5-2022

Beyond a divide: reconceptualizing digital capital and links to academic proficiency.

Ryan Price
University of Louisville

BEYOND A DIVIDE: RECONCEPTUALIZING DIGITAL CAPITAL AND LINKS TO ACADEMIC PROFICIENCY

By:
Ryan Price
B.A., University of Louisville, 2019

A Thesis
submitted to the faculty of the
College of Arts and Science of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Arts
in Sociology

Department of Sociology
University of Louisville
Louisville, Kentucky

May 2022

BEYOND A DIVIDE: RECONCEPTUALIZING DIGITAL CAPITAL AND LINKS TO ACADEMIC PROFICIENCY

By

Ryan Price
B.A., University of Louisville, 2019

A Thesis Approved on
April 7, 2022

By the following Thesis Committee

____

Dr. Melanie Gast, PhD

____

Dr. David Roelfs, PhD

____

Dr. Robert Carini, PhD

ACKNOWLEDGMENTS

I would like to thank my thesis chair, Dr. Melanie Gast, for her incredible support and mentorship throughout this thesis project. I would also like to thank Dr. David Roelfs for being a great teacher and mentor over the past two years. To my third committee member, Dr. Robert Carini: thank you for challenging me to think critically about my research interests, and for sharing my excitement about the project as it was developing. These three committee members made such a difference in my graduate school experience, and I cannot express enough gratitude for them.

I am also incredibly grateful for the relentless love and support from my family. Mom, Dad, Tyler, Jenna, Josh, Lucy and Rigby: I love you all so much! Thanks for all that you do. Finally, I would like to extend a most sincere thank-you to a few others in my support network. Gretchen, Kyle, Courtney, Skylar, Drake, and Olivia: I couldn’t have done this without you all.

ABSTRACT

BEYOND A DIVIDE: RECONCEPTUALIZING DIGITAL CAPITAL AND LINKS TO ACADEMIC PROFICIENCY

Ryan Price

April 7th, 2022

In the digital inequality literature, the popular notion of a "digital divide" is frequently used to discuss digital inequality; however, this framework is overly simplistic and cannot adequately capture the complex nature of digital inequality. Some scholars have adopted a framework of digital capital to attempt a multidimensional approach to this social problem, but the literature lacks consistent, empirical measurements. Using U.S. data from the Programme for International Student Assessment (PISA) 2018 Survey, I seek to propose, and validate, internally consistent principal components of digital capital among 15-year-old high school students in the US. I conduct a Principal Component Analysis (PCA), resulting in five principal components: Academic Digital Usage, Perceived Digital Autonomy, Perceived Digital Competence, Casual Digital Browsing, and Knowledge-Based Digital Leisure. I then examine Chronbach's alpha coefficient for each component to assess reliability. Next, I conduct an Ordinary Least Squares (OLS) regression analysis to assess how the resulting factors might predict mathematics proficiency scores, as measured by PISA, after controlling for several key background variables. In the regression model, I also include three variables related to material digital access, which were not found to be a reliable component of digital capital. The regression results show statistically significant effects on mathematics proficiency scores for each proposed component of digital capital, except for perceived digital competence.

Additionally, the results indicate that home computer access has a significant, positive effect on mathematics proficiency scores. This exploratory study offers a new direction toward empirical measurement for future research on digital capital and inequality.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

INTRODUCTION

In the United States (US), technological adoption has grown substantially over the last two decades, with 93% of American adults using the internet as of 2021 (Perrin & Atske, 2021). For many people in the US, information and communication technologies (ICTs) might seem like a ubiquitous part of life in an increasingly digitalized society. These recent trends of technological adoption might ostensibly suggest dwindling levels of digital inequality, but as the widespread transition to virtual learning following the onset of the COVID-19 pandemic has made plainly visible, technology-related barriers persist as considerable obstacles for many students and families (Graves et al., 2021; Kelley & Sisneros, 2020). Although the notion of a "digital divide" has received a renewed, heightened sense of public awareness in the era of COVID-19, I argue that this concept is overly simplistic, invoking a clear gap between the "information haves" and the "information have-nots" that does not accurately depict the multidimensional reality of digital inequality.

More recently, some scholars have proposed various conceptualizations of *digital capital* as a multidimensional framework for studying digital inequality (Choi et al., 2021; Ollier-Malaterre et al., 2019; Park, 2017; Ragnedda, 2018; Ragnedda et al., 2020; Santillana et al., 2020). The theoretical construct of digital capital appears quite promising for researchers, but as critics of the framework point out, these conceptualizations are inconsistent and lack clear, empirical measures (Puckett, 2020). Moreover, digital capital has remained relatively unexplored within the context of the US education system. In the literature, the attempts to empirically operationalize the construct of digital capital using data from a nationally representative sample of high school students in the US are quite limited.

The aim for the current research project is twofold: first, to propose a parsimonious set of empirical measures of digital capital among 15-year-old students in the US, and second, to assess how these measures might predict learning outcomes. To achieve these research goals, I use data from the 2018 Programme for International Student Assessment (PISA) survey to conduct a Principal Component Analysis (PCA) on a set of 28 potential indicator variables, reflecting elements of digital capital discussed in the existing literature. After assessing the internal reliability of the extracted components, I retain 5 principal components – Academic Digital Usage, Perceived Digital Autonomy, Perceived Digital Competence, Casual Digital Browsing, and Knowledge-Based Digital Leisure – to include in an Ordinary Least Squares (OLS) regression on student mathematics proficiency, estimated from 10 plausible values of math proficiency. I use this regression model to evaluate the predictive validity of components extracted in the PCA. This research adds to the existing literature by consolidating several proposed measures of digital capital into one model, applying an empirical lens to the construct.

In the next section, I begin with a review of the existing literature on the insufficient "digital divide" paradigm, digital inequalities in the US, the theoretical framework of digital capital, and the intersection of digital and academic inequalities. Following, I provide an in-depth, methodological discussion of the data, sample, variables used for analyses, and analytic procedures. Results are then reported, including relevant tables and figures to aid interpretation of findings. I conclude the project with a discussion of results in relation to prior research, limitations, and recommendations for future research.

LITERATURE REVIEW

The "Digital Divide" Framework

Amidst the ongoing COVID-19 pandemic, the widespread transition to remote learning in the education system has created an acute, heightened awareness of digital inequality. Once considered a luxury, internet connectivity is more recently considered a necessity to participate in an increasingly digitalized society (Lai & Widmar, 2021). As the COVID-19 pandemic has revealed, many families in the US lack the ICT resources necessary to shift their education to a virtual environment, and as a potential result, might be negatively impacted in terms of educational outcomes (Graves et al., 2021). That being said, the "digital divide" is not an entirely new topic in political, academic, or popular discourse. Politicians and scholars alike have wrestled with the problem of bridging the digital divide since the 1990's – that is, closing the technology gap between "information haves" and "information have-nots" (Attewell, 2001; McConnaughey et al., 1998; Parker, 2000; Wresch, 1996). This mainstream conceptualization of a dichotomous "digital divide" – where one has access to information and communication technologies (ICTs), or they do not – is prominent in the literature; however, this framework is limited and overly simplistic. To better understand the more nuanced contours of digital inequality, some scholars have argued for a revisualization of digital inequality as a spectrum, rather than a divide (Lenhart & Horrigan, 2003). The purpose of this section, then, is to reconceptualize digital inequality as a complex, multi-faceted social problem.

The term “Digital Divide” was first officially used in a publication by the US Department of Commerce’s National Telecommunications and Information Administration (NTIA, 1999) to describe a gap in access to ICTs. During the early 2000’s, digital divide scholars focused primarily on unequal material access to computer technologies and internet connection (Jurich, 2000; Parker, 2000). These scholars examined the access gap largely in terms of demographic characteristics, including race, socioeconomic status, gender, education, and geographical location (Cullen, 2001; Mossberger et al., 2006). However, material access has become considerably complicated given the rise of broadband and mobile internet connections, smartphone technologies, and other technologies such as tablets and iPads. Consequently, this binary framework – of either having access to ICTs or not – is less useful today since material access exists in a stratified spectrum through differential levels of device quality (i.e., smartphone vs. desktop computer), internet speed, location of digital access, and access maintenance (Scheerder et al., 2017). Moreover, as the internet and computer access gap in the US has narrowed over the last two decades (Cohron, 2015), many scholars have argued to move beyond this “first level divide” of material access to examine a “second level divide” related to unequal patterns of digital usage, knowledge, and skills (DiMaggio et al., 2004; Hargittai & Hinnant, 2008).

While this concept of a “second level divide” is useful for addressing how digital practices might reproduce social inequalities, the term still implies a rather simplistic, binary view of digital inequality. Just as material digital access is complex and multi-faceted, so too is digital usage; ICTs can be used in many differentially advantageous ways and should be considered within a digital spectrum (Lenhart & Horrigan, 2003).

Moreover, many scholars argue that disparities in digital access still warrant academic and political attention (Campos-Castillo, 2015). As the COVID-19 pandemic has made painfully visible, many students lack adequate computer technologies and broadband home internet necessary to support virtual, synchronous learning (Kelley & Sisneros, 2020). While differential patterns of digital usage are an important consideration of digital inequality, material access remains a significant barrier in many institutional settings, including the education system. Thus, I seek to move away from the "digital divide" framework in this project to examine digital inequality as a stratified, hierarchical phenomenon that exists at multiple levels of digital access and usage.

Unequal Digital Access

In the digital inequality literature, the concept of digital access lacks a clear, operational definition. For some scholars, digital access represents a binary classification of who has access to ICTs and who does not (Attewell, 2001). With this broad definition of digital access, one could argue that digital inequality has been greatly diminished as most people can access online spaces, in some capacity, from some location. However, as other scholars argue, this conceptualization obscures some of the more subtle disparities in the context of digital access; issues of time, cost, quality of technological resources, and environment of digital use are also important considerations in defining effective or meaningful digital access (Selwyn, 2004). Instead of a binary conceptualization of access, one must understand digital access in a sense of gradation, considering the differential levels of material access that one might possess, rather than simply have access to. In this section, I seek to examine some of these nuanced contours of digital access as they relate to digital inequality.

Over the last two decades, the US has experienced rapidly increasing rates of technological adoption among all groups (PEW, 2021). Compared to early studies of digital inequality, more recent research is necessarily more complex as mobile technology has become widely adopted in society, creating a conceptual “gray area” of digital access. On the surface, it appears as if digital inequality is dwindling; there are no significant racial or ethnic differences in smartphone ownership, and the majority (roughly 76%) of those who might be considered economically disadvantaged own a smartphone (PEW, 2021). Despite these trends, digital inequality remains a persistent problem particularly for low-income households. In 2021, 92% of households making more than $75,000/year have access to broadband internet in their homes compared to 57% of households making less than $30,000/year; additionally, only 6% of households making more than $75,000/year rely on smartphones for internet connection compared to 27% of households making less than $30,000/year (PEW, 2021). Comparing trends of broadband home internet access versus reliance on mobile internet connection demonstrates a hierarchical continuum of access *quality.* In the context of schooling during the COVID-19 pandemic, adequate computer technology and broadband home internet are necessary to support virtual, synchronous learning. Mobile internet connection via smartphone might facilitate access to a variety of digital spaces, yet this level of connection might be inadequate to facilitate participation in virtual learning environments. At the onset of the COVID-19 pandemic, millions of students in the US lacked access to quality ICT resources, such as broadband home internet and adequate computer devices, that allow for virtual participation in the classroom (Kelley & Sisneros, 2020).

Relatedly, scholars must also consider the location of one's digital access. Having a broadband internet connection at home provides several advantages over relying on public internet connection. In public libraries, one might have limited time to use various digital resources (Watkins & Cho, 2018). Libraries have specific hours of operation and might have certain organizational rules governing how long an individual is allowed to use a computer, or what software and programs one might have access to. Additionally, privacy concerns might be important to consider for some public ICT users. In another context, some researchers have studied how Black and Latino people encounter racialized policing when attempting to use digital resources, such as Wi-Fi, at restaurants and coffee shops in primarily white neighborhoods (Yang et al., 2021). It could be possible that this racialized, digital gatekeeping acts as a barrier in other institutional settings as well. Digital spaces reflect our society, so issues of racialized policing might impact how Black and Latino communities can safely and comfortably access public digital resources. Moreover, public internet access might not be safe or feasible during the ongoing COVID-19 pandemic.

As the above research suggests, digital access is an ongoing process rather than a singular achievement (Gonzales et al., 2020; Ollier-Malaterre et al., 2019). Although one might have access to various technologies, they must also have the resources to maintain that access. According to these scholars, technology maintenance might be a considerable barrier to digital access over time. This concern is one not adequately addressed by the "digital divide" framework, highlighting a complex, longitudinal nature of digital inequality.

Effective digital access, then, includes, but is certainly not limited to, material access. Some scholars argue that material access to technology is meaningless without the technical skills, knowledge, and support to be able to meaningfully use it (Mossberger et al., 2008; Selwyn, 2004). Instead of conceptualizing these technical skills, knowledge, and support as a separate "second level divide", it might be more useful to consider these factors as mediators of effective digital access, acting as potential access barriers.

In sum, research on digital inequality should consider digital access not as a dichotomous "divide", but rather, as a complex, multi-faceted spectrum. This spectrum of digital access is stratified through differential levels of device quality (i.e., smartphone vs. desktop computer), internet speed, location of digital access, and access maintenance. Considering the context of one's digital access facilitates a deeper understanding of digital inequality. Thus, research should address digital disparities in a way that 1) identifies material access to ICTs as a complex spectrum that requires maintenance over time, and 2) acknowledges how this spectrum of access yields differentially advantageous and institutionally rewarded patterns of digital use.

Differential Digital Usage

In the digital inequality literature, some scholars have argued that research should move beyond the "first level divide" (of unequal digital access) to examine a "second level divide" related to differential patterns of digital usage and technological skill (Attewell, 2001; Hargittai & Hinnant, 2008). Certainly, differences in *how* ICT resources are utilized reflect a critical dimension of digital inequality; however, the construct of a "second level divide" might be overly simplistic and limited in its utility for scholars to thoroughly examine digital inequality. In this section, I review the literature to explain how various patterns of digital usage are differentially advantageous and how digital usage is mediated by a complex, stratified spectrum of digital access.

In a systematic review of the English-language digital inequality literature (2011-2016), Scheerder et al. (2017) identified “unequal digital usage” as the most frequent focus of academic research. Most scholars agree that all digital uses aren’t equally advantageous, but how exactly does digital usage create unequal outcomes, and what are these inequalities? According to DiMaggio et al. (2004), three major dimensions of digital inequality related to digital usage involve variations in 1) the *purposes* for which people use ICTs, 2) the degree to which people exercise *autonomy* in their ICT use, and 3) the *skills* that people bring when using ICTs.

In our increasingly digitalized society, the internet (among other ICTs) can be used to engage in a broad array of activities. Thus, it is imperative for research on the topic to distinguish between different types of ICT use and examine their consequences separately (DiMaggio et al., 2004). Some scholars have focused specifically on “capital-enhancing” uses of the web to explore differences in how people utilize ICTs to enhance life chances (Hargittai & Hinnant, 2008). Capital-enhancing uses of ICT might include those that yield increased economic opportunities, political participation, or health information. According to these scholars, capital-enhancing digital practices provide more opportunities for upward mobility than digital leisure activities. Higher socioeconomic status (SES) is associated with capital-enhancing web uses with income and educational attainment being significant predictors of occupational mobility (Zillien & Hargittai, 2009). While capital-enhancing ICT uses might provide more opportunities for upward social mobility, leisure-based digital activities are not without value.

According to the ethnographic work of Rafalow (2020), digitally mediated play is a foundational source of digital skills for today's "digital youth". Moreover, these digital skillsets might potentially serve as an important resource for traditionally underserved communities, including Black, Latino, and lower-SES youth.

In addition to differentiating between type, or purpose, of digital use, research must consider the degree of autonomy an individual can exercise related to their ICT engagement. Digital autonomy reflects the amount of control an individual has over their ICT usage (DiMaggio et al., 2004). Individuals with low digital autonomy might find their digital use controlled by any number of outside factors, whereas high digital autonomy suggests relatively unrestricted digital use. Location of internet access represents an important consideration of digital autonomy; home internet access allows for more flexibility and freedom in digital usage compared to access from a community center or library (Hargittai & Hinnant, 2008). Additionally, even with household internet access, one might also consider how digital usage could potentially be restricted by other family members' needs to use the available ICT resources. Other scholars have provided slightly different conceptualizations of digital autonomy that focus more on independent use. For example, González-Betancor et al. (2021) include the ability to independently troubleshoot problems with ICTs and independently select software to achieve certain tasks in their conceptualization of digital autonomy. Although these definitions are slightly different, the literature seems to be in consensus that higher digital autonomy is associated with better quality digital usage.

Finally, technological skill represents a critical dimension of differential digital usage. Technological skill can be conceptualized as the knowledge and ability to respond “pragmatically and intuitively to challenges and opportunities in a manner that exploits the Internet’s potential and avoids frustration” (DiMaggio et al., 2004). These technological skills, as others refer to as “digital literacy” (Santillana et al., 2020), are a foundational component for meaningful digital use (Selwyn, 2004). This digital “know-how” might represent a way for highly skilled individuals to maximize advantage from various digital uses. Other scholars describe “digital coping skills” as the cognitive and technical abilities to use ICTs in a way that is useful, rather than distracting (Ollier-Malaterre et al., 2019). Technological skills, then, are essential in efficiently navigating the web to find information, utilizing ICT resources to increase life chances, and avoiding frustration and distraction. However, the effect of *perceived* technological skills remains relatively unexplored. Hargittai and Hinnant (2008) found that self-reported online skill (or perceived technological “know-how”) was significantly associated with visiting “capital-enhancing” websites, but there appears to be few studies to support this finding. Other scholars have reported that perceived ICT competence is affected by the quality of ICT access one possesses, SES, and gender (Zhong, 2011).

These three dimensions of differential digital usage (purpose of use, ICT-related autonomy, and technological skill) might be impossible to discuss without addressing concerns of digital access. High-speed home internet access, in the context of the digital access spectrum, facilitates increased participation in political, educational, social, and recreational activities (Morris & Morris, 2013; Watkins & Cho, 2018). Related to the concept of digital autonomy, home internet access necessarily provides a more unrestricted connection, and a larger number of ICT resources (such as desktop computers or laptops) might result in less restricted digital usage by other family members. Moreover, increased digital access could certainly result in extended digital usage to cultivate advantageous technological skills and familiarity (DiMaggio et al., 2004). Because the concepts of digital access and digital usage are inherently intertwined, I argue that research cannot “move beyond” discussions of access to separately examine differential digital usage. Rather, both digital access and usage in unison to gain a deeper understanding of digital inequality.

Theoretical Framework of Digital Capital

In the sociology of education literature, many scholars adopt a Bourdieusian framework to examine the social reproduction of inequality in the education system (Lareau, 2011; Lareau & Weininger, 2003; Paino & Renzulli, 2013; Rafalow, 2020). Bourdieu (1986) proposes a theory of different forms of capital – economic, social, and cultural – that can be converted to other forms of capital, creating an advantage in various institutional settings. The inheritance of capital over many generations reifies and maintains the ideologies and dominant standards of a ruling class. According to Bourdieu, cultural capital, which includes cultural goods as well as the familiarity with the legitimated, dominant culture within a society, is particularly relevant in examining how social classes and privileges are maintained as power is transferred generationally. Some contemporary scholars have expanded on this framework to conceptualize *digital capital* in an increasingly technological society (Ollier-Malaterre et al., 2019; Paino & Renzulli, 2013; Ragnedda, 2018; Santillana et al., 2020). The aim of this section is twofold: 1) to review and critically analyze theoretical conceptualizations of “digital capital” present in the literature, and 2) to situate the larger research question of the current study within this literature.

Some scholars have conceptualized digital (or techno-) capital as the digital skills, knowledge, and capabilities to effectively utilize technologies (Santillana et al., 2020). In their study, Santillana et al. (2020) conduct an exploratory factor analysis to identify 3 factors or "levels" of techno-capital (or digital literacy). While the composite reliability of their factors were quite high, other scholars contend that digital capital includes external, material ICT resources (as *objectified* forms of digital capital) in addition to digital literacy (Ragnedda, 2018). Additionally, with a small sample from only one US City (Austin, TX), the study might not be generalizable to the larger US. While Santillana et al. (2020) provide a detailed factor analysis of digital literacy in Austin, TX, the larger conceptualization of digital capital could be theoretically improved. Comparatively, other scholars provide more "theoretically robust" (i.e., well supported by social theory) conceptualizations of digital capital, but lack reliable and valid empirical measurements to bolster such a theoretical model (Ollier-Malaterre et al., 2019; Ragnedda, 2018).

Alternatively, other scholars have utilized qualitative methods to explore the "digital dimension of cultural capital". Rafalow (2020) provides an in-depth, ethnographic study of the reproduction of digital inequality in the education system. This project was one of the first studies to examine digital capital with rich, qualitative data. According to Rafalow (2020), children acquire similar levels of digital skills through play – digital expressions through social media, video games, and/or digital leisure activities. However, despite comparable levels of digital skill, various racial and class-based stereotypes affect how teachers, as institutional agents, value these skills as cultural capital [See also Paino and Renzulli (2013)]. Digital play, an important source for the acquisition of digital capital, is disciplined when pedagogical approaches in predominantly Latino and Asian schools restrict these opportunities, as teachers view play as either irrelevant or threatening to education. In stark contrast, wealthy white students are often encouraged to engage in digital play, as pedagogical approaches often view play as essential to learning. Despite the comparable levels of digital capital between racial & SES groups, the institutional valuing of wealthy, white digital capital legitimates unequal advantages in the education system.

While studies such as Rafalow's (2020) have made qualitatively powerful contributions to the literature, some scholars critique the framework of digital capital for lacking consistent, clear, and empirical operationalizations (Puckett, 2020). The goal of this study, then, is to begin empirically measuring the theoretically robust construct of digital capital. Drawing on prior research, I seek to conceptualize digital capital as the institutionally valued set of technological skills to autonomously use ICTs in an advantageous way. Additionally, I intend to include material digital access in this conceptualization, particularly because hardware, software, and high-speed (broadband) home internet connection are important *objectified* forms of digital capital that must be considered. Also, a conceptualization of digital capital might include measures of "digital play," or other forms of digital usage that might contribute to the acquisition of digital skills. Rather than approaching digital inequality as a multi-level "digital divide", the conceptualization of digital capital that I propose will seek to understand digital inequality at intersecting locations of unequal digital access and differential digital usage.

Digital Capital and Educational Outcomes

In the digital inequality literature, many scholars have examined how stratified access to and use of ICT resources might exacerbate various social inequalities (Dimaggio et al., 2004; Morris & Morris, 2013; Gonzales et al. 2020). More specifically, in the US education system, digital inequality seems to be a persistent factor in reproducing educational inequalities, potentially contributing to an "achievement gap" (Fairlie, 2012; Lai & Widmar, 2021). Although popular and political discourse has largely returned to the outdated "digital divide" framework during the COVID-19 pandemic to discuss digital inequality in the education system, the theoretical lens of digital capital might provide a better framework to analyze the reproduction of educational inequality. While stratified access to material ICT resources remains a prominent concern, some scholars have acknowledged that examining digital access alone cannot adequately address the complex nature of digital inequality in the reproduction of educational inequalities (Fairlie & Robinson, 2013; Puckett, 2020). Here, I seek to review how unequal digital access, in addition to differential patterns of digital use, specifically contribute to disparities in the education system.

Scholars have examined the effects of unequal material access on various educational outcomes, including high school graduation (Fairlie et al., 2010), GPA (Fairlie, 2012), and standardized test scores (Fairlie & Robinson, 2013). Some studies have found a significant, positive correlation between home computer ownership and high school graduation after controlling for individual and family characteristics (Beltran et al., 2006; Fairlie et al., 2010). Additionally, in higher education, home computers were found to be significantly associated with higher course completion rates and higher course grades for college students of color (Fairlie, 2012). Similar effects might exist for high school students. A 2018 PEW research study found that 17% of teens in the US *often* or *sometimes* can't complete their homework because of a lack of access to a reliable computer or internet connection at home (Anderson & Perrin, 2018). Despite the rapidly increasing rates of technological adoption among all groups in the US, access to ICTs remains a significant problem, particularly for low-income households (PEW, 2021). According to these scholars, digital access has serious implications in the education system, which might be magnified in the context of the ongoing COVID-19 pandemic.

Moreover, other scholars contend that quality of material access is an important predictor of "interest-driven learning" opportunities (Katz et al., 2017). According to Katz and colleagues (2017), Interest-driven learning is "prompted by a child's own interests and curiosities (as opposed to by adults' directives)". This is an important consideration of educational outcomes because interest-driven learning is associated with learning motivation and self-confidence, which might provide a more subtle, indirect link through which inequality related to digital access contributes to educational inequalities.

Although the literature suggests that stratified access to ICTs produces differential educational outcomes, the proposed theoretical conceptualization of digital capital also includes the set of technological skills to autonomously use ICTs in an advantageous way. As in other institutional settings, certain digital skills and uses are differentially valued and rewarded in the education system. Some digital activities, such as using a learning software or researching college or future career opportunities, might be more advantageous than other, more leisure-based, digital activities (such as idly browsing the web). Also, some scholars have found that students from high-resource schools are more likely to use technology for experimental and creative uses than students from low-resource schools (Valadez & Duran, 2007). Middle-class parents with greater time and ICT skills might be better able to actively monitor their child's virtual learning as compared to working-class parents. As a result, access to and employment of digital capital (in multiple forms) is stratified, which could reproduce social inequalities within the education system.

While some scholars argue that future research should move beyond access to study differential patterns of digital use, material access to ICTs remains persistently stratified in the US today. Certainly, research should address how differential patterns in digital use affect various educational outcomes, but barriers to digital access still require academic attention. Using the theoretical perspective of digital capital allows for research to focus on unequal digital access and differential digital usage simultaneously.

DATA AND METHODS

Data

For this study, I use secondary data from the Programme for International Student Assessment (PISA) 2018 survey coordinated by the Organization for Economic Cooperation and Development (OECD, 2019a). Conducted every three years, PISA surveys are administered internationally to a target population of 15-year-old students enrolled in an educational institution at grade 7 or higher. Administered in 79 different nations, PISA 2018 is designed to assess “15-year-olds’ ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges.” (OECD, 2019c). Although the OECD has publicly available data from student, parent, and teacher respondents on their website, my analysis only includes data from the PISA 2018 student-questionnaire dataset. This dataset includes variables that measure student demographic characteristics, student attitudes and behaviors, and ICT availability and usage patterns. Additionally, the dataset provides 10 plausible values (PVs) for each of the subject areas assessed by PISA: mathematics, reading, and science.

After selecting this dataset for analysis, I use a filtering variable (Country ID) to exclude data from outside of the US in the analytic sample. Since my research questions were developed primarily in the context of US-based digital inequality literature, I restrict the analytic sample to reflect contemporary digital trends specific to the US. Because of this methodological decision to only analyze national-level data, the findings from this study are only generalizable to 15-year-old students in the US.

In the US, data collection for PISA 2018 occurred during October and November of 2018 (OECD, 2015b). The PISA sampling report outlines sampling procedures in detail, including information on sample design, response rates, and special school sampling situations (Westat, 2016). Briefly, PISA 2018 was conducted using a two-stage stratified sampling design. In the first stage, PISA-eligible schools were systematically sampled from a comprehensive, national list of schools that serve 15-year-old students. The individual schools within this sampling frame were assigned a probability relative to the number of PISA eligible students enrolled in the school, following a "systematic probability proportional to size" (PPS) sampling technique (Westat, 2016). The second stage of the sampling process randomly selected 15-year-old students within the sampled school units. Robust sampling requirements were enforced to ensure a nationally representative sample of 15-year-old students, and as a result, the estimated exclusion rates for both schools and students were under 5% (OECD, 2015a)

Since the PISA assessment test design randomly assigned students to fill out booklets instead of completing all sets of data, missing values in the dataset were assumed to be "Missing at Random" (Hu & Yu, 2021). Missing data were addressed using listwise deletion to remove cases that have one or more missing values for any of the variables included in the analysis. Some scholars have argued that listwise deletion is a good alternative to Multiple Imputation, particularly because this technique produces relatively unbiased standard errors (Allison, 2001).

After removing cases with missing values, I obtained an analytic sample (n = 3,308) from the full US sample (N = 4,838). However, this amount of lost data cannot be ignored. Upon inspection of the student-questionnaire item compendium (OECD, 2019b), I found that a large portion of the missing data was from the ICT portion of the survey. Item nonresponse was a clear contributor to missing data in this section of the survey. Nonresponse bias can be detrimental for analysis and generalizability; however, in the technical report, PISA describes the implementation of weighting adjustments to offset the impact of nonresponse (OECD, 2019d). Consistent with the PISA data analysis manual (OECD, 2009), I made sure to include the final, trimmed student-weight where appropriate. While the inclusion of weighting variables might help limit nonresponse bias, the amount and nature of missing values are limitations of the study.

Variables

For the Principal Component Analysis (PCA), I use the existing literature to guide variable selection and hypothesize on the following five dimensions of digital capital: material digital access, perceived digital competence, perceived digital autonomy, (home) academic ICT usage, and (home) personal ICT usage. I initially include 28 potential indicator variables that might reliably measure these latent, or unobserved, components of digital capital. Since this analysis is an exploratory technique, components were not specified a priori. Rather, my research aim is to analyze the intercorrelations between these variables to identify a satisfactory component solution.

In Table 1 below, I display descriptive statistics for the 28 indicator variables included in the PCA. I include three variables related to material digital access. Specifically, these variables measure home access to the internet, desktop computers, and internet-connected cellphones. I recoded the measurement of internet and smartphone variables from three categories (“No,” “Yes, but I don’t use it,” and “Yes, and I use it”) to binary measurement of “No” and “Yes” since my goal is to measure access, not necessarily usage, with these variables. In their conceptualization of digital capital, Ragnedda (2018) includes material access as an objectified form of digital capital. Others agree that material access, in addition to autonomy, skill, and purpose of use, differentiate digital usage patterns among digital users (DiMaggio et al., 2004). For other scholars, material access is *not* included in conceptualizations of digital capital (Ollier-Malaterre et al., 2019; Santillana et al., 2020). Although the literature is inconsistent with the inclusion of material access as a component (or factor) of digital capital, I choose to include these three indicator variables in the current PCA to potentially provide support for other studies indicating that material access should be included in measurement of the construct.

Additionally, I include five potential measures of perceived digital competence and five related to perceived digital autonomy as suggested by DiMaggio et al. (2004). These ten variables are ordinally measured on a 4-point Likert scale ("Strongly Disagree" to "Strongly Agree") and measure respondents' beliefs about their digital skills (perceived digital competence) and about their ability to use ICTs independently (perceived digital autonomy). Finally, I identify seven variables that could potentially measure institutionally rewarded, academic digital practices (Hargittai & Hinnant, 2008), as well as eight variables representing a variety of non-academic digital behaviors (Rafalow, 2020; Watkins & Cho, 2018). These 15 variables are ordinally measured on a 5-point scale.

Table 1. Descriptive Statistics – PCA Indicator Variables

Potential Digital Capital Item	Measures	%	% Missing From N (US)*
Available at home: cellphone with internet access	0 = no 1= yes	1.90 98.10	5.19
Available at home: desktop computer for schoolwork	0 = no 1 = yes	10.52 89.48	5.29
Available at home: link to internet	0 = no 1 = yes	3.17 96.83	2.31
Feels comfortable using ICTs not familiar with	1 = SD 2 = D 3 = A 4 = SA	6.77 28.87 50.70 13.66	9.07
Can give advice to others on which ICTs to purchase	1 = SD 2 = D 3 = A 4 = SA	4.81 19.53 56.29 19.38	9.20
Comfortable using ICTs (home)	1 = SD 2 = D 3 = A 4 = SA	1.69 4.93 52.78 40.60	9.65
Overcomes technical difficulties with devices	1 = SD 2 = D 3 = A 4 = SA	2.90 13.45 58.71 24.94	9.09
Helps others with ICTs	1 = SD 2 = D 3 = A 4 = SA	3.87 14.15 58.86 23.13	9.32
Independently installs new software	1 = SD 2 = D 3 = A 4 = SA	11.43 27.36 42.56 18.65	9.18
Reads information about ICTs to be independent user	1 = SD 2 = D 3 = A 4 = SA	7.86 28.99 47.58 15.57	9.47
Uses ICTs how desired	1 = SD 2 = D 3 = A 4 = SA	3.14 13.27 58.89 24.70	9.59
Independently solves problems with ICTs	1 = SD 2 = D 3 = A 4 = SA	4.35 19.26 55.62 20.77	9.20
Selects apps/programs independently	1 = SD 2 = D 3 = A 4 = SA	4.99 21.67 51.78 21.55	9.47

Potential Digital Capital Item	Measures	%	% Missing From N (US)*
Uses ICTs to follow up lesson (ex: for an explanation)	1 = Never/Hardly Ever	15.39	9.16
	2 = 1-2x/month	20.28	
	3 = 1-2x/week	31.02	
	4 = Almost Every day	22.19	
	5 = Every day	11.12	
Emails peers about schoolwork	1 = Never/Hardly Ever	42.50	10.15
	2 = 1-2x/month	18.05	
	3 = 1-2x/week	20.68	
	4 = Almost Every day	11.52	
	5 = Every day	7.26	
Emails teachers about schoolwork	1 = Never/Hardly Ever	24.40	9.59
	2 = 1-2x/month	24.88	
	3 = 1-2x/week	28.23	
	4 = Almost Every day	14.36	
	5 = Every day\	8.13	
Downloads (DL) & uploads (UL) material from school website	1 = Never/Hardly Ever	32.77	9.67
	2 = 1-2x/month	23.79	
	3 = 1-2x/week	21.80	
	4 = Almost Every day	12.82	
	5 = Every day	8.83	
Check school website for announcements	1 = Never/Hardly Ever	42.65	9.47
	2 = 1-2x/month	19.80	
	3 = 1-2x/week	18.08	
	4 = Almost Every day	11.70	
	5 = Every day	7.77	
Uses learning apps on computer	1 = Never/Hardly Ever	34.64	8.63
	2 = 1-2x/month	19.71	
	3 = 1-2x/week	21.52	
	4 = Almost Every day	14.39	
	5 = Every day	9.73	
Uses learning apps on phone	1 = Never/Hardly Ever	35.91	9.38
	2 = 1-2x/month	18.23	
	3 = 1-2x/week	21.19	
	4 = Almost Every day	14.63	
	5 = Every day	10.04	
Online chat	1 = Never/Hardly Ever	14.96	8.54
	2 = 1-2x/month	6.59	
	3 = 1-2x/week	11.73	
	4 = Almost Every day	21.40	
	5 = Every day	45.31	
Uses social media	1 = Never/Hardly Ever	14.18	7.90
	2 = 1-2x/month	6.68	
	3 = 1-2x/week	11.43	
	4 = Almost Every day	23.07	
	5 = Every day	44.65	
Browses internet for fun	1 = Never/Hardly Ever	4.29	8.04
	2 = 1-2x/month	4.53	
	3 = 1-2x/week	13.00	
	4 = Almost Every day	27.15	
	5 = Every day	51.03	

Potential Digital Capital Item	Measures	%	% Missing From N (US)*
Reads news on the internet	1 = Never/Hardly Ever 2 = 1-2x/month 3 = 1-2x/week 4 = Almost Every day 5 = Every day	18.89 16.75 26.06 20.59 17.71	8.06
Obtains practical info from the web (ex: dates, events, locations)	1 = Never/Hardly Ever 2 = 1-2x/month 3 = 1-2x/week 4 = Almost Every day 5 = Every day	13.06 14.57 27.60 24.55 20.22	7.19
Downloads music, software, games, films from the internet	1 = Never/Hardly Ever 2 = 1-2x/month 3 = 1-2x/week 4 = Almost Every day 5 = Every day	16.54 18.44 21.52 21.10 22.40	8.47
Uploads created content for sharing (ex: poetry, music, videos, etc.)	1 = Never/Hardly Ever 2 = 1-2x/month 3 = 1-2x/week 4 = Almost Every day 5 = Every day	49.52 16.87 14.69 9.52 9.40	7.96
Downloads applications on mobile device	1 = Never/Hardly Ever 2 = 1-2x/month 3 = 1-2x/week 4 = Almost Every day 5 = Every day	15.08 38.97 22.40 12.79 10.76	8.21

**n = 3,308; N (US) = 4,838*

In the OLS regression model, I analyze student mathematics proficiency as the outcome variable of interest. In the literature, scholars have identified that digital inequality is closely linked to a variety of academic inequalities (Fairlie, 2012; Fairlie et al., 2010; Fairlie & Robinson, 2013; Gonzales et al., 2020; Katz et al., 2017). Since digital learning strategies (Byun & Joung, 2018) and digital access (Olive et al., 2009) are both positively correlated with mathematics learning outcomes, I select mathematics proficiency as the dependent variable for analysis. Mathematics proficiency is estimated through 10 PVs obtained from Item Response Theory models (OECD, 2009). Item Response Theory, for PISA, involves assigning a set of 10 plausible values to each student based on their responses in the PISA mathematics assessment, representing the plausible distribution of mathematical abilities of respondents.

Components from the PCA were included as independent variables in the regression model if the Cronbach's alpha test statistic for the component was greater than or equal to .70. In addition to these components, several student and family background variables were included as control variables. Specifically, I control for student sex (binary measurement of male/female) since prior research has documented the relationship between sex and mathematics anxiety, which has a negative effect on mathematics performance (Devine et al., 2012). In the regression, male students are treated as the reference group and female students as the focal group. I also include immigration status, which PISA measured in three categories: "native" (reference group), "first-generation immigrant," and "second-generation immigrant". I include this control variable since prior research has identified a significant correlation between immigration status and mathematics test scores (Devine et al., 2012).

Finally, I control for parental occupational status (as an index score) and highest parental educational attainment, which are both indicators of socioeconomic status (SES). The literature identifies SES as a major contributor to many academic inequalities, including the "mathematics achievement gap" (Galindo & Sonnenschein, 2015). The variable for highest parental education is measured by International Standard Classification of Education (ISCED) level instead of years of schooling or degree completed. Briefly, the 6-level measurement for this variable is as follows: (0) "None" [kindergarten or below]; (1) "ISCED level 1" [1^{st}-6^{th} grades]; (2) "ISCED level 2" [7^{th}-9^{th} grades]; (3) "ISCED 3a and/or ISCED 4" [10^{th}-12^{th} grades and/or 1-year vocational program after 12th grade]; (4) “ISCED level 5b” [2- to 4-year vocational program post-secondary]; and (5) “ISCED 5a and/or ISCED 6” [tertiary education below the doctorate level and/or completion of doctorate degree]. More information about ISCED measurements can be found in Annex D of the PISA 2018 Technical Report (OECD, 2019d).

Statistical Methods

Principal Component Analysis

To address my first research aim, I conduct PCA to propose a parsimonious set of latent constructs that best measure the intercorrelations between variables. Since the literature on digital capital lacks clear consensus in operationalization of the construct – and as some scholars have argued, is lacking empirical measures (Puckett, 2020) – PCA is an appropriate method for this project. Generally, statisticians recommend PCA for situations in which there is no strong *a priori* theory (Bandalos & Finney, 2018). In practice, PCA is used to identify and group “clusters” of variables based on their shared variance into descriptive categories, or principal components (Yong & Pearce, 2013).

Since the variables I include in the PCA are either binary or ordinally measured, I begin by creating a polychoric correlation matrix (Table 5, Appendix) to input as data instead of raw data values. This technique is typically suitable when data violate the assumption of interval-level measurement (Rupp et al., 2003). After running the initial PCA, I then decide how many components to extract by considering eigenvalues, scree plots, parallel analyses, and the total variance explained of each additional included component. The *Kaiser Criterion* for factor extraction suggests that only factors (or in this case, principal components) with eigenvalues greater than one should be extracted, as an eigenvalue of 1 indicates that the factor explains as much variance as a single variable (Kaiser & Dickman, 1959). Moreover, I consider a scree plot, which graphically depicts the number of components on the x-axis and eigenvalues on the y-axis, and a parallel analysis. In a parallel analysis, STATA compares the eigenvalues from the PCA against a randomly generated correlation matrix of an identical size over a specified number of replications (10 for the current project). When the eigenvalues of the PCA drop below the eigenvalues for the randomly generated correlation matrix, the components after this point are mostly "random noise" (UCLA, 2016). Finally, I consider the total variance explained to guide factor extraction. Yang (2005) suggests that an Exploratory Factor Analysis (EFA) should retain the factors that can explain at least 60% of the total variance, but others suggest a minimum of 70% of variance explained (Jolliffe & Cadima, 2016).

In PCA, as a variant of Exploratory Factor Analysis, components are typically rotated to provide better interpretation of results and to create a "simple structure" where variables load cleanly onto one principal component (Yong & Pearce, 2013). For this analysis, I use Promax, an oblique rotation method, to rotate component solutions. Oblique rotation methods are typically used when components are considered to be correlated, and Promax is a common oblique rotation technique that usually provides greater correlations among components and achieves a simple structure (Yong & Pearce, 2013). This decision was made based upon the item correlations depicted in the polychoric correlation matrix, in addition to the correlation between the resulting components (See Table 7, Appendix). Tabachnick et al. (2007) suggest that if factor correlations exceed 0.32, an oblique rotation method is justified.

To assess reliability in the PCA model, I consider Cronbach's alpha test statistic (α) for each extracted component. Although there are several heuristics for assessing reliability with the alpha-coefficient, scholars typically agree that $\alpha \geq 0.70$ is preferred, but $\alpha \geq 0.60$ might suffice in an exploratory pilot study (Hair et al., 2006). I decide to reject components below the threshold of 0.70 since there is an, albeit limited, guiding framework of digital capital established in the literature. I then examine reliability through item-to-total correlations to identify any weak items that should be dropped to improve component reliability. Finally, I run a parsimonious version of the PCA model, excluding any variables that did not meet reliability criteria.

OLS Regression

In the linear regression analysis, I analyze the effect of digital capital components on math proficiency. Additionally, I include any individual ICT-related variables from the PCA that did not meet reliability criteria, as a singular component, and student background variables. Before computing final estimates, I begin by running the initial OLS regression model using the first plausible value of student math proficiency to evaluate model assumptions and regression diagnostics. Specifically, I examine the model for data sparseness, multi-collinearity (Table 8, Appendix), non-linearity, outliers, and non-normality. Since the model reports robust standard errors, I assume homoskedasticity. To address data sparseness in the independent variables, I combine the two lowest levels of parental education into one level, "Kindergarten – 6th grade," so that the combined level makes up 1.9% of values. Moreover, I review Lowess scatterplots, and conduct Likelihood Ratio and Box-Tidwell tests, for each independent variable to assess non-linearity. I found parental education to be non-linear, so I use dummy variables in the model to correct for this. Additionally, the Box-Tidwell test identified 3 statistically significant non-linear independent variables. Thus, I add squared variables to the model as a polynomial regression technique. With the inclusion of these adjustments for further analysis, the model meets the assumptions of OLS regression.

After evaluating regression diagnostics, I then incorporated the 9 remaining PVs into the analysis using the STATA *pisareg* command (Jakubowski, 2013). Since PISA has a complex sample design, using common statistical procedures would produce biased estimates of standard error and sampling variance (OECD, 2019d). Alternatively, the *pisareg* command uses the nonresponse adjusted student weight and 80 replicate weights to average estimates over the full set of PVs, producing unbiased standard errors and estimates of sampling variance (Jakubowski, 2013).

RESULTS

Principal Component Analysis

In the initial PCA model, I retain 6 principal components with eigenvalues > 1 and strong face validity. The omnibus Keiser-Meyer-Olkin (KMO) test yielded a test statistic of 0.887, suggesting a "meritorious" (Kaiser, 1974) level of sampling adequacy in the PCA. However, upon inspection of the more granular item KMO test statistics, three items yielded considerably low ratios. Specifically, the item KMO ratios for home internet access (KMO = 0.592), home computer access (KMO = 0.615), and smartphone access (KMO = 0.658) were considerably lower than the rest of the items, which ranged from 0.834 to 0.954. These 3 item KMO ratios raise concern, since Kaiser (1974) refers to these values as "miserable" (0.50) and "mediocre" (0.60s). However, since the item ratios fall into the "acceptable" range, I continue with analysis.

In Figure 7 (Appendix), I include the scree plot and parallel analysis used to guide component extraction in the initial PCA. Although the scree plot alone might suggest a 5-component solution, I decided to retain 6 components since the eigenvalue for this 6^{th} component was larger than 1, and the components made clear conceptual sense. The rotated solution for this initial PCA is displayed in Table 6 (Appendix).

Next, I examine the internal reliability and find that for each component, with the exception of "Material Access," the Cronbach's α test statistic indicated acceptable internal reliability ($\alpha \geq .70$). As measured in this exploratory analysis, Material Access did not appear to be a reliable component of digital capital ($\alpha = .414$). Following the principle of Occam's Razor, I conduct a parsimonious final iteration of the PCA, excluding the Material Access component and its associated indicator variables.

In the final iteration of the PCA, I follow the same analytic procedures used in the initial run. This final PCA model resulted in a slight improvement in sampling adequacy with an Omnibus KMO statistic of 0.908, which is likely a result of dropping the items with the 3 lowest item KMO ratios. The parsimonious model yields 5 components with eigenvalues greater than 1. See Figure 1 below for the scree plot of eigenvalues and parallel analysis. Although the scree plot seems to suggest a 4-component solution, the parallel analysis (indicated by the dashed line) suggests that 5 components might be present. I explored the 4-component solution for the PCA but found that the 5-component solution was satisfactory in maximizing variance explained (0.697) while being conceptually meaningful. This percentage of variance explained is close to the suggested criterion from Jolliffe and Cadima (2016).

Figure 1. Parallel Analysis - Final PCA

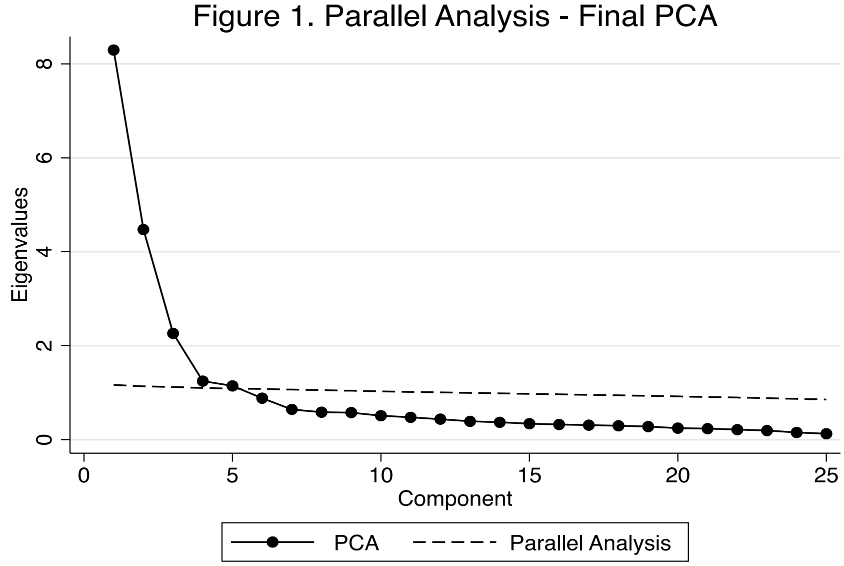


Table 2 reports the rotated 5-component solution for the final PCA iteration. Component loadings $\geq$ .30 are displayed in the table. In PCA, component loadings represent the correlation between the indicator variable and each component. These correlation coefficients provide an idea of how much the variable "contributed" to the component. Typically, component loadings < .30 are considered "weak" in strength (Yong & Pearce, 2013).

Table 2. Final PCA Model

Variable	Component 1	Component 2	Component 3	Component 4	Component 5	Communalities
Use ICTs to follow up lesson	0.358					0.682
Email Peers	0.373					0.677
Email teacher	0.387					0.680
DL & UL from school website	0.370					0.743
Check school school website	0.364					0.690
Use learning app - computer	0.380					0.733
Use learning app - phone	0.367					0.720
Indep. Install new software		0.482				0.762
Read tech info to be indep.		0.472				0.750
Uses ICTs how desired		0.356				0.680
Indep. Solves problems ICT		0.412				0.777
Selects apps/programs indep.		0.446				0.750
Uses unfamiliar ICTs			0.421			0.546
Can give ICT advice-purchase			0.457			0.745
Comfortable home ICT use			0.377			0.711
Overcomes tech. difficulties			0.453			0.809
Helps others with ICTs			0.468			0.793
Online chat				0.390		0.564
Uses social media				0.464		0.629
Browse Internet for fun				0.486		0.682
Read news on internet				0.365		0.548
Practical info from web				0.380		0.637
DL music, SW, games, films					0.460	0.686
UL created content for share					0.577	0.728
DL applications - phone					0.566	0.699
Eigenvalues	8.293	4.473	2.262	1.246	1.145	
Cumulative Variance Exp.	0.332	0.511	0.601	0.651	0.697	
Cronbach's $\boldsymbol{\alpha}$	0.901	0.873	0.844	0.758	0.727	

Only depicts component loadings $\geq$ .30. See Table 2. (Appendix) for variable list.

The first principal component, *Academic Digital Usage*, explains 33.17% of variance in the PCA. This was also the most reliable component with a standardized α-

coefficient of 0.901. The grouped variables in this component measure digital uses related to school or learning. The next two components, *Perceived Digital Autonomy* and *Perceived Digital Competence*, provide an additional 17.89 and 9.05 percent of variance explained, respectively. Perceived Digital Autonomy ($\alpha = 0.873$) and Perceived Digital Competence ($\alpha = 0.844$) also meet reliability criteria. Perceived Digital Autonomy includes 5 variables measuring respondents' self-reported ability to use digital devices independently, while the 5 variables contributing to Perceived Digital Competence measure self-reported beliefs of respondents' knowledge and skills related to ICT usage. The last two components, *Casual Digital Browsing* and *Knowledge-Based Digital Leisure*, explain 4.98 and 4.58 percent of variance, respectively. These components were also considered reliable with $\alpha = 0.755$ for Casual Digital Browsing, and $\alpha = 0.727$ for Knowledge-Based Digital Leisure. Casual Digital Browsing reflects online, perhaps passive digital behaviors that don't require too much effort or skill, while knowledge-based digital leisure involves leisure behaviors that require some level of software, application, or program specific digital knowledge. Upon inspection of the item-to-total correlations related to reliability statistics, none of the items seemed to negatively impact the standardized alpha-coefficients.

OLS Regression

The OLS regression analysis was conducted to assess the predictive validity of the five components extracted from the PCA. Additionally, I include the Material Access variables used in the initial PCA model as three individual variables, rather than as a singular component, to analyze the potential effects of digital access on student math proficiency. Four student/family demographic variables were also included. Descriptive statistics are reported in Table 3 below.

Table 3. Descriptive Statistics – OLS Regression

Variable	Measure / [Range]	% / Mean	% Missing (Dropped)
Plausible Values (Mathematics Proficiency)			
PV1Math	[200.962, 785.222]	488.610	0
PV2Math	[214.949, 771.545]	490.082	0
PV3Math	[226.593, 842.375]	489.349	0
PV4Math	[227.389, 730.334]	489.797	0
PV5Math	[208.304, 754.065]	488.955	0
PV6Math	[215.624, 768.664]	490.241	0
PV7Math	[179.279, 778.881]	489.451	0
PV8Math	[192.603, 755.067]	489.579	0
PV9Math	[213.712, 756.952]	489.919	0
PV10Math	[234.953, 805.569]	489.619	0
Independent Variables			
Academic Digital Usage	[-1.138, 1.955]	0	-
Perceived Digital Autonomy	[-2.360, 1.436]	0	-
Perceived Digital Competence	[-2.796, 1.368]	0	-
Casual Digital Browsing	[-1.992, 1.062]	0	-
Knowledge-Based Digital Leisure	[-1.250, 1.803]	0	-
Home Internet Access	0 = No 1 = Yes	1.90 98.10	5.19
Smartphone Access	0 = No 1 = Yes	3.17 96.83	2.31
Home Computer Access	0 = No 1 = Yes	10.52 89.48	5.29
Control Variables			
Student Sex	0 = Male 1= Female	49.33 50.67	0
Parental Occupational Status	[11.56, 88.96]	56.12	8.43
Parental Education	0 = K-6th grade 1 = 7th-9th grade 2 = 10th-12th; 1-yr voc. 3= 2-4 yr. voc. program 4 = Tertiary (higher ed)	1.90 5.20 25.24 13.18 54.47	1.63
Immigrant Status	0 = Native US Born 1 = Second-Generation 2 = First-Generation	79.38 15.36 5.26	3.14

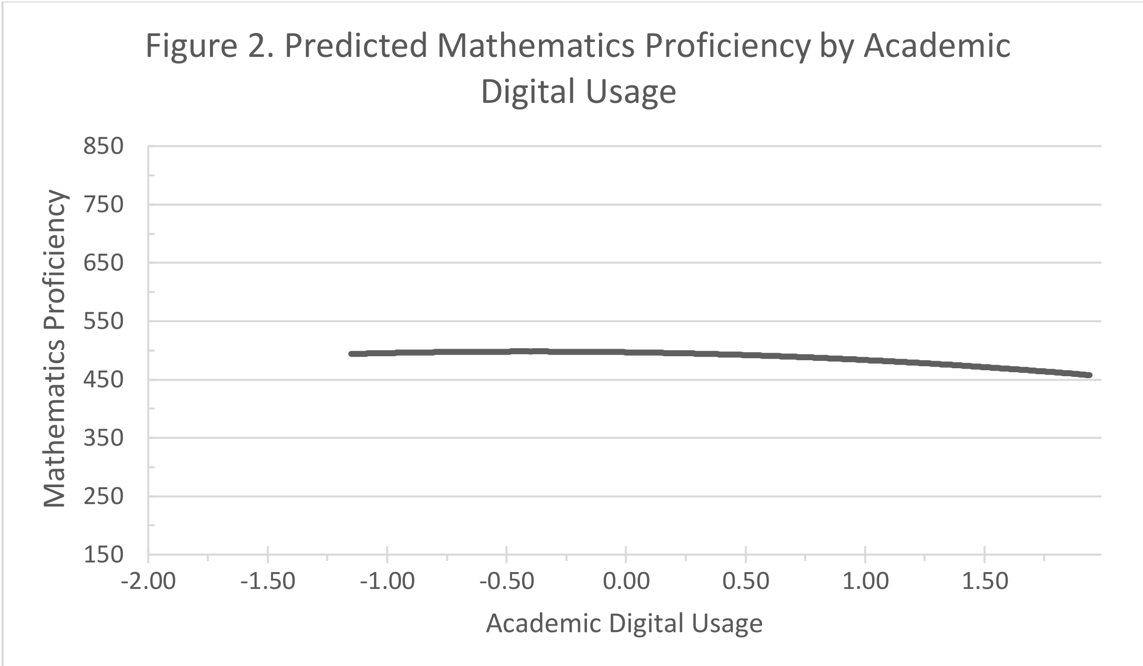
Table 4. OLS Regression Results

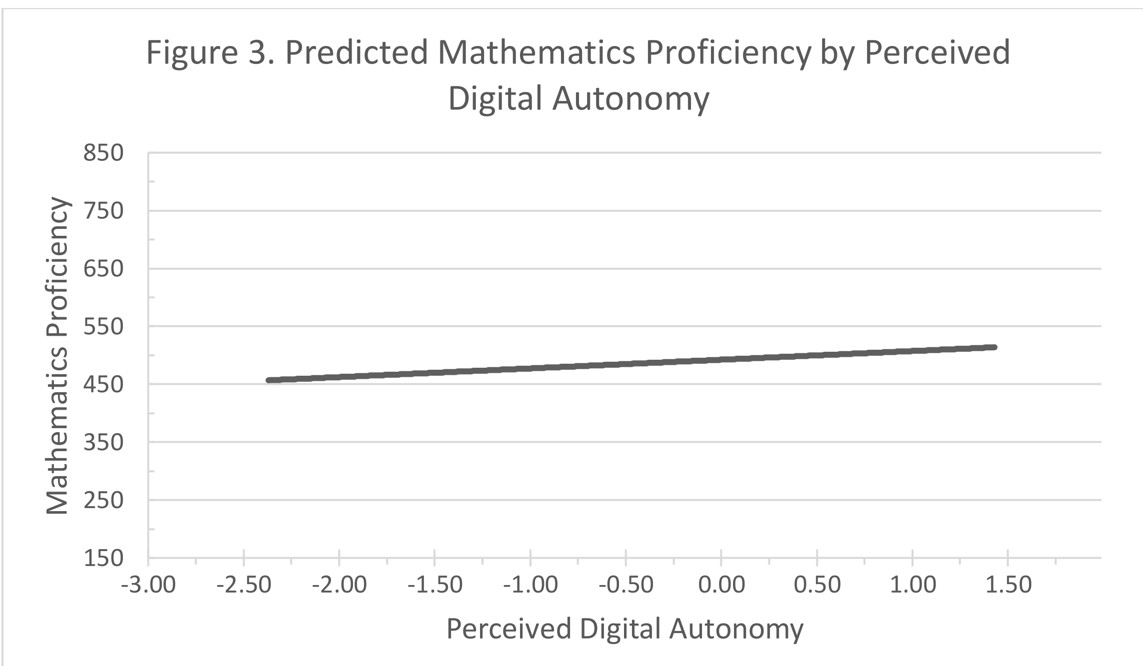
Independent Variable	Coefficient	$R^2 = 0.301$ n = 3,308 [95% Confidence Interval]
Digital Capital Components		
Academic Digital Usage		
Main effect	-5.468*	[-10.566, -0.374]
Squared Effect	-7.885**	[-13.319, -2.461]
Perceived Digital Autonomy	14.894***	[10.010, 19.770]
Perceived Digital Competence		
Main Effect	4.061	[-1.487, 9.607]
Squared Effect	2.785	[-0.836, 6.416]
Casual Digital Browsing		
Main Effect	28.446***	[22.354, 34.546]
Squared Effect	-7.087**	[-12.186, -1.994]
Knowledge-Based Digital Leisure	-35.435***	[-40.046, -30.834]
Material Access Variables		
Home Internet Access	10.090	[-12.391, 32.571]
Smartphone Access	13.694	[-7.458, 34.838]
Home Computer Access	24.860***	[15.511, 34.209]
Control Variables		
Student Sex (Ref: Male)	-10.792**	[-18.493, -3.087]
Parental Occupation Status	0.810***	[0.653, 0.967]
Parental Education (Ref: K-6)		
7^{th}-9^{th} grade	-7.039	[-29.933, 15.853]
10^{th}-12^{th} grade; 1-yr. voc.	20.525	[-2.314, 43.354]
2-4 yr. voc. program	22.687	[-0.928, 46.308]
Tertiary (higher ed.)	41.046***	[19.568, 62.531]
Immigration Status (Ref: Native)		
1st-generation	-7.353	[-20.776, 6.076]
2nd-generation	16.570***	[7.260, 25.880]
Intercept (constant)	380.644	[344.596, 416.584]

***$p < 0.05$; **$p < 0.01$; ***$p < 0.001$**

The final regression results (using the *pisareg* command to apply sampling and replicate weights) are provided in Table 4 above. With an adjusted R^2 of 0.301, the model is statistically significant and explains 30.1% of the variance in student mathematics proficiency scores. This value suggests adequate model fit, particularly for an exploratory study. Components that were identified as significantly ($p < 0.05$) nonlinear in the Box-Tidwell test are reported with two coefficients: one for the "main" component effect and another to account for nonlinearity in the component's relationship with mathematics proficiency.

Of the five digital capital components included in the model, four are identified as statistically significant predictors of mathematics proficiency. More specifically, Perceived Digital Competence was the only component to yield non-significant findings at the 95% confidence level. Figure 2 displays the predicted mathematics proficiency scores by Academic Digital Usage. For this relationship, the regression results indicate a main coefficient (COEF) of -5.468 ($p < 0.05$) and a nonlinear coefficient ($COEF^2$) of -7.885 ($p < 0.01$). As the graph demonstrates, Academic Digital Usage was found to have a nonlinear relationship with mathematics proficiency. At low levels of academic digital usage, the slope of the curve is positive, indicating that a slight increase in academic digital usage would predict a marginally higher score for mathematics proficiency. The peak of the curve is found at Academic Digital Usage = -0.390, which approximately represents the 34^{th} percentile (P_{34}) in the distribution of values for this component. After this point, an increase in the component yields a lower predicted proficiency score, with a minimum predicted value occurring at Academic Digital Usage = 1.940.

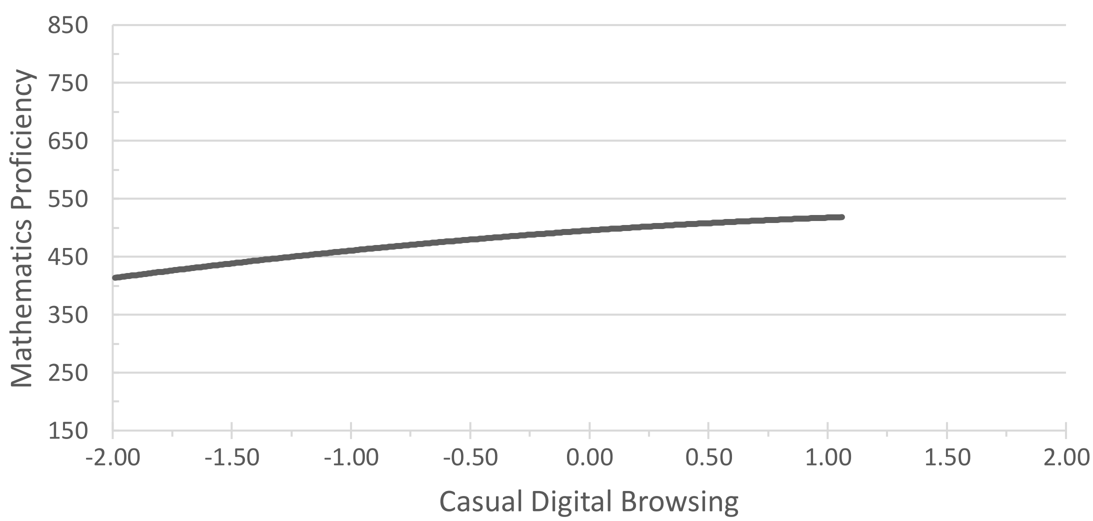




In Figure 3 above, I display the prediction model for Perceived Digital Autonomy. The slope of this line indicates that a one-point increase in Predicted Digital Autonomy is associated with a 14.894-point increase in mathematics proficiency. To provide a more practical interpretation, each one standard deviation (s_x = 0.814) increase in the component is predicted to yield a 12.124-point increase in mathematics proficiency. This model predicts a proficiency score of 456.816 at the minimum value for Perceived Digital Autonomy and 513.854 at the maximum value.

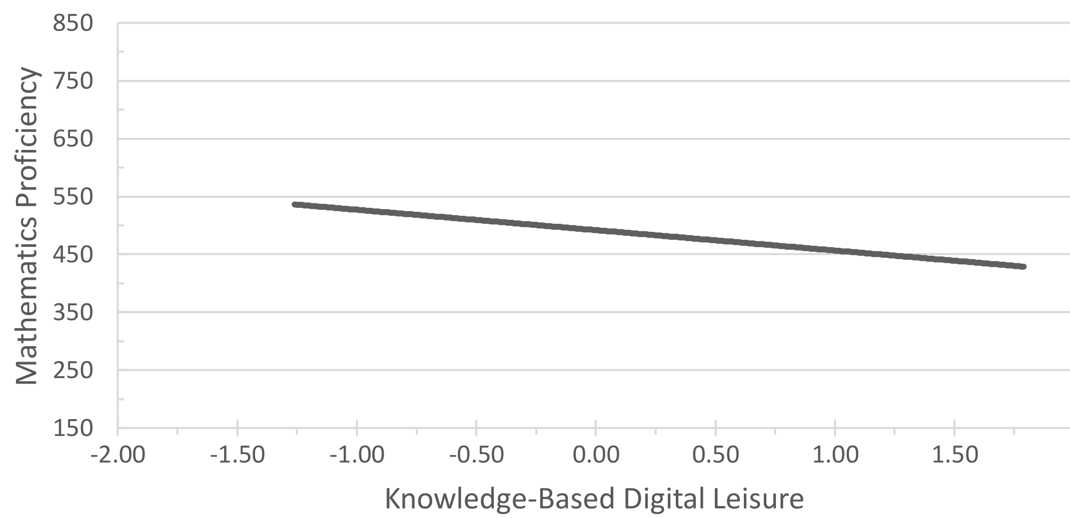
Similarly, Casual Digital Browsing was also found to have a positive effect on mathematics proficiency (COEF = 28.446, $p < 0.001$; COEF^2 = -7.087, $p < 0.01$). Figure 4 graphically displays the prediction model for this relationship. The lowest predicted proficiency score (413.728) occurs at the minimum value of the component (-1.99), while the highest predicted value (518.482) occurs at the maximum component value (1.06). Since the effect is nonlinear, an increase at lower levels of Casual Digital Browsing is predicted to yield a larger increase in mathematics proficiency compared to a similar size increase at the higher levels. To interpret predicted mathematics proficiency in terms of the component's distribution, P_{25} predicts a proficiency score of 481.044, P_{50} = 498.251, and P_{75} = 507.658.

Figure 4. Predicted Mathematics Proficiency by Casual Digital Browsing



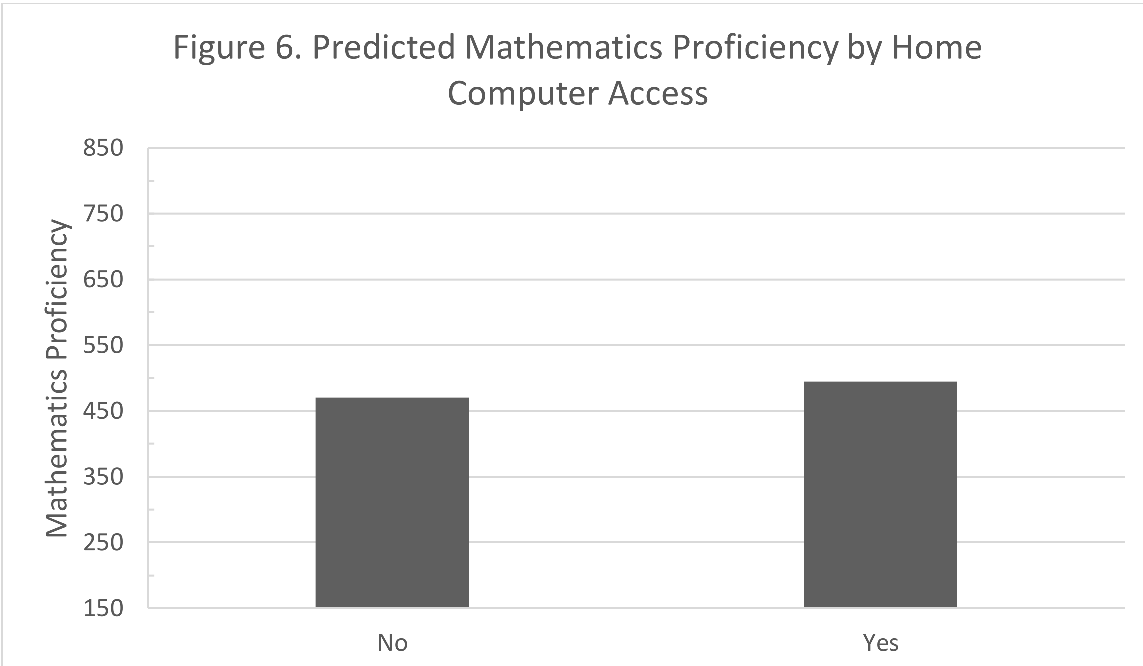
The final statistically significant component, Knowledge-Based Digital Leisure, was found to have a negative effect on mathematics proficiency (COEF = -35.435, $p < 0.001$). Since this effect is linear, the model in Figure 5 predicts that a one standard deviation ($s_x = 0.804$) increase in the component would yield a 28.490-point reduction in mathematics proficiency. The highest predicted proficiency score in this model occurs at the minimum value of Knowledge-Based Digital Leisure (-1.250), while the lowest predicted score occurs at the maximum component value (1.803).

Figure 5. Predicted Mathematics Proficiency by Knowledge-Based Digital Leisure



Although the material access was not retained as a reliable component of digital capital ($\alpha = .414$), I include the 3 variables in the regression analysis as separate independent variables rather than a singular latent component. Notably, home computer access was found to have a significantly positive effect on student mathematics proficiency (COEF = 28.860, $p < 0.001$). Figure 6 displays this relationship, with students who have home computer access being predicted to score 28.860 points higher in mathematics proficiency compared to those who do not have home computer access.

Other variables related to digital access – home internet access and smartphone access – were not identified as statistically significant (at the 95% confidence level) in the analysis.



To summarize the regression effects of the control variables included in the analysis, each variable resulted in at least one significant finding. Parental occupational status was found to have a statistically significant, positive effect on student mathematics proficiency. A one standard deviation (s_x = 21.135) increase in the occupational status index score is associated with a 17.119-point increase in student mathematics proficiency. Parental education level, which was analyzed through the inclusion of dummy variables to address non-linearity, was also found to be statistically significant in the model between the lowest and highest levels of education (COEF = 41.046; $p < .001$). With this coefficient, one might predict students whose parents have tertiary (collegiate) education to score approximately 41 points higher in mathematics proficiency compared to those whose parents have not completed education higher than 6th grade. Likewise, only one of the immigrant status dummy variables was significant at the 95% confidence level; second-generation immigrant status is associated with a 16.570-point higher score compared to those categorized as native US born. Finally, student sex was statistically significant in the model (COEF = -10.79; $p < .001$), indicating that female students were scored 10.79 points lower in mathematics proficiency compared to male students.

DISCUSSION

This study yields several important contributions to the literature. My first research aim, to identify a parsimonious set of latent principal components that best describes how the observed variables are intercorrelated, was achieved using PCA. In prior conceptualizations, the theoretical construct of digital capital was missing clear and consistent operationalization between scholars (Ollier-Malaterre et al., 2019; Ragnedda, 2018; Santillana et al., 2020). As a result, meaningful research was hindered by such a large number of differing measures to consider. The PCA conducted in this study proved to be useful as a data reduction strategy. I used this method to “cluster” and define groups of variables, reducing a total of 25 observed variables into 5 latent components. This smaller set of descriptive variable “clusters” is much easier to work with and allows researchers to take a wholistic, yet simplified, approach in addressing the multi-faceted issue of digital inequality. This also provides an excellent alternative to the narrow scope of the traditional “digital divide” framework, which is limited to addressing inequality in digital access or digital usage in a binary way (i.e. “the information haves” vs. “the information have-nots”).

Material Access

Of the 6 principal components initially proposed in this study, 5 were found to display acceptable internal reliability ($\alpha \geq .70$). The component related to material access, as measured by the 3 indicator variables previously described, was not found to be reliable ($\alpha = .414$). This finding might provide support for other studies that do not include material access in operationalizations of digital capital (Ollier-Malaterre et al., 2019; Santillana et al., 2020). Importantly, the current study uses an exploratory, rather than *confirmatory*, analysis, so the findings do not “confirm” or “disprove” any other perspective of digital capital. Perhaps, material access could be included as a component (or latent factor) of digital capital, but the observed variables included in this study did not support a reliable scale. Also, the findings from this study are only generalizable to 15-year-old students in the US, perhaps representing a considerably different population from those examined in other analyses such as Ragnedda et al. (2020) who consider digital capital among adults (ages 18+) in the United Kingdom.

Moreover, as several scholars have addressed in the literature, “quality of access” remains an important, yet underexplored, consideration related to digital inequality (DiMaggio et al., 2004; Katz et al., 2017; Lai & Widmar, 2021; Selwyn, 2004). The dataset used in this analysis was limited in terms of measurement for digital access variables. Binary measures of access (i.e., “has access” or “does not have access”) are overly simplistic and do not reflect the true nature of digital inequality as a stratified, hierarchical spectrum. Unfortunately, detailed measurement related to quality of digital access (i.e., type or speed of internet connection available to respondents) was unavailable in the PISA dataset, which could be partially responsible for the low alpha coefficient.

Although measures of digital access in this study were not ideal, the OLS regression provided evidence to support that material access, in terms of home computer access, is a significant predictor of student mathematics proficiency (COEF = 28.45; $p < .001$). Contrary to those who argue that research should “move beyond digital access” to address unequal patterns of digital usage (Attewell, 2001; Hargittai, 2001), this study revealed that material access still warrants academic attention. Regardless of whether material access should, or should not, be considered a factor of digital capital, there are still clear implications for addressing material access in relation to academic inequality (Anderson & Perrin, 2018), particularly in the era of COVID-19 (Lai & Widmar, 2021).

Final PCA Model

The 5 principal components that were extracted in the final PCA model – Academic Digital Usage, Perceived Digital Autonomy, Perceived Digital Competence, Casual Digital Browsing, and Knowledge-Based Digital Leisure – provide support for some conceptualizations of digital capital proposed by other scholars. DiMaggio et al. (2004) contend that purposes of digital use, digital autonomy, and digital skills (in addition to other concepts not included in this analysis) differentiate digital usage patterns among users to create some form of institutional advantage, such as in economic, occupational, or political contexts. Likewise, others agree that digital skill is a mediating factor of "capital-enhancing" digital usage (Hargittai & Hinnant, 2008). With the current study, I was able to identify reliable measures that form the principal components of Academic Digital Usage, Perceived Digital Autonomy, and Perceived Digital Competence.

Some scholars have done extensive qualitative research on "digital play" as a foundation for building essential digital skills and capital (Rafalow, 2020). By and large, most of the literature is focused exclusively on the institutionally valued forms of digital usage, which is characterized by academic digital usage in an educational context. Yet, the findings indicate that other forms of digital usage, such as casual digital browsing and knowledge-based digital leisure, might be important to consider as components of digital capital.

OLS Regression

In the second portion of the analysis, I sought to validate the 5 principal components, in terms of predictive validity, by analyzing their effects in an OLS regression on student mathematics proficiency. Additionally, I included the three material access measures as individual, independent variables to evaluate the potential impact that these might have on student mathematics proficiency. In this analysis, I was able to find evidence of predictive validity for the Casual Digital Browsing and Perceived Digital Autonomy components. These components, as independent variables, were statistically significant and predicted student math proficiency in the hypothesized, positive direction. For this reason, I argue that these two components warrant considerable attention in future factor analyses. Although casual digital browsing might not represent the “capital-enhancing” or institutionally valued digital practices that are frequently discussed in the literature (DiMaggio et al., 2004; Hargittai & Hinnant, 2008), this form of digital engagement appears to be positively related to mathematics proficiency. Since casual digital media participation is popular among Generation Z (Watkins & Cho, 2018), it will be important for research to consider this form of digital usage as a potential source of digital capital.

Moreover, the validation finding related to perceived digital autonomy was interesting, as these measures represent one’s *perception* of autonomy in digital usage rather than one’s *actual* digital autonomy, as discussed by DiMaggio et al. (2004). In the current project, perception of digital autonomy seemed to constitute a reliable principal component that displays predictive validity, which was not the case for perceived digital competence. Although the observed indicators of this latent component also measured self-reported beliefs about one's digital usage, the results indicated a non-significant finding. Other studies (Santillana et al., 2020) measure respondents' actual digital skills, which might provide richer detail than the self-reported skills measured by PISA. Although Hargittai and Hinnant (2008) report that self-report measures of online skill were significant predictors of "visiting capital-enhancing websites," perception of one's technical skills were not found to have a significant effect on mathematics proficiency in this study. Perhaps an objective assessment of digital skill (rather than subjective, self-reported measurement) would yield a considerably different outcome in the analysis.

The last two principal components included in the OLS regression, Academic Digital Usage and Knowledge-Based Digital Leisure, both had statistically significant, negative effects on student mathematics proficiency. Academic Digital Usage displayed high internal consistency ($\alpha = .901$), but I was not able to validate the component in this study. It's possible that this might be because certain digital uses related to academics (such as emailing teachers about assignments) might be more common for students who are struggling and need assistance. Thus, the effects of this component appear to be washed out, as the indicator variables likely have both positive and negative effects on mathematics proficiency. Likewise, Knowledge-Based Digital Leisure was not validated in this study. This component only had 3 associated indicator variables, and perhaps a larger number of indicators included in the component would improve its psychometric properties. Additionally, the digital "knowledge" required for these activities, of downloading and uploading digital materials, might be considered fairly basic by other scholars (Choi et al., 2021). Perhaps better measures would include digital leisure activities that require more advanced, technical knowledge, such as designing a personal webpage or using software or programs to create digital art.

Finally, it is possible that the dependent variable in the OLS regression, student mathematics proficiency, might not be the ideal outcome variable to assess predictive validity for measures of digital capital. While knowledge and application of mathematical concepts represents an important learning outcome, this variable does not represent an institutional evaluation. Other academic outcomes, such as student GPA, might yield a more accurate examination of predictive validity in the proposed principal components. Since grades are assigned to students by institutional agents (teachers), digital capital might be differentially rewarded based on students' intersecting identities (Paino & Renzulli, 2013; Rafalow, 2020; Watkins & Cho, 2018). If other academic outcome variables were included in the PISA dataset, I might have found stronger evidence of predictive validity in the PCA.

CONCLUSION

In this study, I reconceptualized the theoretical construct of digital capital using national survey data of 15-year-old US students and identify 5 reliable principal components to guide operationalization in future research: Academic Digital Usage, Perceived Digital Autonomy, Perceived Digital Competence, Casual Digital Browsing, and Knowledge-Based Digital Leisure. Of these 5 principal components, Perceived Digital Autonomy and Casual Digital Browsing displayed evidence of predictive validity in the OLS regression on student mathematics proficiency. Although I was not able to reliably identify a principal component from the 3 indicator variables measuring material access, the OLS regression analysis revealed that home desktop computer access has a statistically significant positive effect on student mathematics proficiency.

These findings make an important contribution to the literature by providing an initial step toward empirically measuring characteristics of digital capital among students in the US. Previously, this population had not been considered in the literature on digital capital, as most research has been limited to adult populations (ages 18+). Also, since the existing literature has theorized on digital capital but lacked consistent empirical measures (Ollier-Malaterre et al., 2019; Paino & Renzulli, 2013; Ragnedda et al., 2020; Santillana et al., 2020), this project provides a clear starting point for future research to accurately and empirically study digital capital. Implications from the present findings will hopefully be used to guide meaningful, systemic change focused on digital equity in the education system.

Although this project makes a considerable contribution to the literature, the limitations of the study must be considered. The generalizability of the study might be weakened by nonresponse bias in the PISA survey. Survey items from the ICT portion of the survey contained a sizable number of missing values. To combat nonresponse bias, I include the appropriate "non-response adjusted" student weight in the analysis, as explained in the PISA Data Analysis Manual (OECD, 2009). However, it is certainly possible that the effect of nonresponse was not totally eliminated, so I consider this as a potential limitation.

Another limitation is that the current project only focused on mathematics proficiency scores as an academic outcome in the regression model. As a supplementary analysis, I substituted reading proficiency as the dependent variable in an OLS regression model, following the same statistical techniques used for the analysis on mathematics proficiency. The results from this supplementary analysis are included in Table 9 in the Appendix. Notably, the results for reading proficiency indicate that Perceived Digital Competence (which was identified as a linear effect by the Box Tidwell Test) has a statistically significant positive effect on reading proficiency (COEF = 6.06, $p < 0.05$). Additionally, I find a significant positive coefficient for smartphone access (COEF = 30.00, $p < 0.05$). This supplementary analysis reveals that the use of another dependent variable could yield different findings. Certainly, future research should examine these findings as they relate to other academic outcomes.

Moreover, upon completion of analysis and writing of results, I found that the National Center for Education Statistics (NCES) has US-specific, supplemental PISA 2018 data available for download on their website (NCES, 2021). With a deadline to meet, I was not able to include this supplemental data, which contains relevant student data on race and ethnicity, as well as a variable measuring "access to high-speed internet" (Kastberg et al., 2021). These variables would have certainly strengthened my analysis, but I was not aware that this supplementary datafile was available when I conducted analysis. Also, since the PISA survey is conducted every three years, PISA 2021 (not released at the time of this project) might provide better insights on digital capital in the era of COVID-19.

Despite these limitations, this project might prove to be particularly useful for future research. Since this study is exploratory, the present findings should be tested in different US samples and contexts. For example, future research might examine the 5 principal components proposed here using another nationally representative sample of 15-year-old students, or perhaps, with a sample of a different age group. If these findings can be replicated among other groups, CFA might then be used to evaluate the strength of these components, as *factors*, in a measurement model. Before conducting research in a confirmatory context though, future research should evaluate: 1) how objective (rather than self-reported) indicators of digital skill and autonomy could improve the model, and 2) how the proposed components might predict other, institutionally evaluated outcomes in the education system.

Finally, as the results from this study indicate, material digital access still warrants academic attention. Future research should not only continue to consider the relationship between digital access and academic inequality, but also operationalize these variables in a way that can adequately capture concerns of "access quality," including access to software applications and programs. Certainly, this project makes important progress in shifting away from the “digital divide” framework, which cannot realistically capture the complex reality of digital inequality today. The theoretical construct of digital capital, as it begins to be strengthened by empirically robust measurement, can and should be used to address the many areas in which digital inequality persists in the education system.

REFERENCES

Allison, P. D. (2001). *Missing data*. Sage publications.

Anderson, M., & Perrin, A. (2018). Nearly one-in-five teens can't always finish their homework because of the digital divide. *Pew Research Center, 26*.

Attewell, P. (2001). Comment: The first and second digital divides. *Sociology of education, 74*(3), 252-259.

Bandalos, D. L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 98-122). Routledge.

Beltran, D., Das, K., & Fairlie, R. W. (2006). Do home computers improve educational outcomes? Evidence from matched current population surveys and the national longitudinal survey of youth 1997.

Bourdieu, P. (1986). The forms of capital.(1986). *Cultural theory: An anthology, 1*, 81-93.

Byun, J., & Joung, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science and Mathematics, 118*(3-4), 113-126.

Campos-Castillo, C. (2015). Revisiting the first-level digital divide in the United States: Gender and race/ethnicity patterns, 2007–2012. *Social Science Computer Review, 33*(4), 423-439.

Choi, J. R., Straubhaar, J., Skouras, M., Park, S., Santillana, M., & Strover, S. (2021). Techno-capital: Theorizing media and information literacy through information technology capabilities. *New media & society, 23*(7), 1989-2011.

Cohron, M. (2015). The continuing digital divide in the United States. *The Serials Librarian, 69*(1), 77-86.

Cullen, R. (2001). Addressing the digital divide. *Online information review*.

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions, 8*(1), 1-9.

DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). Digital inequality: From unequal access to differentiated use. *Social inequality*, 355-400.

Fairlie, R. W. (2012). Academic achievement, technology and race: Experimental evidence. *Economics of Education Review, 31*(5), 663-679.

Fairlie, R. W., Beltran, D. O., & Das, K. K. (2010). Home computers and educational outcomes: Evidence from the NLSY97 and CPS. *Economic inquiry, 48*(3), 771-792.

Fairlie, R. W., & Robinson, J. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics, 5*(3), 211-240.

Galindo, C., & Sonnenschein, S. (2015). Decreasing the SES math achievement gap: Initial math proficiency and home learning environments. *Contemporary Educational Psychology, 43*, 25-38.

Gonzales, A. L., McCrory Calarco, J., & Lynch, T. (2020). Technology problems and student achievement gaps: A validation and extension of the technology maintenance construct. *Communication research, 47*(5), 750-770.

González-Betancor, S. M., López-Puig, A. J., & Cardenal, M. E. (2021). Digital inequality at home. The school as compensatory agent. *Computers & Education, 168*, 104195.

Graves, J., Abshire, D., Amiri, S., & Mackelprang, J. (2021). Disparities in technology and broadband internet access across rurality: Implications for health and education. *Family and Community Health*.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis (Vol. 6).

Hargittai, E. (2001). Second-level digital divide: Mapping differences in people's online skills. *arXiv preprint cs/0109068*.

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication research, 35*(5), 602-621.

Hu, J., & Yu, R. (2021). The effects of ICT-based social media on adolescents' digital reading performance: A longitudinal study of PISA 2009, PISA 2012, PISA 2015 and PISA 2018. *Computers & Education, 175*, 104342.

Jakubowski, M. (2013). PISAREG: Stata module to perform linear regression with PISA data and plausible values.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202.

Jurich, S. (2000). The information revolution and the digital divide: a review of literature. *TeckKnowLogia, 2*, 42-44.

Kaiser, H. F. (1974). An index of factorial simplicity. *psychometrika, 39*(1), 31-36.

Kaiser, H. F., & Dickman, K. W. (1959). Analytic determination of common factors. American Psychologist,

Kastberg, D., Cummings, L., Ferraro, D., & Perkins, R. (2021). *Technical Report and User Guide for the 2018 Program for International Student Assessment (PISA): Data Files and Database with U.S.-Specific Variables*. https://nces.ed.gov/pubs2021/2021011.pdf

Katz, V. S., Gonzalez, C., & Clark, K. (2017). Digital inequality and developmental trajectories of low-income, immigrant, and minority children. *Pediatrics, 140*(Supplement 2), S132-S136.

Kelley, B., & Sisneros, L. (2020). Broadband access and the digital divides. *Education Commission of the United States*.

Lai, J., & Widmar, N. O. (2021). Revisiting the digital divide in the COVID-19 era. *Applied Economic Perspectives and Policy, 43*(1), 458-464.

Lareau, A. (2011). *Unequal childhoods*. University of California Press.

Lareau, A., & Weininger, E. B. (2003). Cultural capital in educational research: A critical assessment. *Theory and society, 32*(5), 567-606.

Lenhart, A., & Horrigan, J. B. (2003). Re-visualizing the digital divide as a digital spectrum. *IT & society, 1*(5), 23-39.

McConnaughey, J., Lader, W., & Chin, R. (1998). ``Falling through the net II: new data on the digital divide'', National Telecommunications and Information Administration. *Department of Commerce, US Government, Washington, DC*.

Morris, D. S., & Morris, J. S. (2013). Digital inequality and participation in the political process: Real or imagined? *Social Science Computer Review, 31*(5), 589-600.

Mossberger, K., Kaplan, D., & Gilbert, M. A. (2008). Going online without easy access: A tale of three cities. *Journal of Urban Affairs, 30*(5), 469-488.

Mossberger, K., Tolbert, C. J., & Gilbert, M. (2006). Race, place, and information technology. *Urban Affairs Review, 41*(5), 583-620.

NCES. (2021). *Program for the International Student Assessment (PISA) 2018 Public Use File (PUF)*. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021019

NTIA. (1999). FALLING THROUGH THE NET: DEFINING THE DIGITAL DIVIDE. https://www.ntia.doc.gov/legacy/ntiahome/fttn99/contents.html

OECD. (2009). *Sample Weights*. https://doi.org/doi:https://doi.org/10.1787/9789264056251-4-en

OECD. (2015a). *PISA 2018 Technical Standards*. http://www.oecd.org/pisa/pisaproducts/PISA-2018-Technical-Standards.pdf.

OECD. (2015b). *PISA 2018 Technical Standards*. http://www.oecd.org/pisa/pisaproducts/PISA-2018-Technical-Standards.pdf

OECD. (2019a). *PISA*

OECD. (2019b). *PISA 2018 Assessment and Analytical Framework*. https://doi.org/doi:https://doi.org/10.1787/b25efab8-en

OECD. (2019c). *PISA 2018 Results (Volume I)*. https://doi.org/doi:https://doi.org/10.1787/5f07c754-en

OECD. (2019d). *PISA 2018 Technical Report*. https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018%20TecReport-Ch-04-Sample-Design.pdf

Olive, J., Makar, K., Hoyos, V., Kor, L. K., Kosheleva, O., & Sträßer, R. (2009). Mathematical knowledge and practices resulting from access to digital technologies. In *Mathematics education and technology-rethinking the terrain* (pp. 133-177). Springer.

Ollier-Malaterre, A., Jacobs, J. A., & Rothbard, N. P. (2019). Technology, work, and family: Digital cultural capital and boundary management. *Annual Review of Sociology, 45*, 425-447.

Paino, M., & Renzulli, L. A. (2013). Digital dimension of cultural capital: The (in) visible advantages for students who exhibit computer skills. *Sociology of education, 86*(2), 124-138.

Park, S. (2017). *Digital capital*. Springer.

Parker, E. B. (2000). Closing the digital divide in rural America. *Telecommunications Policy, 24*(4), 281-290.

Perrin, A., & Atske, S. (2021). *7% of Americans don't use the internet. Who are they?* PEW Research Center. Retrieved 3/21/2022 from https://pewrsr.ch/2GrhLUj

PEW. (2021). *Demographics of Internet and home broadband usage in the United States*. Pew Research Center: Internet, Science & Tech.

Puckett, C. (2020). Digital Adaptability: A New Measure for Digital Inequality Research. *Social Science Computer Review*, 0894439320926087.

Rafalow, M. H. (2020). *Digital Divisions*. University of Chicago Press.

Ragnedda, M. (2018). Conceptualizing digital capital. *Telematics and informatics, 35*(8), 2366-2375.

Ragnedda, M., Ruiu, M. L., & Addeo, F. (2020). Measuring digital capital: An empirical investigation. *New media & society, 22*(5), 793-816.

Rupp, A., Koh, K., Zumbo, B. D., Zumbo, I. B., & Ochieng, C. (2003). What is the impact on exploratory factor analysis results of a polychoric correlation matrix from LISREL/PRELIS and EQS when some respondents are not able to follow the rating scale. annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois,

Santillana, M., Sraubhaar, J., Schrubbe, A., Choi, J., & Strover, S. (2020). Digital inequalities: Homework gap and techno-capital in Austin, Texas. *First Monday*.

Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics, 34*(8), 1607-1624.

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New media & society, 6*(3), 341-362.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). pearson Boston, MA.

UCLA. (2016). *Introduction to SAS. UCLA: Academic technology services, statistical consulting group*. https://stats.oarc.ucla.edu/stata/faq/how-to-do-parallel-analysis-for-pca-or-factor-analysis-in-stata/

Watkins, S. C., & Cho, A. (2018). *The Digital Edge*. New York University Press.

Westat. (2016). *SAMPLING IN PISA*.

Wresch, W. (1996). *Disconnected: Haves and have-nots in the information age*. Rutgers University Press.

Yang, B. (2005). Factor analysis methods. *Research in organizations: Foundations and methods of inquiry*(181-199).

Yang, T., Ticona, J., & Lelkes, Y. (2021). Policing the Digital Divide: Institutional Gate-keeping & Criminalizing Digital Inclusion. *Journal of Communication, 71*(4), 572-597.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology, 9*(2), 79-94.

Zhong, Z.-J. (2011). From access to usage: The divide of self-reported digital skills among adolescents. *Computers & Education, 56*(3), 736-746.

Zillien, N., & Hargittai, E. (2009). Digital distinction: Status-specific types of internet usage. *Social Science Quarterly, 90*(2), 274-291.

APPENDIX

Table 5. Polychoric Correlation Matrix

		Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20	Var21	Var22	Var23	Var24	Var25
Var1	Use ICTs to follow up lesson	1.000																								
Var2	Use ICTs to follow up lesson	0.583	1.000																							
Var3	Email teacher	0.596	0.726	1.000																						
Var4	DL & UL from school website	0.600	0.655	0.632	1.000																					
Var5	Check school school website	0.577	0.627	0.585	0.772	1.000																				
Var6	Use learning app - computer	0.617	0.588	0.608	0.663	0.636	1.000																			
Var7	Use learning app - phone	0.605	0.601	0.584	0.649	0.627	0.862	1.000																		
Var8	Indep. Install new software	0.167	0.099	0.124	0.148	0.119	0.127	0.117	1.000																	
Var9	Read tech info to be indep.	0.198	0.153	0.142	0.169	0.159	0.164	0.167	0.734	1.000																
Var10	Uses ICTs how desired	0.140	0.022	0.111	0.062	0.049	0.085	0.084	0.575	0.562	1.000															
Var11	Indep. Solves problems ICT	0.150	0.070	0.134	0.104	0.081	0.113	0.106	0.674	0.677	0.685	1.000														
Var12	Selects apps/programs indep.	0.146	0.096	0.137	0.143	0.105	0.138	0.124	0.695	0.632	0.675	0.735	1.000													
Var13	Online chat	0.190	0.137	0.111	0.125	0.125	0.133	0.156	0.195	0.167	0.233	0.212	0.198	1.000												
Var14	Use social media	0.233	0.128	0.163	0.151	0.140	0.161	0.181	0.113	0.094	0.239	0.190	0.161	0.612	1.000											
Var15	Browse Internet for fun	0.237	0.045	0.113	0.078	0.093	0.086	0.084	0.233	0.186	0.336	0.254	0.257	0.522	0.532	1.000										
Var16	Read news on internet	0.378	0.253	0.291	0.309	0.311	0.318	0.308	0.224	0.240	0.225	0.249	0.221	0.314	0.377	0.437	1.000									
Var17	Practical info from web	0.444	0.273	0.323	0.353	0.350	0.365	0.337	0.259	0.278	0.278	0.289	0.275	0.354	0.393	0.479	0.707	1.000								
Var18	Uses unfamiliar ICTs	0.138	0.089	0.134	0.129	0.127	0.137	0.124	0.408	0.392	0.397	0.432	0.411	0.157	0.125	0.170	0.190	0.190	1.000							
Var19	Can give ICT advice-purchase	0.213	0.129	0.157	0.118	0.132	0.162	0.162	0.494	0.504	0.474	0.536	0.505	0.251	0.218	0.274	0.252	0.260	0.583	1.000						
Var20	Comfortable home ICT use	0.151	-0.056	0.073	-0.025	0.012	0.045	0.043	0.374	0.322	0.601	0.490	0.447	0.282	0.277	0.383	0.228	0.279	0.442	0.546	1.000					
Var21	Overcomes tech. difficulties	0.186	0.108	0.161	0.140	0.119	0.171	0.147	0.505	0.512	0.542	0.650	0.555	0.231	0.211	0.263	0.264	0.286	0.541	0.702	0.661	1.000				
Var22	Helps others with ICTs	0.198	0.159	0.182	0.181	0.160	0.187	0.186	0.499	0.513	0.475	0.599	0.522	0.241	0.204	0.228	0.268	0.260	0.523	0.745	0.567	0.831	1.000			
Var23	DL music, SW, games, films	0.259	0.355	0.291	0.425	0.378	0.379	0.403	0.194	0.223	0.104	0.159	0.179	0.291	0.203	0.180	0.338	0.376	0.123	0.185	-0.004	0.182	0.214	1.000		
Var24	UL created content for share	0.267	0.315	0.252	0.372	0.348	0.349	0.387	0.157	0.186	0.113	0.115	0.173	0.289	0.301	0.301	0.283	0.315	0.163	0.227	0.090	0.198	0.227	0.553	1.000	
Var25	DL applications - phone	0.277	0.243	0.215	0.290	0.267	0.282	0.294	0.259	0.266	0.240	0.260	0.263	0.414	0.387	0.462	0.425	0.482	0.216	0.321	0.218	0.303	0.322	0.528	0.551	1.000

Table 6. Initial PCA

Variable	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Communalities
Use ICTs to follow up lesson	0.355						0.675
Email Peers	0.378						0.682
Email Teacher	0.386						0.685
Check school school website	0.371						0.744
DL & UL from school website	0.367						0.692
Use learning app - computer	0.376						0.739
Use learning app - phone	0.360						0.718
Indep. Install new software		0.473					0.760
Read tech info to be indep.		0.467					0.748
Uses ICTs how desired		0.362					0.688
Indep. Solves problems ICT		0.405					0.771
Selects apps/programs indep.		0.439					0.748
Uses unfamiliar ICTs			0.417				0.555
Can give ICT advice-purchase			0.462				0.751
Comfortable home ICT use			0.372				0.705
Overcomes tech. difficulties			0.448				0.806
Helps others with ICTs			0.467				0.795
Online chat				0.439			0.589
Use social media				0.477			0.647
Browse Internet for fun				0.478			0.680
Read news on internet				0.334			0.535
Practical info from web				0.360			0.626
DL music, SW, games, films					0.449		0.680
UL created content for share					0.554		0.706
DL applications - phone					0.557		0.702
internet						0.634	0.805
homecomp						0.493	0.621
smartphone						0.569	0.720

**Only depicts component loadings* $\geq$ *.30.*

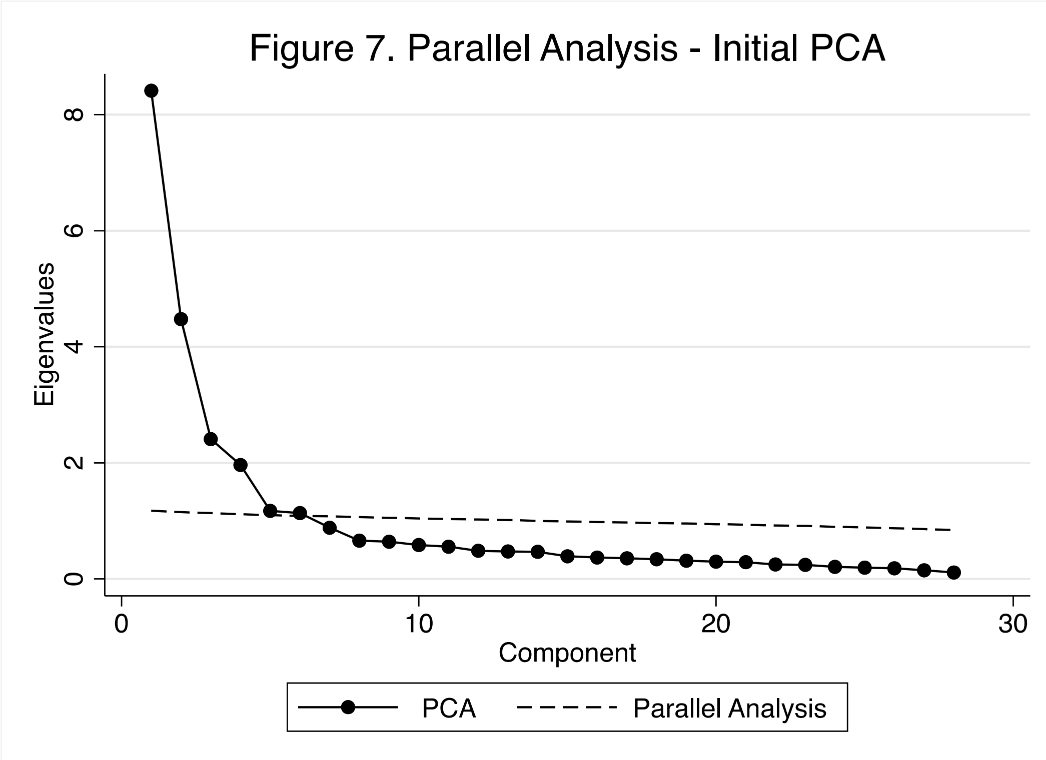


Table 7. Factor Correlation Matrix – Final PCA

	Academic Dig. Use	Perceived Dig. Autonomy	Perceived Dig. Competence	Casual Browsing	KB Dig Leisure
Academic Dig. Use	1.000				
Digital Autonomy	0.162	1.000			
Perceived Dig. Competence	0.177	0.630	1.000		
Casual Browsing	0.336	0.314	0.342	1.000	
KB Dig Leisure	0.432	0.240	0.259	0.485	1.000

Table 8. Multicollinearity Statistics

Variable	VIF	1/VIF
Perceived Digital Autonomy	1.82	0.550136
Perceived Digital Competence	1.77	0.56476
Knowledge-Based Digital Leisure	1.56	0.642577
Casual Digital Browsing	1.5	0.667067
Parental Education	1.41	0.710707
Parental Occupational Status	1.33	0.749636
Academic Digital Usage	1.32	0.759891
Home Computer Access	1.12	0.891479
2^{nd} generation immigrant	1.12	0.896377
Home Internet Access	1.11	0.898773
Student Sex	1.08	0.927048
Smartphone Access	1.07	0.938501
1st generation immigrant	1.05	0.956294
Mean VIF	1.33	

Table 9. Supplementary Analysis – OLS Regression on Reading Proficiency

Independent Variable	Coefficient	[95% Confidence Interval]
		R^2 = 0.301
		n = 3,308
Digital Capital Components		
Academic Digital Usage		
Main effect	-10.91***	[-16.67, -5.15]
Squared Effect	-10.08***	[-14.88, -5.28]
Perceived Digital Autonomy	12.91***	[6.97, 18.85]
Perceived Digital Competence	6.06*	[0.12, 12.00]
Casual Digital Browsing		
Main Effect	35.30***	[28.64, 41.96]
Squared Effect	-11.22***	[-16.92, --5.52]
Knowledge-Based Digital Leisure	-39.67***	[-44.51, -34.83]
Material Access Variables		
Home Internet Access	4.32	[-13.38, 22.02]
Smartphone Access	30.00*	[3.72, 56.28]
Home Computer Access	21.87***	[10.99, 32.75]
Control Variables		
Student Sex (Ref: Male)	19.85***	[12.44, 27.26]
Parental Occupation Status	0.88***	[0.70, 1.06]
Parental Education (Ref: K-6)		
7^{th}-9^{th} grade	-10.03	[-33.37, 13.31]
10^{th}-12^{th} grade; 1-yr. voc.	21.13*	[0.61, 41.65]
2-4 yr. voc. program	18.50	[-4.55, 41.55]
Tertiary (higher ed.)	37.05**	[13.29, 60.81]
Immigration Status (Ref: Native)		
1st-generation	-18.18*	[-33.08, -3.28]
2nd-generation	12.95**	[3.97, 21.93]
Intercept (constant)	395.23	[356.79, 433.67]

p < 0.05; **p < 0.01; ***p < 0.001

CURRICULUM VITA

Ryan James Price

EDUCATION

M.A. (May 2022)
Sociology, University of Louisville. Louisville, KY

B.A. (2019)
Sociology, University of Louisville. Louisville, KY
Cum Laude graduate

PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant in Sociology, University of Louisville
Race in the US; Race & Ethnicity
Supervising Faculty: Oliver Rollins, PhD
(August 2020 – May 2021)

Discussion Section Leader, University of Louisville
Introduction to Sociology
Supervising Faculty: Mark Austin, PhD
(August 2021 – May 2022)

Graduate Research Assistant, University of Louisville
Supervising Faculty: Melanie Gast, PhD
(August 2021 – May 2022)

RESEARCH CONFERENCE PRESENTATIONS

University of Louisville's Graduate Student Regional Research Conference
"Beyond a Divide: Reconceptualizing Digital Capital and Links to Academic Proficiency"

SERVICES TO THE UNIVERSITY & DEPARTMENT

Graduate Student Council Representative (August 2020-Present)
Faculty Awards Committee, Sociology (January 2020 – April 2020)
Graduate Representative for Sociology Department Faculty Meetings (August 2021-Present)