provided an emotional touchstone, to my husband, sister, and best friends who walked with me every step of the way, and most of all to my daughter – to whom I hope to be a role model and constant source of love, compassion, and support.

Acknowledgements

This dissertation would not have possible without the exceptional mentorship and guidance provided by Dr. Carolyn Heinrich. She has been instrumental in my development as a researcher, facilitating access to research partners, collaborators, and data. She encouraged me to pursue and take ownership of my own research, valued my contributions, and protected my time. She served as a role model and provided flexibility and unwavering support as I navigated the dissertation process while pregnant and parenting. She exemplifies excellence as a researcher, collaborator, and advisor.

I also thank Jason Grissom, Gary Henry, and Mark Warschauer for their invaluable feedback and support throughout the dissertation process; Annalee Good for her insight and instruction that enhanced my technical and conceptual research skills as well as my ability to build mutually beneficial relationships with school district partners; Mason Shero, Hillary Bendert, and Jacqueline Aboulafia for their excellent research assistance; Vendor researchers for their support in accessing data from the online instructional system; District administrators and educators for their assistance in providing student record data and support of the project data analysis and fieldwork; Vanderbilt University and the Wisconsin Evaluation Collaborative at the Wisconsin Center for Education Research, University of Wisconsin-Madison, for ongoing support of this initiative; and the American Educational Research Association, National Science Foundation, and William T. Grant Foundation for generous funding of this research.

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those AERA or NSF.

Table of Contents

List of Tables

Table Page

List of Figures

CHAPTER 1

Introduction

As part of a larger trend towards the privatization of education, school districts are increasingly outsourcing the education of the nation's most marginalized students to third-party vendors (Boninger, Molnar, & Murray, 2017; Burch & Good, 2014; Selwyn, 2016). One such strategy, online education, has been adopted rapidly with approximately 1 in 7 public school students at the secondary level enrolled in online courses[1] in 2015 (Gemin, Pape, Vashaw, & Watson, 2015). In many urban, high-poverty school districts, the students targeted for online courses are disproportionately students at risk of course failure or with disabilities — the very students who are most in need of in-person support and personalized learning (Heinrich, Darling-Aduana, Good, & Cheng, 2019). Yet, these students may be least likely to possess the self-regulation skills that are a prerequisite to accessing the primary benefits of online learning, such as self-pacing and anytime-anywhere access to content, facilitated by the modal, typically standardized, online course curriculum (Burch, Good, & Heinrich, 2016; Darling-Aduana, Good, & Heinrich, 2019b; Heinrich et al., 2019; Smith & Basham, 2014). Thus, the widespread reliance on and increasing use of online courses within schools serving students from marginalized backgrounds raises equity concerns regarding the potential exacerbation of socioeconomic, race, and disability-based opportunity gaps (Ahn & McEachin, 2017; Darling-Aduana et al., 2019b; Dynarski, 2018; Heinrich et al., 2019; Heppen et al., 2017). While these disparities are well documented in traditional classroom settings, I extend current literature by examining the extent to which these inequities may be magnified within the growing online learning sector.

[1] Online education refers to the delivery of content via the Internet. It may occur synchronously or asynchronously and may or may not incorporate in-person (i.e., blended) learning experiences or interactions with instructors.

The federal government, foundations, and private sector companies have subsidized the rapid growth of education technology to simultaneously meet and encourage demand from school districts looking for educational efficiencies in the face of regular budget cuts and demands for increased fiscal accountability and transparency (Bakia, Shear, Toyama, & Lasseter, 2012; Burch & Good, 2014). With support across stakeholder groups and sectors, as well as the importance of technology use in daily life, technology-based educational programs like online courses are likely to be a core component of schooling in the future. To minimize negative equity implications, a greater understanding of the underlying processes and student experiences within these programs is essential, as problematic content can impact millions of students. Yet, for similar reasons, advances in online courses could contribute to vast improvements in learning opportunities.

Further, the expansion of online courses in the United States has increased access to detailed information on students' educational experiences. From these data, it is possible to extend knowledge about student learning and engagement in online settings (Angeli, Howard, Ma, Yang, & Kirschner, 2017; Bienkowski, Feng, & Means, 2012; Jacob, Berger, Hart, & Loeb, 2016). Some findings may have applicability to traditional, face-to-face classrooms as well. For instance, standardized instructional delivery across students in many online courses allows for the isolation of the effects of varying factors, such as any benefits or challenges associated with instructional techniques that facilitate higher-order thinking or provide real-world relevance. In the face of ever-expanding reliance on educational technology in schools, leveraging detailed data on student course-taking behaviors and performance in online instructional settings may provide a key to expanding understanding of student learning processes and interactions with various online curricula and instructional techniques.

The purpose of this dissertation was to examine how students interact with and learn from online structures, course content, and instructor identity with a focus on implications for historically marginalized student populations. I used both nationally representative data and approximately one million data points collected on over 30,000 high school students between 2016 and 2018 from a large, urban district in the Midwest. Within this district, I leveraged classroom observations, interviews, and access to online course videos, activities, and assessments as well as microdata from every online student login. By integrating results from these data sources, I add to the empirical literature on both national trends and discuss mechanisms within a single district using an online course system that has partnered with over 16,000 schools, including eight of the 10 largest districts in the United States.

The first essay investigated student access to online courses nationally and the student outcomes associated with that access through an examination of the following research questions. To what extent are online courses (and for whom are) associated with changes in students' academic trajectories nationally? To accomplish this, I explored associations between attending a school offering online courses and students' high school credit accumulation and graduation as well as their postsecondary enrollment, persistence, and college selectivity. In answering this question, I was particularly interested in disparities in school-level access to online courses associated with more positive academic trajectories between students belonging to various socioeconomic, racial, and ethnic subgroups. I examined these questions through the implementation of inverse probability-weighted regression adjustment (IPWRA) and an instrumental variable approach on nationally representative data from the High School Longitudinal Study of 2009 (HSLS:09).

The second essay used a sequential mixed method design to establish the prevalence of authentic work in popular online courses and identify how students respond differentially based on variations in the level of authenticity present across lessons within a course. In the online context, authentic work consists of the opportunity to use higher-order thinking on topics with real world relevance. This analysis was conducted using courses developed by one of the largest online course providers in the nation, whose courses are used in over 16,000 schools. Specifically, I examined the following research questions. What is the prevalence of authentic work in popular online courses designed by one of the largest online course vendors in the nation? To what extent do students demonstrate different levels of engaged behaviors and achievement when working on online course lessons that ask them to engage in authentic work? By how much do students respond differentially to authentic work based on prior achievement and socio-demographic identities? This is the first study of which I am aware to systematically examine the content of online secondary courses in this manner.

Finally, in the third essay I employed a student-by-course fixed effect to explore by how much students' lesson outcomes are shaped by racial, ethnic, and gender congruence with the remote, pre-recorded instructors delivering content within the online courses. Toward that end, I investigated the following research questions. First, what is the distribution of remote and live (in-person) instructor identities in the thirty most enrolled in online courses in a large, urban district? Secondly, to what extent is being taught by an instructor of the same race/ethnicity or gender associated with course outcomes when educational content is delivered through a standardized, prerecorded video? And finally, what insight do the patterns observed when examining race/ethnicity and gender congruence in an online setting provide about the mechanisms through which positive associations in traditional, face-to-face classrooms are likely realized?

Findings from all three essays deepen understanding of the factors influencing student access, learning, and outcomes in an instructional medium with rapid adoption across the United States and draw out implications for policy and practice at the local, state, and federal level. By examining student use at this level of detail, I also aim to identify potential levers for improving curricular content, design, and delivery in online classroom settings.

CHAPTER 2

Long-Term Outcomes Associated with Online Course Access in High School: An Efficient Substitute for Face-to-Face Instruction or a Mechanism for Social Reproduction?

Approximately 75 percent of school districts in the United States offer blended or online learning, primarily to increase course offerings or for credit recovery (Powell, Roberts, & Patrick, 2015). Further, around three-fourths of the 3.8 million online courses enrolled in by public school students are in core subject areas (i.e., math, science, social studies, or language arts) (Gemin et al., 2015). Despite over 85 percent of high schools offering high school courses, little is known about student enrollment and performance in these contexts (Stevens, Frazelle, Bisht, & Hamilton, 2016; Viano, 2018). Plausibly causal research on this topic is limited to findings from specific school districts (Heinrich et al., 2019; Heinrich & Darling-Aduana, 2020; Heissel, 2016) and states (Ahn & McEachin, 2017; Cavalluzzo, Deborah, Mokher, & Fan, 2012; Viano, 2018). Subsequent paragraphs detail what is known about equity concerns in student access, performance, and learning experiences in online high school courses and highlight the gaps in current literature that this paper aims to address.

Equity Implications of Online Courses

Federal programs, such as eRate and the use of Title I funds to purchase educational technology, minimized early equity concerns regarding the digital divide – differential access to digital tools and the Internet - in schools (Hohlfeld, Ritzhaupt, Dawson, & Wilson, 2017). However, disparities persist in how those digital tools are used in schools serving more versus less affluent student populations (Warschauer, 2006; Warschauer, Knobel, & Stone, 2004), with enrollment in some online course appearing to further disadvantage lower-achieving, minority students (Hohlfeld et al., 2017; Xu & Jaggars, 2014). From an instructional standpoint, students enrolled in online courses demonstrate more engaged behaviors when technology allows students to actively manipulate their learning environment and participate in problem-solving and higher-order thinking-based activities (Darling-Aduana & Heinrich, 2018; Means, Toyama, Murphy, Bakia, & Jones, 2009; Warschauer, 2006). Unfortunately, teachers of lower-income students appear less likely to integrate these types of student-controlled, technology-based activities (Hohlfeld et al., 2017; Warschauer, 2006; Warschauer et al., 2004). Further, success in online learning environments often requires more self-regulation and discipline than learning in a traditional, face-to-face classroom setting due to features such as self-pacing (Broadbent & Poon, 2015; Kizilec, Pérez-Sanagustín, & Maldonado, 2017).

Most rigorous studies on online course-taking found that students enrolled in online courses at the post-secondary level experienced lower rates of achievement and persistence (Alpert, Couch, & Harmon, 2016; Bettinger, Fox, Loeb, & Taylor, 2017; Fischer, Xu, Rodriguez, Denaro, & Warschauer, 2020; Xu & Jaggars, 2013, 2014). The limited prior research on online course outcomes at the secondary level also indicates that there may be reasons for concern, with low rates of course completion of around 30-55 percent (Roblyer, 2006; Stevens et al., 2016; Viano, 2018). Using a comparative interrupted time-series, Viano (2018) found that beginning to offer online credit recovery courses in high school was associated with a decline in student test scores and graduation rates. Similarly, the one peer-reviewed experimental study conducted at the secondary level found that 10 percent more students earned course credit in Chicago Public Schools when randomly assigned to retake algebra one in-person versus online (Heppen et al., 2017). A second experimental study conducted on Kentucky Virtual Schools found no significant difference in outcomes between students that participated in a blended algebra one course that included online instruction for 40 percent of course content (Cavalluzzo et al., 2012).

However, there are mixed findings on longer-term outcomes. Researchers found that students randomly assigned to online or face-to-face credit recovery experienced no significant long-term effects related to high school graduation (Rickles, Heppen, Allensworth, Sorensen, & Walters, 2018). In contrast, quasi-experimental studies have found that students enrolled in online courses for credit recovery were more likely to graduate high school (Heinrich & Darling-Aduana, 2020; Hart, Berger, Jacob, Loeb, & Hill, 2019). In addition, there is evidence that online courses may be effective in increasing access to advanced courses (Heppen et al., 2012) and increasing college graduation rates when offered at the college-level (i.e., Fischer et al., 2020; Goodman, Melkers, & Pallais, 2019). For instance, a recent study identified an increased likelihood of graduating in four years and faster time-to-degree completion among college students who enrolled in a major-required course online (Fischer et al., 2020).

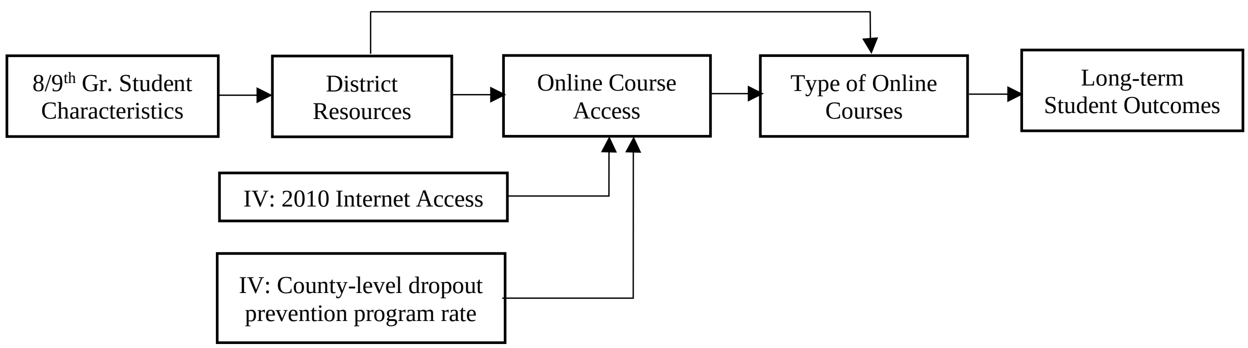
Rigorous, empirical research on the long-term ramifications of online high-school courses has yet to be conducted at the national level, while the documentation of concerning disparities in student access to high-quality online courses at the district and state level indicates the need for continued attention to this topic. Within research on education technology, there is also a need to move beyond questions of whether educational technology has the potential to enhance test scores for the modal student and whether it does so under ideal circumstances (Warschauer & Matuchniak, 2010). Consequently, the following study examines whether and for whom online courses are associated with changes in students' academic trajectories nationally using HSLS:09. Specifically, I examine associations between attending a school offering online courses and students' high school credit accumulation and graduation as well as their postsecondary enrollment, persistence, and college selectivity. Because of the potential for increased technology exposure when attending a high school offering online course-taking, I also examined associations with whether students chose a postsecondary major related to science, technology, mathematics, or engineering (STEM). I describe the theoretical framework, data and sample, analytic strategy, contributions, and limitations below.

Theoretical Framework

The achievement gap between high- and low-income families has increased over the last fifty years despite a relatively stable relationship between parental education level and student achievement (Reardon, 2011). Reardon (2011) hypothesized that a likely reason for this is that financial resources have been leveraged more over time to provide differential access to educational opportunities. Other factors such as growing degree inflation and increasing technical demands in the labor market, and thus the need for advanced degrees, may also contribute (Carnevale, Smith, & Strohl, 2013; Fuller & Raman, 2017).

Another important mediator appears to be the increasing residential segregation of families by income since the 1970s (Owens, Reardon, & Jencks, 2016). As local property taxes account for nearly half of K-12 funding, this increasing segregation places greater financial constraints on school districts serving lower-income populations, creating differential access to intellectual property based on financial resources (Ladson-Billings & Tate, 1995). Districts educating the nation's most marginalized student populations are thus more likely to be swayed by promises of increased efficiency made by technology vendors and opt for technology-based solutions that are more cost-effective but may also be of lower quality (Bakia et al., 2012). Therefore, I expect to find systematic differences not only in the extent to which schools serving less affluent student populations offer online courses but also in the effectiveness of the online courses provided to support favorable educational outcomes, as shown in the theory of action below. Because socioeconomic differences are associated with and magnified by other factors such as race, ethnicity, and English language status, I also expect to observe differences across these subgroups.

Figure 1. Online Course Access Theory of Action



My primary treatment variable of interest was student online course access during their high school experience from 2009 through 2012. However, I subsequently examined the extent to which district resources were associated with differential student outcomes based on the type of online courses offered (i.e., online courses for high achievers versus credit recovery) and the intensity of online course-taking in the school. For the intensity of use, I examined student outcomes associated with attending a school where 11 percent or more of students enrolled in any type of course online. I was interested in high-intensity use because nationally 14 percent of high school students enroll in an online course (Gemin et al., 2015). Thus, examining outcomes for students in this high-intensity group may provide helpful information for understanding the likely outcomes of students currently enrolled in online courses.

It is important to reiterate that all treatment variables are at the school level and indicate access to online courses (or a type of online course) versus online course-taking. Although this attenuates any effects observed among individual students enrolled in online courses, examining access is relevant for policy conversations on topics such as whether to invest in technology infrastructure and support state-level virtual schools. Further, the examination of access within a given school prevents the conflation of individual student-level characteristics that might be associated with assignment to online courses from driving results. This strategy also allows school-based discretion in student assignment to be considered part of the treatment, mirroring the likely implementation of any education policies related to online learning. Lastly, while there are certainly potential equity concerns in student assignment to online course-taking within schools, those topics could be better examined using different types and levels of data. The examination of access to online courses within a students' school, on the other hand, examines more directly possible opportunity gaps in access to quality educational opportunities based on disparities in school resources.

Data and Sample

This study used restricted HSLS:09 data from a cohort of ninth-grade students beginning in 2009 through 2016. Data were collected using a stratified, random sampling design; the resulting data are nationally representative. In total, 13,280 students participated through the 2016 wave of data collection, with 61 percent of those students attending schools responding to pertinent questions on student access to online courses and other covariates included in the main model specifications. I applied survey weights using the Taylor series linearization method to adjust standard errors for clustering, oversampling, and nonresponse. I also integrated information on Internet access from the National Broadband Map Data (see Figure 2), which provided information on Internet download and upload speeds by county.

Figure 2. Maximum Internet Download and Upload Speed by County (2010)

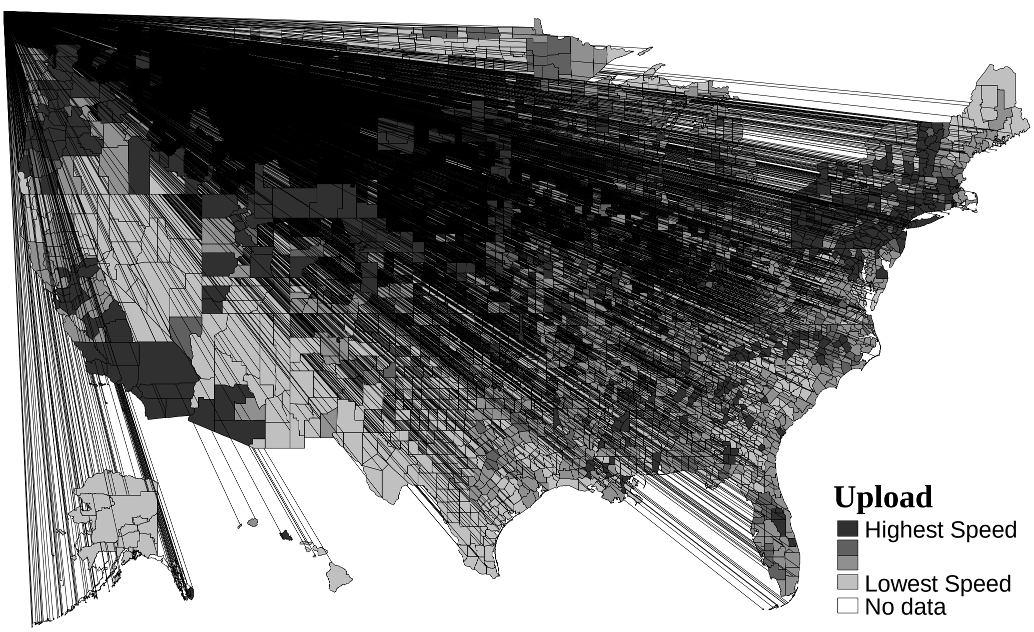


Table 1. Student Characteristics and T-test Estimates by Online Course Access[2]

		T-test Estimates			
			Type of Online Course Access		
	All Students	Any Access	Credit Recovery	High Achieving	High Intensity[3]
Female	0.494	0.486	1.179	-0.200	0.120
	(0.500)	(0.013)	(0.011)	(0.010)	(0.013)
Amer. Indian/Alaska Native	0.007	0.534	-1.392	-0.949	-3.023**
	(0.082)	(0.002)	(0.002)	(0.002)	(0.002)
Asian	0.035	1.095	0.867	0.324	-2.662**
	(0.183)	(0.007)	(0.006)	(0.005)	(0.007)
Black/African-American	0.113	1.084	2.014*	-1.303	-3.198**
	(0.317)	(0.007)	(0.006)	(0.005)	(0.007)
Hispanic	0.204	4.308***	2.180*	3.487***	-3.426***
	(0.403)	(0.009)	(0.008)	(0.007)	(0.010)
More than One Race	0.083	0.612	1.063	0.001	0.713
	(0.276)	(0.007)	(0.006)	(0.005)	(0.007)
Native Hawaiian/Pacific	0.006	0.376	1.724	-0.155	-1.523
Islander	(0.077)	(0.002)	(0.002)	(0.001)	(0.002)
White	0.553	-4.867***	-3.796***	-1.837	5.997***
	(0.497)	(0.013)	(0.011)	(0.009)	(0.013)
Non-English First Language	0.163	3.965***	2.482*	1.755	-6.579***
	(0.370)	(0.009)	(0.009)	(0.007)	(0.010)
Individualized Education Plan	0.100	-2.198*	1.537	-2.871**	-1.804
	(0.300)	(0.007)	(0.006)	(0.005)	(0.007)
Socioeconomic Status (Std.)	-0.055	0.340	-7.090***	5.722***	-0.410
	(0.752)	(0.020)	(0.017)	(0.015)	(0.021)
Ninth-grade Math Test Score	50.688	0.131	-3.660***	2.509*	-1.248
	(9.769)	(0.256)	(0.227)	(0.192)	(0.264)
Ninth-grade GPA	2.721	-3.100**	-2.536*	3.344***	-0.456
	(0.869)	(0.023)	(0.020)	(0.016)	(0.023)
Observations[4]	8,160	7,070	4,430	5,070	1,150
N (in thousands)	2,481.74	2,112.43	1,305.18	1,493.59	337.80

Standard deviations in parenthesis in the first column. All subsequent columns include standard errors from t-tests in the parenthesis. Asterisks indicate p-values from a t-test comparing to students without the type of access indicated, * p<0.10 ** p<0.05 *** p<0.01.

[2] This analysis used analytic weights and was limited to the analytic sample.
[3] High intensity is defined as 11 percent or more students enrolled in an online course within the students' school.
[4] I rounded all counts to the nearest ten based on NCES guidelines for using restricted data.

There was minimal evidence of systematic missingness based on observable characteristics between students with and without information on online course access among students with valid survey weights through the 2016 wave of data collection. The only significant observed differences included that students with information on online course access earned math test scores that were 0.15 higher on a 100 point-scale than students without information, $t=2.36$, $p=0.019$. Further, public schools were 11.7 percent more likely to provide information on online course access, $t=3.10$, $p=0.002$.

Among students in the analytic sample, 85.1 percent attended a high school that offered online courses, while only 18.8 percent attended a school that enrolled 11 percent of more of their students in online courses. Schools were coded as offering online courses if online courses were included in the schools' course catalogs at any point during the students' high school career. Although there was limited data available on online course availability during the 2008-09 and 2011-12 school years, the 85.1 percent of students identified through HSLS that attended a school with online course access is consistent with an NCES report reporting at least 75 percent of districts offered online courses (Queen & Lewis, 2011).

I chose to define high-intensity use for the purpose of this study as any school in which 11 percent or more students completed an online or distance learning course for credit for two reasons. First, the 11-24 percent range was the closest categorical response option to the current national average of online course-taking (14 percent). Second, the 11 percent cutoff was the highest categorical value on the HSLS questionnaire with a meaningful number of students attending schools at or above the cutoff (19 percent) compared to the next largest category (25 percent or more), which included only six percent of the sample. While an imperfect measure, the intensity of use in a school is important to understand underlying school processes that may be associated with either assignment to online course-taking or the quality of the online courses to which students may be assigned.

Table 2. Descriptive Statistics of Dependent Variables by Online Course Access

	High School Credits	High School Grad	Enrolled in Post-Secondary	Earned 2-Year Degree	Still Enrolled	Selective College	STEM Major
No online course access	24.705 (7.117)	0.945 (0.229)	0.861 (0.346)	0.051 (0.221)	0.566 (0.496)	0.347 (0.476)	0.147 (0.355)
Online course access	25.163 (5.904)	0.971 (0.167)	0.878 (0.327)	0.044 (0.204)	0.567 (0.495)	0.313 (0.464)	0.173 (0.378)
No online credit recovery courses	25.388 (5.455)	0.977 (0.149)	0.885 (0.319)	0.042 (0.120)	0.561 (0.496)	0.285 (0.451)	0.169 (0.375)
Credit recovery courses	25.506 (5.494)	0.975 (0.157)	0.882 (0.323)	0.045 (0.208)	0.583 (0.493)	0.328 (0.470)	0.169 (0.374)
No online high achieving courses	26.105 (5.369)	0.976 (0.153)	0.887 (0.316)	0.047 (0.212)	0.592 (0.492)	0.340 (0.474)	0.158 (0.365)
High achieving courses	25.482 (5.395)	0.978 (0.147)	0.892 (0.310)	0.044 (0.204)	0.587 (0.492)	0.334 (0.472)	0.181 (0.385)
Low intensity (< 11%)	25.697 (5.379)	0.981 (0.138)	0.900 (0.301)	0.043 (0.202)	0.597 (0.490)	0.339 (0.473)	0.177 (0.382)
High intensity (11% +)	25.420 (5.408)	0.983 (0.128)	0.906 (0.292)	0.046 (0.210)	0.600 (0.490)	0.345 (0.476)	0.175 (0.380)

Standard deviations in parenthesis.

Students were slightly less likely to attend high schools offering online courses specifically for credit recovery (M=66.2) or high achieving students (M=63.2). Students enroll in credit recovery courses to retake a course they previously failed in order to earn back course credit. In contrast, online courses for high achieving students included AP/IB and other types of enrichment courses. As shown in Table 1, the schools attended by students with higher socioeconomic status were more likely to offer online courses for high achieving students while also less likely to offer online courses for credit recovery. Similar patterns emerged by students' ninth-grade math test scores and GPA. Further, students identified as Hispanic attended schools offering online courses, including online courses for credit recovery and high achieving students, more often than other students, while students identified as White attended schools offering online courses, specifically online credit recovery courses, less often than other students. However, these trends shifted when examining student attendance at high-intensity schools, where 11 percent or more of the school population enrolled in an online course. Students identified as White attended high-intensity schools more often than the general population, while students identified as American Indian or Alaskan Native, Black, Hispanic, or with a first language other than English were less likely to attend such schools. Despite these demographic differences, I observed little variation in secondary or post-secondary outcomes of interest by online course access, as summarized in Table 2.

Analytic Strategy

Under ideal conditions, I would leverage experimental data generated through random assignment to estimate treatment effects associated with attending a school offering online courses. As these data are not available, I am concerned about several potential sources of endogeneity in my estimates, including those related to selection and simultaneity. Other than when influenced by institutional constraints or incentives, school leaders self-select into offering online courses, presumably based on beliefs of the likelihood that online courses will meet a need in the school. These needs may be related to the belief that online courses may improve student outcomes, contribute to organizational efficiency, and/or provide increased legitimacy. Concerns regarding endogeneity from these sources can be minimized by controlling for confounding variables associated with both self-selection into treatment and the academic outcomes examined. However, not all pertinent data, such as the beliefs of school decision-makers regarding the likelihood that online courses will improve student outcomes or parental pressures to integrate technology, are available. This raises concerns regarding omitted variable bias in instances where conditioning on observables may not adequately control for unobservables. Lastly, the institutional constraints, incentives, and school characteristics that contributed to self-selection into offering online courses may also be associated with the differential likelihood of participating in other types of educational programs that may also be associated with student outcomes. For instance, schools with the technology infrastructure to support online course-taking may also support blended learning opportunities in traditional, face-to-face classrooms or be more likely to use online data dashboards.

To minimize concerns associated with these potential sources of endogeneity, I first employed inverse probability-weighted regression adjustment (IPWRA) to control for the extent to which there were systematic differences in whether a students' school offered online courses based on both pre-treatment and fixed school characteristics.

$$\Pr(z_i = 1 | X_i) = \frac{1}{1+\exp[-(\boldsymbol{X}\beta)]} \tag{1}$$

As shown above, I estimated the likelihood of receiving treatment (i.e. attending a school that offers online courses) using logistic regression conditioned on a vector of baseline school-level covariates ($\boldsymbol{X}\beta$) including information about the geography and student body of each students' high school (Rosenbaum & Rubin, 1983). I also included several school-level variables associated with the likelihood of schools adopting online courses, which are used primarily for credit recovery or advanced courses. The school attendance rate and the proportion of school facility capacity filled by enrolled students represented structural limitations and potential motivations for offering online courses. Whether the school has an on-site dropout prevention program, the percentage of students who transferred to alternative school settings, and the percentage of ninth-grade students who repeated the grade served as indicators of the need for credit recovery in the school. Whether the school offered AP and IB courses represented whether the school had an emphasis on advanced course-taking. It was often only particularly low resourced schools that did not offer these advanced course options.

Table 3 shows bivariate correlation coefficients between matching and treatment variables. Notable relationships included a greater likelihood of public schools offering any online courses (ρ=0.142) and offering online courses for high achieving students (ρ=0.157). Additionally, schools with more students relative to the facility capacity (ρ=0.137) and offering IB courses (ρ=0.144) were more likely to have high intensity online course-taking.

Propensity scores were calculated using radius matching with a 0.01 caliper with no replacement. This method resulted in a 55 to 65 percent reduction in bias depending on the treatment measure and dropped no more than four percent of the analytic sample due to being outside the range of common support. This bias reduction compares favorably to the most restricted matching conducted (nearest-neighbor with a 0.001 caliper with no replacement) that reduced bias based on observables by at least 79 percent but also dropped approximately 72 percent of the sample due to lack of common support. Nearest-neighbor matching reduced the sample to such as large extent when predicting access to online course-taking due to the large percentage of schools that offered these courses, meaning there were insufficient cases in the comparison group to provide matches for all treatment cases. I also implemented radius matching at the .001 caliper with no replacement. Due to the higher rate of cases dropped due to being outside the range of common support (24 to 30 percent), the more restrictive matching procedure resulted in less bias reduction (21 to 56 percent bias reduction depending on the treatment measure).

Table 3. Predicting Treatment Variables from Matching School-Level Covariates

	Treatment Variables			
	Any Access	Credit Recovery	High Achieving	High Intensity
Public school	0.142***	-0.057***	0.157***	0.046***
School location				
- Urban	-0.072***	0.002	-0.053***	-0.016**
- Suburban	0.037***	-0.050***	-0.027***	0.016**
- Town	-0.006	-0.017**	-0.026***	-0.054***
School percentage FRL				
- Under 20% FRL	-0.041***	0.080***	-0.073***	-0.012
- 20-40% FRL	0.048***	-0.058***	0.070***	0.006
- 40-60% FRL	0.030***	0.059***	0.031***	-0.018**
- 60-80% FRL	-0.042***	-0.069***	-0.044***	-0.002
- Over 80% FRL	-0.001	-0.057***	0.030***	0.058***
On-site dropout prevention	0.057***	0.037***	0.077***	0.059***
Attendance rate	-0.021***	-0.029***	0.009	-0.005
Rate transfer to alternative	-0.005	-0.085***	0.004	-0.052***
Capacity filled	-0.021***	0.019**	-0.017**	0.137***
9^{th} repeat rate	0.047***	-0.066***	0.090***	0.064***
Offers AP	-0.020***	0.085***	-0.033***	-0.044***
Offers IB	0.067***	-0.067***	0.096***	0.144***
Number of cases	8,160	6,130	7,850	6,030

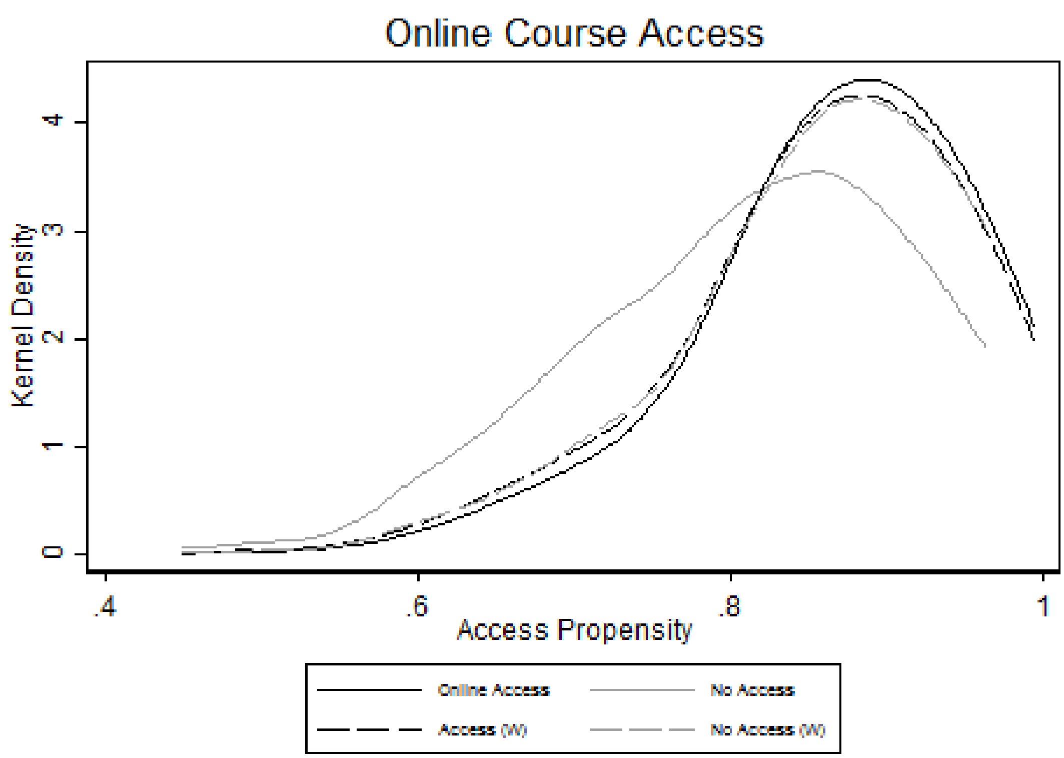
* p<0.10 ** p<0.05 *** p<0.01. Each cell represents a beta coefficient from a separate model with the school covariate indicate entered in an equation predicting the indicated treatment variable.

The radius matching procedure implemented resulted in improved balance between treatment and comparison groups on several key covariates when predicting online course access, as shown in Table 4. I also identified substantial overlap in propensity scores between treatment and control groups, as shown by the kernel density plot for online access in Figure 3. Similar plots are available for the secondary treatment variables in Appendix A. As shown, the weighted propensity score distributions contributed to additional improvements in the distribution of propensity scores between groups, which is required to minimize bias (Murnane & Willett, 2011).

Table 4. Propensity Score Matching Reduction in the Bias

		LR chi2	Mean Bias	B	R	% Var	Off Support
Any Access	Unmatched	661.77	9.7	66.9	0.97	75	---
	Matched	546.51	4.1	32.2	1.22	75	251
Credit Recovery	Unmatched	704.22	9.2	57.7	1.35	75	---
	Matched	235.12	4.1	27.1	1.02	50	415
High Achieving	Unmatched	1007.95	9.8	62.9	1.10	75	---
	Matched	254.34	3.4	25.8	0.63	88	258
High Intensity	Unmatched	561.85	8.8	63.7	0.99	63	---
	Matched	65.78	3.6	28.1	1.02	75	87

Figure 3. Probability of Online Course Access by Weighted versus Unweighted Propensity Score Distributions



After estimating the likelihood of receiving each treatment, I used the resulting propensity score to fit a weighted regression model with covariate adjustment, as shown in the second equation.

$$y_{is} = \beta_0 + \beta_1 online\_access_{is} + \boldsymbol{X_s\beta} + \boldsymbol{A_{is}\beta} + \varepsilon_{is} \tag{2}$$

This model accounts for variation in each dependent variable (y) for a given student (i) who attended their assigned high school (s) using the same covariates included in the propensity score calculation ($\boldsymbol{X}$) along with student-level covariates ($\boldsymbol{A}$). These student-level covariates include student gender, race, ethnicity, socioeconomic status, ninth-grade GPA, and ninth-grade math standardized test score as well as the student's English Language Learner and Individualized Education Plan status. More specifically, when assigning weights, treatment participants received a weight of $1/(\hat{\rho})$, while control participants received a weight of $1/(1-\hat{\rho})$ (Imbens & Wooldridge, 2009; Murnane & Willett, 2011). The advantage of this doubly robust method is that the estimates will be unbiased if either the propensity score or regression adjusted models are correctly specified. In addition to the base model described above, I examined potential heterogeneous treatment effects by interacting variables indicating whether students belonged to various historically marginalized or advantaged student populations with the treatment variable.

I also generated treatment estimates using a two-stage least squares (2SLS) instrumental variables (IV) approach to mitigate potential endogeneity in estimates calculated using IPWRA due to systematic difference in online course access as specified in the following equations (Angrist & Pischke, 2009).

$$online\_access_{is} = \gamma_0 + \gamma_1 download_s + \gamma_2 upload_s + \gamma_3 county\_dropout_s + \mathbf{X_s}\boldsymbol{\gamma} + \boldsymbol{A_{is}\gamma} + r \tag{3}$$

$$y_{is} = \beta_0 + \beta_1 \widehat{online\_access}_s + \boldsymbol{X_s\beta} + \boldsymbol{A_{is}\beta} + \varepsilon_{is} \tag{4}$$

The first stage equation predicted whether a student's school offered online courses based on a vector of the same student ($\boldsymbol{A}$) and school characteristics ($\boldsymbol{X}$) included above and by leveraging differences in Internet upload and download speeds as well as the county-level rate of on-site dropout prevention programs as excluded instruments. Then, I used the predicted estimate of online course access to predict each dependent variable of interest, controlling for the same vectors of student and school characteristics.

I entered information on the maximum download and upload speeds available in each county during 2010, the first year in which the National Broadband Map published data, as separate variables in the first-stage equation. The maximum download and upload speeds available in each county represented institutional constraints, a common source for excluded variables because these constraints are generally unrelated to an individual students' innate ability or motivation (Angrist & Pischke, 2009). There is also precedence. Skinner (2019) successfully leveraged variation in local Internet access as an exogenous predictor of enrollment in online courses at the post-secondary level. In addition, I included an aggregated measure of the proportion of schools in the county that offered on-site dropout prevention programs as a measure of the need for credit recovery, one of the most common uses of online course-taking at the secondary level (Powell et al., 2015; Queen & Lewis, 2011). (Refer to Appendix B for a map demonstrating variability across counties for those counties where one or more schools participated in the HSLS:09.) Aggregate measures such as this are often useful excluded instruments because they represent institutional constraints without being conflated with school-level decision-making, which also may be associated with access to treatment (Boozer & Rouse, 2001).

Table 5. Tests of the Exclusion Restriction for High School Credits

	Online access	Credit recovery	Higher achieving	High intensity (11% +)
High School Credits				
First-stage R-squared	0.200	0.245	Exog.	Exog.
First-stage F-statistic	14.559	17.754		
2SLS size of 5% Wald test	10-15%	10-15%		
Sargan Overidentification Test	2.427	3.355		
	(0.297)	(0.187)		
High School Graduation				
	Exog.	Exog.	Exog.	Exog.
Enrolled in College				
	Exog.	Exog.	Exog.	Exog.
Earned 2-Yr Degree				
First-stage R-squared	Exog.	Exog.	0.260	Exog.
First-stage F-statistic			22.605	
2SLS size of 5% Wald test			5-10%	
Sargan Overidentification Test			0.669	
			(0.716)	
Still Enrolled				
First-stage R-squared	Exog.	Exog.	0.257	0.221
First-stage F-statistic			22.964	9.714
2SLS size of 5% Wald test			5-10%	15-20%
Sargan Overidentification Test			4.054	5.186
			(0.132)	(0.075)
College Selectivity				
	Exog.	Exog.	Exog.	Exog.
STEM Major				
	Exog.	Exog.	Exog.	Exog.

P-values in parenthesis. Exogeneous indicates that assignment to treatment is independent and thus use of OLS versus IV is preferred due to its greater efficiency. Exogeneity of indicated treatment variables established using the Wu-Hausman test statistic (p>.05).

Post-estimation tests, summarized in Table 5, demonstrated that the excluded variables predicted assignment into treatment, when endogenous. In instances where assignment into treatment was identified as exogenous, I did not report estimates from the IV models, as the IPWRA results were more efficient. For models where the use of IVs was required to produce unbiased estimates, the Sargan Overidentification test results provided support that the exclusion restriction likely held. The estimates presented in Table 5 do not include survey weights or robust standard errors, as the post-estimation tests used do not accommodate these adjustments. However, I did use these adjustments in the main IV estimates.

Contributions and limitations. This study provides the first nationwide examination of the equity implications of high school student access to online courses. The use of HSLS allows for an examination of long-term trajectories but also captures access to online courses in a policy context where online course-taking was slightly less prevalent. For instance, since 2009, federal funding through programs such as eRate has increased access to the Internet in schools. Thus, current trends may vary slightly, and additional research is necessary to determine the extent to which these findings hold today. Additionally, while the study provides evidence for policy questions regarding whether providing access to online courses influences students' academic trajectories, results do not speak to any potential systematic assignment mechanism within schools. Further, the examination of access to treatment at the school-level likely results in attenuation bias, with effects of treatment on the students treated likely higher than those identified. However, alternatively, the identification of treatment at the school-level also allows estimates to capture any spillover effects for students attending the school who did not enroll in online courses.

The methodological rigor of the study is another important contribution. One of the primary advantages of IPWRA is that it limits the analysis to students within the range of common support, excluding students whose likelihood of access to online courses is so high (or low) that there are no viable comparison cases. When matching variables account for assignment into treatment, the use of propensity scores also minimizes baseline differences between treatment and comparison cases that may introduce bias in estimates of the treatment effect (Imbens & Wooldridge, 2009; Rosenbaum & Rubin, 1983). The use of an IV approach further minimizes bias due to endogeneity by only using information from the exogenous variation in the excluded instruments to estimate the effect of treatment (Wooldridge, 2013). This reduction in bias requires that the potential excluded instruments both predict assignment to treatment and only influence the outcome indirectly through assignment to treatment. Valid excluded instruments would meet these criteria by influencing district policy decisions but not individual student outcomes within those districts except through online course access. I provided evidence that the excluded instruments should not be included in the second-stage equation using the Sargan overidentification test.

From a methodological standpoint, although the use of IPWRA is an improvement on naive OLS estimates, it is unlikely to remove all potential sources of endogeneity. Matching only supports causal inference when the variables in the first-stage model fully account for selection, which I am unable to fully accomplish in this study. While the IV approach has the potential to address bias due to endogeneity, resulting estimates are causal only for the population of students for whom access to online course-taking was associated with one or more of the excluded instruments, meaning that estimates may not generalize to students who would have always or never have had access regardless of local Internet access and county-level dropout prevention. Further, the exclusion restriction required for unbiased IV estimates is untestable, and thus I can present only circumstantial evidence that the conditions required for unbiased causal estimates hold. Lastly, IV is a less efficient estimator than OLS, increasing finite sample bias, particularly in instances where there may be concern regarding a weak instrument and as the number of instruments increases (Angrist & Pischke, 2009). While this is less of a concern due to the large sample size as well as the strength and the low number of excluded instruments, it merits noting. However, if estimates are consistent, the use of both IPWRA and IV approaches provide evidence of the extent to which the above stated limitations inherent to each model may have biased estimates and thus in tandem provide greater support for the plausibility of causality.

Results

I first examined the extent to which attending a school with online course access was associated with high school and post-secondary outcomes, as shown in Table 6. On average, students who attended a school that offered online courses earned a comparable number of high school credits. Although IPWRA estimates indicated that students attending a school offering online courses earned 1.2 more high school credits ($\rho<0.05$), the IV estimates identified a negative (versus positive) relationship. The coefficients from the more rigorous, IV model indicated that students earned 2.4 fewer high school credits when they attended a school offering online courses. While the IV estimate was not significant, they indicate that positive associations identified in other model specifications may be due to endogenous factors.

Students were also more likely to graduate high school, enroll in college, and major in a STEM field when they attended a school offering online courses. According to the IPWRA model, students with online course access were 3.4 percent more likely to graduate high school, 5.5 percent more likely to enroll in college, and 4.6 percent more likely to enroll in a STEM major. Since assignment to treatment was identified as exogenous in these models, IPWRA results are preferred to those from the IV model due to the increased efficiency of IPWRA. Despite these positive associations, students with online course access in high school were no more likely to earn a two-year degree, be still enrolled in college in February 2016, or attend a more selective college.

Table 6. Propensity Score Matching and Instrumental Variable Results, Dependent Variable: Online Access

	OLS with Survey Weights	IPWRA	IV with Covariates
High School Credits	0.688	1.218**	-2.362
	(0.584)	(0.520)	(4.007)
High School Graduation	0.043*	0.034***	Exog.
	(0.023)	(0.012)	
Enrolled in College	0.062**	0.055***	Exog.
	(0.026)	(0.018)	
Earned 2-Yr Degree	-0.020	-0.014	Exog.
	(0.016)	(0.011)	
Still Enrolled	0.020	0.017	Exog.
	(0.033)	(0.029)	
College Selectivity	0.000	0.006	Exog.
	(0.032)	(0.029)	
STEM Major	0.035*	0.046***	Exog.
	(0.019)	(0.017)	

* p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

In contrast, estimates indicated mostly less favorable outcomes associated with attending a school that offered online courses for credit recovery (see Table 7). I identified no relationship between online credit recovery access and credits earned, high school graduation, college enrollment, two-year degree attainment, being still enrolled in February 2016, or STEM major. However, according to IPWRA estimates, students who attended a high school that offered online courses for credit recovery were 3.9 percent more likely to attend a moderately or highly selective college. Since I identified no positive associations with college enrollment, this increase in college selectivity may indicate a possible spillover effect for students who already had a high probability of attending college when their classmates requiring credit recovery were placed in an alternative, online classroom setting.

Table 7. Propensity Score Matching and Instrumental Variable Results, Dependent Variable: Credit Recovery

	OLS with Survey Weights	IPWRA	IV with Covariates
High School Credits	-0.343	-0.270	-5.786
	(0.370)	(0.375)	(5.184)
High School Graduation	0.005	0.003	Exog.
	(0.007)	(0.007)	
Enrolled in College	0.001	-0.004	Exog.
	(0.015)	(0.014)	
Earned 2-Yr Degree	-0.001	-0.000	Exog.
	(0.010)	(0.010)	
Still Enrolled	0.030*	0.025	Exog.
	(0.018)	(0.018)	
College Selectivity	0.040**	0.039**	Exog.
	(0.017)	(0.017)	
STEM Major	-0.004	-0.001	Exog.
	(0.014)	(0.014)	

* p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

Table 8. Propensity Score Matching and Instrumental Variable Results, Dependent Variable: High Achieving

	OLS with Survey Weights	IPWRA	IV with Covariates
High School Credits	-0.967***	-0.912**	Exog.
	(0.361)	(0.400)	
High School Graduation	-0.006	-0.005	Exog.
	(0.010)	(0.011)	
Enrolled in College	0.005	0.006	Exog.
	(0.015)	(0.015)	
Earned 2-Yr Degree	-0.005	-0.001	-0.289
	(0.009)	(0.008)	(0.193)
Still Enrolled	-0.024	-0.023	-0.704
	(0.021)	(0.019)	(0.446)
College Selectivity	-0.022	-0.023	Exog.
	(0.018)	(0.016)	
STEM Major	0.008	0.009	Exog.
	(0.015)	(0.015)	

* p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

Students whose schools offered online courses for high achieving students earned approximately one fewer high school credit (see Table 8). However, there were no significant associations identified between access to online courses for higher achieving students and longer-term outcomes. Similarly, I identified no significant, differential student outcomes associated with attending a school with high-intensity online course enrollment (see Table 9), defined as 11 percent or more students enrolled online, by model specification.

Table 9. Propensity Score Matching and Instrumental Variable Results, Dependent Variable: High Intensity

	OLS with Survey Weights	IPWRA	IV with Covariates
High School Credits	-0.070	0.017	Exog.
	(0.438)	(0.470)	
High School Graduation	0.007	0.005	Exog.
	(0.008)	(0.010)	
Enrolled in College	0.023	0.020	Exog.
	(0.016)	(0.017)	
Earned 2-Yr Degree	-0.002	-0.000	Exog.
	(0.010)	(0.010)	
Still Enrolled	0.022	0.018	-0.953
	(0.028)	(0.029)	(0.986)
College Selectivity	0.020	0.018	Exog.
	(0.023)	(0.024)	
STEM Major	-0.013	-0.018	Exog.
	(0.017)	(0.017)	

* p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

Heterogeneous effects. Because my theoretical framing of online course access indicated that students from lower-income families, minoritized backgrounds, and disadvantaged geographic areas were likely to be exposed to online courses of differential quality, I next examined heterogeneous effects by interacting student characteristics with the primary treatment estimate. The estimated effects from these interaction terms were presented in Table 10. Because results were qualitatively similar across model specifications and the IPWRA estimates were more efficient than the IV estimates, I presented only results from the IPWRA models. (Results from the interaction between online course access for students by gender, other racial/ethnic groups, ELL and IEP status, and ninth grade GPA are available in Appendix C.) In addition, due to the smaller sample size for the high intensity online course-taking variable, I only present results for the other three treatment variables to avoid drawing conclusions from the experiences of only a small number of individuals belonging to a certain subgroup attending schools with high intensity online course-taking.

I observed few differences by student socioeconomic status apart from students attending a high school offering online courses for credit recovery being 2.9 percent more likely to enroll in college for each one standard deviation increase in socioeconomic status. Results varied more widely by student ethnicity. The largest differential effect occurred for students identified as Hispanic who were 11.4 percent less likely to attend a selective college when attending a school offering any time of online courses and 5.4 percent less likely to earn a two-year degree when attending a school offering online courses for higher achieving students. The negative interaction between access to online courses for higher achieving students and earning a two-year degree among Hispanic students persisted when examining models employing instrumental variables. These heterogeneous effects highlight how online course access appears to provide greater benefits to students from historically advantaged groups, such as students with higher SES. In contrast, students identified as Hispanic were less likely to benefit.

Table 10. Heterogeneous Effects, Interaction between Student Characteristic and Access to Online Course Treatment (IPWRA Estimates)

	Dependent variable	Online access		Credit recovery		Higher achieving	
Student Characteristics							
Black/African-American	H.S. Credits	-0.995	(1.620)	-0.366	(0.821)	-0.174	(0.739)
	H.S. Grad	-0.019	(0.029)	-0.029	(0.028)	0.057	(0.049)
	Enrolled	0.076	(0.071)	0.042	(0.033)	-0.015	(0.035)
	Earned 2-Year	0.035	(0.024)	-0.001	(0.021)	0.000	(0.021)
	Still Enrolled	-0.165	(0.137)	-0.010	(0.063)	-0.112	(0.079)
	Selectivity	-0.125	(0.132)	0.094	(0.071)	-0.095	(0.058)
	STEM Major	0.054	(0.052)	-0.008	(0.055)	0.011	(0.053)
Hispanic	H.S. Credits	2.181	(2.086)	-0.269	(0.658)	-0.355	(0.599)
	H.S. Grad	0.019	(0.030)	0.019	(0.016)	-0.002	(0.014)
	Enrolled	-0.060	(0.045)	-0.023	(0.049)	-0.024	(0.038)
	Earned 2-Year	-0.014	(0.026)	-0.033	(0.026)	-0.054**	(0.023)
	Still Enrolled	0.174	(0.123)	0.020	(0.089)	0.010	(0.080)
	Selectivity	-0.114**	(0.054)	-0.056	(0.043)	-0.036	(0.047)
	STEM Major	0.032	(0.056)	-0.011	(0.045)	-0.057	(0.045)
SES	H.S. Credits	-0.866	(0.619)	0.415	(0.275)	-0.035	(0.271)
	H.S. Grad	-0.029	(0.022)	-0.003	(0.008)	-0.018	(0.011)
	Enrolled	0.002	(0.030)	0.029*	(0.016)	0.014	(0.016)
	Earned 2-Year	-0.028	(0.021)	-0.002	(0.014)	-0.009	(0.010)
	Still Enrolled	-0.018	(0.037)	0.014	(0.025)	-0.003	(0.024)
	Selectivity	0.041	(0.030)	0.000	(0.023)	0.031	(0.022)
	STEM Major	-0.001	(0.031)	-0.003	(0.024)	0.003	(0.022)
School Characteristics							
Rural	H.S. Credits	-0.884	(1.007)	0.558	(1.123)	-0.346	(0.906)
	H.S. Grad	-0.040	(0.027)	-0.034**	(0.013)	-0.015	(0.015)
	Enrolled	-0.019	(0.048)	-0.064**	(0.029)	0.009	(0.041)
	Earned 2-Year	0.006	(0.027)	0.001	(0.020)	0.008	(0.017)
	Still Enrolled	-0.003	(0.063)	-0.035	(0.051)	0.029	(0.044)
	Selectivity	0.028	(0.052)	-0.007	(0.040)	-0.039	(0.036)
	STEM Major	0.046	(0.032)	-0.008	(0.037)	0.070**	(0.030)
High Poverty (70% or More FRPL)	H.S. Credits	3.424*	(1.756)	-0.005	(1.562)	-3.209***	(1.183)
	H.S. Grad	0.144*	(0.073)	0.066	(0.073)	-0.062	(0.080)
	Enrolled	0.162**	(0.066)	0.082	(0.125)	0.027	(0.044)
	Earned 2-Year	0.056*	(0.031)	-0.011	(0.048)	0.046*	(0.027)
	Still Enrolled	0.112	(0.113)	-0.014	(0.090)	-0.051	(0.091)
	Selectivity	-0.122	(0.078)	0.044	(0.071)	-0.091	(0.063)
	STEM Major	0.084	(0.064)	0.038	(0.063)	0.045	(0.069)

* p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

I also examined heterogeneous effects by whether the school a student attended was in a rural area or served a population where 70 percent or more of students qualified for free or reduced-price lunch (FRPL). I hypothesized that because access to different types and quantities of resources varied in each setting, I would see differential associations with student outcomes. Students attending schools where 70 percent or more students qualified for FRPL experienced an additional benefit to attending a school that offered any online course. In these cases, students were 16.2 percent more likely to enroll in college. Positive associations were also observed for this group in terms of high school credits, high school graduation, and earning a two-year degree, although these estimates were not significant at the 0.05 level. However, students attending high-poverty schools also earned 3.2 fewer high school credits, on average, when attending a school offering online courses for high achieving students. In contrast, students attending a rural school were seven percent more likely to major in a STEM field when their high school offered online courses for high achieving students. Students attending rural schools were also less 3.4 less likely to graduate high school and 6.4 percent less likely to enroll in college when their school offered online courses specifically for credit-recovery.

Discussion

Starting at the most immediate outcome of online course-taking in high school, students attending a high school that offered online courses earned an equivalent number of high school credits. While not necessarily contradictory, this finding should be placed in the context of other studies identifying low rates of online course completion (Heppen et al., 2017; Roblyer, 2006; Stevens et al., 2016; Viano, 2018). This means that if reforms were enacted to improve the rate of online course completion, the increase in the number of high school credits earned could potentially be higher. Examining variation in credits earned also focuses, by its very nature, predominately on students who might not otherwise have earned enough credits to graduate without online course access, since few students are likely to enroll in more courses than required for high school graduation.

Students enrolled in a high school that offered online courses were also 3.4 percent more likely to graduate high school, 5.5 percent more likely to enroll in college, and 4.6 percent more likely to major in a STEM field. These findings are consistent with longitudinal studies conducted at the district level that identified improved rates of high school credit accumulation, high school graduation, and post-secondary enrollment (Heinrich et al., 2019; Heinrich & Darling-Aduana, 2020). Importantly, however, any benefits did not appear to persist beyond college enrollment or major selection into post-secondary persistence or attainment (Heinrich & Darling-Aduana, 2020).

In contrast, students attending high schools offering online courses specifically for credit recovery or high achieving students or enrolling 11 percent or more students in online courses did not appear to benefit in the same manner as students attending high schools offering any online course access did. The one possible exception to this statement was a positive association between college selectivity and attending a high school offering online credit recovery, which may indicate a possible positive spillover benefit in terms of college selectivity for students already likely to attend college when classmates requiring credit recovery were assigned to an online-based course. Null associations between online credit recovery courses and high school graduation were consistent with research from Chicago Public Schools showing that students randomly assigned to an online versus face-to-face course to retake Algebra 1 were no more likely to graduate high school (Rickles et al., 2018). The lack of associations between online credit recovery and student outcomes varies from the decline in graduation rate that Viano

(2018) identified in North Carolina but could be explained by the national versus state focus and varied time frame.

The identification of positive associations between long-term outcomes and general online course access - but not when online courses were used for credit recovery, high achieving students, or in high-intensity - offers suggestive evidence that students may benefit more from the option of online course-taking when targeted more selectively for enrollment (versus high-intensity use) and for students requiring less personalization (i.e., NOT students requiring extensive academic or self-regulatory assistance or conversely enrichment). Non-significant associations, however, do not necessarily preclude the use of online courses in these settings if there are other non-instructional benefits to either the student or school. For instance, if students are likely to achieve comparable outcomes, they may prefer the flexibility of anytime, anywhere access some online courses may offer (Jaggars, 2014). Alternatively, a school may find this instructional method preferable if it delivers comparable student outcomes more efficiently in terms of the financial or human capital inputs required (Bakia et al., 2012).

One of the most important contributions of this study, however, may be the identifications of differential associations by student and school characteristics. Negative differential associations among students identified as Hispanic compared to positive differential associations among students of higher SES demonstrate that online course access has the potential to magnify existing educational opportunity gaps. In contrast, with potentially positive equity implications, students attending schools where 70 percent or more students qualified for FRPL benefited more academically when attending a school offering any online course. Also, students attending a rural school were seven percent more likely to major in a STEM field if their school offered online course-taking for high-achieving students. Advanced, online courses may be particularly beneficial in this setting due to less general knowledge regarding STEM fields and access to advanced coursework in STEM among students residing in rural settings.

Implications for Policy and Practice

One of the challenges of conducting longitudinal research, particularly on topics related to technology, is how quickly educational practices change. Study findings highlight the experiences of a nationally representative group of students who attended high school from the 2008-2009 through 2011-2012 school years (assuming a four-year high school experience) through their secondary and post-secondary educational experiences. By 2012, many states allowed, and some even provided state-sponsored, online courses, with approximately 275,000 students attending fully online schools and an estimated five percent of K-12 students participating in some form of online or blended learning program each year (Watson, Murin, Vashaw, Gemin, & Rapp, 2012). Since 2012, there has been an expansion in broadband access due to programs such as eRate and an overall increase in the use of online courses across the nation (Digital Learning Collaborative, 2019; Hohlfeld et al., 2017). By 2019, approximately 310,000 students attended fully online schools, with over 4.5 million students (or around eight percent of K-12 students) enrolling in one (or more) supplemental online course through their state virtual school or a district-based digital learning program (Digital Learning Collaborative, 2019; Gemin et al., 2015).

Because of the changing online learning landscape, it is important to look at larger trends versus specifics when using findings to inform policy and practice. First, this study - and related studies using more recent, district and state-level data - demonstrate that access to online courses has the potential to improve student outcomes such as high school graduation and college attendance (Hart et al., 2019; Heinrich & Darling-Aduana, 2020; Heinrich et al., 2019; Rickles et al., 2018). This indicates there is likely a place for technology-based courses in high schools. What appears to reduce the magnitude of any possible gains to student attainment is the use of online courses to serve students in need of more personalized instructional experiences (i.e., students in need of credit recovery or advanced coursework). Until online course developers improve the ability of course software to differentiate learning to students who fall outside the *modal user* online systems are designed for in terms of aptitude, prior knowledge, self-regulatory skills, and cultural background, policy-makers and practitioners should exercise caution in assigning students to online courses who may be unlikely to benefit (Cottom, 2014; Heinrich et al., 2019; Selwyn, 2016). Further, potentially for similar reasons, the use of online courses at high levels of intensity within a school may minimize potential benefits. As high-intensity online course-taking was defined in this study as 11 percent of more students enrolled online, this high-intensity group has policy relevance, since currently 14 percent of high school students nationally enroll online. This study demonstrates the potential promise to the use of online, high school courses in improving educational attainment in most settings when used judiciously. To the extent that online course developers and schools continue to improve the digital learning tools and systems to support their use, online courses have the potential to become an important tool in the educational tool kit used to improve student learning and educational attainment.

CHAPTER 3

Authenticity, Engagement, and Performance in Online High School Courses: Insights from Micro-Interactional Data

High school students are increasingly enrolling in online courses to fulfill graduation requirements (Gemin et al., 2015). Yet, the often dry, direct-lecture heavy instruction observed in the most popular online course systems appears designed to meet seat-time requirements without supporting deep, or even sometimes surface-level, learning (Heinrich et al., 2019; Darling-Aduana et al., 2020). With credit recovery one of the most common reasons for enrollment, online course-taking has the potential to perpetuate and even magnify the social reproduction that occurs through the enactment of the hidden curriculum, the tacit teaching of norms, values, and dispositions in educational institutions that often disadvantage students from marginalized backgrounds (Anyon, 1980; Apple, 2004).

The purpose of this study is to examine how students interact with and learn from online course content and instructional tasks with a focus on implications for students belonging to marginalized subgroups. I focused my examination on authentic work due to concerns raised by Haberman (2010) and others that students from predominately low-income backgrounds are exposed to a *pedagogy of poverty* focused primarily on remembering and recitation versus critical thinking and application (Anyon, 1980; Darling-Hammond, 2001; Oakes, 2005) with negative ramifications for student engagement, learning, and well-being (Apple, 2004; Au, 2012; Bernstein, 1975; Solorzano & Yosso, 2001). A research team and I quantified the extent to which curricular content and instructional activities in course videos and assignments facilitated authentic work through the development and use of the Authentic Online Work Rubric. The rubric asked raters to evaluate the extent to which courses provided opportunities for higher-order thinking and real-world relevance, two primary components of authentic work (Marks, 2000; Newmann, Marks, & Gamoran, 1996; Reeves, Herrington, & Oliver, 2002).

The analysis relies on data collected from approximately one million student logins made by over 25,000 high school students enrolled in online courses during the 2016-17 and 2017-18 school years. All students' primary enrollment locations were schools in a large, urban district in the Midwest. The most common reason for online course-taking in the district was credit recovery of a course previously failed by the student. I leveraged access to online course videos, activities, and assessments as well as microdata from every online student login for the study. These data sources allowed me to examine micro-interactional processes between students and the online course interface within a single district using one of the largest online course systems in the nation. The online course vendor studied has partnered with over 16,000 schools, including schools in all 50 states and within eight of the 10 largest districts in the United States. Further, as the district serves a predominately low-income, minoritized student population, I was able to focus my analysis on the learning and engagement implications for students belonging to marginalized subgroups.

Applying a sequential mixed method design, I used ordinary least squares (OLS) regression with student-by-course fixed effects combined with descriptive analysis to determine in what ways and by how much students' in-course learning behaviors were related to variations in access to authentic work between course lessons[5] within a single district. Specifically, I examined the following research questions.

1. What is the prevalence of authentic work in popular online courses designed by one of the largest online course vendors in the nation?

[5] Lessons are subcomponents of the larger course that contain three to six tasks around a single subtopic. Lesson tasks commonly include watching an instructional video, completing practice problems, and submitting a quiz.

2. To what extent do students demonstrate different levels of engaged behaviors and achievement when working on online course lessons that ask them to engage in authentic work?
3. By how much do students respond differentially to authentic work based on prior achievement and socio-demographic identities?

This study represents an examination of the content of online secondary courses at a level of detail not currently available in the research literature. Attention to these practices in both online and face-to-face classrooms can support the achievement and well-being of high school students who have been marginalized and/or alienated by current institutional structures (Au, 2012; Yair, 2000).

Authentic Work in Face-to-Face and Online Settings

Authentic work allows students to create their own meaning and connect course material to their lives (Apple, 2004; Herrington & Herrington, 2007; Lebow & Wager, 1994). Students with access to these types of enriching experiences invest more effort and demonstrate higher rates of engagement in learning (Marks, 2000; Newmann, Wehlage, & Lamborn, 1992). Providing opportunities for students to engage in higher-order thinking with real-world relevance is also associated with improved academic performance, critical skills, thoughtfulness, and in-depth knowledge (Apple, 2004; Means et al., 2009; Newmann, 1992; Newmann et al., 1996). For instance, using a regression discontinuity design, Dee and Penner (2017) found increases in student attendance (by 21 percent), GPA (1.4 points), and credits earned (23 credits) when students enrolled in an ethnic studies course that encouraged critical thinking and resonated with students' lived experiences by focusing on themes such as social justice, discrimination, stereotypes, and social movements. Further, authentic work may serve as a protective factor in the prevention of alienation, particularly for students at greater risk of disengaging from instruction (Au, 2012; Yair, 2000).

Specific to technology-based tasks, students learn more when completing activities that prompted them to reflect, think critically, and provide explanations for work (Bixler, 2008; Golanics & Nussbaum, 2008; Nelson, 2007; Saito & Miwa, 2007; Suh, 2006). Crippen and Earl (2007) examined student performance and self-efficacy when randomly assigned to receive guided examples and prompts to talk aloud to themselves to explain their reasoning while problem solving. Compared to students who completed the online activity without additional supports, those who were assigned to the experimental condition achieved higher grades on both multiple choice and problem-based assessment questions (Crippen & Earl, 2007). An experiment by Gao and Lehman (2003) found that students achieved higher scores on end-of-lesson true/false, multiple choice, and word problems when randomly assigned to receive immediate feedback on responses during the lesson and generate their own examples using course content when reviewing a webpage. Both strategies are designed to trigger more higher-order thinking. Students in the experimental groups in this study also spent more time on-task than students in the control group (Gao & Lehman, 2003). Further, two separate studies found that when students learning a programming language or statistics had access to a more interactive tool for compiling, saving, and running code or analyzing data, students achieved higher levels of competency as measured by midterm, final, and end-of-course grades (Cavus, Uzonbovlu, & Ibrahim, 2007; Dinov, Sanchez, & Christou, 2008). These benefits may be due to more opportunities for student-generated (versus teacher-generated) knowledge and clearer alignment between the school-based skills being developed and real-world applications.

A smaller number of studies, all conducted at the postsecondary level, identified only a partial, or no clear, relationship between student learning and similar instructional activities also designed to facilitate authentic work, such as participant inquiry, active responses, and access to an organizer tool (Chen, 2007; Cook, Gelula, Dupras, & Schwartz, 2007). Chen (2007) found that students learned better when randomly assigned to use a concept map but not when randomly assigned to use a text outline, likely due to the higher level of higher-order thinking required when creating a concept map. There was also evidence of heterogeneous effects, with both treatments benefiting students with lower prior academic track records more (Chen, 2007). This finding may explain some of the null results identified in other studies, such as the Cook et al. (2007) study, which used medical students as participants. As medical students have generally experienced high rates of prior academic success, they may learn well regardless of the level of authenticity present in instructional tasks. Although there is mixed evidence of the impact of authentic work at the post-secondary level, the relationship identified in the comparatively fewer studies at the secondary level appears generally positive. There remains a need to establish the prevalence of authentic work in online courses, including the extent to which authentic work is feasible in these settings, and how students completing online courses predominately for credit recovery respond to authentic work when present.

Defining Authentic Work

Several researchers have defined authentic work. Marks (2000) conceptualized authentic work as (a) asking students to solve new and interesting questions, (b) prioritizing deep dives into a single topic, (c) applying content to situations outside of school, and (d) communicating ideas with others. The related Framework for Authentic Intellectual Work emphasizes the importance of student construction of knowledge through higher-order thinking, disciplined inquiry, and value beyond school (Newmann et al., 1996). Disciplined inquiry requires students to demonstrate and communicate deep understanding by building upon prior knowledge, while value beyond school necessitates an application outside of the school context (Newmann et al., 1996). Similarly, Reeves and colleagues (2002) defined authentic work in an online context as consisting of complex tasks that were open to multiple interpretations to allow for competing solutions and a diversity of outcomes. Development of these complex tasks is often accomplished by integrating real-world examples that draw on students' existing *funds of knowledge*, the knowledge students gain through participation in daily familial and community life (Lebow & Wager, 1994; Moll & González, 2004).

I consolidated these definitions for the purpose of this study, conceptualizing authentic online work as requiring both opportunities for higher-order thinking and real-world relevance. An important caveat is that many conceptualizations of authentic work require students to "engage in extended conversational exchanges with the teacher and/or their peers about subject matter in a way that builds an improved and shared understanding of ideas or topics" (Newmann et al., 1996). The online courses observed did not have systems in place to facilitate these types of interactive, social, and communication-based instructional activities. Because I intended to leverage variation in instructional activities, I did not include this construct in the Authentic Online Work Rubric generated for this study. Below, I define higher-order thinking and real-world relevance in greater detail and summarize how they have been operationalized in prior research.

Newmann (1992) conceptualized higher-order thinking in the classroom as requiring the posing of challenging questions or tasks, sustained examination of a few related topics, appropriate time to think, and expectations for reasoned communication (also Hiebert et al., 2005; Stein, Grover, & Henningsen, 1996). Munter, Stein, and Smith (2015) extended upon this work through the definition of an instructional model that emphasized student generation of knowledge (versus direct instruction) using dialogue, collaborative work, real-time feedback, and student ownership. The facilitation of higher-order thinking, and student-generated knowledge more specifically, therefore requires open-ended (rather than closed-ended) assessment questions, practice problems, and other instructional tasks that allow for multiple correct answers based on how students choose to define the various tasks required to complete an assignment (Gamoran & Nystrand, 1992; Land, Bartell, Drake, Foote, Roth McDuffie, Turner, & Aguirre, 2018; Lebow & Wager, 1992; Reeves et al., 2002; Stein et al., 1996). This type of instructional activity also supports deeper understanding of underlying processes by allowing students to examine content from multiple perspectives (Hill & Hannafin, 2001; Reeves et al., 2002; Reeves & Reeves, 1997).

Real-world relevance likewise takes a variety of forms. At the most basic level, real-world relevance involves embedding instructional tasks in a meaningful context (Hiebert et al., 2015; Hunsader et al., 2013; Lebow & Wager, 1994; Newmann, 1992). For instance, opportunities for students to address a social problem encourage critical thought while providing information and skills essential for civil discourse and action in a democratic society (Apple, 2004; Au, 2012; Griner & Stewart, 2013). Providing the opportunity for students to create meaningful work product can also give classroom work meaning outside of educational contexts (Au, 1998; Brown, Collins, & Duguid, 1989; Reeves et al., 2002).

Theoretical Framework

Despite the academic and psychological benefits of authentic work, access is provided disproportionately to students from more affluent backgrounds, while students from less affluent backgrounds are more often taught in a manner that emphasizes rote memorization and respect for authority (Anyon, 1980; Darling-Hammond, 2001; Oakes, 2005). Thus, students from less affluent backgrounds are often subjected to what Haberman (2010) termed a *pedagogy of poverty*, which includes teachers "giving information, asking questions, giving directions, making assignments, monitoring seatwork, reviewing assignments, giving tests, reviewing tests, assigning homework, reviewing homework, settling disputes, punishing noncompliance, marking papers, and giving grades" (p. 82). The result of these educational processes is often a sense of alienation among students, both from school and society more broadly (Au, 2012).

Apple (2004) asserted that this dichotomy, termed the *hidden curriculum*, occurs through "the tacit teaching to students of norms, values, and dispositions that goes on simply by their living in and coping with the institutional expectations and routines of school" (p. 13). Educators, often unconsciously, integrate the hidden curriculum because schools are organized to provide the skills their students are perceived to need later in life, which may include intellectual flexibility for students who may one day serve as managers versus habit formation and compliance for their employees (Anyon, 1980; Apple, 2004). The implicit assumption is that students attending more affluent schools, or tracked into higher level courses, are more likely to serve as managers upon graduation, while students serving in less affluent schools, or tracked into lower level courses, are more likely to serve in positions with limited autonomy. However, it is in part through the implementation of the hidden curriculum that these self-fulfilling prophecies occur, as many students internalize evidence of academic gaps as personal failures and "evidence of their own innate intellectual deficits" (Apple, 2004, p. 90; Bernstein, 1975; Solorzano & Yosso, 2001).

The hidden curriculum is further reinforced within and across schools by curricular standardization, such as that facilitated through the incursion of private vendors into education through online courses (Au, 2012). Since differentiation is costly, most online course developers design content for the modal student, “disembodied from place, culture, history, markets, and inequality regimes” (Cottom, 2014, p. 45). The result is online courses that do not cater to the academic and personal realities of students whose lived experiences lie outside the White, middle class norms of American society (Darling-Aduana et al., 2020). Further, without opportunities for teachers to adapt content to local contexts, online course content may represent a greater encroachment of dominant values, norms, and expectations into schooling, potentially exacerbating the alienation of students from lower-income and minoritized backgrounds.

Although these patterns appear across schools, they are also replicated within schools, where students enrolled in more advanced courses tend to be exposed to more authentic work (Gamoran & Nystrand, 1992; Newmann, 1992; Oakes, 2005). As students often demonstrate initial resistance when exposed to authentic work due to less defined metrics for success and the need for new skill acquisition (Gamoran & Nystrand, 1992; Hoffman & Ritchie, 1997), initial attempts at increasing authentic work may appear to reinforce beliefs that certain students may not benefit from this type of instruction. In these ways, the beliefs underpinning the hidden curriculum are reinforced, legitimizing the practices that contributed to the systemic inequality in the first place.

In fact, some researchers caution that an emphasis on authentic work may inadvertently harm students belonging to marginalized subgroups. Kirschner and colleagues (2006) argued that students are only likely to successfully complete the more autonomous, open-ended assignments that facilitate authentic work with sufficiently high prior knowledge and self-regulatory skills.

Up until that point, students benefit from more structured, direct-instruction (Kirschner et al., 2006). Although only correlational, the OECD (2016) identified a positive relationship between direct instruction and science achievement in all but three countries. More recently, a meta-analysis found that although English Language Learners (ELLs) made learning gains when taught using an inquiry science instruction model, which incorporates techniques consistent with authentic work, non-ELL students made larger gains (Estrella, Jeaggi, & Collins, 2018). This raises concerns that increasing expectations of students to complete authentic work may exacerbate rather than minimize existing achievement disparities.

Despite the importance of authentic work and prevalence of online, high school courses, there is a dearth of information on the micro-interactional process between students and online course content that form the building blocks of knowledge acquisition in online instructional settings. Drawing on prior research within the larger framework of critical curriculum studies, I intend to contribute to understanding of how students shift their behavioral patterns based on exposure to online curricular material and instructional tasks that reflect various levels of authenticity. I subsequently examine the ramifications of variation in authentic work for the academic achievement of students belonging to historically marginalized populations enrolled in online courses. I focus within this analysis specifically on those courses to which students were regularly assigned versus all available courses.

Hypothesis. When students are exposed to higher rates of authentic work, I expect to observe higher rates of behavioral engagement, captured through an increase in active time and decrease in idle time logged in the online course-taking system. Increased behavioral engagement may translate into higher achievement if students have prior experience with authentic work that will allow them to successfully complete more intellectually challenging content. Thus, I anticipate observing heterogeneous associations with student performance conditional on students' prior academic experiences, which I use as a proxy for prior exposure to authentic work based on the work of Oakes (2005) and others on tracking. More specifically, I hypothesize that students who have been more academically successful in the past are better able to translate access to authentic work into improved learning. I also expect to see differential associations by student identity (i.e., whether students belong to dominant or minoritized subgroups), with some attempts at real-world relevance – particularly those that assume White, middle class life experiences – more likely to engage students who identify with dominant cultural groups (see also Darling-Aduana et al., 2020).

Methods

The study incorporated a sequential mixed method design, whereby initial, inductive qualitative analysis informed research question and scale development (Creswell & Clark, 2011). I developed and validated a reliable measure of authentic online work that was used in the coding of approximately 200 hours of online course content. A descriptive analysis of the qualitative themes and patterns observed within lessons was then corroborated and supplemented with statistical analysis that leveraged student-by-course fixed effects to identify variation in student performance and behaviors when exposed to different levels of higher-order thinking and real-world relevance within the same course. This method allowed for the exclusion of endogenous variation in student achievement and engagement associated with student and course-specific information that remained constant during the semester when students enrolled in the course. An overview of each data collection and analysis stage is shown in Table 11, with additional information provided in the following section.

Table 11. Research Design Overview

Research Stage	Accompanying Tasks
Research question generation	The research team and I drafted, analyzed, and discussed analytic memos from interviews with in-person lab instructors and observations of the in-district labs where students completed online courses to generate questions and hypotheses. These data were collected as part of a larger study on online learning.
Rubric development	I developed the original Authentic Online Work Rubric, generating items based on a review of literature and pre-existing scales. Items were added, dropped, and refined based on feedback from experts in online learning and authentic work as well as initial pilot coding.
Lesson observations	Using the newly developed rubric, I trained and established interrater reliability with three research assistants to help rate each online lesson identified for review on items related to higher-order thinking and real-world relevance. Raters recorded information on topics such as the instructor and order of thinking instructional tasks required. The rubric also asked raters to generate narrative vignettes of the lesson content, instruction, and assessment.
Scale validation	Throughout the lesson coding process, my research assistants and I discussed the clarity and overlap between items. We refined (and recoded) items as needed and ultimately decided to drop unclear or repetitive items. I then used confirmatory and exploratory factor analysis to better understand the relationship between items and to affirm the appropriate items to consolidate into each scale. Ultimately, I generated two scales using IRT, one for higher-order thinking and one for real-world relevance.
Descriptive analysis	I summarized data collected using the Authentic Online Work Rubric including the frequency that lessons required students to use various orders of thinking and the relationships between the two subscales identified. I also used rating scale information to identify patterns in typical and exemplary models of higher-order thinking and real-world relevance observed across lessons.
Statistical analysis	I used rating scale and course vendor data to identify variations in student behavior and learning when exposed to different levels of higher-order thinking and real-world relevance across lessons within the same course in aggregate, and by subgroup.

Data and sample. The quantitative component of this study relied on administrative data provided by a large, urban school district in the Midwest that serves a predominately low-income, minoritized student population. As shown in Table 12, 76 percent of students in the district qualified for free or reduced-price lunch, 61 percent identified as Black, and 21 percent identified as Hispanic during the study period. Data were provided for the 2016-17 and 2017-18 school years for all ninth through twelfth-grade students. For each student in a given year, the district provided data on achievement and sociodemographic variables. Among students enrolled in online courses, I also had access to information on course-taking behaviors including how many online courses in which the students enrolled, how many sessions the students logged, and any assessment scores associated with each online login.

Over the two-year period for which data were provided, a total of 376,206 student-lesson observations were available for analysis with anywhere from around 20 to 60 lessons per course. Each lesson was highly structured, aligned with standards where applicable, and often built on content taught in previous lessons within the same course. The typical lesson contained an approximately 20 minute lecture where students were presented with slides that were read and expounded upon by an instructor asynchronously. Students saw this prerecorded instructor on the top right-hand side of the screen. Slides often contained bullet-pointed information. Other times the slides provided interactive tools, such as a statistics calculator, or showed instructors writing and talking through a practice math problem. After the lecture, students completed an activity, most often multiple choice or true/false practice problems. Other times, students completed a worksheet that required the review and analysis of outside resources or the answering of open-ended short answer questions. Every lesson ended with an assessment of student knowledge acquisition, which contained entirely, or predominately, forced-response type questions.

Table 12. Student Demographic and Academic Characteristics by Online Course-Taking Status (2016-17 through 2017-18 School Years)

	All High School Students	All Online Course-Taker	Analytic Sample
Female	0.494	0.455	0.440
	(0.500)	(0.498)	(0.496)
Black	0.612	0.683	0.707
	(0.487)	(0.465)	(0.455)
Hispanic	0.211	0.212	0.204
	(0.408)	(0.409)	(0.403)
White	0.098	0.069	0.060
	(0.298)	(0.253)	(0.237)
English Language Learner (ELL)	0.157	0.122	0.114
	(0.364)	(0.328)	(0.317)
Free/Reduced Price Lunch Eligible (FRL)	0.757	0.795	0.814
	(0.429)	(0.403)	(0.389)
Special Education Eligible (SPED)	0.226	0.235	0.271
	(0.418)	(0.424)	(0.444)
Prior Year GPA	1.926	1.444	1.273
	(1.046)	(0.842)	(0.806)
Failed Course(s) in Prior Year	0.591	0.835	0.873
	(0.492)	(0.371)	(0.333)
Percent Absent	0.222	0.300	0.338
	(0.231)	0.239	(0.244)
Mean Pre-test Score (if pre-tested)	----	56.028	57.779
		(30.203)	(30.340)
Number of Student-Year Observations	26,428	7,110	5,337

Standard deviations in parenthesis

I collected additional data by reviewing and coding lessons according to the Online Authentic Work Rubric using system access provided by the district to the actual online course lectures, activities, and assessments. I limited this data collection, and thus the subsequent analysis, to the 10 courses in which the most students enrolled over the two-year study period based on course-taking data provided by the online course vendor. Where more than one semester of a course fell into the top 10, I selected the semester in which more students enrolled for analysis after spot checking course content to establish similar levels of authenticity between semesters. For instance, the algebra 1 course was a yearlong course that required the completion of both algebra 1 semester one and algebra 1 semester two. Since slightly more students enrolled in the first semester of the course, in order to maximize the diversity of the 10 courses reviewed, I only watched and coded lessons in the first semester algebra 1 course.

In total, these courses represented 60 percent of all online course enrollments in the district during the study period. Courses include (in descending order of enrollment) career planning and development, citizenship, ninth grade English/language arts (ELA), algebra I, personal finance, healthy living, geometry, U.S. history, world history, and physical science. Of the ten courses, one (the algebra course) was marked as specifically for credit recovery, with a separate algebra course available for non-credit recovery purposes. The other nine could be taken for either credit recovery or acceleration purposes. However, conversations with district administrators indicate that the vast majority of students enrolled in online courses in the district for credit recovery. The analytic sample limited to these ten courses contained 134,272 student-lesson observations, representing 4,790 unique students (with 5,337 student-by-year observations). As shown in Table 12, student characteristics in the analytic sample mirrored those observed in the larger population of online course-takers within the district. However, I observed lower rates of prior academic performance and attendance among online course-takers compared to the larger student population, with nearly 84 percent of online course-takers having failed a course in the prior school year.

Outcomes of interest included the grade students earned on their first attempt of the end-of-lesson assessment (from zero to 100 percent), whether students earned at least a 60 percent on their first attempt of the lesson assessment (the district threshold to avoid having to retake a lesson to earn credit), and the natural logarithm of active and idle time students spent interacting with the course system when completing a lesson. This time excludes minutes spent watching lectures, which should be consistent across students but varied due to challenges associated with identifying whether students were actively watching lecture videos. Active time measured the time logged into a course where the student was completing course content as measured by the online course system. For instance, progressing through assessment questions or completing a virtual lab were considered active engagement, whereas staying on an assessment slide beyond the time required to complete the section or letting the computer go to sleep would be logged as idle. Thus, active time, versus the total course duration which included idle time, provides an imperfect but nevertheless helpful measure of students' behavioral engagement. The use of this vendor-captured data is particularly important due to the high percentage of off-task time noted during classroom observations. On average, I noted during in-person observations that students were off-task 48 percent of the time. However, within the analytic sample, the system only identified an average of 27 percent of student time in the system as idle (SD=32). This discrepancy can be explained by differences in how in-person raters measured off-task time versus how idle time was logged in the course system. For instance, an in-person rater seeing a student playing on his or her phone while clicking through an assignment haphazardly would log this time as off-task. In contrast, the online course system would log this time as active regardless of the reduced quality of the student's interaction with the assignment. Despite this limitation, student-by-lesson level data on active and idle time provide important signals of engagement not normally available at this level of detail.

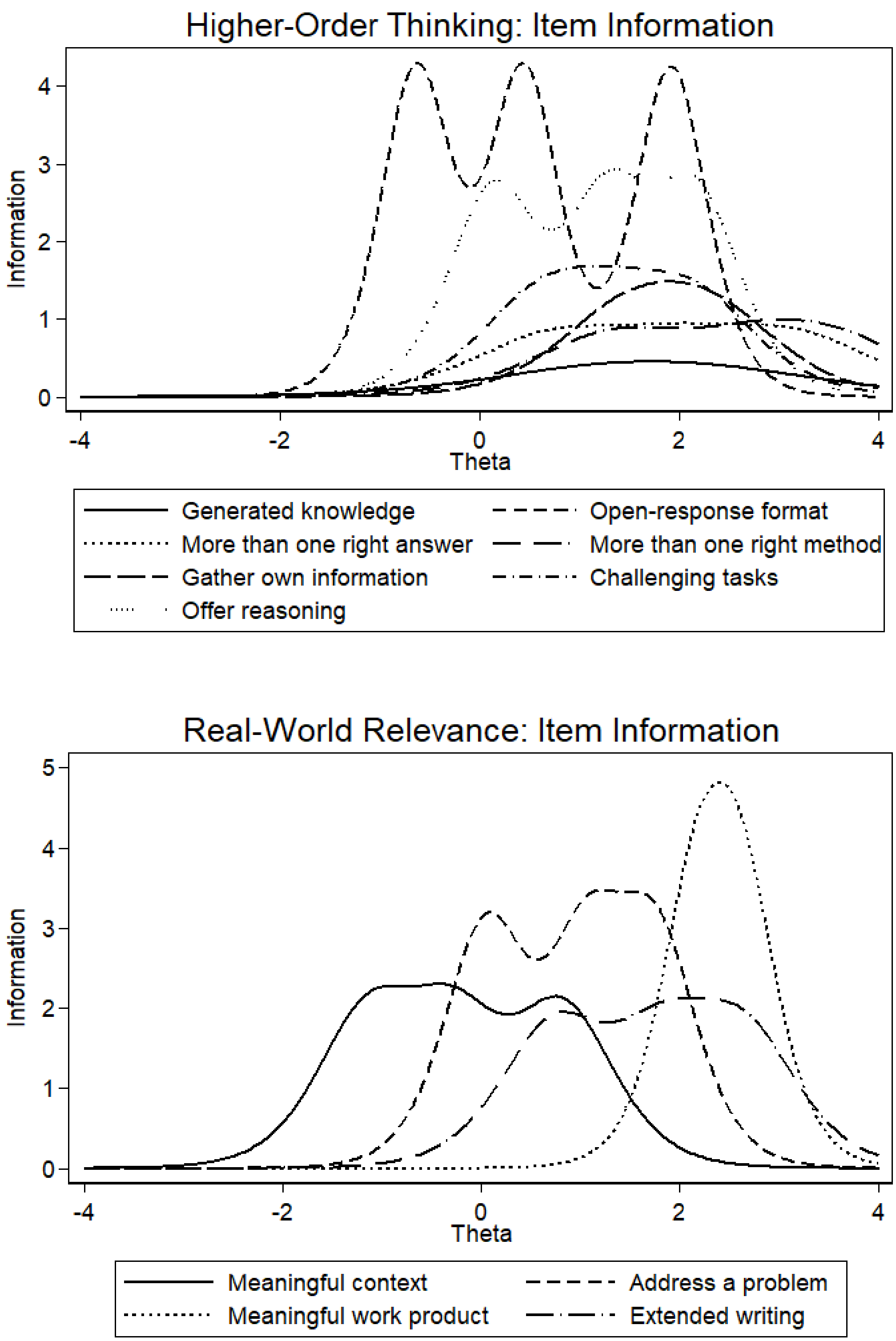
The average score earned on students' first end-of-lesson assessment attempt was 75 percent. Scores ranged from zero to 100 percent. There was a slight left skew to the data, with students scoring 60 or above on the end-of-lesson assessment on the first attempt in 81 percent of student-lesson observations. Most items on the assessments were forced response questions presented in either a multiple choice or true/false format with the occasional open-ended response required. The vast majority, if not all, items in each assessment required remembering lecture content versus applying or creating new content. Students spent a median of 2.19 active hours and a median of 0.39 idle hours engaged with the course system for each lesson, excluding time devoted to watching the lecture video. Among cases where students passed the end-of-lesson assessment on their first attempt, 31 percent of lessons were completed in one hour or less.

Empirical Strategy

Scale development. First, I developed an original rubric to measure authentic work in online contexts (see Appendix D for the final rubric). I relied on a review of prior literature and preexisting instruments on authentic work, which tended to be more theoretical than psychometrically validated, to define and operationalize the constructs of interest (i.e., Au, 2012; Bidwell, Frank, & Quiroz, 1997; Hiebert et al., 2005; Newmann, 1992; Newmann et al., 1996; Reeves et al., 2002; Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016; Stein et al., 1996). The higher-order thinking scale was designed to measure the extent to which students were asked to think deeply and critically about course content, often requiring students to generate new knowledge. The real-world relevance scale was created to identify the extent to which course content resonated with or was applicable to students' lives, interests, and/or aspirations. The rubric was then refined based on feedback from content experts and pilot coding. I trained three additional raters using the rubric, establishing interrater reliability at the beginning of and throughout the coding process. Each online lesson was then evaluated on the extent to which higher-order thinking and real-world relevance were present. All responses were entered in Qualtrics for analysis. There was a primary rater assigned to each course who rated every lesson in the course. Others coded a few lessons from that course to establish interrater reliability and ensure consistent rubric interpretation. Before reconciliation, raters assigned a rating within one point of each other on the four-point Likert-type scale items in 93 percent of cases. To minimize concerns regarding variability in ratings based on the rater, I used only ratings from the primary rater of each course to ensure each lesson within a given course was rated by the same person.

Throughout the coding process and after all courses were rated, the research team and I discussed any discrepancies in or confusion regarding the interpretation of items. We revised or dropped these items and culled items whose meanings overlapped substantially with other items. For instance, the original rubric asked raters to evaluate the extent to which lessons "asked students to communicate responses verbally or in written form" and "asked to offer reasoning to support responses." The first item was removed, because it provided no additional information after accounting for the second item. After coding, I use item response theory (IRT) rating scale models to place the extent to which higher-order thinking and real-world relevance were present in each lesson on standardized, continuous scales with a mean of zero and standard deviation of one. The Cronbach's alpha for the higher-order thinking scale was 0.82, while the Cronbach's alpha for the real-world relevance scale was 0.74. Conventions in the social sciences identified the internal consistency of the real-world relevance scale as acceptable and the internal consistency of the higher-order thinking scale as good (DeVellis, 2016). The two scales represented two distinct but correlated constructs, $r=0.384$, $p<0.001$.

Figure 4. Higher-Order Thinking and Real-World Relevance Item Information Functions



As shown in Figure 4, seven items loaded onto the higher-order thinking scale. Of those, the extent to which the lesson asked students to respond in an open-response format and offer reasoning to support their assertions provided the most influential information for scale development. Of the four items that loaded onto the real-world relevance scale, not providing meaningful context for lesson content was most influential on the low end of the scale. Whether students were asked to evaluate, apply, or synthesize complex information to solve a problem or issue provided the most information in the middle range of the scale, while whether students were asked to create work product with meaning outside of a school context distinguished the lessons with the highest level of real-world relevance. The resulting scales had close to a normal distribution, as shown in Figure 5, apart from a second peak on the higher-order thinking scale at the extreme low end of the distribution.

Figure 5. Distribution of Higher-Order Thinking and Real-World Relevance Scales

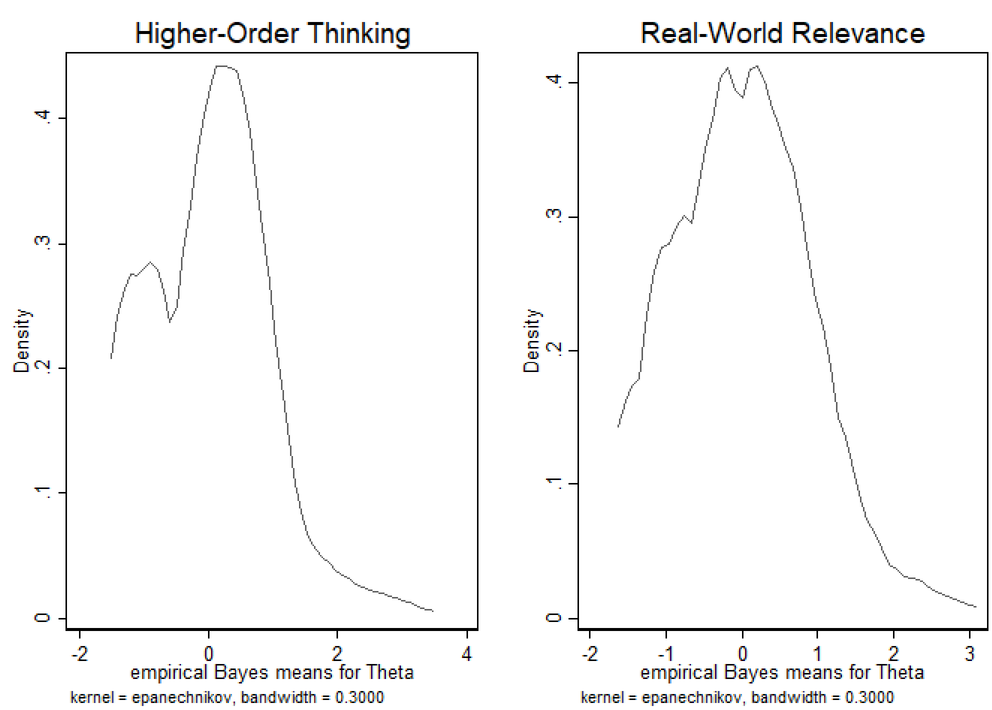


Table 13. Correlations between Subscales, Rubric Ratings, and Course Components

	Higher-Order Thinking	Real-World Relevance
Higher-Order Thinking	1.000	
Real-World Relevance	0.384***	1.000
Rubric Ratings		
Proportion Skill Introduction	-0.069	0.014
Interactive Task(s)	0.369***	0.284***
Reading Task(s)	-0.107**	0.053
Writing Task(s)	0.480***	0.182***
Recite Task(s)	-0.074	0.158***
Demonstrate Task(s)	0.206***	0.238***
Critical Thinking Task(s)	0.244***	0.442***
Application Task(s)	0.184***	0.424***
Evaluation Task(s)	0.210***	0.498***
Synthesis Task(s)	0.235***	0.609***
Creation Task(s)	0.479***	0.425***
Vendor-Provided Course Components		
Assignment	-0.066	0.097**
Lab	0.073*	0.189***
Material Title	0.053	0.129***
Online Resource	-0.116***	-0.026
Summary	-0.073*	0.109***
Vocabulary	-0.167***	-0.151***
Warm-up	-0.030	0.161***

To evaluate convergent validity, I examined correlations between each scale and the types of tasks raters identified as present within each lesson. As shown in Table 13, lessons that required more higher-order thinking were also more likely to include student-directed tasks that required interactivity and writing, while lessons that demonstrated more real-world relevance were more likely to require the evaluation and synthesis of ideas. Both higher-order thinking and real-world relevance were likely to be present in lessons requiring students to create work product. Work product in this context refers to any output created by completing instructional activities, including but not limited to an essay, multimedia presentation, business plan, or family budget. High correlations between real-world relevance and critical thinking, application, and evaluation tasks reinforce observational findings that integrating real-world examples were one of the most common means used in the online courses to facilitate these processes. However, the higher-order thinking scale was better at distinguishing between the inclusion of recitation tasks (which were correlated with real-world relevance but not high-order thinking) and more complex tasks (which higher-order thinking was correlated with).

Nonetheless, lower correlations between the higher-order thinking scale and tasks requiring critical thinking, application, and evaluation indicate an important distinction between some measures of higher-order thinking and this scale, in that this scale prioritizes processes that require students to take ownership of learning processes and generate their own knowledge. For example, a math problem that required students to solve an equation might require critical thinking or the application of recently introduced skills to a new context, but would not meet the higher bar for this higher-order thinking scale, since students were expected to replicate a process to determine the solution, which had only one correct answer. However, an in-depth worksheet on budgeting that asked students to research trends in household expenses in the United States and apply that knowledge along with their mathematical skills to develop current and future personal budgets was rated highly on higher-order thinking (as well as real-world relevance).

There was comparatively less association between vendor-provided information on course components and the higher-order thinking and real-world relevance scales. Notably, the inclusion of additional activities (i.e., assignments, labs, material titles) in addition to direct instruction by the vendor when designing lessons was generally associated with more real-world relevance. This makes sense because the additional activities often provided a more in-depth example, with warm-up and summary components often focusing specifically on framing the content the lecture will introduce in terms of real-world applicability. In contrast, the inclusion of additional vendor-developed activities such as assignments, labs, or material titles did not appear to be associated with higher-order thinking. However, lessons that included more technology-directed, non-interactive features (i.e., vocabulary, online resources) were often rated lower in higher-order thinking.

Descriptive analysis. To demonstrate the prevalence of authentic work in the online courses examined, I presented in the Authentic Work in Online Courses section descriptive statistics of rubric findings to illustrate course content and instructional delivery. I systematically examined patterns in lessons demonstrating high, medium, and low rates of higher-order thinking and real-world relevance. I supplemented this analysis with vignettes of course lessons to provide context, clarify classroom practices, and assist in the interpretation of the higher-order thinking and real-world relevance scales. I ensured the reliability and validity of these qualitative findings by establishing saturation, triangulating findings across courses, and searching for alternative interpretations (Huberman & Miles, 2002). This additional analysis was intended to classify and provide description of the type of authentic work present in the online lessons as well as the contexts in which those tasks appeared.

Statistical analysis. Next, I used a student-by-course (α_{ic}) fixed effects model to examine whether exposure to higher-order thinking or real-world relevance were separately or jointly associated with variation in student course-taking behaviors and achievement. When estimating OLS regression without fixed effects, selection bias would occur if students more likely to perform higher (or be more engaged) were systematically more likely to be assigned to courses with higher (or lower) rates of authentic instruction. The student-by-course fixed effect strategy allowed for the removal of all variation that was constant within the student and courses during the semester that a student completed a course. This means attributes such as a students' innate ability can no longer bias estimates. Student-by-course fixed effects are effective but computationally intensive, as multiple observations of the same individual in the same course are required. The unique characteristics of the data collected by the online course system combined with ratings collected using the authentic online work protocol allowed for the implementation of this rigorous method.

I estimated the equation leveraging variation across lessons (*l*), as specified in the following equation, for each student (i) and course (c), where I inputted each higher-order thinking and real-world relevance subscale individually (as $authentic\_work_{cl}$) and then together. When examining the role of higher-order thinking and real-world relevance in the same equation, I also included an interaction between the two terms. The interaction represented differences in students' responses that only occurred when both higher-order thinking and real-world relevance were present at higher (or lower) levels in the same lesson.

$$y_{icl} = \beta_0 + \beta_1 authentic\_work_{cl} + \alpha_{ic} + \boldsymbol{X_{icl}\beta} + \boldsymbol{A_{cl}\beta} + \delta_{cl} + \varepsilon_1 \quad (1)$$

Dependent variables included the students' first score on the end-of-lesson assessment, a binary measure of whether the student passed the assessment the first time by scoring above the threshold required to earn credit (i.e., avoided having to retake the lesson), and the natural logarithm of active and idle time. I focused on a students' first score on the end-of-lesson assessment to provide the most direct comparison across scores, since students were only required to retake the assessment if they scored below a certain threshold, and based on observations that instructor or research-based assessment assistance was more common on subsequent assessment attempts. Thus, the first score on the end-of-lesson assessment was most likely to reflect true learning. Beyond achievement measures, active and idle time provided insight into behavioral engagement.

In equation one, I controlled for student-by-lesson level ($\boldsymbol{X_{icl}}$) covariates including whether the student took the pretest for a given lesson and, if they did so, his or her pretest score.[6] This accounted for prior knowledge students might have on lesson content, where available, to more fully isolate learning due specifically to the lesson being examined. When predicting active and idle time, I also included a series of binary variables indicating the number of times a student retook the lesson, since it was not possible to isolate the active time spent on only the first time a student completed the course. At the lesson level ($\boldsymbol{A_{cl}}$), I controlled for the total time raters estimated it would take to complete a given lesson and the order the lesson appeared in the course. This variable was collected by raters when watching and completing online course content. These measures controlled for differences in base course length (without controlling for differences in the type of activities students were asked to do during the lesson) and any differences in student achievement or behavior associated with increasing familiarity with the lesson format as they progress through the course. I also used teacher fixed effects (δ_{cl}) to condition on differences in the online teacher quality identified through descriptive analyses. Information on which teacher taught each lesson was collected at the same time raters watched lesson lectures to evaluate the level of authentic work using the Online Authentic Work Rubric. The inclusion of teacher fixed effects accounted for differences in instructional delivery when different teachers taught different lessons in the same course to better isolate student

[6] If a student did not take a pretest score, they were coded as achieving a zero. The binary variable indicating whether a student took a pretest or not accounted for mean differences associated with whether a student took a pretest or not. As a sensitivity test, I estimated the main model excluding pretest score, and coefficients remained identical when rounded to the hundredth decimal place.

responses to instructional tasks. All models used student-level robust, clustered standard errors to account for correlated errors.

Next, I examined potential heterogeneous treatment effects. First, I ran separate regressions, splitting the dataset into four quantiles based on students' prior year GPA and running equation one separately for each quantile. I also implemented equation two, shown below, by controlling for and interacting with the authentic work term a binary indicator identifying whether students belonged to historically marginalized student subgroups. I examined these differential effects by estimating separate regressions for each of the following subpopulations: students who qualified for Free/Reduced-Priced Lunch (FRL) or receive special education services as well as students identified as Black, Hispanic, or English Language Learners. The coefficient of interest in these estimations is β_3, which captures differences in students' responses to authentic work for each subgroup.

$$y_{icl} = \beta_0 + \beta_1 authentic_{work_{cl}} + \beta_2 subgroup_i + \beta_3 authentic_{work_{cl}} * subgroup_i + \alpha_{ic} + \boldsymbol{X_{icl}\beta} + \boldsymbol{A_{cl}\beta} + \delta_{cl} + \varepsilon_1 \qquad (2)$$

Sensitivity tests. One of the potential sources of endogeneity I was concerned with was that there could be lesson-specific characteristics that might be associated with, but not directly contributing to, authentic work and student outcomes. Thus, as a sensitivity test, I added a series of binary variables indicating the type of activities present in each lesson to the vector of lesson-level covariates ($\boldsymbol{A_{cl}\beta}$), excluding lecture and assessment components that were present in all lessons. Examples included the need to review linked material (such as a news article or book chapter), virtual labs, or online activities. The type of activities was important to control for because they might require different time commitments. I did not include this information in the main estimates due to concern that certain types of activities might be associated with the presence of authentic work. For instance, lessons that included extended writing might be more likely to facilitate authentic work than lessons that did not. However, I identified minimal associations between these variables and the extent to which raters observed either higher-order thinking or real-world relevance ($-0.221 \leq r \leq 0.167$), indicating that it was rarely the activity itself but how it was realized that influenced the level of authentic work present.

I was also concerned that results might be biased due to differences in students' prior academic experiences and orientations to learning that might be associated with both which courses they are assigned to and the engagement and academic outcomes examined. The use of student-by-course fixed effects minimized concerns associated with these factors when invariant over time. However, as a sensitivity check in a second model, I controlled for a vector of time-varying student covariates ($\boldsymbol{S_i\beta}$) to account for any extraneous variation in the dependent measures.

$$y_{icl} = \beta_0 + \beta_1 authentic\_work_{icl} + \alpha_{ic} + \boldsymbol{X_{icl}\beta} + \boldsymbol{A_{cl}\beta} + +\boldsymbol{S_i\beta} + \delta_{cl} + \varepsilon_2 \qquad (3)$$

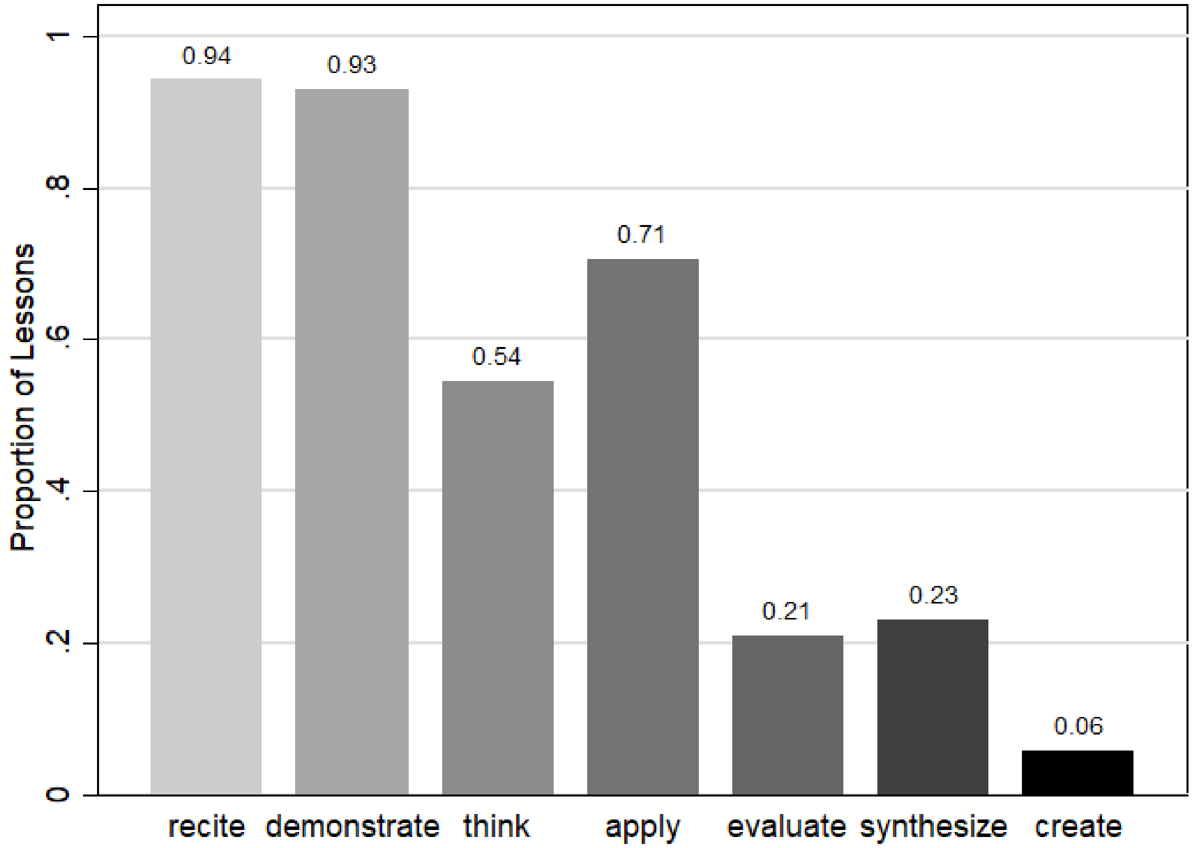
These variables included student English Language Learner status, days of school absent, prior year GPA, and prior year number of courses failed as well as whether the student qualified for FRL or special education services. Each of these variables was captured at the student-year level.

Further, since course content was only available from the 2017-18 school year, I calculated alternative estimates from models limited to that single school year to demonstrate the extent to which findings were sensitive to the assumption of lesson stability across school years. One last challenge was that the use of course fixed effects required an assumption of homogeneous treatment effects to generalize beyond students enrolled in the courses examined. I demonstrated in the Sensitivity Tests section the extent to which this assumption held by estimating models excluding course fixed effects to present evidence on the plausibility of assuming homogenous treatment effects across courses. Lastly, to determine sensitivity to the exclusion of a measure of reasoned communication, I estimated the main model including a binary variable indicating whether writing tasks were incorporated in the lesson (along with interactions between writing and the higher-order thinking and real-world relevance subscale).

Results

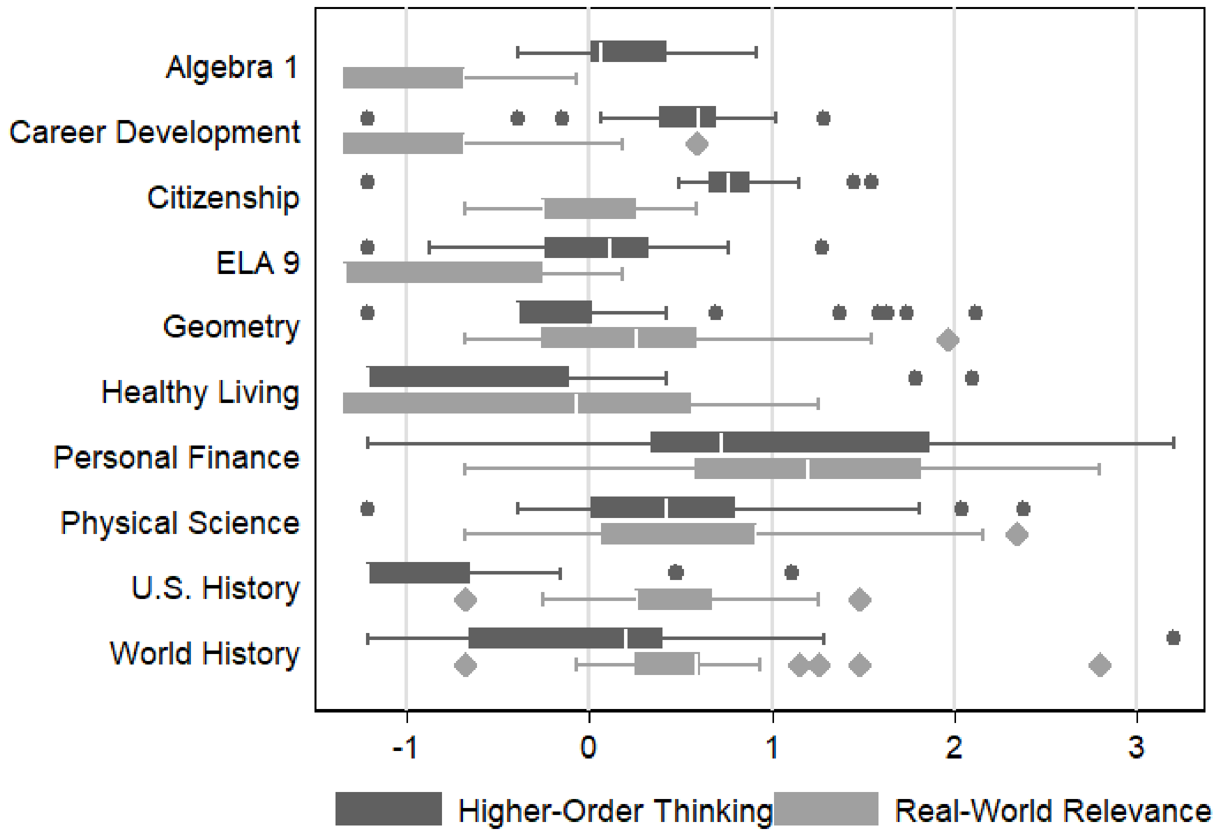
Authentic work in online courses. I summarized the proportion of online lessons observed that asked students to engage in each level of Bloom's taxonomy in Figure 6, where tasks generally increase (from left to right) in the extent to which they require students to engage in higher-order thinking. Bloom's taxonomy is meant to assist teachers in designing lessons, where an ideal lesson would include tasks across the entire taxonomy. This allows tasks requiring greater higher-order thinking to be scaffolded on more direct forms of instruction. The typical online lesson required students to recite lecture information, demonstrate understanding by answering problems, and apply what they learned to new contexts. About half of all lessons (54 percent) required students to think critically (labeled *think*). Generally, application is considered a higher level of thinking. However, in the case of most online lessons, the application was surface-level and thus did not always require critical thinking. For instance, students might be asked to apply new mathematical knowledge to a word problem. While the situation may be new, the process might not be, limiting the amount of critical thinking required. Fewer than one-quarter of lessons required students to evaluate, synthesize, or create. These measures are related but separate from the authentic work subscales. Since the higher-order thinking and real-world relevance scales were standardized, it's important to place subsequent findings in context of the modal lesson to which students were exposed.

Figure 6. Presence of Tasks Requiring Various Levels of Thinking by Online Lesson



There was wide variation in the presence of both higher-order thinking and real-world relevance within as well as across courses, as shown in Figure 7. When comparing courses, citizenship and career development lessons were rated consistently above average on higher-order thinking, while the United States history class was rated consistently below average. The personal finance course provided a high level of higher-order thinking in many (but not all) lessons. The algebra one and career development courses were rated the lowest on the real-world scale, while the personal finance and physical science courses had some extremely highly rated lessons on real-world relevance.

Figure 7. Higher-Order Thinking and Real-World Relevance by Course



The most common reason lessons were rated low on both higher-order thinking and real-world was an overemphasis on direct instruction that provided only surface-level information or focused on abstract versus applied motivations for understanding lesson content. Several lessons across courses epitomized this type of one-dimensional instruction. One lesson regarding the Abolitionist Movement presented a minimum level of information, such as listing jobs of enslaved workers and providing an overview of the Quakers, skimming over content that might make students feel uncomfortable and not requiring them to think deeply or analyze content. An ELA lesson on how to use a variety of sentence types to add interest to writing provided only one opportunity for students to edit a passage on their own. Instead, most tasks consisted of watching a lecture or the practice of each sentence skill separately on multiple choice questions. In a citizenship course, the instructor listed traits of good citizens, “responsible, helpful, curious, honest, altruistic, humble, knowledgeable, respectful,” without any acknowledgement of different interpretations or realization of these qualities. The presentation of absolute, vague personality traits avoided addressing the complexities of citizenship, such as when embodying one ideal might contradict another, which could have encouraged higher-order thinking or provided a measure of real-world relevance.

Rated higher on at least one dimension, the following descriptions provide examples of lessons with a range of higher-order thinking and real-world relevance scores. These examples demonstrate the types of authentic work that are feasible within an online setting and help to differentiate between the two authentic work components. For instance, the following notes from a ninth grade ELA lesson were rated highly on higher-order thinking but below average on real-world relevance.

> This lesson took a step back from the Odyssey to cultivate student skills in utilizing dictionaries and thesauruses. This lecture was particularly important since Greek mythology and medieval literature are hard to read, so this content tried to instill independence within students to look up words they do not know while reading. The lesson question was, how do reference resources help improve vocabulary? The skill focus was to develop reading comprehension and writing skills to improve vocabulary. Students practiced using a dictionary, identifying synonyms and antonyms in a thesaurus, and choosing correct homophones for given contexts. The instructor modelled how to do each skill, where student practiced directly following the introduction of each skill. Instructional tasks were all multiple choice except for one short answer that required students to practice the same skills as in multiple choice questions. These questions had students identify facts about words (syllables, origin) from a dictionary, use context clues to choose best definition of words with multiple definitions, and choose the best synonym for a word in context. This last example requires students to follow multiple steps: understand if word is noun, adjective, or verb, then go to thesaurus. The lecture content used less advanced words such as abstract and culture. The quiz questions had students practice these skills using more advanced words such as lavish, lore, and crude.

The above lesson introduced students to a range of skills, allowing them to select from multiple methods to answer problems with more than one right answer. These tasks required students to gather their own information, facilitating student-generated (versus instructor-directed) knowledge. However, this lesson was not rated highly on the real-world relevance scale, because the instructor failed to ground the skills learned in a meaningful context outside of academic work, such as by addressing a real-world problem.

In contrast, the following U.S. History lesson provided high levels of real-world relevance with only average requirements for higher-order thinking.

> The instructor explained that the lesson will describe the lives of workers during industrialization and help students understand the impact of industrialization during the Gilded Age. She reviewed the concept of mass production, talking about how it revolutionized consumption before introducing Jan Matzeliger, who invented a machine to mechanize a part of the shoemaking process. The instructor used him as an example of how mass production applied to many industries. She then talked about the dangerous conditions for workers. She accomplished this by showing old newspaper headlines that highlighted factory accidents and showing a graph of the number of deaths from industrial accidents in railroads and mines from 1900 to 1940. She began the section on child labor… mentioning that it is important to understand that child labor still exists today, and many of our consumer goods are the result of child labor in developing countries. The practice problem students were expected to answer after this section were open-ended. The first question showed a picture of child laborers and asked what dangers these child laborers were exposed to. Next, the instructor explained the impact of industrialization on women. She highlighted a graph of women's labor in 1900, emphasizing that their options for work were limited and their wages lower than men. The question following this part of the lecture was open-ended and asked how industrialization affected women, children, and families. She then briefly discussed how these factors led to the reforms that began in the Gilded Age.

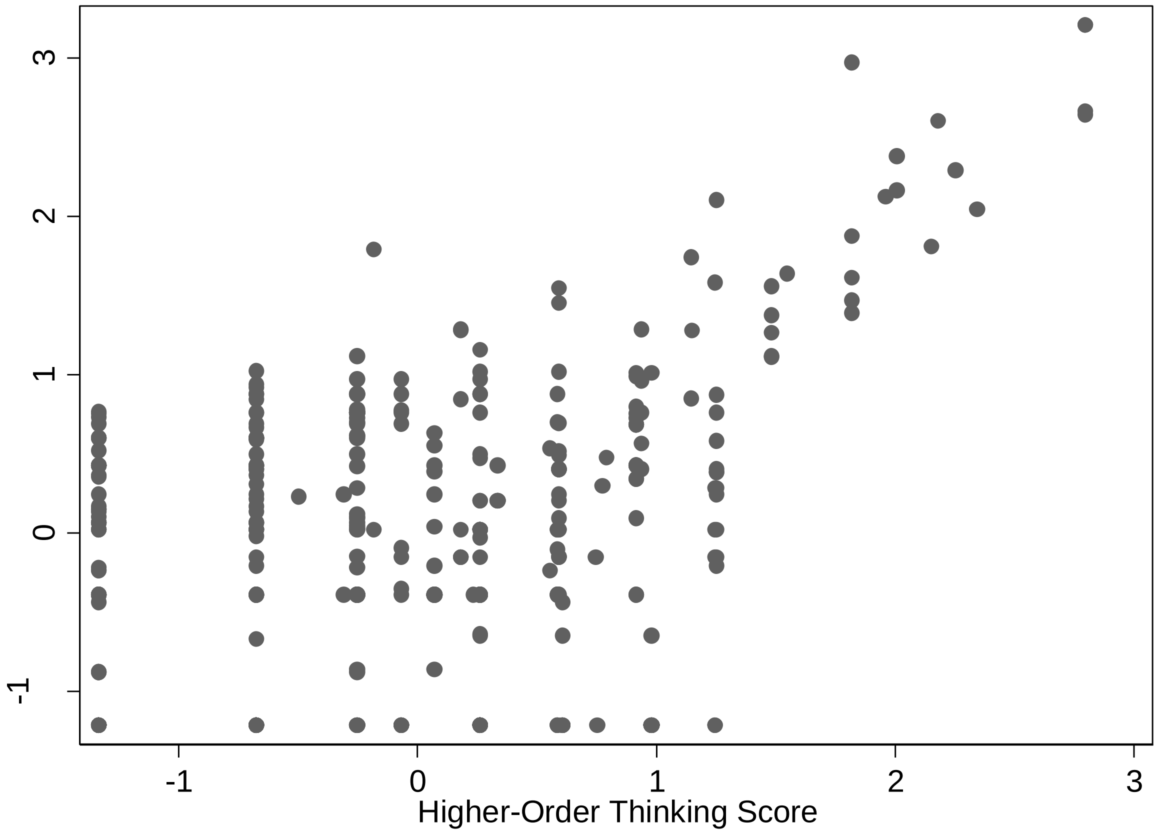
The instructor established a meaningful context for lesson content by highlighting case studies, integrating archival documents, and linking historical content to modern life. Successful completion of the lesson did not require students to create work product with meaning outside an academic context, but there were opportunities for extended writing about social problems that reinforced the meaningful application of the lesson content introduced. Despite these strengths, the lesson did not require much higher-order thinking. Instead, students could earn credit for completion of the lesson by only remembering and reciting lecture content introduced by the instructor regarding the dangers of child labor and how industrialization impacted women.

While there were examples of lessons rated favorably in only one of the two authentic work components, the lessons rated most highly in one area tended to be rated equally highly in the other (see Figure 8). These patterns provide some suggestive evidence that integrating real-world relevance might help facilitate higher-order thinking and vice versa. Many lessons in the personal finance course exemplified this intersection. For instance, in a lesson in the personal finance course that received the highest possible score on both higher-order thinking and real-world relevance, the instructor outlined how to research careers.

> The instructor laid out information concisely in a conversational tone. He also provided detail about where the proper resources were to do research about jobs, including counselors, libraries, and the US Bureau of Labor Statistics. In choosing a career, he said you should know about employee benefits, comparing job offers, and lifetime income… Overall, he gave generic but helpful advice to make sure students at least have a baseline of information to think about when considering career options. The practice question asked students to identify Julio's best job offer out of three based on location, salary, benefits, and average monthly rent. The instructor discussed in-depth lifetime income and the importance of the different facets of it. In the summary, the instructor compared the career of a doctor and mechanic, considering interests and skills, education, job outlook and income, and potential lifetime income, but didn't come to any conclusions about what the best choice might be. The assignment asked students to research two careers of their choice and show the results in a multimedia presentation. They were given resources to assess their interests, as well as websites to search occupations. Students were also given an assessment consisting of nine multiple choice questions that applied definitions and concepts from the lesson, along with one open response asking students to outline steps in choosing a career.

This lesson was notable for the open-ended nature of assignments, the integration of a research-based activity that allowed student-generated knowledge, and the emphasis on multiple "right" choices depending on a student's situation and preferences. These attributes required higher-order thinking to complete the lesson successfully, but the same elements – such as the student-directed nature of the assignment and the element of choice – also personalized content in a way that enhanced real-world relevance.

Figure 8: Associations between Higher-Order Thinking and Real-World Relevance Scales



The following excerpt also earned a perfect score on both scales, but this Word History lesson used writing versus research to facilitate higher-order thinking and real-world relevance.

> The instructor asked students to write an informative essay on a turning point in history. She split the lesson up into three parts: research, draft, revise. She used the fall of Constantinople as an example of a historical turning point, as well as Martin Luther's 95 Theses, the fall of the Roman Empire, and the Renaissance. She looked at the students' writing prompt, emphasizing that students could choose any turning point they want. The first instruction section looked at the lesson question and writing prompt… The next section walked through the prewriting process: identify a topic, research, and organize an outline. The instructor told students what to think about while completing this before going over examples of reputable sources. She emphasized the importance of choosing significant and relevant facts. She then described how to create an outline in three steps (introduction, body, conclusion)… She showed an example of a strong conclusion and then went over the revision process, looking at two strategies: a clear topic and effective language.

This lesson shared many of the strengths identified in the personal finance course. Students were required to complete an intensive, multi-step, student-directed activity on a topic of the students' choosing. The World History lesson further supported higher-order thinking by scaffolding expectations, new skill introduction, and review into the lecture component to improve the probability that students could complete what might be a challenging task in a way that reinforced the skills and knowledge acquisition processes the lesson was designed to teach.

This last example of a highly rated lesson on both dimensions described a virtual lab that students were expected to complete as part of the physical science course.

> The instructor stated that physics and chemistry could be quite colorful, showing a cabbage juice reaction to different solutions. He said we will investigate pH in the lab. He provided the lab goals by dividing them into four parts. He went over the vocabulary words and reviewed acids and bases, as well as pH. He translated the scale in to mol/L. Instruction began and he again showed various cabbage juice solutions. Students came up with a research question for the lab based on a brief overview, as well as a hypothesis. He provided a summary of the four parts of the lab. The student guide for the lab was then provided, going through 18 detailed steps of the procedure and providing blank data tables for students to fill in. The instructor then went through the data tables and demonstrated how to fill them out. He suggested using colored pencils to match the colors they observe. Students answered questions identifying the independent and dependent variables. The instructor then introduced serial dilutions so that students understand how they will be completing their lab. A few questions check for understanding. Then, he talked about qualitative measurements of pH with test strips and indicator solutions, after which he explained quantitative measurement of pH using a pH meter. Safety procedures were reviewed. Students could choose whether to complete the lab in real life or virtually. The virtual lab consisted of dragging materials into beakers on the screen and recording the results shown digitally. Students measured the pH of different substances and then common household items. The assignment asked students questions to ensure they understand what was done in the lab. Finally, students completed a detailed lab report to illustrate their conclusions. The rubric and guide for writing this was very detailed and structured.

The three types of assignments highlighted as facilitating high ratings on both scales (research, writing, and interactive activities) were common components in the most authentic lessons. As highlighted when discussing the research and writing-based tasks, the virtual lab also required interactivity, application, and critical thinking. Specifically, the emphasis on student-generated knowledge through data collection and analysis, instead of solely remembering and reciting information from an instructor-driven lecture, provided the framework for higher-order thinking and a meaningful context with real-world applicability.

Students' learning and behavioral responses to authentic online work. The previous section identified variation both across and within courses in the level of authentic work presented in each lesson. I also provided examples of both typical and exceptional levels of authenticity feasible within an asynchronous, online course system. In this section, I describe how students responded and performed differently in response to the level of higher-order thinking and real-world relevance observed by examining variation across lessons within the same course. In doing so, I used a teacher fixed effect in an effort to control for potential differences in instructor approach observed among different instructors teaching different lessons within the same course.

When completing the end-of-lesson assessments of lessons that were one standard deviation higher on the higher-order thinking scale, students scored one percent lower than on other assessments within the same course, as seen in the first column of Table 14. Associations were consistent across the end-of-lesson quiz distribution when conducting a quantile regression providing evidence of a homogeneous treatment effect. The small, decrease in scores associated with one standard deviation more higher-order thinking corresponded to being three percent more likely to pass the quiz the first time. The difference in directionality between lower scores and higher pass rates might hint at heterogeneous effects, which I explored in subsequent analyses. Also, I observed evidence of gaming behavior at the threshold required to earn lesson credit (refer to Appendix E for more detail), leading me to be more confident in the small,

Table 14. Lesson Outcomes as a Function of Variation in Higher-Order Thinking and Real-World Relevance

	(1) Thinking	(2) Relevance	(3) Interaction	(4) W/ Lesson Activities	(5) W/ Student Covariates[7]	(6) 2017-18 Only[8]	(7) Student FE Only	(8) Controlling for Writing
Dependent Variable: First Score (0-100%) (N=128,463)								
Higher-Order Thinking	-0.007***		-0.010***	-0.009***	-0.009***	-0.009***	-0.003***	0.000
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Real-World Relevance		-0.000	0.004***	0.003***	0.003***	0.004**	-0.001	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Higher-Order Thinking			0.001*	0.001	0.001	0.001	-0.004***	0.007***
* Real-World Relevance			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Dependent Variable: No Retake, Earned At Least 60% on First Score (N=134,272)								
Higher-Order Thinking	0.030***		0.006***	-0.012***	0.009***	0.001	0.019***	0.027***
	(0.002)		(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Real-World Relevance		0.029***	0.015***	0.007***	0.014***	0.012***	0.002	0.055***
		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Higher-Order Thinking			0.026***	-0.000	0.024***	0.020***	0.015***	0.081***
* Real-World Relevance			(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Dependent Variable: Log of Active Time (N=134,272)								
Higher-Order Thinking	-0.009		-0.029	-0.030	-0.023	-0.022	-0.149***	-0.052**
	(0.018)		(0.022)	(0.022)	(0.020)	(0.029)	(0.010)	(0.026)
Real-World Relevance		0.022	0.032	0.024	0.055***	0.044	0.028***	0.047
		(0.020)	(0.023)	(0.023)	(0.020)	(0.029)	(0.010)	(0.029)
Higher-Order Thinking			0.011	0.005	0.011	0.021	-0.026***	0.052**
* Real-World Relevance			(0.013)	(0.014)	(0.012)	(0.017)	(0.006)	(0.022)
Dependent Variable: Log of Idle Time (N=134,272)								
Higher-Order Thinking	0.032**		0.038**	0.041**	0.043***	0.033***	0.044***	0.027
	(0.013)		(0.016)	(0.016)	(0.012)	(0.012)	(0.009)	(0.019)
Real-World Relevance		0.009	-0.008	-0.012	0.001	-0.005	-0.014	-0.027
		(0.014)	(0.016)	(0.016)	(0.012)	(0.012)	(0.009)	(0.021)
Higher-Order Thinking			-0.005	0.002	-0.004	-0.001	0.020***	-0.029*
* Real-World Relevance			(0.010)	(0.010)	(0.007)	(0.007)	(0.006)	(0.016)

Each column within each dependent variable section represents a different mode

[7] The inclusion of student covariates reduces the sample to 97,807 cases for first score and 102,279 for the log of active and idle time.
[8] The examination of only cases from the 2017-18 school year reduced the sample to 62,704 cases for first score and 64,893 for the log of active and idle time.

negative association observed between higher rates of higher-order thinking and quiz scores than between higher rates of higher-order thinking and higher first-time lesson pass rates.

In addition, I observed no change in the active time logged and a three percent increase in the amount of idle time logged for each one standard deviation more higher-order thinking observed. Because I took the natural logarithm of active and idle time, each one point increase in higher-order thinking or real-world relevance is associated with a percent increase in the dependent variable equivalent to 100 times the estimated coefficient. For a student who logged the median number of idle hours on a lesson (0.39), a three percent increase in idle time would increase the total idle time to 0.45 hours. These results remained consistent when controlling for real-world relevance, shown in column three, apart from a portion of the increase in never retaking being attributed to the interaction between the two components of authentic work.

However, associations between real world relevance and student outcomes and behaviors shifted when accounting for the level of high-order thinking present. Without controlling for higher-order thinking, students exposed to one standard deviation more real-world relevance within a given course scored no higher on the end-of-lesson assessment and logged no additional active or idle time. Conditioning on the presence of higher-order thinking, as shown in column three, students scored half a percent higher on the end-of lesson assessment.

Sensitivity tests. Estimates were qualitatively similar when controlling for lesson activities and student covariates, as shown in columns four and five, with a few exceptions. First, the positive association between no retake and higher-order thinking reversed when accounting for lesson activities. However, this discrepancy represents a change in the estimate of only 0.018 points, thus potentially speaking more to the precision of the estimates than a large practical change in the directionality of the estimate. Second, controlling for student covariates resulted in the identification of a positive association between active time logged and real-world relevance. The assumption that lessons observed for the 2017-18 school year with the same course and lesson names in 2016-17 were identical appears plausible based on the estimates in column six that show consistent results when limiting the analysis to only cases from the 2017-18 school year when compared to the main estimates presented in column three. However, significant differences between models with and without course fixed effects (in columns three and seven respectively) demonstrated that the assumption of homogenous treatment effects likely does not hold, and course-specific differences must be accounted for when estimating student responses to authentic work.

Lastly, because I was concerned that students might not experience benefits without the presence of reasoned communication, the third component of authentic work, I ran the same models conditioning on whether the lesson included a writing-based activity as a proxy for reasoned communication (shown in column eight). I also included interaction terms between writing and the higher-order thinking and real-world relevance scales. Although only a proxy, qualitative analysis indicated that writing was one the most successful strategies to support high levels of authentic work. Movement in estimates when controlling for the presence of writing in a lesson might indicate that main model estimates were biased by not being able to measure and account for opportunities for reasoned communication. Instead of observing this, estimates remained comparable (although of slightly larger magnitude in some instances) when accounting for whether writing was integrated in the lesson apart from a switch from a negative association between higher-order thinking and quiz scores to no relationship. In this model, students only scored lower when more higher-order thinking in a lesson was accompanied by writing tasks (β= -0.020, SE= -0.002).

Table 15. Examination by Prior Year GPA

	Lowest Quartile Q1	Q2	Q3	Highest Quartile Q4
	Dependent Variable: First Score (0-100%)			
Higher-Order Thinking	-0.008***	-0.010***	-0.008***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
Real-World Relevance	0.004*	0.003	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Higher-Order Thinking	0.001	0.002	0.001	-0.001
* Real-World Relevance	(0.002)	(0.001)	(0.001)	(0.001)
	Dependent Variable: No Retake, 60% Threshold			
Higher-Order Thinking	0.006	0.004	0.012**	0.011***
	(0.005)	(0.005)	(0.005)	(0.005)
Real-World Relevance	0.015***	0.016***	0.015***	0.012**
	(0.005)	(0.006)	(0.005)	(0.005)
Higher-Order Thinking	0.026**	0.027***	0.025***	0.018***
* Real-World Relevance	(0.003)	(0.003)	(0.003)	(0.003)
	Dependent Variable: Log of Active Time			
Higher-Order Thinking	-0.032	-0.057*	0.020	-0.003
	(0.028)	(0.033)	(0.029)	(0.020)
Real-World Relevance	0.038	0.069**	0.029	0.018
	(0.029)	(0.034)	(0.029)	(0.020)
Higher-Order Thinking	0.029	0.014	-0.025	-0.009
* Real-World Relevance	(0.018)	(0.020)	(0.017)	(0.012)
	Dependent Variable: Log of Idle Time			
Higher-Order Thinking	0.048**	0.034	0.054***	0.005
	(0.021)	(0.022)	(0.021)	(0.021)
Real-World Relevance	0.015	0.038	0.010	-0.020
	(0.022)	(0.023)	(0.021)	(0.020)
Higher-Order Thinking	-0.012	-0.014	0.010	0.023**
* Real-World Relevance	(0.014)	(0.013)	(0.012)	(0.012)
N	25,289	24,996	26,023	25,971

Each column within each dependent variable section represents a different model.

Heterogenous effects. Prior research indicated that students would likely respond differently to authentic work based on prior academic experiences. I was also concerned that assignment to online courses might differentially disadvantage students with lower rates of prior academic achievement who were less likely to have developed the prerequisite self-regulatory skills and academic knowledge needed to fully access and learn in an online instructional environment. Using the model that included both higher-order thinking, real-world relevance, and an interaction between the two variables, I examined estimates by prior GPA quantiles to represent prior academic experiences. This examination showed relatively consistent associations between access to authentic work and both end-of-lesson assessment scores and avoiding the need to retake the lesson, as shown in the top half of Table 15, with one exception. Students with GPAs in the top half of the distribution, but not students in the bottom half of the distribution, were one percent more likely to meet the 60 percent threshold when the lesson employed one standard deviation more higher-order thinking. This coefficient is statistically significant but relatively small in magnitude.

When examining time logged, students in the highest quartile of prior year GPAs appeared less likely to log additional idle time and no more likely to log active time when exposed to higher rates of higher-order thinking. This might be because students with higher prior academic performance may be more accustomed to and more likely to know how to successfully complete assignments requiring higher-order thinking. I also observed an outlier in the percent of active time students logged among students exposed to higher rates of higher-order thinking among students with prior year GPAs in the second quartile. Students in this quartile may not have developed the self-regulation to force themselves to spend more time on assignments requiring higher-order thinking compared to students with higher prior year GPAs. The lack of a similar drop in active time logged among students in the first quartile complicates the interpretation of this coefficient. However, descriptive statistics showed a u-shaped relationship between the amount of active time logged and students’ prior year GPA, with students with extremely low and extremely high prior year GPAs logging the least amount of active time. Likely students with higher rates of prior academic achievement completed the courses more quickly because of better preparation, while students with the lowest rates of prior academic achievement might be more likely to disengage in an environment with little external monitoring. Thus, the lack of a negative coefficient on higher-order thinking when predicting active time logged among students in the first quartile might be a result of low overall engagement, while students in the second quartile might be particularly susceptible to lesson differences in authentic work. This could explain why students in the second quartile were more likely to disengage (logging less active time) when the lesson requires higher rates of higher-order thinking and reengage (logging more active time) when the lesson provided more real-world relevance.

For similar reasons regarding the equity implications of assignment to online courses and subsequent access to authentic work, I also examined differential associations by student characteristics. As shown in Table 16, the largest magnitude differences included more active time logged among students who qualified for FRL, received special education services, and identified as Black when exposed to lessons that required more higher-order thinking. For instance, students who qualified for FRL spent an average of 23 percent more time on lessons requiring one standard deviation higher-order thinking holding all else constant. Like the above explanation for differences in active time logged between quartiles, it is possible that these students had lower levels of academic preparedness requiring them to spend additional time on lessons requiring higher-order thinking due to less familiarity and comfort with these types of activities (Gamoran & Nystrand, 1992; Hoffman & Ritchie, 1997).

Other notable differential effects included lower levels of active time logged among students who qualified for FRL when exposed to lessons with more real-world relevant content. Lesson observations indicated this response might be a due to the type of real-world examples

Table 16. Heterogeneous Effects of Higher-Order Thinking and Real-World Relevance

	FRL	SPED	ELL	Black	Hispanic
	Dependent Variable: First Score (0-100%) (N=128,463)				
Higher-Order Thinking * Group	-0.005**	-0.002	0.001	0.001	-0.002
	(0.009)	(0.002)	(0.003)	(0.002)	(0.002)
Real-World Relevance * Group	-0.003	-0.009***	-0.003	-0.008***	0.005**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Higher-Order Thinking * Real-World Relevance * Group	0.003*	0.001	0.001	0.000	0.001
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
	Dependent Variable: No Retake (N=134,272)				
Higher-Order Thinking * Group	0.001	-0.002	0.013**	0.002	-0.001
	(0.005)	(0.004)	(0.006)	(0.004)	(0.005)
Real-World Relevance * Group	-0.003	-0.008*	-0.010*	-0.017***	0.010**
	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)
Higher-Order Thinking * Real-World Relevance * Group	0.009***	0.006*	0.003	0.000	0.006**
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
	Dependent Variable: Log of Active Time (N=134,272)				
Higher-Order Thinking * Group	0.228***	0.124***	-0.057	0.118***	0.017
	(0.038)	(0.036)	(0.049)	(0.033)	(0.038)
Real-World Relevance * Group	-0.207***	-0.103***	0.094*	-0.012	-0.020
	(0.040)	(0.038)	(0.049)	(0.034)	(0.039)
Higher-Order Thinking * Real-World Relevance * Group	-0.080***	0.054**	0.016	0.014	-0.003
	(0.025)	(0.026)	(0.032)	(0.022)	(0.025)
	Dependent Variable: Log of Idle Time (N=134,272)				
Higher-Order Thinking * Group	0.010	0.036*	0.097***	-0.019	-0.067***
	(0.020)	(0.019)	(0.025)	(0.018)	(0.020)
Real-World Relevance * Group	0.024	0.014	-0.057**	0.026	0.053**
	(0.021)	(0.020)	(0.026)	(0.018)	(0.021)
Higher-Order Thinking * Real-World Relevance * Group	-0.005	-0.028**	-0.037**	-0.009	-0.002
	(0.013)	(0.013)	(0.017)	(0.012)	(0.013)

Each column within each dependent variable section represents a different model.

integrated. Specifically, many surface-level real-world applications emphasized normative life experiences more likely to resonate with students belonging to dominant cultural groups. Examples of these normative life experiences include New England-style town meetings, parents buying expensive electronics and vehicles, and international travel.

Contributions and limitations. Available data allowed for the linking of detailed information from classroom observations and course content (pulled directly from the videos, activities, and assessments accessed by students) with microdata on student course-taking behaviors and achievement, providing a level of depth to my analysis that is rarely available to researchers. I further leveraged these data using student-by-course fixed effects, minimizing the extent to which confounding variables that were fixed during the semester the students completed a given course, such as a students' innate ability or consistent course attributes, might influence results (Wooldridge, 2013). When correctly specified, quasi-experimental studies such as these are often as good as random assignment, particularly when models integrate covariates that explain assignment into and response to treatment (Shadish, Clark, & Steiner, 2008). Specifically, innate ability, prior year course failure, and general orientations to learning cannot or are unlikely to change over the course of a single semester.

Claims of generalizability are strengthened by the examination of content developed by one of the largest online course vendors in the United States. That said, the use of courses by only one vendor prevents drawing conclusions regarding the level of authenticity in online courses developed by other vendors. Relatedly, the focus on a single online course vendor likely limited the potential variability in authentic online work observed. Further, evidence of heterogenous effects among students with different prior levels of achievement and belonging to different subgroups demonstrates that students may respond differentially to authentic online content based on background and prior educational experiences.

Another limitation of the study is the challenge of operationalizing authenticity. Although the rubric was based on a thorough review of the literature, the subjective nature of developing the scales and rating online videos on each construct needs to be recognized in the interpretation of findings. Relatedly, I was unable to examine components of authentic work, such as communication and interaction with peers, that were not facilitated by the online course system evaluated. It is also possible that benefits to authentic work are not (fully) realized unless all components are present, and thus findings from the current study might not reflect what was feasible or typical student outcomes when exposed to authentic work that incorporates opportunities for reasoned communication along with higher-order thinking and real-world relevance. It is also possible that the predominately remember and recite type, multiple choice assessment items might not have captured the greater depth of learning achieved when exposed to lessons that required more higher-order thinking (Warschauer, 2006). However, the examination of differences in active and idle time showing that the typical student logged four percent more idle time but no more active time for each one standard deviation increase in higher-order thinking does not appear to support this hypothesis.

I also only have access to course content for the 2017-18 school year, although I attempted to minimize this concern by demonstrating that estimates were not sensitive to restricting the data to only the 2017-18 school year. Further, when years were included in course names or “current” events alluded to in lectures, most indicated that the most recent major update to courses occurred in 2016 or earlier. Lastly, I lacked systematic information on authentic work within traditional, face-to-face classrooms. Observations of traditional, face-to-face classrooms in the district indicated wide variability in authenticity across classrooms (Darling-Aduana & Shero, 2020), which is consistent with prior studies in similar settings that raised question regarding the assumption that because teachers in traditional, face-to-face classrooms can provide opportunities for more interactive, authentic work that they are so inclined and possess the expertise to successfully implement these strategies (i.e., Darling-Aduana & Shero, 2020; Diamond, 2007; Gamoran, Secada, & Marrett, 2000). Despite these limitations, this study makes an important contribution by delving into the black box of vendor-developed online curriculum and instruction. Findings also drew new, direct links between critical curriculum studies and emerging contexts in online learning to further understanding of patterns in high school student engagement and learning based on exposure to authentic online work.

Discussion

Existing research tells us that students are more engaged and learn better when exposed to authentic work, which includes opportunities for higher-order thinking with real-world applications (Dee & Penner, 2017; Marks, 2000). Yet, students belonging to marginalized subgroups are systematically less likely to have access to this type of instruction (Darling-Hammond, 2001; Haberman, 2010). The increased standardization and profit-driven motives of vendors developing online high school courses raise concerns regarding the extent to which the incursion of online courses into the secondary education sector may magnify existing inequities in access to quality, authentic learning experiences (Boninger, Molnar, & Murray, 2017).

This study represents the first large-scale attempt to document the prevalence of authentic work in popular online courses designed by one of the largest online course vendors in the United States. Subsequent analyses demonstrated differences in student engagement patterns and performance in lessons with more opportunities for authentic work within the same course in aggregate, by prior academic performance, and by subgroup. As a secondary contribution, my analytic process required the development and validation of the Authentic Online Work Rubric, which is now freely available for additional testing and research use (see Appendix D).

Consistent with concerns that online course-taking may exemplify the *pedagogy of poverty* for the predominately low-income, minoritized student population enrolled in online courses in this study, the courses were most likely to require students to recite information and demonstrate understanding on multiple choice or true/false questions versus provide opportunities for authentic work (Haberman, 2010). This means that the widespread integration of online courses, particularly among schools serving marginalized populations, might exacerbate existing educational opportunity gaps conditional on the quality of educational opportunities available in alternative, face-to-face instructional settings. At the same time, qualitative analysis demonstrated that some types of authentic work were possible in online learning settings and that differential student responsiveness to authentic work existed in online, as well as face-to-face, instructional environments. Notably, I highlighted activities – including student-directed research, writing, and interactive assignments – that could facilitate high levels of higher-order thinking and real-world relevance when integrated in the standardized, asynchronous online course environment examined (Cavus et al., 2007; Dinov et al., 2008; Gao & Lehman, 2003).

The strategies rated highest on higher-order thinking and real-world relevance occurred in tandem, indicating that both components may be necessary for the highest quality learning experiences (Hiebert et al., 2005; Lebow & Wager, 1994; Moll & González, 2004). However, this is particularly concerning when considered in conjunction with research on the hidden curriculum, as it indicates that the lack of opportunities for higher-order thinking observed in many classrooms serving students from marginalized populations likely also limits the level of real-world relevance integrated (Au, 2012; Haberman, 2010). This supports and provides additional information on a possible mechanism to explain higher rates of disengagement and alienation among students with less access to authentic curriculum and instruction year after year (Apple, 2004; Au, 2012; Bernstein, 1975; Solorzano & Yosso, 2001; Yair, 2000).

Next, I identified that students exposed to lessons requiring more higher-order thinking tended to score lower on the end-of-lesson assessment and logged more idle time, while those exposed to lessons with more real-world relevance tended to score higher on the end-of-lesson assessment. Differential associations by prior GPA along with lower performance and behavioral engagement patterns associated with higher rates of higher-order thinking indicated that many students enrolled in online courses within the district studied might have been unprepared to successfully complete these assignments. However, many of those same students achieved higher rates of learning when lessons demonstrated more real-world relevance. These findings regarding changes in student behaviors and performance associated with differential exposure to higher-order thinking and real-world relevance were robust to the inclusion of information on lesson activities and student covariates, with generally consistent patterns observed across prior GPA quartiles. Possible exceptions included less variation in active and idle time logged by students in the highest GPA quartile when completing lessons requiring more higher-order thinking, likely due to great familiarity and comfort with this type of content (Oakes, 2005, Newmann, 1992). Concerningly, students who qualified for FRL appeared to benefit less from lessons with real-world relevance, possibly due to the lack of applicability of some surface-level applications to these student groups (Darling-Aduana et al., 2020).

While these associations identify a significant relationship, the magnitude of estimates should not be overstated. On average, students exposed to a lesson facilitating one standard deviation more higher-order thinking scored one percent lower on the end-of-lesson quiz. This slight dip in score is unlikely to delay student course completion and may be deemed an acceptable outcome in exchange for exposing students to more authentic work. Similarly, although in the opposite direction, students exposed to a lesson facilitating one standard deviation more real-world relevance scored only, on average, half a percent higher on the end-of-lesson quiz. As these findings were based on standardized, "inauthentic" quizzes, results might have been more pronounced if authentic assessment were administered, as the greater depth of understanding that authentic work facilitates are unlikely to help students as much when answering multiple choice questions. However, the magnitude of the estimate identified in this study demonstrates that increasing real-world relevance on its own is unlikely to have a profound effect on reducing any sort of educational achievement gap. Instead, the most important implications of this research are in establishing the extent to which these type of widespread, standardized online course systems may be contributing to an expanding opportunity gap for students most at risk of not graduating high schools who are the student group most likely to be assigned to online, high school courses (Powell et al., 2015) and students attending low resourced schools who likely to be swayed out of necessity by promises of increased efficiency (Heinrich et al., 2019).

Implications for Policy and Practice

Policy-makers, educators, and online course vendors can leverage these findings to identify instructional practices likely to improve access to authentic work in online instructional settings (see also Cavus et al., 2007; Crippen & Earl, 2007; Dinov et al., 2008; Gao & Lehman, 2003). The inclusion of research, writing, and interactive activities like those highlighted here are also likely to improve access to authentic work when implemented in face-to-face classrooms. While there is room to improve access to authentic work in online courses, the most transformative restructuring of learning for students belonging to marginalized groups requires more collaborative, interactive instructional environments not available in the most popular, standardized online course systems currently on the market. To combat this, improvements could be made to the systems affordable to low resourced schools, including integrating more opportunities for synchronous and blended instructional techniques, or additional resources could be allocated to schools for higher quality instructional resources.

It is also important to determine in a given school, program, and student context the possible merits of any increased expectations of higher-order thinking considering what seems to be a need for more support and attention for students to learn from these types of activities. This is not to say that opportunities for higher-order thinking should not be integrated but rather that its integration may require larger systemic changes, such as intensive opportunities for student-teacher interactions to support the scaffolding and alignment of tasks requiring higher-order thinking, to overcome the initial resistance and confusion students experience when exposed to these tasks for the first time (Hoffman & Ritchie, 1997). For these reasons, students may benefit more from the integration of higher-order thinking in face-to-face or other types of synchronous instructional settings. The demonstration of real-world relevance appears a more easily implementable leverage point, yet there is a ceiling on the extent to which this can be achieved without also requiring additional higher-order thinking (Hiebert et al., 2005; Lebow & Wager, 1994; Moll & González, 2004).

While students belonging to historically marginalized groups may have the most to gain by access to authentic work (Au, 2012; Yair, 2000), exposure in a single online course in high school may be too little, too late to have a profound effect on their academic trajectories. Thus, the possible benefits identified in this and prior research (i.e., Dee & Penner, 2017; Marks, 2000; Newmann, 1992; Yair, 2000) must be weighed and considered within larger institutional structures and constraints to prevent the implementation of strategies to foster authentic work from harming students. For instance, higher instructional expectations may prevent some students from meeting minimum competencies required to obtain needed educational credentials or other prerequisites for post-secondary educational, labor market, or personal success. Access to authentic work is a powerful piece, but still only one part, of dismantling systemic inequalities in educational opportunities.

CHAPTER 4

A Remote Instructor Like Me: Student-Teacher Congruence in Online, High School Courses

Researchers have established positive impacts for citizens served by government bureaucrats of the same race/ethnicity and gender across policy areas, including education (Grissom, Kern, & Rodriguez, 2015; Hindera, 1993; Keiser, Wilkins, Meier, & Holland, 2002; Meier & Nicholson-Crotty, 2006; Selden, 1997; Wilkins & Keiser, 2004). There are two pathways through which bureaucratic representation occurs: symbolic and active. Symbolic representation occurs when bureaucrats, such as teachers, share demographic characteristics with their constituents, which may result in more positive interactions due to shared expectations and norms or differences in how constituents interpret bureaucrat behavior (i.e., through a reduction in stereotype threat) (Grissom et al., 2015). Active representation occurs in schools when sharing demographic characteristics results in an increased willingness by teachers to make decisions favorable to in-group members (Nicholson-Crotty, Grissom, & Nicholson-Crotty, 2011).

Based on evidence of positive associations between student-teacher racial/ethnic and gender congruence and academic outcomes, researchers have proposed policy interventions that aim to increase the diversity of the teacher workforce, integrate culturally responsive teaching practices, and minimize teacher implicit bias with the goal of improving the academic experiences of both elementary and secondary students belonging to historically marginalized groups (Dee, 2005; Egalite et al., 2015; Gershenson et al., 2017; Gershenson, Holt, & Papageorge, 2016; Grissom et al., 2015; Irizarry, 2015; Joshi, Doan, & Springer, 2018; Lindsay & Hart, 2017). However, research has yet to identify the extent to which the mechanisms that support improved academic outcomes translate to online instructional settings and whether the most effective policy interventions should address the symbolic versus active mechanisms through which bureaucratic representation may act. For these reasons, I examined the following research questions.

1. What is the distribution of remote and live (in-person) instructor identities in the thirty most enrolled in online courses in a large, urban district?
2. To what extent is being taught by an instructor of the same race/ethnicity or gender associated with course outcomes when educational content is delivered through a standardized, prerecorded video?
3. What insight do the patterns observed when examining race/ethnicity and gender congruence in an online setting provide about the mechanisms through which positive associations in traditional, face-to-face classrooms are likely realized?

Theoretical Framework

Building upon prior literature on bureaucratic representation, this study examines student interactions with remote instructors to determine the extent to which symbolic bureaucratic representation improves course outcomes when educational content is delivered through a standardized, prerecorded video where active mechanisms are not possible. In traditional school settings, it is all but impossible to determine the extent to which symbolic versus active bureaucratic representation shapes the advantage students experience when taught by a racial/ethnic or gender congruent teacher (Dee, 2004, 2005; Egalite et al. 2015; Gershenson et al., 2016, 2017; Grissom et al., 2015; Joshi et al., 2018; Lindsay & Hart, 2017). However, the extent to which the academic benefits identified in prior studies can be realized more widely through policy intervention depends on whether the mechanisms behind the gains are due to symbolic versus active forces. Currently, researchers propose improving both active and symbolic forms of representation (i.e., Gershenson et al., 2016). Determining the importance of each can support more targeted, and thus ultimately more effective policy interventions.

Online course systems provide a means by which to separate active from symbolic representation as a potential mechanism for increased achievement, as all students receive standardized instructional material and delivery from the same remote instructors. Since the discretion and ability to respond differently to different students required for active representation is not feasible for the instructor delivering prerecorded content in these settings, results from this study can be more fully attributed to symbolic bureaucratic representation. Examining bureaucratic representation in this type of asynchronous, online course setting also removes the possibility of parental coproduction through increased involvement in their children's schooling, another hypothesized mechanism of representation in traditional, face-to-face classroom settings (Markowitz, Bassok, & Grissom, 2020; Vinopal, 2017).

Instead, mechanisms related to symbolic representation, such as stereotype threat (Steele, 1997; Steele & Aronson, 1995) and role model effect (Morgenroth, Ryan, & Peters, 2015), are most likely to drive any associations observed in the asynchronous, pre-recorded online setting studied. Stereotype threat describes a phenomenon whereby students perform lower when a negative stereotype about their racial/ethnic or gender identity is activated due to concern that they might confirm that negative stereotype (Steele, 1997; Steele & Aronson, 1995). Steele (1997) established that achievement can be improved when students are in a setting where they are not concerned about these negative stereotypes (also Borman & Pyne, 2016). Relatedly, being taught my an instructor of the same identity might activate a role model effect, whereby the racial/ethnic or gender congruent instructors act "as behavioral models, representing the possible, and being inspirational" as a means to improve student motivation and self-efficacy along with subsequent performance (Morgenroth et al., 2015, pp. 465).

To the extent that the positive associations identified in prior literature between student-teacher congruence and achievement are due to these types of symbolic mechanisms, implementing policies such as increasing the diversity of the teaching workforce could on its own have a large, positive impact on closing the achievement gap. While the supply of minority teachers might limit the immediacy with which a reform like this could be implemented in traditional, face-to-face classrooms, the asynchronous, prerecorded structure of online courses could provide relatively scalable access to instructors with racial/ethnic and gender congruence. However, this type of intervention would be less effective if the primary mechanisms through which representation improved student achievement were active.

Bureaucratic Representation in Education

In Dee's (2004) seminal study, White and Black elementary students of both genders achieved three to six percent higher reading scores when randomly assigned to a same race teacher as part of the Tennessee Project STAR class size experiment, with larger benefits observed among students who received Free or Reduced-Price Lunch (FRL). Using the National Education Longitudinal Study of 1988 (NELS:88), Dee (2005) demonstrated that the benefits of race and gender congruence generalized to the national context. However, the most profound benefits were observed again among students from low socioeconomic status backgrounds and those residing in the South. In a more recent study using Florida administrative data, Egalite and colleagues (2015) found that after controlling for teacher quality, students with own-race teachers for one year achieved 0.001 standard deviations higher test scores in reading and 0.008 standard deviations higher test scores in math using a student fixed effect strategy. The advantage was greater among elementary grade students, students identified as Black or White, and students with lower prior track records of achievement (Egalite et al., 2015). Similarly, Joshi and colleagues (2018) identified the largest effects of student-teacher race congruence in Tennessee among students identified as Black and with lower prior achievement levels.

The benefits of race congruence persist to high school graduation and college attendance, particularly for low-income, Black, male students (Gershenson et al., 2017). Among this population, access to one Black, male teacher in third through fifth grade reduced the probability of dropping out of high school by seven percent and intent to earn a four-year college degree by eight percent (Gershenson et al., 2017). Low-income, Black, female students also reported intent to earn a four-year college degree at higher rates when exposed to at least one Black teacher in third through fifth grade (Gershenson et al., 2017).

Beyond test scores, Grissom et al. (2015) identified more favorable disciplinary, special education, and gifted referrals for minority students when taught by same race teachers. More specifically, a recent study found that Black students received fewer behavioral referrals on average by Black teachers, estimating that a 25 percent increase in the number of Black teachers would result in four percent fewer behavioral referrals for Black students (Lindsay & Hart, 2017). Another study using nationally representative survey data identified lower rates of teacher-reported externalized behavior among kindergarten students when taught by a race congruent teacher but no benefits to internalizing behaviors, interpersonal skills, or approaches to learning (Wright, Gottfried, & Le, 2017).

Researchers have identified similar effects among students who experienced gender congruence with their teachers in STEM fields (Keiser et al., 2002; Xu & Li, 2018). Bureaucratic representation is particularly salient in these subjects for female students, as society imposes norms of math as male-dominated increasingly as children progress through elementary into secondary school (Solanki & Xu, 2018; Keiser et al., 2002; Xu & Li, 2018). In a seminal study, eighth-grade students who identified as female achieved higher math scores when taught by a female math teacher (Keiser et al., 2002). Similarly, Xu and Li (2018) observed higher female math performance and ratings of both student-teacher interactions and self-perceived math ability after randomly assigning Chinese, junior high students to female teachers. Solanki and Xu (2018) demonstrated that positive effects in STEM fields are due at least in part to a role model effect that results in reduced identity incongruence for female students when taught by a female teacher. As a result, female students report being as likely to ask for help and attend lectures as male students when taught by a female teacher (Solanki & Xu, 2018). Similar associations were not observed to the same extent in other subjects, such as literature, where gender was less relevant to the measured outcomes (Carlana, 2019; Keiser et al., 2002; Solanki & Xu, 2018; Xu & Li, 2018).

Disentangling symbolic and active mechanisms. Many studies identified potential active mechanisms between student outcomes and race/ethnicity and gender congruence. In schools, teachers act as street-level bureaucrats that interact directly with students and parents with a high level of autonomy and discretion (Grissom et al., 2015; Lipsky, 1980). This discretion allows for the personal preferences, values, and beliefs of each teacher to pervade their classroom learning environment (Grissom et al., 2015; Lipsky, 1980; Meier, 1993). Race/ethnicity and gender - as key components of identity in the contemporary United States - shape teachers' prior experiences. Subsequently developed worldviews inform how teachers interact with the students in their classroom based on each students' racial/ethnic and gender identity (i.e., Dee, 2004, 2005; Egalite, Kisida, & Winters, 2015; Gershenson, Hart, Lindsay, & Papageorge, 2017; Keiser et al., 2002).

For instance, Dee (2005) documented a less favorable classroom environment and fewer educational opportunities for students taught by a non-race congruent teacher based on data collected as part of the National Education Longitudinal Study of 1988. Researchers also identified disparate teacher ratings of student performance by student race or ethnicity (i.e., Gershenson et al., 2016; Irizarry, 2015). These disparate ratings often translate into less positive student-teacher interactions that shape achievement and students' orientations toward schooling (Burgess & Greaves, 2013; Dee, 2015; Sorhagen, 2013). For instance, non-race congruent teachers were less likely to place minority students in gifted programs (Fish, 2017; Grissom et al., 2015) and more likely to refer minority students for disciplinary action (Fish, 2017; Grissom, Nicholson-Crotty, and Nicholson-Crotty, 2009; Grissom et al., 2015; Lindsay & Hart, 2017).

Using Measures of Effective Teaching (MET) data, Cherng and Halpin (2016) found that students reported more favorable perceptions of minority versus White teachers, which could indicate either symbolic or active mechanisms depending on whether the perceptions are based on differential teacher practices and interactions or based solely on identity. However, the researchers offered suggestive evidence that students might rate minority teachers more favorably because those teachers tended to be more multiculturally aware – an active mechanism (Cherng & Halpin, 2016). Further, Lindsay and Hart (2017) observed the largest decrease in disciplinary referrals for Black students by Black teachers regarding subjective offenses, which the authors interpreted as indicating differences in teacher responsiveness or classroom management strategies versus changes in student behaviors as the primary mechanism contributing to lower rates of disciplinary referrals. Other examples of possible active mechanisms through which bureaucratic representation might provide benefits to demographically congruent students include investing increased time instructing these students (Grissom et al., 2015).

In contrast, symbolic mechanisms include changes in how students perceive their teachers and themselves when taught by a demographically congruent teacher (Grissom et al., 2015). For instance, students might be less likely to experience stereotype threat when taught by a teacher of the same race/ethnicity or gender (Dee, 2005; Steele, 1997; Steele & Aronson, 1995; Xu & Li, 2018). Same race/ethnicity and gender teachers also might serve as positive role models to students with similar identities, who subsequently engage more in educational tasks (Xu & Li, 2018). Similarly, there is evidence that teachers belonging to minoritized groups are more likely to induce positive behaviors in their students, termed coproduction inducement (Lim, 2006; Lindsay & Hart, 2017). While coproduction inducement is a type of role model effect that is purely symbolic, it is not always clear whether role model effects are passive or active. In other words, this study can help disentangle whether students respond differently based solely on the identity of the instructor – a passive mechanism - or whether the identity of the instructor is associated with variations in implicit bias or teacher stereotyping – suggesting an active mechanism – to which students responded (Carlana, 2019). Lastly, students might respond more positively to instructors who share the same identity due to shared culture (Kao, 2004). For instance, students could feel more comfortable or better understand teachers with certain accents or who use language a certain way even though this is not something a teacher consciously chooses or varies in their interactions between students.

Based on studies isolating the effect of bureaucratic representation in schools, researchers have offered tentative policy interventions. Many asserted that expanding the diversity of the teaching pool might have a positive effect on student achievement (Dee, 2005; Egalite et al., 2015; Gershenson et al., 2016; Joshi et al., 2018; Lindsay & Hart, 2017). Gershenson and colleagues (2017) advocated for purposeful matching of minority students to one or more same race teacher during elementary school grades. Researchers have also called for professional development on culturally responsive practices to minimize deficit thinking and teacher bias (Gershenson et al., 2016; Irizarry, 2015), while others suggested examining and potentially replicating differential classroom strategies, including how students are identified for gifted programs or disciplinary referrals (Grissom et al., 2015; Lindsay & Hart, 2017). One enactment of many of these recommendations indicates promise. In Oakland, assigning Black, male students to be taught by Black, male teachers in culturally relevant courses was associated with a three percent decrease in the dropout rate (Dee & Penner, 2019). However, it is unclear which mechanism(s) are contributing to this decrease.

Each policy recommendation has political, resource, and opportunity cost implications. To the extent that most gains in student outcomes are due to active representation, new policy goals such as professional development to help teachers better handle discretionary decisions become more desirable. Whereas, if symbolic bureaucratic representation accounts for a larger proportion of identified benefits, policy interventions that fail to increase the diversity of the teaching workforce or provide teachers the tools to connect with students based on shared values, assumptions, and experiences are unlikely to have the desired effect.

Bureaucratic representation within online courses. The above studies examining bureaucratic representation in schools focused on traditional, face-to-face classrooms. As 14 percent of secondary school students enroll in online courses nationally, research is needed to extend understanding of possible benefits to this rapidly expanding instructional medium (Gemin et al., 2015). Further, as online-based courses change the role of teachers in the classroom, the effects of representation may be realized differently.

To understand the potential mechanisms underlying representation in online classrooms, it is first necessary to understand what student-teacher interactions look like in these contexts. Online courses have the potential to integrate elements of personalized learning by identifying, targeting, and providing real-time feedback relevant to students' academic strengths and needs, allowing students to proceed through course content at the pace required for content mastery (Archambault et al., 2010; Bakia, Mislevy, Heying, Patton, Singleton, & Krumm, 2013; Graesser, Conley, & Olney, 2012; Powell et al., 2015). However, most online courses, including those integrated within the district examined in this study, tend toward reductionism, or the packaging of easily digestible facts into lecture-based instruction with frequent multiple choice assessments (Heinrich, Darling-Aduana, Good, & Cheng, 2019; Herrington, Oliver, & Reeves, 2006).

In order to mass market courses at a cost affordable for low resourced schools, online course vendors rely on a standardized curriculum, which therefore cannot address or respond to local contextual factors or identities (Au, 2012; McLoughlin, & Oliver, 2000). The use of a common video with the same remote instructor using a single script means that the only variation in instructional content and delivery emerges from how students interpret and respond to the online course, as teachers cannot adjust actions for each student. This increased standardization also means that remote instructors, regardless of their personal opinions or preferences, have less discretion in providing instructional delivery that meets the needs of students belonging to historically marginalized groups, which is a precondition for active representation.

Despite the lack of instructor discretion feasible in the online environment studied, it is possible that remote instructors who appear to belong to racial/ethnic or gender-based minorities may on average use different linguistic or interactional styles that resonate with students who share aspects of that instructors' perceived identity (Goffman 1983; Kao, 2004; Rawls 1987, 2000). In this manner, study findings may identify some effects of bureaucratic representation through remote instructor actions that, while provided to all students, have differential effects on students with particular identities. However, because this type of discrepancy does not rely on teacher discretion, which is not feasible in the online context study, the policy recommendations in this situation would align with those designed to engage forms of symbolic representation.

Lastly, it is possible that because students know the instructor cannot view or respond to them that stereotype threat is not activated (Steele, 1997; Steele & Aronson, 1995). While this may result in more positive learning outcomes in online learning contexts it is less likely to generalize to face-to-face contexts, where stereotype threat may be more present. Therefore, I explicitly examine and discuss the implications of this possibility to supplement my main analysis.

Gaps in the literature. Building upon prior literature on bureaucratic representation in schools, this study expands understanding of the relevance and implications of student-teacher congruence in online classroom settings. I further leveraged the standardization of curricular content and instructional delivery inherent in the online courses studied to better understand the underlying mechanisms through which bureaucratic representation operates. These contributions have wide-ranging implications for how online course vendors design and school districts select online courses, as well as what policy interventions researchers, policy-makers, and practitioners recommend and implement to improve the academic experiences and outcomes of students belonging to historically marginalized populations. I describe the data and sample, analytic strategy, contributions, limitations, and findings below.

Methods

Data and setting. This study relies on administrative data provided by a large, urban school district in the Midwest that serves a predominately low-income, minority student population. As shown in Table 17, 81 percent of students in the analytic sample qualified for FRL, 69 percent identified as Black, and 21 percent identified as Hispanic, which largely reflects the demographics of the district population. Data were provided for the 2016-17 and 2017-18 school years for all ninth through twelfth-grade students. For each student in a given year, the district provided data on achievement and sociodemographic variables. Among students enrolled in online courses, I also have access to information on course-taking behaviors, including how many online courses in which a student enrolled, how many sessions a student logged, and any assessment scores associated with each online login. Over the two-year period for which data were provided, 300,809 student-lesson observations (representing 7,328 unique students) were available for analysis with anywhere from around 20 to 60 lessons per courses. The vast majority of students enrolled in these online courses for the purpose of credit recovery, providing students a second chance to earn course credit required for high school graduation after previously failing the course.

The courses studied were developed by one of the largest online course vendors in the United States. In communications with the vendor, they emphasized that all courses were developed by teams of experienced educators and instructional designers in alignment with research-based best practices including the iNACOL National Standards for Quality Online

Table 17. Descriptive Statistics of Students in Each Analytic Sample

	Enrolled in Top 30 Course(s)	Instructor Gender Diversity	Black Instructor	Hispanic Instructor	Cases with Pretest
Female	0.448	0.451	0.435	0.458	0.448
	(0.497)	(0.498)	(0.496)	(0.498)	(0.497)
Black	0.689	0.688	0.692	0.679	0.720
	(0.463)	(0.463)	(0.462)	(0.467)	(0.449)
Hispanic	0.210	0.213	0.215	0.221	0.190
	(0.407)	(0.410)	(0.411)	(0.415)	(0.392)
White	0.068	0.067	0.065	0.065	0.064
	(0.252)	(0.250)	(0.246)	(0.247)	(0.244)
English Language Learner (ELL)	0.121	0.122	0.125	0.128	0.103
	(0.326)	(0.328)	(0.331)	(0.334)	(0.303)
Free/Reduced Price Lunch Eligible (FRL)	0.812	0.815	0.824	0.825	0.824
	(0.391)	(0.389)	(0.381)	(0.380)	(0.381)
Special Education Eligible (SPED)	0.251	0.256	0.285	0.245	0.241
	(0.433)	(0.436)	(0.452)	(0.430)	(0.428)
Prior Year GPA	1.429	1.435	1.341	1.440	1.398
	(0.795)	(0.780)	(0.771)	(0.810)	(0.770)
Failed Course(s) in Prior Year	0.867	0.875	0.890	0.851	0.865
	(0.339)	(0.331)	(0.313)	(0.356)	(0.331)
Pretested for One or More Lesson	0.497	0.484	0.462	0.611	1.000
	(0.500)	(0.500)	(0.499)	(0.488)	(0.000)
Number of Students	7,328	6,129	4,278	2,663	3,575
Number of Student-Lesson Observations	300,809	205,520	109,670	67,417	67,637

Courses. When asked about how much discretion instructors are allowed, a spokesperson responded, "[The vendor]'s content development team provides clear guidelines, scripts, and talking points for online instructors. Because [the vendor]'s online instructors are highly qualified and state-certified, the company respects instructors' experience and knowledge to make edits that ensure effective and natural lesson delivery. A team of editors and content experts review all online material before it is released" (personal communication, July 8, 2019). This leaves open the possibility that instructors may use different conversational styles or highlight different types of examples when teaching depending on their background. However, the online instructors were provided scripts and talking points in addition to the lecture slides that dictated lesson content and structure.

I limited the analytic sample to the 30 courses in which the most students enrolled over the study period, which represented 80 percent of all online courses enrolled in during the study period. All but four of the 30 courses fulfilled core subject graduation requirements. Examples of required courses in non-core content areas include personal finance and healthy living. Of the 30 online courses, I identified variation in the race/ethnicity of the teacher in 14 courses and variation in the gender of instructors in 22 courses. The race/ethnicity and gender of remote instructors were based on the assessment of two raters based on characteristics observable in the lecture videos. I refer to the raters' evaluations as presenting identity because it reflects how students were likely to categorize the instructor's identity but may not in all cases reflect how the instructor would have self-identified. Disagreements between raters were resolved through discussion and the assistance of a third rater. I was able to match ratings to 95 percent of lessons in the top 30 courses identified. Since live lab monitors had the option to go into the system and add or modify a lesson, resulting in the system generating a lesson with a new name, this match rate represents a high degree of accuracy. For instance, live lab monitors reported using this option to provide accommodations for students in accordance with their Individualized Education Plans (IEPs).

While descriptive statistics summarized data from all 30 courses, analyses requiring teacher racial/ethnic or gender diversity within the same course were limited to the courses that met the diversity precondition. Thus, analysis examining student-teacher racial congruence among students identified as Black was limited to 11 courses. Analysis examining student-teacher racial congruence among students identified as Hispanic was limited to five courses. Additional information on racial/ethnic and gender diversity by course is available in Table 18. Across the 14 courses with any variation in teacher racial/ethnic diversity, there were 159,239 student-lesson observations representing 5,191 unique students during the study period. This included 109,670 observations of 4,278 students enrolled in a course with at least one instructor presenting as Black and 67,417 observations of 2,663 students enrolled in a course with at least one instructor presenting as Hispanic. There were 205,520 student-lesson observations representing 6,129 unique students within the 22 courses with variation in teacher gender.

In addition, I used data collected from 210 classroom observations of the physical computer lab settings where students had access to computers during the school day to complete their assigned online courses. (Refer to Appendix F for a description of the dimensions rated on the research-based instrument used for data collection.) I used these data to provide descriptive findings related to the race/ethnicity and gender distribution of the instructors in the physical computer lab settings, as well as student-teacher congruence between students and the live lab monitors. These data were also used in sensitivity tests, which I discuss in greater detail below.

Table 18. Gender and Racial/Ethnic Diversity by Course

Course[9]	Gender Diversity	Racial/Ethnic Diversity	Description
Algebra 1 A	Yes	Yes	1 Black Instructor
Algebra 1 B	Yes	---	
Algebra II A	Yes	Yes	1 Hispanic Instructor
Algebra II B	Yes	---	
Biology A	Yes	Yes	1 Black Instructor
Biology B	Yes	---	
Career Development	---	---	
Chemistry A	Yes	---	
Chemistry B	Yes	---	
Citizenship A	---	---	
Citizenship B	Yes	---	
ELA 9 A	Yes	Yes	1 Black Instructor
ELA 9 B	Yes	Yes	2 Black Instructors, 1 Hispanic Instructor
ELA 10 A	---	---	
ELA 10 B	---	---	
ELA 11 A	Yes	---	
ELA 11 B	Yes	Yes	1 Black Instructor
ELA 12 A	---	---	
ELA 12 B	---	---	
Geometry A	Yes	Yes	1 Hispanic Instructor
Geometry B	Yes	---	
Healthy Living	---	Yes	1 Hispanic Instructor
Entrepreneurship	---	---	
Personal Finance	Yes	Yes	1 Black Instructor, 1 Hispanic Instructor
Physical Science A	Yes	Yes	1 Black Instructor
Physical Science B	Yes	---	
US History A	Yes	Yes	2 Black Instructors
US History B	Yes	Yes	3 Black Instructors
Word History A	Yes	Yes	3 Black Instructors
World History B	Yes	Yes	1 Black Instructor

[9] For year-long courses, “A” signifies the first semester and “B” signifies the second semester of a course.

Classroom observations of the physical computer lab settings were collected during the 2014-2015 through the 2018-2019 school years. Of those, 149 observations examined the experiences of individual students within the classroom, while 61 examined whole-class dynamics. Schools with a higher proportion of students enrolled in online courses were purposefully oversampled for classroom observations. Due to concern that in-person dynamics might shape student interactions with instructors in the online courses, I aggregated the information on student-live teacher congruence and time spent interacting with a live lab monitor to the school-by-year level and merged it with the primary student-lesson level data file to test the sensitivity of results to their inclusion as potential confounding influences. I was able to merge observation data with 75 percent of the student-lesson cases in the analytic sample.

Analytic strategy. Using a student-by-course fixed effect strategy, I examined associations between whether students experienced racial/ethnicity or gender congruence with the remote instructor for a given lesson and lesson outcomes. This strategy allowed for the exclusion of all variance associated with student and course-specific information that was constant during the semester that the students completed the course. Thus, the use of student-by-course fixed effects provided a more plausibly causal estimate than possible using covariate-adjusted OLS due to the removal of student and course characteristics that are fixed over time and might be related to both lesson participation and outcomes.

Dependent variables included the first score the student earned on the end-of-lesson quiz, whether the student had to retake the end-of-lesson quiz, and active time within a given lesson. Students were required to retake a quiz if they scored below the threshold required to earn course credit, which likely reflects less interest in or comprehension of course content, while active time provides a measure of behavioral engagement in addition to the achievement measures examined.

The estimation strategy is summarized in the equation below. I estimated the model separately for each racial/ethnic and gender group, limiting the analytic sample to students who met the condition of interest enrolled in courses with variation in teachers based on that same condition.

$$\gamma_{icl} = \beta_0 + \beta_1 inst\_identity_{icl} + \boldsymbol{S_i \beta} + \alpha_{ic} + \varepsilon_{icl} \qquad (1)$$

The use of fixed effects allowed for the examination of associations with student-teacher racial/ethnicity or gender congruence for a given student from lesson to lesson (*i*) within a given course (*c*) for each course lesson (*l*). The equation controls for a vector of student (*S*) covariates, including the lesson order in the course and type of activities included in the lesson. These activities included a quiz (observed in 95 percent of lessons), warm-up exercise (56 percent), assignment (53 percent), review slides (51 percent), learning definitions of keywords (29 percent), using an outside electronic resource (4 percent), or reading an accompanying text (4 percent). These activities serve as a proxy for lesson expectations and difficulty that are likely to be related to both end-of-lesson quiz scores and the amount of time required to complete lesson content. I controlled for these activities, as curriculum experts – and not the remote instructors – choose how to design the course and which activities to include. I also controlled for whether the student took the pretest for a given lesson, which provided the opportunity to test out of some lesson content, as well as the score received on the pretest where applicable. All models used robust standard errors, clustered at the student-by-course level, to account for correlations between the error terms.

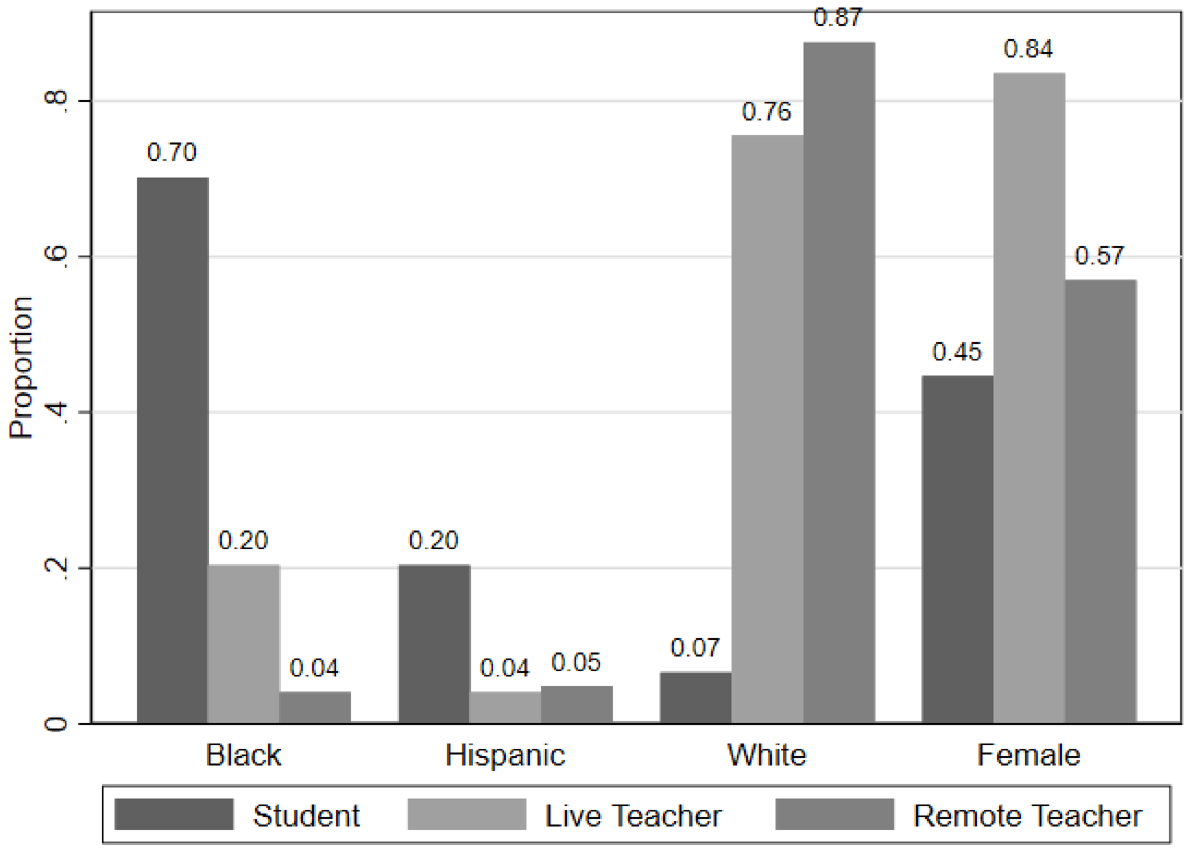
Next, I examined heterogeneous effects using equation one by limiting the sample to each subject or subgroup of interest. Based on my review of prior research, I examined estimates for each of the four core subjects (ELA, social studies, math, and science), as well as for students who qualified for FRL and students whose prior year GPA was below a 2.0 (indicating a C average). To examine intersectionality, I also explored racial/ethnic congruence, limiting the sample to first female and then male students.

Lastly, I ran several sensitivity tests to demonstrate robustness to alternative model specifications and samples (summarized in Table 21). I tested whether the inclusion of school or grade fixed effects resulted in qualitatively similar estimates to determine whether any identified effects might be context specific. I also examined whether the number of lessons within a course that the remote teacher taught might be confounded with exposure to a given teacher and student outcomes. An additional concern might be that the race/ethnicity and gender of live computer lab monitors might confound the effect of any student congruence with remote instructors. Online students received an average of only two minutes of interactions with their lab instructor in classroom observations. Nonetheless, I integrated information on live computer lab monitors' presenting race/ethnicity and gender collected from a purposeful sample of classroom observations to examine the validity of this potential confounder as a sensitivity test. Next, I controlled for median course length, since particularly engagement (and through engagement achievement) might be impacted by lesson length. I also accounted for the extent to which each lesson facilitated authentic work (i.e., asked students to engage in higher-order thinking and provided real-world relevance). Both lesson length and authenticity are dictated by the curriculum developers and not the remote instructor, so accounting for these characteristics allows for an examination of whether any previously identified effects might be due to systematic assignment of teachers of a given race/ethnicity or gender to different types of lessons. This sensitivity test controlling for authentic work restricted the sample to the top ten courses with the largest student enrollment. The last two sensitivity tests also restricted the analytic sample, first to only lessons where students had pretest scores and second to only students who completed all lessons in the course. Although the samples are not expected to be equivalent, the inclusion of estimates from these two analytic samples provides estimates (for students with available information) that more fully account for prior knowledge and are not influenced by potential variability due to associations between lesson order and course retention.

Results

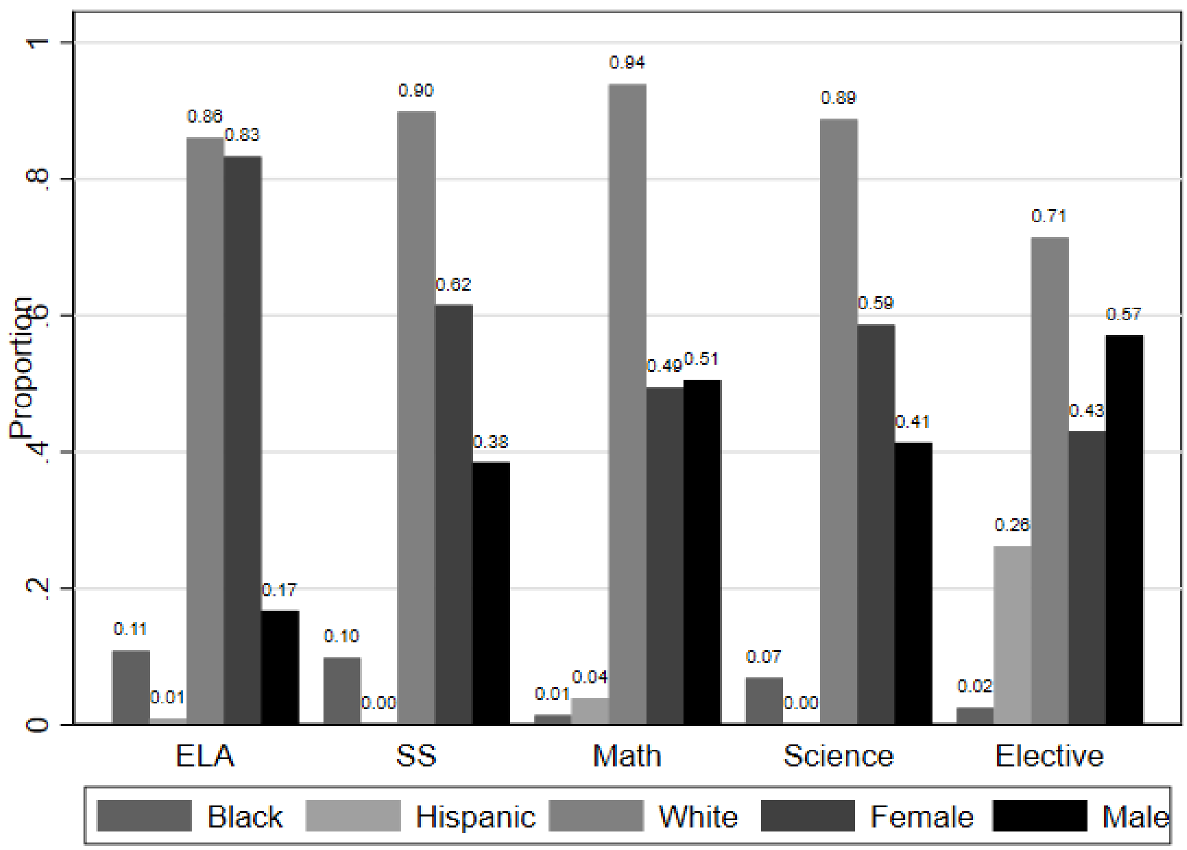
Student and instructor demographic identities. As shown in Figure 9, seven percent of students in the analytic sample identified as White and 45 percent as female compared to 87 percent of remote instructors who presented as White and 57 percent that presented as female. In instances where the percentage of students or teachers do not add up to 100 percent, the additional percentages represent individuals identified as a race/ethnicity not listed, such as Asian or Pacific Islander. The live lab instructors hired by the school district more closely mirrored the student population in terms of racial and ethnic diversity, although discrepancies remained. Consequently, students identified as Black and Hispanic were each taught by a remote instructor of the same race/ethnicity in only five percent of student-lesson observations. In contrast, students identified as White were taught by an instructor of the same race in 87 percent of observations.

Figure 9. Student and Teacher Racial/Ethnic and Gender Proportions



The proportion of instructors who identified as Black or Hispanic varied slightly by subject, as shown in Figure 10. I identified the largest percent in elective, ELA, and social studies courses (28, 12, and 10 percent of lessons respectively) compared to only five percent of instructors in math and seven percent in science. The gender distribution of instructors also varied by subject. In ELA, 83 percent of instructors presented as female, with 43 to 62 percent of instructors presenting as female in other subject areas.

Figure 10. Remote Instructor Racial/Ethnic and Gender Proportions by Subject



Student-live instructor interactions. Due to concern that dynamics with the live lab monitor might influence the course outcomes of interest, I first examined how individual student experiences varied by the majority race/ethnicity or gender of the live lab monitor(s) and whether there was student-teacher identity congruence, as shown in Figure 11 and Figure 12. Chi-squared tests indicated that there were no significant differences in ratings of instructor engagement or interactions by either race/ethnicity or gender congruence between students and live lab monitors. Ratings of instructor engagement, interactions, and presenting race/ethnicity and gender were generated by the research team conducting observations in the school-based computer labs and classrooms where students were assigned to complete online lessons during the school day. Refer to Appendix F for more information on how each construct was defined.

Figure 11. Instructor Engagement by Race/Ethnicity and Gender Match

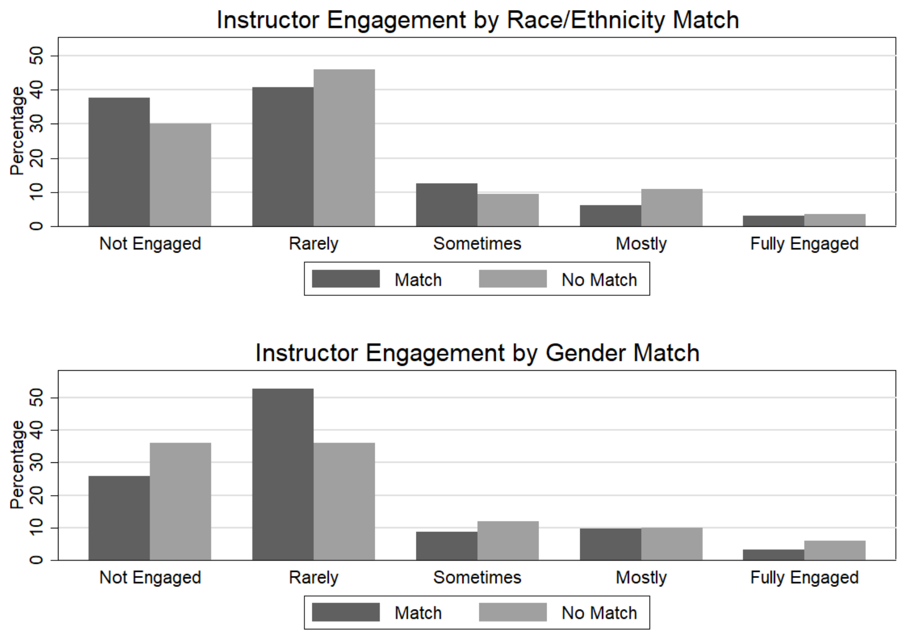
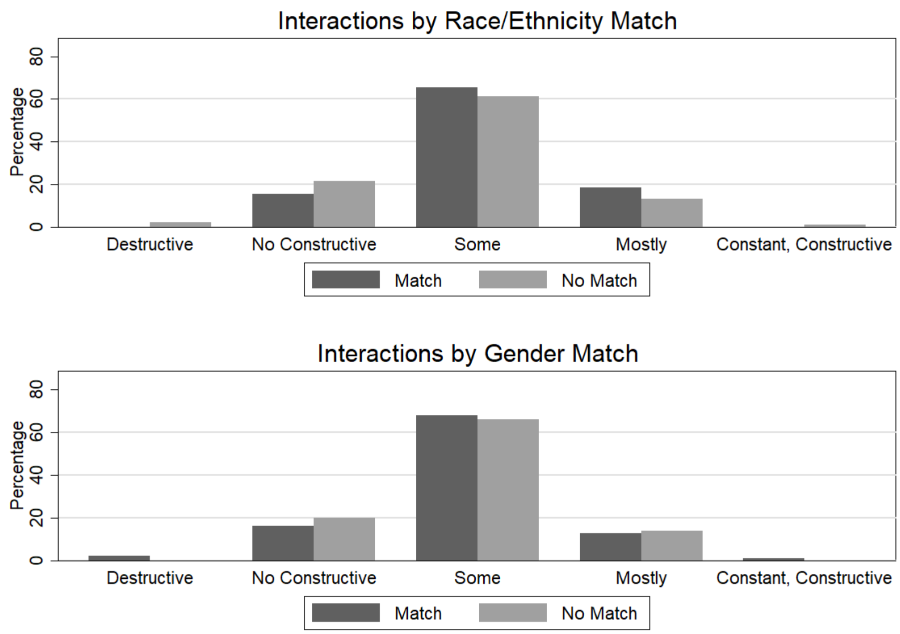


Figure 12. Interactions by Race/Ethnicity and Gender Match



Beyond the lack of a significant relationship between student-teacher congruence and classroom experiences, limited interactions between students and live lab monitors further minimized concerns regarding the feasibility that live lab monitors might confound the main estimates. Individual students enrolled in online courses spent an average of 4.74 percent (std. dev. = 7.97 percent) of classroom time interacting with live lab instructors. For a forty-minute class period, that translates into approximately two minutes of interaction. Further, most of these interactions were administrative versus instructional. The nature of these interactions was described as providing minimal constructive interactions with the modal student (in 67 percent of observations). Further, the most common rating of instructor engagement (in 46 percent of observations) indicated that instructors were rarely engaged in instruction.

Student-remote instructor identity congruence. Next, I examined lesson outcomes by remote instructor characteristics using a student-by-course fixed effect, as summarized in Table 19. I observed no significant associations between student performance or engagement and racial/ethnic or gender congruence with the remote instructor. In addition to the lack of significance, the magnitude of estimates was small for students belonging to minoritized groups (less than or equal to 1.1 percent on the end-of-less quiz, 0.4 percent for never retaking, and 1.6 percent for active time). The standard errors on these estimates indicates that a value of approximately one to two percent would have been identified as significant in most cases, demonstrating a high level of precision.

Some courses have only one or two lessons taught by an instructor who belonged to a minoritized group. Due to concerns that this might mean that the treatment intensity was too small in some courses, I next limited the sample to courses where an instructor of the same race/ethnicity taught at least 10 percent of lessons. The estimates from these models were consistent in direction, magnitude, and significance from the main estimates apart from a significant association between never having to retake a quiz to earn credit and being taught by an instructor presenting as Black for students identified as Black (β=-0.007, SE=0.003). Although significant, this coefficient is small in magnitude, indicating that students identified as Black were less than one percent less likely to pass the end-of-lesson exam the first time if taught by an instructor of the same race.

Table 19. Student Performance and Behavior on Online Lessons When Taught by a Same race/Ethnicity and Gender Instructor

	Score (Std.)	Never Retake	Active Time (Log)
Black student-instructor match	-0.011	-0.004	-0.009
	(0.012)	(0.003)	(0.012)
Hispanic student-instructor match	0.009	-0.001	0.016
	(0.027)	(0.005)	(0.025)
White student-instructor match	-0.028	-0.003	0.011
	(0.027)	(0.006)	(0.025)
Male student-instructor match	0.006	0.002	-0.008
	(0.008)	(0.002)	(0.008)
Female student-instructor match	-0.005	-0.003	-0.003
	(0.008)	(0.002)	(0.009)

* 0.10 ** 0.05 *** 0.01 significance level. Each cell represents estimates from a different model.

I also was concerned that the main estimates might have been biased if vendor hiring practices resulted in differential instructor quality by race/ethnicity or gender. For that reason, I also provided estimates from models that controlled for instructor race/ethnicity or gender for all students separately (refer to Appendix G). In these models, the main effect is the interaction between a student and an instructor of the same race/ethnicity or gender. I identified a 0.8 percent decreased likelihood of never retaking among students identified as Black when taught by a remote instructor presenting as Black, but this association was not significant at the 0.05

level. Further, estimates remained qualitatively similar in terms of directionality and magnitude in most instances; the few instances where the estimate shifted in directionality were non-significant. However, it is also notable that excluding any association due to racial or ethnic congruence, students logged approximately four percent less time in a lesson when the remote instructor presented as Hispanic and approximately two percent more time in lessons when the remote instructor presented as White. Students (regardless of gender match) were 0.4 percent less likely to pass the end-of-less quiz on the first attempt when taught by a female versus male remote instructor. Despite significance, the magnitude of this last association is so small as to have little practical meaning.

Variation by subject and subgroup. Prior literature in face-to-face instructional settings indicated that students were likely to respond differently to racial/ethnic or gender congruence with their instructor depending on the course subject and student characteristics. I examined whether there was evidence of these heterogeneous effects in online courses in Table 20. Consistent with the estimates from the main model specifications, few of the subgroup analyses identified significant associations apart from a few subject-specific patterns. In science, both students presenting as Black and those presenting as female were one to two percent less likely to pass the end-of-lesson quiz on their first attempt when taught by a same race or same gender remote instructor. In math, students identified as Hispanic scored 12.4 percent lower when taught by an instructor presenting as Hispanic. This reflects the experiences of the 677 students enrolled in a math course with one or more Hispanic instructors. Further, students identified as male scored 2.6 percent higher in social studies when taught by an instructor presenting as male. I was

Table 20. Student-Teacher Congruence by Subject and Subgroup

	ELA	SS	Math	Science	FRL	< 2.0 Prior GPA	Female	Male
				Dependent Variable: First Score (Std.)				
Black match	-0.028	-0.002	-0.049	0.016	-0.016	-0.018	-0.014	-0.008
	(0.021)	(0.020)	(0.053)	(0.027)	(0.013)	(0.014)	(0.017)	(0.018)
Hispanic match	0.062	----	-0.124**	----	-0.006	0.051*	-0.010	0.020
	(0.112)		(0.052)		(0.030)	(0.028)	(0.053)	(0.028)
White match	-0.013	-0.029	0.064	0.073	0.008	-0.023	-0.006	-0.042
	(0.047)	(0.068)	(0.109)	(0.087)	(0.033)	(0.038)	(0.050)	(0.031)
Male match	0.026	0.026**	-0.024	0.003	-0.001	0.006	----	----
	(0.017)	(0.013)	(0.016)	(0.024)	(0.009)	(0.009)		
Female match	0.002	0.004	-0.005	-0.030	-0.006	-0.004	----	----
	(0.018)	(0.015)	(0.017)	(0.027)	(0.010)	(0.010)		
				Dependent Variable: Never Retake				
Black match	-0.006	-0.004	0.029	-0.014**	-0.004	-0.006	-0.002	-0.005
	(0.005)	(0.005)	(0.018)	(0.006)	(0.003)	(0.003)	(0.004)	(0.004)
Hispanic match	-0.005	----	-0.000	----	-0.001	0.001	-0.012	0.005
	(0.021)		(0.016)		(0.006)	(0.006)	(0.009)	(0.006)
White match	-0.003	0.011	-0.020	-0.010	-0.004	-0.006	-0.010	0.002
	(0.014)	(0.014)	(0.021)	(0.012)	(0.007)	(0.009)	(0.009)	(0.007)
Male match	-0.000	-0.002	0.001	0.006	0.002	0.002	----	----
	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)	(0.002)		
Female match	-0.004	0.001	-0.003	-0.017***	-0.001	-0.003	----	----
	(0.006)	(0.004)	(0.004)	(0.005)	(0.002)	(0.003)		
				Dependent Variable: Active Time (Log)				
Black match	-0.014	-0.031*	0.059	-0.003	-0.001	-0.011	0.000	-0.004
	(0.024)	(0.017)	(0.048)	(0.031)	(0.013)	(0.015)	(0.018)	(0.017)
Hispanic match	0.071	----	0.031	----	0.009	-0.001	0.043	0.000
	(0.049)		(0.066)		(0.026)	(0.024)	(0.040)	(0.032)
White match	0.043	-0.014	0.014	-0.061	-0.030	0.043	0.035	-0.000
	(0.056)	(0.055)	(0.088)	(0.068)	(0.031)	(0.034)	(0.043)	(0.031)
Male match	-0.018	-0.023	-0.003	0.016	-0.007	-0.006	----	----
	(0.013)	(0.016)	(0.015)	(0.021)	(0.009)	(0.009)		
Female match	0.028	0.001	-0.007	0.008	-0.003	-0.008	----	----
	(0.018)	(0.015)	(0.015)	(0.027)	(0.009)	(0.011)		

* 0.10 ** 0.05 *** 0.01 significance level. Each cell represents estimates from a different model. No remote instructors of math or science courses were identified as Hispanic.

not able to examine differential effects for students identified as Hispanic in some subjects, as most instructors presenting as Hispanic taught in math or elective courses.

Sensitivity tests. In Table 21, I demonstrated that estimates examining student-teacher congruence by race/ethnicity and gender were robust to the inclusion of school fixed effects, grade fixed effects, remote instructor frequency, live lab monitor engagement and identity characteristics, median lesson length, and lesson authenticity with one exception. I also estimated models for two separate, reduced analytic samples by limiting cases to only students with pretest scores for a given the lesson and then to only students who completed all lessons in their course.

The estimation of qualitatively similar estimates when including a school fixed effect (see Table 21, Column 1) indicates that students did not appear to respond differently based on their school setting or student characteristics associated with school assignment that were not already accounted for within the base model. This lack of significant differences was expected, since students in each school setting accessed the same lessons in very similar manners with generally little interaction with live lab monitors. Qualitatively similar estimates when including grade fixed effects (see Table 21, Column 2) demonstrates that results were unlikely to be biased by variation in students' grade-level or other school experiences associated with their grade. Next, I controlled for the number of lessons in a course each remote instructor taught to test whether any identified estimates might instead be attributed to greater familiarity with the remote instructor (see Table 21, Column 3). The inclusion of this covariate resulted in no significant differences in the estimates, indicating that exposure to a particular remote instructor did not appear to be associated with changes in students' responsiveness to racial/ethnic or gender congruence.

Next, I merged school-by-year level observation data on live lab monitor engagement, race/ethnicity, and gender. Including these variables in the models allowed me to examine the

Table 21. Sensitivity Tests for Student-Teacher Congruence Models

	(1) School Fixed Effect	(2) Grade Fixed Effect	(3) Remote Teacher Frequency	(4) Live Teacher Match	(5) Live Teacher Engagement	(6) Course Length (Median)	(7) Authentic Work in Lesson	(8) Only Pretested Lessons	(9) Only Completed Courses
Dependent Variable: First Score (Std.)									
Black student-instructor match	-0.011	-0.011	-0.008	-0.013	-0.006	-0.011	-0.018	-0.093	0.035
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.017)	(0.076)	(0.022)
Hispanic student-instructor match	0.013	0.009	0.012	0.004	0.004	0.009	0.002	0.089	0.131
	(0.027)	(0.027)	(0.028)	(0.029)	(0.029)	(0.027)	(0.046)	(0.127)	(0.107)
White student-instructor match	-0.028	-0.029	-0.026	-0.025	-0.025	-0.028	0.010	-0.057	0.040
	(0.027)	(0.027)	(0.028)	(0.029)	(0.029)	(0.027)	(0.039)	(0.145)	(0.069)
Male student-instructor match	0.006	0.007	0.014*	0.005	0.004	0.006	-0.001	0.046*	-0.023
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.025)	(0.014)
Female student-instructor match	-0.006	-0.005	-0.011	-0.010	-0.010	-0.005	-0.008	-0.045*	-0.012
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.012)	(0.025)	(0.016)
Dependent Variable: Never Retake									
Black student-instructor match	-0.004	-0.004	-0.005	-0.001	-0.002	-0.004	-0.001	-0.001	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.023)	(0.006)
Hispanic student-instructor match	-0.001	-0.000	0.000	-0.001	-0.001	-0.001	-0.014	-0.035	-0.051***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.011)	(0.026)	(0.018)
White student-instructor match	-0.003	-0.004	-0.003	-0.005	-0.004	-0.003	0.002	-0.073**	0.017
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.029)	(0.012)
Male student-instructor match	0.001	0.001	-0.000	0.001	0.001	0.002	-0.004	-0.005	-0.008**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.007)	(0.004)
Female student-instructor match	-0.003	-0.003	-0.002	-0.002	-0.002	-0.003	0.000	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.007)	(0.004)

	Dependent Variable: Active Time (Log)								
Black student-instructor match	-0.006	-0.008	-0.008	-0.011	-0.013	-0.009	0.045***	0.041	-0.011
	(0.012)	(0.012)	(0.012)	(0.014)	(0.013)	(0.012)	(0.017)	(0.061)	(0.027)
Hispanic student-instructor match	0.016	0.018	-0.001	0.021	0.022	0.016	0.009	0.154	0.091
	(0.025)	(0.025)	(0.026)	(0.027)	(0.027)	(0.025)	(0.037)	(0.103)	(0.057)
White student-instructor match	0.012	0.013	0.009	0.031	0.030	0.011	-0.011	-0.043	-0.093
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.041)	(0.086)	(0.066)
Male student-instructor match	-0.007	-0.009	-0.002	-0.003	-0.003	-0.008	0.005	-0.040**	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.018)	(0.016)
Female student-instructor match	-0.000	-0.003	-0.006	-0.001	-0.001	-0.003	-0.004	0.065***	0.013
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)	(0.020)	(0.019)

* 0.10 ** 0.05 *** 0.01 significance level. Each cell represents estimates from a different model.

extent to which any variance in student outcomes attributed to the online instructors could be explained by variation in the presenting identity (see Table 21, Column 4) or interactions with the students' live lab monitor (see Table 21, Column 5). As expected, due to minimal interactions recorded between students and live lab monitors in the lab settings, the inclusion of these characteristics did not significantly change the estimated coefficients associated with student-remote instructor racial/ethnic congruence.

Estimates when controlling for median course length (see Table 21, column 6) were similar in terms of magnitude and significance, indicating that identified estimates could not be attributed to differences in course length. Results also remained comparable when accounting for authentic work – the extent to which the lesson facilitated higher-order thinking and provided real-world relevance – with one exception (see Table 21, Column 7). After accounting for the extent to which students were provided opportunities for authentic work in a lesson, a positive association between student-instructor racial congruence emerged for students identified as Black when examining active time. This estimate indicated that conditional on the level of authenticity in the lesson, students identified as Black logged 4.5 percent more active time when taught by a remote instructor presenting as Black.

I next estimated models limiting the analytic sample to cases where the student completed a pretest. Pretests were lesson specific. Over 3,500 students completed one or more pretest. However, only about half of the student-lesson observations included pretest scores, as taking a pretest was optional. When accounting for students' pretest scores, I identified that students identified as White were 7.3 percent less likely to pass the end-of-lesson quiz the first time when taught by a same race instructor. Students identified as female also logged 6.5 percent more minutes when taught by a same gender instructor, while students identified as male logged

4.0 percent *fewer* minutes when taught by a same gender instructor. I identified no significant associations among students identified as Black or Hispanic when taught by a remote instructor of the same race/ethnicity. Differences between samples may account for at least some of the discrepancies identified. For instance, students were most likely to take a pretest for Career Planning (77 percent of student-lesson observations included a pretest), Citizenship (75-77 percent), and Geometry (71-72 percent) courses compared to other courses where students took a pretest in fewer than 17 percent of student-lesson observations. However, differences in student characteristics between student-lesson observations with and without pretest scores were relatively small in magnitude.

Similarly, I estimated models limiting the analytic sample to students who completed all lessons in the course (see Table 21, Column 9) to examine whether selection bias might be a concern, since students who only completed some lessons were more likely to be exposed to only those instructors in the first couple of lessons. Thus, if instructors with a certain identity disproportionately taught lessons early or late in a course, estimates could be biased by the students who accessed only earlier versus all lessons. In total, 1,621 students met the criteria for this restricted sample, indicating a substantially reduced sample. When limiting the analytic sample in this way, students identified as Hispanic were 5.1 percent less likely to need to retake an end-of-lesson quiz among students when taught by a remote instructor presenting as Hispanic. However, this finding only reflected the experiences of 36 students identified as Hispanic, due to the restricted nature of the same. On the other hand, the negative association between never retaking and gender-congruence among students identified as male reflected the experiences of 594 students identified as male.

A contextual exploration of mechanisms. The above analysis provides useful information on the minimal extent to which students engaged with or performed differently when taught by a remote instructor of the same race/ethnicity or gender in a prerecorded, standardized learning environment. These estimates provide evidence regarding the limitations of symbolic interaction in the online context studied. However, there are still several possible mechanisms that could explain null effects in an online context that would not necessarily generalize to a traditional, face-to-face classroom. Specifically, students might not benefit from symbolic representation in an online instruction setting because stereotype threat or role model effects were less likely to be triggered when interacting with an online versus face-to-face instructional system.

In order to examine the plausibility that the predominately null estimates observed might be in response to the different relevance of identity in the asynchronous online classroom context, I compared differences in student performance on the end-of-lesson quiz in the online courses by race/ethnicity and gender to differences in the same students' performance on the pretest for the same lesson, where available. This fully descriptive examination represents only a rough approximation versus the more rigorous estimates generated through the primary study models. Results should be interpreted with this caveat in mind.

As shown in Table 17, the characteristics of the 49 percent of students who took at least one pretest were qualitatively similar to the larger analytic sample. Among student-lesson observations with both pretest and posttest scores, students identified as Black scored on average two percent higher on the pretest (a difference of approximately 0.05 of a standard deviation) but the same on the posttest as students identified as White. Similarly, students identified as Hispanic scored an average of one percent higher on the pretest and two percent lower on the posttest than students identified as White. The slight drop in the relative performance of students belonging minoritized groups between the pretest (when the student had only been exposed to a teacher within a traditional, face-to-face classroom) to the posttest (when the student was taught using the online course system) does not appear to support the hypothesis that the asynchronous, online course structure removed barriers (such as those associated with stereotype threat or teacher implicit bias) that might be activated in a traditional, face-to-face setting. The drop in performance among students belonging to minoritized groups from pretest to posttest instead provides suggestive evidence that these students might experience a relative advantage from aspects of learning in a traditional, face-to-face environment (versus the asynchronous, online system examined) compared to students belonging to dominant groups.

A different pattern emerged when conducting the same examination by gender with students identified as male scoring an average of 11 points lower on the pretest than students identified as female, a difference that was eliminated on the posttest. The magnitude of the difference remained constant when controlling for student- and lesson-level covariates. When examining across core subjects, students identified as male gained on average nine and 14 percent more between the pre- and posttests than students identified as female in ELA and social studies respectively, which literature has established as being considered more traditionally feminine (Solanki & Xu, 2018; Keiser et al., 2002; Xu & Li, 2018). In contrast, students identified as male gained an average of only one and four percent more between the pre- and posttests in math and science respectively, which are historically considered more male-dominated fields.

In contrast to the analysis of achievement gaps among students by race and ethnicity, this analysis by gender provides suggestive evidence that male students might benefit more from the online instructional setting, particularly in fields considered more female-dominated such as ELA and social studies. These findings are also consistent with the positive association identified between end-of-lesson quiz scores and being taught by an instructor of the same gender for male students in the subgroup analysis, with estimates of a similar magnitude also observed in ELA. These findings by gender demonstrate that where there was evidence of reduced barriers in the online instructional setting (i.e., stereotype threat was not triggered or teacher implicit bias didn't shape interactions), students also benefited from being taught by an instructor of the same identity.

Contributions and limitations. This is one of the first studies to examine the characteristics of online course instructors. As such, the reporting of summary statistics represents a contribution to the literature on vendor-developed courses, particularly in light of the vast disparities in how often students experienced racial/ethnic congruence by identity documented by this study. Beyond establishing demographics, the unique nature of online course-taking allows for greater separation of any possible educational benefits of symbolic versus active bureaucratic representation, which would be more challenging in a traditional, face-to-face instructional setting. Findings inform policy discussions regarding which strategies associated with symbolic and active representation are most likely to improve the educational experiences and achievement of historically marginalized students. Another strength of this study is the relatively large sample size and detailed available data that allow for the use of student-by-course fixed effects. Further, while concerns regarding generalizability may remain to other online contexts, student populations, and to online courses in which no racial or ethnic variation in instructor identity was observed, claims of generalizability are strengthened by the examination of content developed by one of the largest online course vendors in the United States.

Despite these strengths, there are some limitations to the study design that should be considered when interpreting results. Observers rated instructor race/ethnicity and gender. Thus, these labels represent the instructors' presenting identity and not necessarily how the instructors themselves would identify. While there is merit to self-identification, presenting identity is likely more relevant to student perceptions of an instructor's identity. Perhaps more importantly, the low proportions of instructors presenting as a racial/ethnic minority reduced my effective sample size and potentially represented too small an intervention to induce a student response. I attempted to test whether this might be influencing my results by examining associations with student-teacher congruence when limiting the analytic sample to only those courses where 10 percent or more (up to 57 percent) of lessons were taught by an instructor of the same race/ethnicity. I identified generally consistent estimates between this restricted and the non-restricted samples.

Further, the predominately null effects identified do not preclude the possibility that students might benefit from student-instructor congruence on non-measured outcomes. It is also not possible to directly compare effect sizes to those identified in prior studies, as most prior studies used general content area standardized assessments instead of the more closely aligned content-specific end-of-lesson quizzes examined in this study. Lastly, the sample for this study included students enrolled in online courses predominately for the purpose of credit recovery within a large, urban, low-resourced school district. The experiences of these students are of great importance to researchers and practitioners interested in educational equity. At the same time, this constraint should be recognized when interpreting and attempting to generalize any findings to a wider population of students.

There are also some limitations to the sensitivity tests. There may remain associations between instructor race/ethnicity or gender and instructional content relevant to student outcomes that I was not able to identify and control for. Also, the tests that controlled for live lab monitor engagement, race/ethnicity, and gender were limited to only those schools observed in a given year. Thus, these analyses were limited to 75 percent of the larger analytic sample. Since schools were purposefully sampled for in-person classroom observations based on the number of students enrolled online, results may not hold for students attending schools with fewer students taking online courses.

Discussion

Within the large, urban school district in which the study occurred, students completing online courses in school-based computer labs, primarily for the purpose of credit recovery, interacted infrequently with live lab monitors. Of the approximately two minutes of interaction between the lab instructor and students during a typical 40-minute class period, most involvement was administrative versus instructional. While the racial/ethnic distribution of the live lab monitors more closely mirrored the student population than the distribution of remote teachers, I observed no significant difference in interactions or instructor engagement by student-live lab monitor racial/ethnic or gender congruence.

When examining the extent to which being taught by a remote instructor of the same race/ethnicity or gender was associated with course outcomes, I identified few significant associations. These results were robust to several sensitivity tests, including those accounting for remote teacher frequency and live lab monitor characteristics and interactions. In my examination by subject and subgroup (where the sample size was not severely restricted), I found that male students scored slightly higher in social studies when taught by a same gender teacher, while students identified as Black and female were 1.4 and 1.7 percent more likely, respectively, to score so low on their first attempt that they had to retake the end-of-lesson quiz when taught by a same race or same gender instructor.

The relatively precise null findings from the main analyses (and for most subgroups) are notable, as researchers have identified a generally positive association between student-teacher congruence and outcomes for students from minoritized groups in traditional, face-to-face classrooms where both symbolic and active forms of representation are feasible (Dee, 2004, 2005; Egalite et al., 2015; Gershenson et al., 2017; Keiser et al., 2002). Although less rigorous, the supplemental analysis showing lower relative achievement among students belonging to a minoritized group on posttest versus pretest scores appears to contradict concerns that student-instructor congruence failed to induce positive achievement in the asynchronous, online setting because students didn't experience the same barriers (i.e., stereotype threat, teacher implicit bias) online. Thus, the finding that the mechanisms through which representation could benefit students in a standardized online course context were not associated with increased engagement or achievement indicates that other mechanisms, notably active forms of representation, likely contribute a larger portion of the benefit of representation in the classroom.

These results could also indicate that symbolic representation isn't triggered in impersonal settings where bureaucrat and client don't interact. However, this appears less likely, since students identified as male did achieve at higher levels when taught by a male instructor in more traditionally female-dominated fields (Solanki & Xu, 2018; Keiser et al., 2002; Xu & Li, 2018). Instead, the lack of a positive association for students from minoritized groups might be because the intervention – being taught by an instructor of the same race/ethnicity for as few as one in 45 lessons – was too small to induce an effect. However, the identification of generally consistent estimates when restricting the sample to courses where an instructor belonging to a minorized group taught at least 10 percent of lessons does not support this hypothesis.

Implications for Research, Policy, and Practice

This study contributes to the larger literature on representation within education for minorized student populations (i.e., Egalite et al., 2015; Gershenson et al., 2017; Grissom et al., 2015). Compared to staffing traditional classrooms, it is relatively easier to provide students access to instructors belonging to minorized racial/ethnic groups through online courses due to the standardized nature of the lesson. Instead of positively influencing a single class of 30 students, increasing the diversity of instructors in even a single course developed by the online course vendor studied has the potential to impact the academic experiences of tens of thousands of students across the country during any given year. Despite this potential, I identified less racial/ethnic diversity among remote instructors within the most frequently accessed courses than among live lab monitors within the district studied.

Specifically, the use of online session-level course data allowed for the identification of within student and course variation by student-teacher congruence for courses in which the race/ethnicity or gender of the instructor varied by lesson, demonstrating that positive associations identified in prior literature conducted in traditional, face-to-face settings examining racial/ethnic congruence were more likely associated with teacher discretion, a key prerequisite to active forms of representation in the classroom which was not present in the asynchronous, online courses. Students identified as male, however, did achieve higher performance when taught by an instructor of the same gender in female-dominated fields. As such, online learning in its current form, particularly when the instructor is of the same gender in female-dominated fields may be particularly beneficial for students identified as male requiring credit-recovery. Continuing to expand representation in online courses is also merited, particularly as benefits for students from minoritized groups may similarly be realized if the proportion of minoritized teachers increased beyond the low levels currently observed. Beyond improved achievement, normalizing instructors of different identities as experts across subject areas could have broader societal benefits for all students.

The authors of prior studies identifying positive student outcomes associated with student-teacher congruence called for a range of policy interventions from increasing the diversity of the teacher workforce to minimizing teacher implicit bias (Dee, 2005; Egalite et al., 2015; Gershenson et al., 2016, 2017; Grissom et al., 2015; Irizarry, 2015; Joshi et al., 2018; Lindsay & Hart, 2017). The findings of this study allow policy-makers and practitioners to develop more effective, targeted policy interventions. Suggestive evidence of the importance of active representation for students by both race/ethnicity and gender means that research and policy targeting how teachers make discretionary decisions are likely to be particularly fruitful. Possible areas for examination include minimizing the role of implicit bias in the classroom, integrating culturally relevant pedagogy, and re-evaluating how students are identified for gifted coursework, special education services, and disciplinary referrals (Burgess & Greaves, 2013; Carlana, 2019; Cherng & Halpin, 2016; Dee & Penner, 2019; Fish, 2017; Grissom et al., 2009, 2015; Lindsay & Hart, 2017). This also means that providing professional development that changes how teachers understand and respond to these types of discretionary decisions is likely to have a positive impact on subsequent student experience and outcomes. Future research should continue to disentangle the effects of these active mechanisms associated with teacher discretion to provide more nuanced and causal evidence regarding which policies are most likely to replicate the positive academic outcomes associated with exposure to teachers belonging to minoritized racial and ethnic groups in traditional, face-to-face settings.

References

Ahn, J., & McEachin, A. (2017). Student enrollment patterns and achievement in Ohio's online charter schools. *Educational Researcher, 46*(1), 44-57.

Alpert, W. T., Couch, K. A., & Harmon, O. R. (2016). A randomized assessment of online learning. *American Economic Review, 106*(5), 378-82.

Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? *Computers & Education, 113*, 226-242.

Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's guide.* Princeton, NJ: Princeton University Press.

Anyon, J. (1980). Social class and the hidden curriculum of work. *Journal of Education*, 67-92.

Apple, M. W. (2004). *Ideology and curriculum.* New York, NY: Routledge.

Au, K. H. (1998). Social constructivism and the school literacy learning of students of diverse backgrounds. *Journal of Literacy Research, 30*(2), 297-319.

Au, W. (2012). *Critical curriculum studies: Education, consciousness, and the politics of knowing.* Routledge.

Bakia, M., Mislevy, J., Heying, E., Patton, C., Singleton, C., Krumm, A. (2013). *Supporting K–12 students in online learning: A review of online Algebra I courses.* Menlo Park, CA: SRI International.

Bakia, M., Shear, L., Toyama, Y., Lasseter, A. (2012). *Understanding the implications of online learning for educational productivity.* Washington, D. C.: U.S. Department of Education.

Bernstein, B. (1975). Class and pedagogies: Visible and invisible. *Educational Studies, 1*(1), 23-41.

Bettinger, E. P., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review, 107*(9), 2855-75.

Bidwell, C. E., Frank, K. A., & Quiroz, P. A. (1997). Teacher types, workplace controls, and the organization of schools. *Sociology of Education*, *70*(4), 285-307.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.* Washington, D.C.: US Department of Education, Office of Educational Technology.

Bixler, B. A. 2008. *The effects of scaffolding student's problem-solving process via question prompts on problem solving and intrinsic motivation in an online learning environment.* PhD diss., The Pennsylvania State University, State College, Penn.

Boozer, M., & Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. *Journal of Urban Economics, 50*(1), 163-189.

Boninger, F., Molnar, A., & Murray, K. (2017). *Asleep at the switch: Schoolhouse commercialism, student privacy, and the failure of policymaking.* Boulder, CO: National Education Policy Center.

Borman, G. D., & Pyne, J. (2016). What if Coleman had known about stereotype threat? How social-psychological theory can help mitigate educational inequality. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(5), 164-185.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Burch, P., & Good, A. G. (2014). *Equal scrutiny: Privatization and accountability in digital education.* Cambridge, MA: Harvard Education Press.

Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics, 31*(3), 535-576.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics, 134*(3), 1163-1224.

Carnevale, A. P., Smith, N., & Strohl, J. (2013). *Recovery: Job growth and education requirements through 2020.* Washington D. C.: Georgetown University Center on Education and the Workforce.

Cavalluzzo, L., Lowther, D. L., Mokher, C., & Fan, X. (2012). *Effects of the Kentucky Virtual Schools' Hybrid Program for Algebra I on grade 9 student math achievement* (NCEE 2012-4020). Washington, D.C.: National Center for Education Evaluation and Regional Assistance.

Cavus, N., H. Uzonboylu, and D. Ibrahim. 2007. Assessing the success rate of students using a learning management system together with a collaborative tool in Web-based teaching of programming languages. *Journal of Educational Computing Research 36* (3):301–21.

Chen, B. 2007. Effects of advance organizers on learning and retention from a fully Web-based class. PhD diss., University of Central Florida, Orlando.

Cherng, H. Y. S., & Halpin, P. F. (2016). The importance of minority teachers: Student perceptions of minority versus White teachers. *Educational Researcher, 45*(7), 407-420.

Cook, D. A., M. H. Gelula, D. M. Dupras, and A. Schwartz. 2007. Instructional methods and cognitive and learning styles in Web-based learning: Report of two randomised trials. *Medical Education, 41*(9), 897–905.

Cottom, T. M. (2014). *Democratizing ideologies and inequality regimes in digital domains.* Cambridge, MA: Berkman Center for Internet & Society Series, Harvard University.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Crippen, K. J., and B. L. Earl. 2007. The impact of Web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. Computers & Education 49 (3):809–21.

Darling-Aduana, J., Good, A., & Geraghty, E. (2020). *Curricular cultural responsiveness and relevance in online high school courses.* Manuscript in preparation, Department of Leadership, Policy, and Organizations, Vanderbilt University, Nashville, Tennessee.

Darling-Aduana, J., Good, A., & Heinrich, C. J. (2019b). Mapping the inequity implications of help-seeking in online credit-recovery classrooms. *Teachers College Record, 121(11).*

Darling-Aduana, J., & Heinrich, C. J. (2018). The role of teacher capacity and instructional practice in the integration of educational technology for emerging bilingual students. *Computers & Education*, *126,* 417-432.

Darling-Aduana, J., & Shero, M. (2020). *The enactment of Common Core State Standards in an asynchronous, online algebra 1 course.* Manuscript in preparation, Department of Leadership, Policy, and Organizations, Vanderbilt University, Nashville, Tennessee.

Darling-Hammond, L. (2001). *The right to learn: A blueprint for creating schools that work.* San Francisco: Jossey-Bass.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics, 86*(1), 195-210.

Dee, T. S. (2005). A teacher like me: Does race/ethnicity or gender matter? *American Economic Review, 95*(2), 158–165.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness, 8*(2), 149-168.

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal, 56*(1), 127-166.

Dee, T.S., & Penner, E. (2019). *My brother's keeper? The impact of targeted educational supports* (CEPA Working Paper No.19-07). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp19-07.

DeVellis, R. F. (2016). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications.

Diamond, J. B. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, *80*(4), 285-313.

Digital Learning Collaborative. (2019). *Snapshot 2019: A review of K-12 online, blended, and digital learning.* Durango, CO: Evergreen Education Group.

Dinov, I. D., J. Sanchez, and N. Christou. 2008. Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. Computers & Education 50 (1):284–300.

Dynarski, S. (2018, January 19). Online courses are harming the students who need the most help. *New York Times*. Retrieved from http://www.nytimes.com

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review, 45*, 44-52.

Estrella, G., Au, J., Jaeggi, S. M., & Collins, P. (2018). Is inquiry science instruction effective for English Language Learners? A meta-analytic review. *AERA Open, 4*(2).

Fischer, C., Xu, D., Rodriguez, F., Denaro, K., & Warschauer, M. (2020). Effects of course modality in summer session: Enrollment patterns and student performance in face-to-face and online classes. *The Internet and Higher Education, 45*.

Fish, R. E. (2017). The racialized construction of exceptionality: Experimental evidence of race/ethnicity effects on teachers' interventions. *Social Science Research, 62*, 317-334.

Fuller, J. B., & Raman, M. (2017). *Dismissed by degrees: How degree inflation is undermining US competitiveness and hurting America's middle class.* Cambridge, MA: Harvard Business School.

Gamoran, A., & Nystrand, M. (1992). Taking students seriously. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 40-61). New York, NY: Teachers College Press.

Gamoran, A., Secada, W. G., & Marrett, C. B. (2000). The organizational context of teaching and learning: Changing theoretical perspectives. In M. T. Hallinan (Ed.), *Handbook of the Sociology of Education* (pp. 37-63). New York, NY: Kluwer Academic/Plenum.

Gao, T., & Lehman, J. D. (2003). The effects of different levels of interaction on the achievement and motivational perceptions of college students in a web-based learning environment. *Journal of Interactive Learning Research, 14*(4), 367.

Gemin, B., Paper, L., Vashaw, L., & Watson, J. (2015). *2015 keeping pace with k-12 digital learning: An annual review of policy and practice*. Durango, CO: Evergreen Education Group.

Gershenson, S., Hart, C. M. D., Lindsay, C. A. Papageorge, N. W. (2017). The long-run impacts of same race teachers (Discussion Papers, No. 10630). Bonn, Germany: Institute of Labor Economics (IZA).

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209-224.

Goodman, J., Melkers, J., & Pallais, A. (2019). Can online delivery increase access to education?. *Journal of Labor Economics*, *37*(1), 1-34.

Goffman, E. (1983). The interaction order: American Sociological Association, 1982 presidential address. *American Sociological Review, 48*(1), 1-17.

Golanics, J. D., and E. M. Nussbaum. 2008. Enhancing online collaborative argumentation through question elaboration and goal instructions. Journal of Computer Assisted Learning 24 (3):167–80.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook, Vol. 3. Application to learning and teaching* (pp. 451-473). Washington, DC, US: American Psychological Association.

Griner, A. C., & Stewart, M. L. (2013). Addressing the achievement gap and disproportionality through the use of culturally responsive teaching practices. *Urban Education, 48*(4), 585-621.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher, 44*(3), 185-192.

Grissom, J. A., Nicholson-Crotty, J., & Nicholson-Crotty, S. (2009). Race, region, and representative bureaucracy. *Public Administration Review, 69*(5), 911-919.

Haberman, M. (2010). The pedagogy of poverty versus good teaching. *Phi Delta Kappan, 92*(2), 81-87.

Hart, C. M., Berger, D., Jacob, B., Loeb, S., and Hill, M. (2019). Online learning, offline outcomes: Online course taking and high school student performance. *AERA Open*, *5*(1). DOI: 10.1177/2332858419832852.

Heinrich, C. J., & Darling-Aduana, J. (2020). Does online course-taking increase high school completion and open pathways to postsecondary education opportunities? *Educational Evaluation and Policy Analysis.*

Heinrich, C. J., Darling-Aduana, J., Good, A., & Cheng, H. (2019). A look inside online educational settings in high school: Promise and pitfalls for improving educational opportunities and outcomes. *American Educational Research Journal, 56*(6), 2147-2188.

Heissel, J. (2016). The relative benefits of live versus online delivery: Evidence from virtual algebra 1 in North Carolina. *Economics of Education Review, 53*, 99-115.

Heppen, J. B., Sorensen, N., Allensworth, E., Walters, K., Rickles, J., Taylor, S. S., & Michelman, V. (2017). The struggle to pass algebra: Online vs. face-to-face credit recovery for at-risk urban students. *Journal of Research on Educational Effectiveness*, *10*(2), 272-296.

Heppen, J.B., Walters, K., Clements, M., Faria, A., Tobey, C., Sorensen, N., and Culp, K. (2012). *Access to Algebra I: The effects of online mathematics for grade 8 students.* (NCEE 2012–4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Herrington, A. J., & Herrington, J. A. (2007). What is an authentic learning environment?. In L. A. Tomei (Eds.), *Online and distance learning: Concepts, methodologies, tools, and applications* (pp. 68-77). Hershey, PA: Information Science Reference.

Herrington, J., Oliver, R. & Reeves, T.C. (2006). Authentic tasks online: A synergy among learner, task and technology. *Distance Education*, *27*(2), 233-248

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., ... & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis, 27*(2), 111-132.

Hindera, J. J. (1993). Representative bureaucracy: Further evidence of active representation in the EEOC district offices. *Journal of Public Administration Research and Theory*, *3*(4), 415-429.

Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37-52.

Hoffman, B., & Ritchie, D. (1997). Using multimedia to overcome the problems with problem based learning. *Instructional Science, 25,* 97-115.

Hohlfeld, T. N., Ritzhaupt, A. D., Dawson, K., & Wilson, M. L. (2017). An examination of seven years of technology integration in Florida schools: Through the lens of the Levels of Digital Divide in Schools. *Computers & Education, 113*, 135-161.

Huberman, M., & Miles, M. B. (2002). *The qualitative researcher's companion.* Thousand Oaks, CA: Sage Publications.

Hunsader, P. D., Thompson, D. R., Zorin, B., Mohn, A. L., Zakrzewski, J., Karadeniz, I., ... & MacDonald, G. (2014). Assessments accompanying published textbooks: the extent to which mathematical processes are evident. *ZDM*, *46*(5), 797-813.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86.

Irizarry, Y. (2015). Selling students short: Racial differences in teachers' evaluations of high, average, and low performing students. *Social Science Research, 52*, 522-538.

Jacob, B., Berger, D., Hart, C., & Loeb, S. (2016). Can technology help promote equality of educational opportunities?. *Russell Sage Foundation Journal of the Social Sciences, 2*(5), 242-271.

Jaggars, S. S. (2014). Choosing between online and face-to-face courses: Community college student voices. *American Journal of Distance Education, 28*(1), 27-38.

Joshi, E., Doan, S., & Springer, M. G. (2018). Student-teacher race congruence: New evidence and insight From Tennessee. *AERA Open, 4*(4), 1-25.

Kao, G. (2004). Social capital and its relevance to minority and immigrant populations. *Sociology of Education, 77*(2), 172-175.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record, 97*(1), 47.

Land, T. J., Bartell, T. G., Drake, C., Foote, M. Q., Roth McDuffie, A., Turner, E. E., & Aguirre, J. M. (2018). Curriculum spaces for connecting to children's multiple mathematical knowledge bases. *Journal of Curriculum Studies*, 1-23.

Lebow, D. G., & Wager, W. W. (1994). Authentic activity as a model for appropriate learning activity: Implications for emerging instructional technologies. *Canadian Journal of Educational Communication, 23*, 231-231.

Lim, H. H. (2006). Representative bureaucracy: Rethinking substantive effects and active representation. *Public Administration Review, 66*(2), 193-204.

Lindsay, C. A., & Hart, C. M. (2017). Exposure to same race teachers and student disciplinary outcomes for black students in North Carolina. *Educational Evaluation and Policy Analysis, 39*(3), 485-510.

Markowitz, A. J., Bassok, D., & Grissom, J. A. (2020). Teacher-child racial/ethnic match and parental engagement with Head Start. *American Educational Research Journal*.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153-184.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

McLoughlin, C., & Oliver, R. (2000). Designing learning environments for cultural inclusivity: A case study of indigenous online learning at tertiary level. *Australasian Journal of Educational Technology, 16*(1).

Meier, K. J. (1993). Latinos and representative bureaucracy testing the Thompson and Henderson hypotheses. *Journal of Public Administration Research and Theory, 3*(4), 393-414.

Meier, K. J., & Nicholson-Crotty, J. (2006). Gender, representative bureaucracy, and law enforcement: The case of sexual assault. *Public Administration Review, 66*(6), 850-860.

Moll, L. C., & González, N. (2004). Engaging life: A funds-of-knowledge approach to multicultural education. In J. A. Banks (Ed.), *Handbook of research on multicultural education* (pp. 699-715). San Francisco, CA: Jossey-Bass.

Morgenroth, T., Ryan, M. K., & Peters, K. (2015). The motivational theory of role modeling: How role models influence role aspirants' goals. *Review of General Psychology, 19*(4), 465-483.

Munter, C., Stein, M. K., & Smith, M. A. (2015). Dialogic and direct instruction: Two distinct models of mathematics instruction and the debate(s) surrounding them. *Teachers College Record, 117*(11), 1-32.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York, NY: Oxford University Press.

Nelson, B. C. 2007. Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. Journal of Science Education and Technology 16 (1):83–97.

Newmann, F. M. (1992). Higher-order thinking and prospects for classroom thoughtfulness. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 62-91). New York, NY: Teachers College Press.

Newmann, F.M., Marks, H.M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education 104*(4): 280–312.

Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11-39). New York, NY: Teachers College Press.

Nicholson-Crotty, J., Grissom, J. A., & Nicholson-Crotty, S. (2011). Bureaucratic representation, distributional equity, and democratic values in the administration of public programs. *Journal of Politics, 73*(2), 582-596.

Oakes, J. (2005). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

OECD. (2016). *PISA 2015 results (Volume II): Policies and practices for successful schools.* Paris, France: OECD Publishing.

Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal, 53(*4), 1159-1197.

Powell, A., Roberts, V., & Patrick, S. (2015). *Using online learning for credit recovery: Getting Back on track to graduation.* Vienna, VA: International Association for K-12 Online Learning (iNACOL).

Queen, B., & Lewis, L. (2011). *Distance education courses for public elementary and secondary school students: 2009-10.* (NCES 2012-008). National Center for Education Statistics.

Rawls, A. W. (1987). The interaction order sui generis: Goffman's contribution to social theory. *Sociological Theory*, *5*(2), 136-149.

Rawls, A. W. (2000). “Race” as an Interaction Order phenomenon: W.E.B. Du Bois’s “Double Consciousness” thesis revisited. *Sociological Theory, 18*(2), 241-274.

Reardon, S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity* (pp. 91-116). New York: Russell Sage.

Reeves, T.C., Herrington, J., & Oliver, R. (2002). Authentic activities and online learning. In A. Goody, J. Herrington, & M. Northcote (Eds.), *Quality conversations: Research and Development in Higher Education*, Volume 25 (pp. 562-567). Hammondville, NSW: HERDSA.

Reeves, T. C., & Reeves, P. M. (1997). Effective dimensions of interactive learning on the World Wide Web. In B. H. Khan (Ed.), *Web-based instruction* (pp. 59-66).

Rickles, J., Heppen, J. B., Allensworth, E., Sorensen, N., & Walters, K. (2018). Online credit recovery and the path to on-time high school graduation. *Educational Researcher, 47*(8), 481-491.

Roblyer, M. D. (2006). Virtually successful: Defeating the dropout problem through online school programs. *Phi Delta Kappa, 88*(1), 31–36.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41.

Saito, H., and K. Miwa. 2007. Construction of a learning environment supporting learners’ reflection: A case of information seeking on the Web. Computers & Education 49 (2):214–29.

Selden, S. C. (1997). Representative bureaucracy: Examining the linkage between symbolic and active representation in the Farmers Home Administration. *American Review of Public Administration, 27*(1), 22-42.

Selwyn, N. (2016). *Is technology good for education?* Cambridge, UK: Polity Press.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*(484), 1334-1344.

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past–A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review, 19*, 58-84.

Skinner, B. T. (2019). Making the connection: Broadband access and online course enrollment at public open admissions institutions. *Research in Higher Education, 60*(7), 960-999.

Solanki, S. M., & Xu, D. (2018). Looking beyond academic performance: The influence of instructor gender on student motivation in STEM fields. *American Educational Research Journal, 55*(4), 801-835.

Solorzano, D. G., & Yosso, T. J. (2001). From racial stereotyping and deficit discourse toward a critical race theory in teacher education. *Multicultural Education, 9*(1), 2.

Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology, 105*(2), 465.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488.

Stevens, D., Frazelle, S., Bisht, B., & Hamilton, R. (2016). *Online credit recovery: Enrollment and passing patterns in Montana Digital Academy courses.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest.

Suh, S. 2006. The effect of using guided questions and collaborative groups for complex problem solving on performance and attitude in a Web-enhanced learning environment. PhD diss., Florida State University, Tallahassee.

Viano, S. (2018). *Online learning as a remedy for course failure: An assessment of credit recovery as an intervention to earn credits and graduate from high school* (Doctoral dissertation). Retrieved from Vanderbilt University Electronic Thesis and Dissertations (URN: etd-05312018-143058).

Vinopal, K. (2017). Understanding individual and organizational level representation: The case of parental involvement in schools. *Journal of Public Administration Research and Theory, 28*(10), 1-15.

Watson, J., Murin, A., Vashaw, L., Gemin, B., & Rapp, C. (2012). *2012 keeping pace with K-12 online and blended learning: An annual review of policy and practice.* Durango, CO: Evergreen Education Group.

Warschauer, M. (2006). *Laptops and literacy: Learning in the wireless classroom.* New York, NY: Teachers College Press.

Warschauer, M., Knobel, M., & Stone, L. (2004). Technology and equity in schooling: Deconstructing the digital divide. *Educational Policy, 18*(4), 562-588.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education, 34*, 179-225.

Wilkins, V. M., & Keiser, L. R. (2004). Linking symbolic and active representation by gender: The case of child support agencies. *Journal of Public Administration Research and Theory, 16*(1), 87-102.

Wright, A., Gottfried, M. A., & Le, V. N. (2017). A kindergarten teacher like me: The role of student-teacher race in social-emotional development. *American Educational Research Journal, 54*(1), 78S-101S.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th Ed). Mason, OH: South-Western CENGAGE Learning.

Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review, 37*, 46-57.

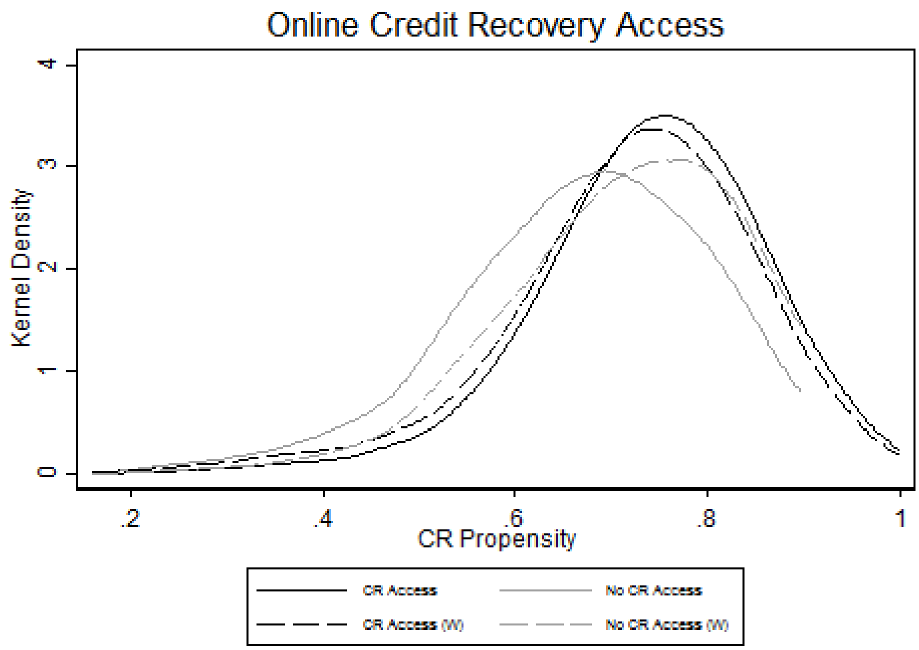
Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *Journal of Higher Education, 85*(5), 633-659.

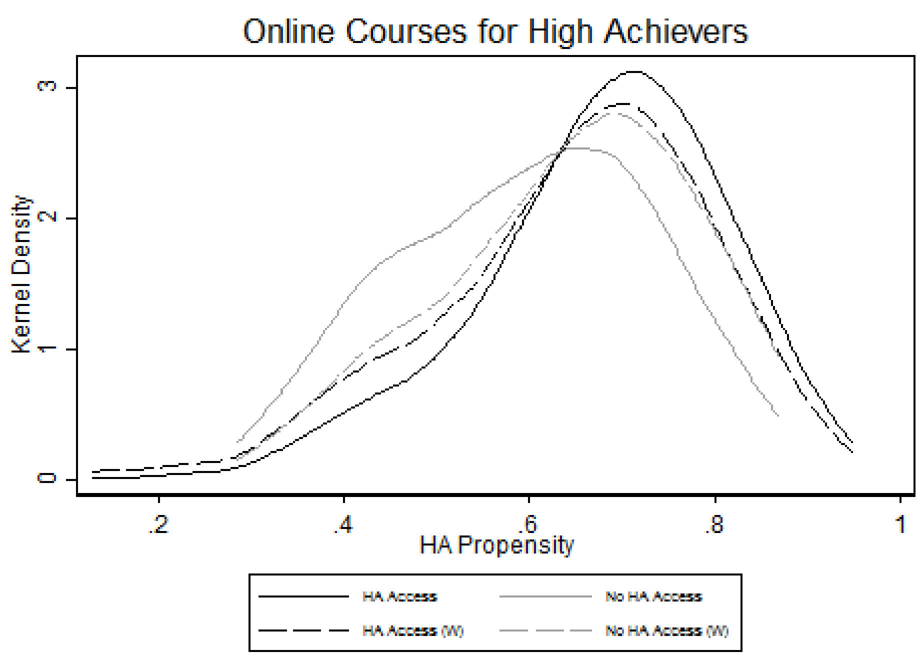
Xu, D., & Li, Q. (2018). Gender achievement gaps among Chinese middle school students and the role of teachers' gender. *Economics of Education Review, 67*, 82-93.

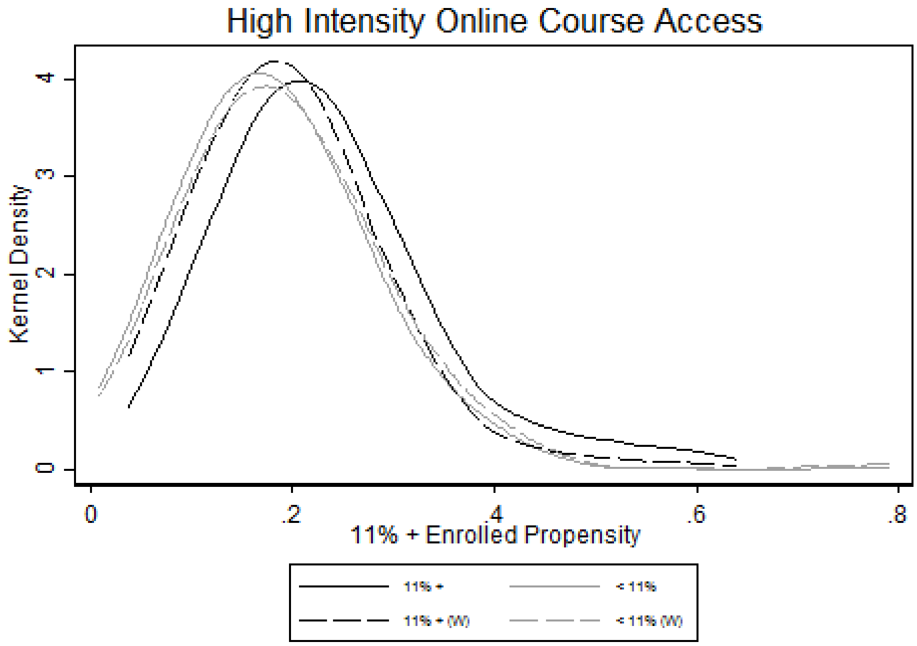
Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education, 73*(4), 247-269.

Appendix

Appendix A: Probability of Various Types of Online Course Access by Weighted versus Unweighted Propensity Score Distributions

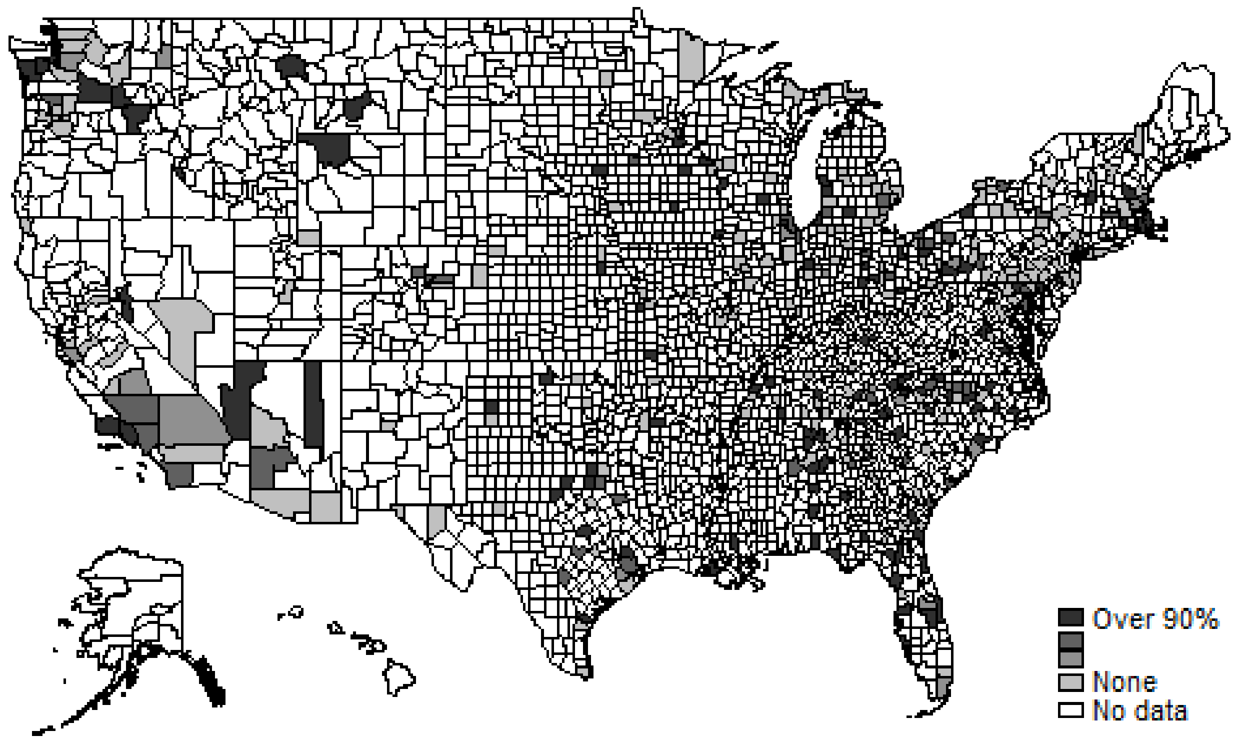






SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-Up, Previously unpublished tabulation (June 2018).

Appendix B. Proportion of Schools with Dropout Prevention Programs by County



SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-Up, Previously unpublished tabulation (June 2018).

Appendix C. Additional Heterogeneous Effects, Interaction between Student Characteristic and Access to Online Course Treatment (IPWRA Estimates)

	Dependent variable	Online access		Credit recovery		Higher achieving		High intensity	
Student Characteristics									
Female	H.S. Credits	-1.295	(1.108)	0.355	(0.432)	0.029	(0.482)	0.356	(0.572)
	H.S. Graduation	-0.040	(0.034)	-0.004	(0.014)	0.003	(0.019)	0.007	(0.013)
	Enrolled in College	-0.028	(0.033)	0.042	(0.026)	0.004	(0.032)	-0.007	(0.035)
	Earned 2-Year Degree	-0.014	(0.018)	0.021	(0.020)	0.004	(0.014)	0.002	(0.027)
	Still Enrolled	0.036	(0.051)	0.038	(0.041)	0.036	(0.045)	0.022	(0.080)
	College Selectivity	0.075	(0.056)	0.028	(0.028)	-0.001	(0.036)	0.018	(0.036)
	STEM Major	-0.013	(0.058)	0.044	(0.028)	-0.049*	(0.027)	0.005	(0.057)
Asian	H.S. Credits	-2.499***	(0.655)	-0.647	(0.580)	0.032	(0.597)	-0.498	(0.756)
	H.S. Graduation	-0.031	(0.028)	0.008	(0.011)	-0.001	(0.012)	-0.003	(0.015)
	Enrolled in College	-0.067	(0.048)	0.001	(0.044)	-0.065**	(0.028)	-0.086	(0.128)
	Earned 2-Year Degree	0.035	(0.027)	-0.019	(0.032)	0.052**	(0.026)	0.009	(0.096)
	Still Enrolled	-0.097	(0.103)	-0.004	(0.130)	0.001	(0.114)	0.008	(0.177)
	College Selectivity	0.064	(0.118)	0.117	(0.077)	-0.053	(0.091)	-0.018	(0.123)
	STEM Major	-0.027	(0.205)	0.070	(0.067)	-0.032	(0.088)	0.021	(0.107)
More than one race	H.S. Credits	0.123	(2.058)	-0.227	(0.559)	0.485	(0.682)	-0.575	(1.401)
	H.S. Graduation	-0.017	(0.021)	0.008	(0.015)	0.001	(0.019)	0.014	(0.019)
	Enrolled in College	0.059	(0.083)	-0.023	(0.047)	-0.028	(0.054)	-0.058	(0.068)
	Earned 2-Year Degree	0.046	(0.032)	-0.001	(0.046)	0.025	(0.025)	0.045	(0.033)
	Still Enrolled	-0.035	(0.081)	0.064	(0.062)	-0.125	(0.083)	0.083	(0.110)
	College Selectivity	0.005	(0.109)	0.086	(0.052)	-0.072	(0.062)	0.076	(0.086)
	STEM Major	-0.107	(0.115)	-0.045	(0.069)	-0.045	(0.036)	-0.026	(0.043)
ELL	H.S. Credits	1.777	(2.108)	0.416	(0.529)	0.761	(0.658)	0.769	(1.249)
	H.S. Graduation	-0.017	(0.029)	0.033*	(0.017)	0.008	(0.015)	0.014	(0.015)
	Enrolled in College	-0.025	(0.057)	-0.007	(0.039)	0.012	(0.036)	0.065*	(0.036)

	Earned 2-Year Degree	0.051**	(0.023)	-0.024	(0.016)	0.025	(0.015)	-0.021	(0.019)
	Still Enrolled	0.058	(0.101)	-0.095	(0.064)	0.045	(0.082)	0.086	(0.086)
	College Selectivity	-0.128	(0.083)	-0.016	(0.038)	-0.006	(0.063)	-0.002	(0.053)
	STEM Major	0.079	(0.053)	0.073	(0.059)	0.003	(0.062)	-0.048	(0.070)
IEP	H.S. Credits	0.163	(1.085)	-0.988	(0.679)	0.364	(2.692)	0.781	(1.395)
	H.S. Graduation	0.037	(0.045)	-0.009	(0.023)	0.060	(0.087)	0.012	(0.028)
	Enrolled in College	-0.020	(0.090)	-0.059	(0.067)	-0.006	(0.071)	0.044	(0.066)
	Earned 2-Year Degree	0.033	(0.048)	0.030	(0.032)	0.038*	(0.022)	0.037	(0.035)
	Still Enrolled	0.065	(0.106)	0.040	(0.061)	0.094*	(0.051)	0.133	(0.101)
	College Selectivity	-0.039	(0.055)	-0.083*	(0.044)	0.004	(0.044)	0.032	(0.062)
	STEM Major	-0.107*	(0.057)	-0.080**	(0.037)	-0.074	(0.050)	0.011	(0.036)
9th Grade GPA	H.S. Credits	-0.728	(0.711)	0.368	(0.405)	0.434	(0.417)	-0.198	(0.716)
	H.S. Graduation	-0.016	(0.024)	0.011	(0.012)	-0.004	(0.015)	-0.009	(0.015)
	Enrolled in College	0.002	(0.037)	0.024	(0.029)	-0.006	(0.026)	-0.017	(0.030)
	Earned 2-Year Degree	-0.007	(0.010)	0.011	(0.010)	-0.003	(0.009)	0.004	(0.011)
	Still Enrolled	0.009	(0.032)	0.000	(0.024)	-0.003	(0.024)	-0.011	(0.035)
	College Selectivity	0.005	(0.034)	0.004	(0.023)	0.034	(0.022)	0.010	(0.030)
	STEM Major	0.043	(0.047)	0.017	(0.015)	0.012	(0.015)	0.029*	(0.017)

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-Up, Previously unpublished tabulation (June 2018). * p<0.10 ** p<0.05 *** p<0.01. Each cell represents the average treatment effect from a different model.

Appendix D: Authentic Online Work Rubric

LESSON INFORMATION
Assign lesson id and instructor id using a 00 format, where the first lesson and instructor are assigned a 01 id and the second lesson and instructor are assigned a 02.

Observer Name:
Course Name:
Lesson Name:
Lesson Id:
Instructor Id:

Which of the following components are included in the lesson?

- ☐ Warm-up
- ☐ Lecture
- ☐ Practice
- ☐ Assessment
- ☐ Writing
- ☐ Interactive (i.e., lab, performance)
- ☐ Other (please describe)

Total number of minutes required to watch any lecture component to the lesson ((round to the nearest minute): ___

Number of minutes spent related to particular instructional expectations (You may allocate the same minute to more than one instructional expectation. The total number of minutes will likely exceed the total lecture length):

___ Skill introduction
___ Drilling/practice
___ Review
___ Assessment
___ Games
___ Enrichment/accelerated instruction
___ Other (please describe)

Which of the following orders of thinking are required to complete instructional tasks? (Refer to Bloom's Taxonomy (Bloom, 1984) for definitions and examples of each component.

- ☐ Listen
- ☐ Recite/remember
- ☐ Demonstrate
- ☐ Think critically
- ☐ Apply
- ☐ Synthesize
- ☐ Evaluate
- ☐ Create

HIGHER-ORDER THINKING

*Rate each item, where **rarely** indicates the item occurred once or twice during the lesson and **often** indicates that the item occurred all but once or twice during the lesson.*

	Never	Rarely	Sometimes	Often
Students spent instructional time generating knowledge (versus direct instruction).	1	2	3	4
Assessment questions, practice problems, and other instructional tasks were delivered in an open-response format (i.e., NOT multiple choice or true/false).	1	2	3	4
Assessment questions, practice problems, and other instructional tasks allow for various correct responses (i.e., open response questions that allow students to apply concepts to a topic of their choosing).	1	2	3	4
There was more than one method for generating an acceptable response.	1	2	3	4
Assignments required students to gather information on their own.	1	2	3	4
Students were asked challenging questions and/or to perform challenging tasks (such as those requiring extensive prior content knowledge, multiple steps, or the application of multiple concepts.)	1	2	3	4
Students were asked to offer reasoning to support responses.	1	2	3	4

REAL-WORLD RELEVANCE

*Rate each item, where **rarely** indicates the item occurred once or twice during the lesson and **often** indicates that the item occurred all but once or twice during the lesson.*

	Never	Rarely	Sometimes	Often
Assessment or instructional tasks were embedding in a specific, meaningful context.	1	2	3	4
Assessment or instructional tasks asked students to synthesize, interpret, explain or evaluate complex information in addressing a concept, problem, or issue.	1	2	3	4
Students were asked to create work product that had value in its own right outside of the school setting.	1	2	3	4
Assessment or instructional tasks asked students to elaborate their understanding, explanations, or conclusions through extended writing.	1	2	3	4

Describe and include personal reflections on the content, skill focus, and instructional tasks included in this lesson. Also describe any implicit (or explicit) values, expectations, norms, or beliefs expressed by the instructor or course content.

Appendix E: Retake Threshold Exploration

There was a discrepancy between the district-specified threshold for passing (60 percent) and the online course system threshold (70 percent). I predicted the district-specified threshold as my main estimates due to concerns regarding gaming behavior identified in classroom observations that was likely to result in the most manipulation at the 70 percent threshold. Specifically, to reduce the fail rate, teachers gave students “checks” before they submitted their assessments (Darling-Aduana et al., 2020; Heinrich et al., 2019). The goal of this policy at the district-level was to encourage student-teacher interactions and initiate the reteaching of content students didn’t understand. In all but a handful of classrooms, this intention morphed into telling students verbally or providing students with a slip of paper to indicate the (predominately multiple choice) numbers of the questions they answered incorrectly. As students sometimes asked for checks more than once before submitting, some could determine the correct answer through process of elimination versus understanding content. Students were also seen googling and otherwise using the Internet to find answers (Darling-Aduana et al., 2019b). Due to the course system incentive to earn above a 70 percent, these types of manipulations are most likely to occur near the 70 percent threshold.

As shown in Appendix Table E1, associations with no retake were generally positive when examined at the 60 percent threshold. However, when examining the 70 percent minimum score threshold students were required to meet to receive credit for completing the lesson within the course system, I identified negative associations with both measures of authentic work. This contradicted the higher average scores observed among students completing lessons with higher rates of real-world relevance. To better understand this contradiction and differences by threshold, and based on gaming behavior seen in classroom observations, I also estimated models predicting whether students were differentially likely to achieve an arbitrarily set 80 percent cutoff thresholds. There was no reason for students to artificially exceed this threshold, or the 60 percent threshold, because there was no external accountability measure associated with achieving over a 59 versus 60 or a 79 versus 80 percent score. Thus, I had more confidence that increases at the 60 or 80 percent cutoff would represent real increases in learning versus increased gaming behaviors, such as googling correct answers.

Appendix Table E1. Retake Threshold Exploration (N=134,272)

		District Cutoff - 60%	
Higher-Order Thinking	0.030***		0.006***
	(0.002)		(0.002)
Real-World Relevance		0.029***	0.015***
		(0.002)	(0.002)
Higher-Order Thinking			0.026***
# Real-World Relevance			(0.001)
		System Cutoff - 70%	
Higher-Order Thinking	-0.043***		-0.033***
	(0.002)		(0.002)
Real-World Relevance		-0.046***	-0.030***
		(0.002)	(0.002)
Higher-Order Thinking			0.004***
* Real-World Relevance			(0.001)
		Arbitrary Cutoff - 80%	
Higher-Order Thinking	0.050***		0.021***
	(0.002)		(0.002)
Real-World Relevance		0.041***	0.016***
		(0.002)	(0.002)
Higher-Order Thinking			0.033***
* Real-World Relevance			(0.001)

Each column within each dependent variable section represents a different model.

When examining the alternative, arbitrarily-chosen cutoffs, I observed that students were more likely to pass the assessment on the first attempt when lessons integrated higher rates of authentic work, meaning they would have avoided the need to retake the lesson if the cutoff was set at that level. The similarity of estimates at the 60 and 80 percent threshold, as well as consistency with estimates predicting end-of-lesson score and classroom observations, lead me to be more confident that changes observed at the arbitrarily identified cutoffs reflect student learning than the actual 70 percent cutoff that students had an incentive to manipulate. This evidence of potential gaming behavior may also contribute to attenuation bias by introducing noise not associated with student learning in end-of-lesson assessment scores.

Appendix F: Dimensions of Digital and Blended Instruction Rated in Observations

The observation protocol directed observers to rate the following dimensions of digital and blended instruction.

- ***Physical environment***: How and where students access the instructional setting, including the technological setting and any associated limitations, and who else in the same physical environment as the student could assist with technological problems and support learning;
- ***Technology and digital tools:*** How students access instruction, including internet connectivity, hardware, and software in use, and the safety, operability, and accessibility of the technology;
- ***Curricular content and structure***: Content and skill focus, who developed it and where it is located (e.g., software loaded onto a tablet, paper workbook), stated learning objectives, sequence and structure, level of rigor or intellectual challenge, and ability to meet and adapt curricular content to student needs;
- ***Instructional model and tasks***: Role of instructor and software in instruction (what drives instruction); purpose or target of instruction; student/instructor ratio and grouping patterns, multimodal instruction; order of thinking required and application of technology in instructional tasks, and ability to meet/adapt instructional model and tasks to student needs;
- ***Interaction:*** How much interaction with a live person, and does the technology affect the ability of the instructor or student to positively interact with one another and the instructional resources?
- ***Digital citizenship:*** Are students using the technology as intended by the instructor and/or instructional program?
- ***Student engagement:*** Overall student engagement levels, level of student self-regulation and persistence, and level of community within the instructional setting;
- ***Instructor engagement:*** Overall instructor engagement levels (symbolic or active) and instructor efforts to encourage engagement;
- ***Assessment/feedback***: Who develops and manages the assessment (instructor, provider via software), structure, and whether it is individualized to student learning and relevant to stated learning goals.

Appendix G: Student Performance and Behavior on Online Lessons When Taught by a Same race/Ethnicity and Gender Instructor (With Teacher Race as a Control)

	Score (Std.)	Never Retake	Active Time (Log)
Black instructor	0.014	0.003	-0.010
	(0.020)	(0.004)	(0.019)
Black student	0.010	0.002	-0.030
	(0.033)	(0.006)	(0.059)
Black instructor # Black student	-0.026	-0.008*	-0.003
	(0.024)	(0.005)	(0.025)
Hispanic instructor	-0.007	0.003	-0.041**
	(0.016)	(0.003)	(0.017)
Hispanic student	-0.014	0.004	-0.003
	(0.047)	(0.006)	(0.062)
Hispanic instructor # Hispanic student	0.022	-0.010	0.052
	(0.038)	(0.006)	(0.043)
White instructor	0.002	0.001	0.023***
	(0.008)	(0.002)	(0.008)
White student	0.033	0.000	0.060
	(0.058)	(0.009)	(0.095)
White instructor # White student	-0.036	-0.009	-0.053
	(0.035)	(0.006)	(0.041)
	Score (Std.)	Never Retake	Active Time (Log)
Male instructor	-0.001	0.003	0.001
	(0.009)	(0.002)	(0.010)
Male student	-0.046**	-0.012***	-0.053
	(0.023)	(0.004)	(0.033)
Male instructor # Male student	0.012	0.001	-0.005
	(0.014)	(0.003)	(0.014)
Female instructor	-0.011	-0.004*	0.004
	(0.008)	(0.002)	(0.009)
Female student	0.034	0.011***	0.058*
	(0.024)	(0.004)	(0.034)
Female instructor # Female student	0.012	0.001	-0.005
	(0.014)	(0.003)	(0.014)

Each column within each section represents estimates from a different model.