Rethinking Equity and Innovation in Education: A Systematic Approach to Mapping and Evaluating Innovative Educational Initiatives

By

Changhee Lee

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Educational Leadership and Policy Analysis)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 7/28/2022

The dissertation is approved by the following members of the Final Oral Committee:

Carolyn J. Kelley, Professor, Educational Leadership and Policy Analysis Xueli Wang, Professor, Professor, Educational Leadership and Policy Analysis Matthew T. Hora, Associate Professor, Educational Policy Studies Eric M. Camburn, Professor, School of Education, University of Missouri-Kansas City © Copyright by Changhee Lee 2022

All Rights Reserved

TABLE OF CONTENTS

LIST OF TABLES	III
LIST OF FIGURES	IV
DEDICATION	V
ABSTRACT	VI
CHAPTER I. INTRODUCTION	1
References	6
CHAPTER 2. WHEN SCHOOL TURNAROUND MEETS DECENTRALIZA	ATON:
STATE-LEVEL POLICY INNOVATION FOR LOW-PERFORMING SCHO	OOLS 8
INTRODUCTION	
LITERATURE REVIEW	
DATA AND METHODS	
RESULTS	
DISCUSSION	
LIMITATIONS	53
CONCLUSION	54
References	57
CHAPTER 3. A CASE OF EXPLOITATIVE INNOVATION: DECENTRAL	IZATION,
SCHOOL TURNAROUND, AND STUDENTS' ACADEMIC ACHIEVEMEN	NT 66
INTRODUCTION	66
CONCEPTUAL FRAMEWORK	68

DATA AND METHODS	
RESULTS	
DISCUSSION AND CONCLUSION	
LIMITATIONS	
References	
CHAPTER 4. A CASE OF EXPLORATIVE INNO	OVATION: A VIRTUAL INTERNSHIP
AS AN ALTERNATIVE WORK-BASED LEARN	ING SPACE TO IMPROVE EQUITY
INTRODUCTION	
LITERATURE REVIEW	
DATA AND METHODS	
RESULTS	
DISCUSSION AND CONCLUSION	
LIMITATIONS	
References	
CHAPTER 5. CONCLUSION	
References	

List of Tables

Table 1. Examples of document list for analysis	. 29
Table 2. Four domains of education function	. 31
Table 3. Codebook to Guide Data Coding	. 32
Table 4. Examples of interventions for low-performing schools by education function	. 36
Table 5. Frequency of Innovation in Florida's Interventions for Low-performing Schools Base	ed
on the Florida State Education Agency (SEA)'s and the Federal Government's Perspective	. 38
Table 6. Descriptive Statistics	. 76
Table 7. Test of mean covariates balance at the cutoff	. 86
Table 8. Estimates of effect of comprehensive intervention on student achievement (h=24)	. 92
Table 9. Sensitivity analysis on student achievement (2 SLS estimates)	. 94
Table 10. Heterogeneous effect by the type of intervention: Prevent 2 schools	. 97
Table 11. Heterogeneous effect by the type of intervention: Correct 2 and Intervene schools	. 98
Table 12. Descriptive statistics of the sample	124
Table 13. Multinomial logistic regression estimates on access to an internship	138
Table 14. Goodness-of-fit Statistics for Cluster Models of Interns	139
Table 15. Means and standard deviations of all internship quality indicators by latent profiles	144
Table 16. Distribution of virtual or in-person interns by profiles	146
Table 17. Means and relative risk ratio of predictors on latent profiles	148
Table 18. Hierarchical regression for internship academic developmental outcome	152
Table 19. Hierarchical regression for internship career developmental outcome	153
Table 20. Hierarchical regression for internship satisfaction	154
Table 21. Covariate adjusted internship outcomes by profile membership	155

List of Figures

Figure 1. Conceptual link between decentralization and policy innovation	11
Figure 2. Conceptual framework on policy innovation	23
Figure 3. Quadrant mapping of innovation in Florida's school turnaround interventions base	ed on
the perspective of Florida ED and federal government	39
Figure 4. Distribution of Students and Schools	85
Figure 5. McCrary manipulation test	85
Figure 6. Probability of receiving comprehensive intervention in by SG points	88
Figure 7. Mean student outcomes by SG points and estimated ITT discontinuity	91
Figure 8. Proportion of interns reporting professional networking opportunities by intern pro	ofile
	142
Figure 9. Latent profiles based on the internship quality indicators	143

DEDICATION

To my beloved husband and parents

ABSTRACT

Equity in education matters for all. It affects every aspect of our daily lives, from having access to a seat on a bus that brings us to school to an opportunity to learn a challenging curriculum from strong teachers there, as well as to the test scores, college degree, and soft skills that may lead to decent jobs and wellbeing after school. Many innovative policies and practices have been established to disrupt the devastating history of inequities enshrined in every stage of education. Yet, ever-changing social, economic, and environmental circumstances bring another layer of tensions and complexities to these educational efforts, often exacerbating preexisting disparities in our lives. Then, how can we change policies and practices in ways that advance equity in education, especially when everything around us comes with more challenges, uncertainties, and risks?

This three-article dissertation attempts to tackle this problem by providing conceptual research base on educational innovation and evaluating the equity potential of two innovative efforts for traditionally disadvantaged students. Specifically, my dissertation focuses on the following key questions: 1) What does innovation mean in the education field conceptually and how is innovativeness in policy operationalized empirically? 2) How does decentralization affect innovativeness of a state's equity-focused policy and students' academic outcomes? and 3) Can a technological intervention such as a virtual internship provide an equitable work-based learning space for traditionally marginalized college students with respect to its access, quality, and outcomes?

To address these questions, I draw on two cases of equity-oriented innovative initiatives—the school turnaround policy of Florida in the K-12 setting and virtual internships under Covid-19 in higher education. In Chapter 2, qualitative content analysis and quadrant

mapping were used to measure and provide visibility of the innovativeness of the school turnaround policy. The analysis revealed that the Florida turnaround model is generally positioned in the exploitation zone, suggesting a transfer of greater authority back to SEAs alone is unlikely to yield significant transformation in policy designs targeting traditionally underserved students. In addition, the quadrant mapping showed conflicting views surrounding the innovativeness of the new model—cognitive tension between the federal and state authorities. In Chapter 3, the regression discontinuity analysis finds no evidence that the Floridabrand school turnaround generated substantial differences in student test scores, probability of meeting high standards, or making gains in math and reading. Results of subgroup analyses exhibit the largest positive impacts across all outcomes for students at schools with the most intensive level of support, yet only a little or no effect for those students at schools that received less intensive support. Using a series of multinomial logistic regressions and latent profile analyses, Chapter 4 uncovered that a virtual internship is a mixed bag of progress and challenges towards the goal of creating an equitable and inclusive career development pathway for young students.

The contributions of this dissertation are threefold. First, this study improves our understanding of educational innovation through research-based conceptualization and an integrative conceptual framework that lays the groundwork for unearthing complex patterns of equity-oriented educational efforts at the system level. This study also supports theoretical grounds for conditions under which exploitative innovation and explorative innovation process arise in education policymaking and practices. Second, my findings provide empirical evidence on whether two discrete types of innovative programs in different educational environments— one in K-12 settings and the other in higher education settings—make changes in the equitable

access, quality learning opportunities, and outcomes of those who were traditionally underserved in each context. Further, this series of inquiries extends our focus on the equitable educational system to the social structures, policy contexts, and design features of policies and practices that may widen or reduce such gaps.

CHAPTER I. INTRODUCTION

As the long history of education reforms has shown us, the education community in the United States has emphasized that thinking anew and acting anew are essential to support some of the vulnerable students in the nation. Central to the theory of action underlying these efforts is the longstanding reflection that there is a lack of policies or practices that may effectively challenge the status quo; the same status quo that spawned the inequitable access, opportunities, and outcomes for historically underserved students in the first place. Even if they ever worked, profound changes in social, economic, technological, and environmental contexts make some of these past approaches no longer appropriate (Cohen et al., 2018). These changes have resulted in increasing calls for bringing about innovation, or *a novel, yet potentially beneficial approach departing from conventional practice* (Lubienski, 2004), into the education system by exploring new strategies and exploiting existing solutions across the K-20 and workforce continuum. Indeed, these efforts have led us to search for innovative approaches, from market-based reforms such as school choice and charter schools to technology-based innovations including virtual schools and connected learning to ensure all children reach their potential and success¹.

While proposals for individual innovation efforts (e.g., school choice) may abound (Finn & Steiner, 2019), this field suffers from several problems. One concern is that research on equity-oriented educational innovation is in dire need of scholarly literature base and systematic analysis to disentangle complex dynamics through which multiple actors interact with contexts to deal with evolving, interrelated problems in education (Cohen & Ball, 2007). Organizations, for example, often lose creative tensions partly due to challenges stemming from organizational

¹ The term 'reform' here is defined as a type of innovations developed by top-down, external processes. The term 'initiative' is used to indicate both policy and practice in this study.

contexts, such as a combination of limited resources, cultural heterogeneity, and multiple stakeholders with diverging interests (Evans, 1996). Equity-oriented educational reforms in such organizations are likely to come more in the form of repeating strategies that worked in the past with slight modifications, focusing on quick, yet unsustainable fixes such as recruiting new staff or closing a failing school, rather than attending to what is at the root of problems (Elmore, 1996; Myers & Smylie, 2017). Complexities also revolve around the management of knowledge and information that play a key role in the creation and diffusion of innovation, not due to the dearth or richness of knowledge available, but to the stickiness of 'local' knowledge, different sensemaking process of new knowledge, and a lack of effective interactions and capacities to translate them into concrete strategies in a particular context to address their own needs (Bryk et al., 2010). In an education arena where designers of initiatives (e.g., state or federal government or external research organizations) usually do not match up with utilizers of those initiatives (e.g., local school districts or school teachers) (Cohen & Ball, 2007), reform efforts that worked in one context may neither be effective in other contexts nor be sustained over time partly for these reasons.

Notwithstanding the multifaceted complexity of innovation surrounding the equity agenda, innovative educational efforts are more often examined with an implementation lens, with a focus on understanding the variation in their fidelity of enactment or effects of discrete innovative cases, rather than a design lens (Fishman et al., 2013; Mintrop, 2020). This stream of research is undoubtedly important as it will provide policy makers with rationales for or against a given equity-focused educational reform on the one hand, and researchers with ground for developing theory of implementation on the other. Nevertheless, the benefits of implementation studies—"what works where, when, and for whom"—centered on disaggregated innovation cases (Means & Penuel, 2005) will be limited without parallel research on initiative design that offers a systematic description on when and how such innovation emerges, spreads, and is adopted across different institutional settings and who will be involved in the change process and what their roles are (Cohen & Ball, 2007).

The absence of comprehensive research on educational innovation may lead to a reinforcement of existing inequality through less effective resource allocation for experimentation and change (Fuhrman et al., 1991). A study by Tipton et al. (2021) looking at the distribution of randomized controlled trials in the U.S. provides a good example of this claim. They found that opportunities for educational experimentation and change are often unevenly split across regions, excluding disproportionately a greater number of small rural schools, compared to large urban schools. It could be somewhat ironic to observe some of the schools that need fundamental improvement the most have limited opportunities for changes and resources, whether successful or not. In addition, how can we reasonably expect organizations to embrace innovative strategies and achieve the intended goals if they were designed to operate fundamentally in different contexts?

In this dissertation, I take up these challenges and attempt to address some of the conceptual and empirical gaps in the educational innovation literature. Innovation certainly is not a concept that can be narrowly defined into a single dimension. As such, this dissertation starts with conceptualizing innovation in educational space based on March's (1991) organizational learning framework. Then, two empirical cases of educational innovation are examined. Specifically, in the first paper, I develop a conceptual framework that defines innovation and operationalizes it as two different, yet interrelated forms—*exploitation and exploration*—by synthesizing the literature on organizational learning. After that, I develop a methodological tool

to measure the innovativeness of a given approach, followed by the quantification and presentation of descriptive contours of educational innovation based on qualitative content analysis. With this tool, I next seek to theorize a relationship between policy context and the emergence of educational innovation by looking at whether decentralization may serve as a policy lever to facilitate equity-oriented policy innovation in education. My study also sheds light on the perception gap in educational innovation by looking at a state's effort to turnaround low-performing schools both from the vantage point of federal and the state government and comparing how they differ.

Shifting a focus of the remaining two chapters from conceptualization and initiative design to implementation, I examine two individual innovative policies and practices—one in the K-12 setting and the other in the postsecondary setting—to evaluate whether these new efforts improve equitable access, opportunities, and outcomes of traditionally marginalized students in each setting. The choice of two distinctive educational innovations is informed by innovation literature that theorizes the policymaking process. Florida's new school turnaround policy in Chapter 3, a makeover of nation's low-performing schools serving concentrations of high-needs students, represents an incremental innovation case with small accommodations (i.e., exploitative innovation) triggered by the decentralization. Chapter 3 focuses on the causal impacts of Florida's school turnaround policy on students' academic achievement, taking into account the dynamics of withdrawal of strict regulatory expectations by federal government and state activism. To this end, regression discontinuity is used for full sample analysis, and a difference-in-difference approach is used for subgroup analysis.

Chapter 4 examines a virtual internship for college students. While virtual internships existed before the pandemic (Feldman, 2021), this is an example of a disruptive initiative that

emerge to accommodate environmental changes caused by the global pandemic. Specifically, I explore whether a virtual internship increases access to beneficial work-learning opportunities to college students from disadvantaged socio-economic, racial/ethnic, gender, and geographical backgrounds by using a multinomial logistic regression. The next part examines whether virtual internship programs vary by their patterns of learning and support compared to in-person internships and whether there are different developmental outcomes for underrepresented students participating in an online or in-person internship opportunity. Latent profile analysis is utilized to identify different groups of internships by quality. Multivariate regressions are used to predict the membership and its relationship with the outcomes, respectively.

In the last chapter, I conclude with summary of findings and develop more comprehensive explanations of under which conditions, which types of innovation happen, and how such significant changes may be an equalizer or reinforce the status quo.

References

- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in sociology* of education (Vol. 1, pp. 127–162). New York, NY: Springer.
- Cohen, D. K., & Ball, D. L. (2007). Educational innovation and the problem of scale. In Schneider, B. & McDonald, S. (Eds.), *Scale- up in education* (Vol. I, pp. 19-36). Lanham, MD: Rowman & Littlefield.
- Cohen, D. K., Spillane, J. P., & Peurach, D. J. (2018). The dilemmas of educational reform. *Educational Researcher*, 47(3), 204-212.
- Dahl, R. A., & Lindblom, C. E. (1953). Politics, economics, and welfare. Chicago: University of Chicago Press.
- Elmore, R. (1996). Getting to scale with good educational practice. *Harvard educational review*, *66*(1), 1-27.
- Evans, R. (1996). *The human side of change: Reform, resistance, and real-life problems of innovation.* San Francisco: Jossey-Bass.
- Feldman, E. (2021). Virtual Internships During the COVID-19 Pandemic and Beyond. *New Horizons in Adult Education and Human Resource Development, 33*(2), 46-51.
- Finn C. E. & Steiner, D. (2019, March 6). *Is education reform possible in America?*. Thomas B. Fordham Institute. <u>https://fordhaminstitute.org/national/commentary/education-reform-possible-america</u>
- Fishman, B., Penuel, W., Allen, A. R., Cheng, B., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *Teachers College Record*, 115(14), 136-156.
- Fuhrman, S., Clune, W., & Elmore, R. (1991). Research on education reform: Lessons on the implementation of policy. In A. Odden (Ed.), *Education policy implementation* (pp. 197-218). Albany: State University of New York Press.
- Lubienski, C. (2004). Charter school innovation in theory and practice: Autonomy, R&D, and curricular conformity. *Taking account of charter schools: What's happened and what's next*, 72-90.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, *2*(1), 71-87.
- Means, B., & Penuel, W. R. (2005). Research to support scaling up technology-based innovations. In C. Dede, J. Honan, & L. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 176-197). New York, NY: Jessey-Bass.
- Meyers, C. V., & Smylie, M. A. (2017). Five myths of school turnaround policy and practice. *Leadership and Policy in Schools*, *16*(3), 502-523.
- Mintrop, R. (2020). *Design-based school improvement: A practical guide for education leaders*. Harvard Education Press.
- Schneider, A., & Sydney, M. (2009). What is next for policy design and social construction theory?. *Policy Studies Journal*, *37*(1), 103–119.

Tipton, E., Spybrook, J., Fitzgerald, K. G., Wang, Q., & Davidson, C. (2021). Toward a System of Evidence for All: Current Practices and Future Opportunities in 37 Randomized Trials. *Educational Researcher*, *50*(3), 145-156.

CHAPTER 2. WHEN SCHOOL TURNAROUND MEETS DECENTRALIZATON: STATE-LEVEL POLICY INNOVATION FOR LOW-PERFORMING SCHOOLS Introduction

One of the longest standing policy debates in U.S. education history is what roles the federal versus state government play to reduce inequality in student outcomes (Hess & Eden, 2017). While state governments have long shouldered primary education responsibilities (McDonnell, 2005), researchers have recognized the increasing federal efforts to support schools serving a large group of low-achieving students from disadvantaged backgrounds since Title I (Cohen D. & Moffitt, 2010; Grissom & Herrington, 2012). At the same time, heavy criticism arose toward federal interventions, especially surrounding national school turnaround initiatives, which include a set of comprehensive interventions that aim at making quick progress to increase student achievement in underperforming schools (Duke et al., 2012; Herman et al., 2008). The main argument against these federal-led interventions is that the unique, chronic challenges of low-performing schools are hardly fixed by the generic 'one-size-fits-all' type of federal services like No Child Left Behind (NCLB) (Finn & Hess, 2004; Forte, 2010; Sunderman et al., 2005).

Indeed, there is a growing attention that innovative school reform paths are required to support some of the most vulnerable schools to achieve a dramatic boost in student performance. Recent efforts from the NCLB Waiver to the Every Student Succeeds Act (ESSA) have sought to prompt innovation for these schools through decentralization. For instance, Tennessee Senator Lamar Alexander commented when ESSA was passed, "It [ESSA] will unleash a flood of excitement and innovation." Hence, one of the key assumptions behind the authority shift is that the devolved governing units will take initiatives to experiment with new approaches, while being more accountable for their local priorities and outcomes. Such assumption has been long

recognized as an important theme in education research on charter schools (e.g., Lake, 2008; Lubienski, 2003; Preston et al., 2012; Wohlstetter et al., 1995) or school-based management (e.g., Robertson et al., 1995). Yet, little empirical analysis supports whether states with different challenges and dynamics of change would make a substantial shift in the status quo as a reaction to the decentralization. Only recently have several studies found that states have uneven tendencies when it comes to leveraging the given flexibility to redesign more reliable, innovative statewide accountability systems in a broad sense (Aldeman et al., 2017; McGuinn, 2019; Polikoff et al., 2014), which is associated with the financial capacity and political conditions of each state (Wrabel et al., 2018). While these findings contribute to the research on decentralized intergovernmental dynamics, much more remains to be understood about how the dynamic elements of decentralization function to formulate creative, statewide systems of support in a sophisticated school turnaround setting.

This paper aims to extend the understanding of the critical link between decentralization and policy innovation for educational equality by addressing unexplored avenues in existing studies—ones focused on the adoption and diffusion of specific policy innovations. This study instead considers a state's relative orientation for changes through a more elaborated conceptual framework based on March's (1991) organizational learning perspective—the exploitation of existing knowledge and exploration of new knowledge. Scholars, policymakers, and practitioners would benefit from employing such an integrated framework that not only enables them to disentangle complex, heterogeneous patterns of innovation in the state school improvement policies (Cohen D. & Ball, 2006), but also offers an opportunity to reconsider the alignment of policy rationales, policy goals, and policy instruments at the system level (McDonnell & Elmore, 1987). Using a qualitative content analysis approach, I seek to answer the following two research questions:

RQ1. How does the Florida state education agency develop policy innovation for low-performing schools when the federal regulatory pressures are withdrawn? RQ2. What does innovation look like by education functions from the perspective of the state and federal government?

To this end, a conceptual framework is developed that begins to quantify the innovativeness of Florida's school turnaround initiatives under decentralization along the temporal and cognitive dimensions. Next, the estimated results are visualized using a quadrant mapping by education functions and perspectives.

Literature review

Conceptual link between decentralization and policy innovation

The conventional wisdom about decentralization suggests this plausible theory of change: deregulation from bureaucratic mandates and external authorities opens up "legitimate" opportunities for lower levels of actors to initiate new solutions (Mintrop & Trujillo, 2005). These are solutions that are otherwise hardly ensued under the 'one-size-fits-all' type of rules mandated by non-local levels of government. However, voluminous literature in education policy, organizational science, economics, and political science suggests that such link may be mediated by several factors which will be discussed below. Although illustrating the plausible consequences of decentralization may carry with it the danger of some oversimplification, Figure 1 presents the conceptual link between decentralization and policy innovation.

Chances for policy innovation for low-performing schools

The first pathway to substantive changes is through the reduction of organizational inertia.



Figure 1. Conceptual link between decentralization and policy innovation

Generally, major changes remain hard to be achieved due to the organization's path dependency to keep the trajectory of resource investment (i.e., resource rigidity) (Christensen, 1997; Gilbert, 2005) and to allocate resources in established routines (i.e., routine rigidity) (Gilbert, 2005). Increased latitude may contribute to solving this problem by providing a momentum for an organization to make sharp departures from past centralized practices. In other words, it may trigger an organization to relax such rigidity and inertia (West, 2017; Wöbmann et al., 2007) embedded in their predisposition in terms of whether and where to allocate available resources as well as how to manage such process (Gilbert, 2005; Zhou & Wu, 2010).

A second pathway to significant change is through stronger congruence between policy and local needs (Fuhrman & Elmore, 1990; Honig & Rainey, 2012; King & Ozler, 1998; Loeb & Hough, 2016; Schoen & Fusarelli, 2008; Weiss & McGuinn, 2016). Given that actors at the lower level are generally better positioned to understand the local conditions than distant national legislators are, it is possible that they make policy more relevant to their own circumstances. Understanding the local conditions is critical for states to better support underperforming schools because they perform poorly for numerous reasons, stemming from the interaction of limited resources and conflicting needs among multi-actors, and constantly evolving contexts in which such interactions occur. Thus, blueprinted types of generic interventions cannot easily address unique tensions that each low performing school face. Insights into why each school is struggling to begin with, what one or more combinations of unique challenges are, and what worked well or poorly in the specific school environment may help states to design more attuned educational infrastructures in ways that support teaching and learning in the classroom (Meyers & Smylie, 2017).

Moreover, these routes would be further facilitated by elected officials and their staffs who are politically more accountable to the realistic preferences of local communities, pursuing appreciable changes to win reelection (Berry & Berry, 1992; Walker, 1969). With the combinative function of better access to neighborhood information and greater incentives for responsiveness, subnational units are likely to arrange and monitor the allocation of local resources to solve the problems more appropriately (Bardhan, 2002; Faguet, 2004), thereby resulting in increased efficiency. This will in turn encourage lower authorities to look beyond mere compliance and to experiment with more divergent, creative paths. Multiple concurrent policy experiments by local governments may lead to greater breakthroughs across the country (Strumpf, 2002).

State innovations could be also inspired by the contextual turbulence such as a restoration of state educational authority. One recent study, using a microeconomic modeling approach, found that the movement of administrative controls back-and-forth between the federal and state government creates conditions that promote greater efficiency of the policy experimentations than when left to states on their own devices (Callander & Harstad, 2015). Extended to the current education policy context, states that have been encumbered by federal regulations could be more motivated to enact stark changes from the mandated rules soon after the end of massive federal control like NCLB.

Risks for policy innovation for low-performing schools

In contrast, some researchers express slight skepticism, but not overall disbelief, about the promising impacts of the retrenchment of federal authority and augmentation in state leadership in school turnaround (Aldeman, 2017; Goertz, 2005; Jochim & Murphy, 2013; Polikoff et al., 2014; Sunderman et al., 2005; Shober, 2017; Weiss & McGuinn, 2016; Wrabel et al., 2018). One of these concerns is that not all the SEAs are well-prepared in terms of expert knowledge and financial resources (Childs & Russell, 2017; Cohen D. & Moffitt, 2010; Duke, 2006; Kober & Rentner, 2011; McGuinn, 2012; Sunderman & Kim, 2007; Wong, 2015) to make new investments in experimenting with alternative turnaround strategies. This lack of organizational capacity causes states to keep the status quo or make only piecemeal changes. These arguments are largely relevant to principles suggested in organizational learning (Cohen W. & Levinthal, 1990), punctuated equilibrium theory (Romanelli & Tushman, 1994), and incrementalism (Quinn, 1980) if primary actors are viewed as state government and information is viewed as any types of knowledge and resources. At the common theoretical center of these perspectives is the notion that one's decision-making may be bounded by one's existing ability to process new information acquired inward and outward (Simon, 1991). Interpreted in our context, even if the federal government devolves appreciable authority to states over crafting structures, process, and programs for schools serving minority, high-poverty students, if a state has a limited absorptive capacity to leverage the leeway (Cohen W. & Levinthal, 1990), then a state may make the least progress in overcoming the inertia to deploy the resources or manage processes that

enable resource allocation. This is because vertical power sharing in and of itself does not spawn capacities concerning how to identify fundamental causes of failure, hit on creative solutions to fix unique problems, and provide technical assistance with the lower levels.

Another line of research also presents a contradictory angle to proponents of flexibility as they find that context in general, and political dynamics in particular, take strong roles in policy formulation in which few meaningful accommodations are favored by states' decision-makers (e.g., see Lieberman & Shaw, 2000; Manna & Ryan, 2011; Mintrom, 1997; Peterson, 1995; Wong, 2015; Wrabel et al., 2018). For example, Manna and Ryan (2011) suggest that a mismatch between a state governor's political tastes and the federal administration's move may discourage state administers to embark on new reform efforts with unknown payoffs due to the absence of political coverage to advance them. Wrabel et al. (2018) show that states with a majority of citizens and political leaders with conservative foundations are significantly disinclined to adopt non-proficient super subgroups for accountability determination of schools, which deviates from the conventional school performance rating matrix grounded in the individual subgroups. More broadly, like several studies of decentralization in welfare policy reveal (e.g., see Lieberman & Shaw, 2000; Peterson & Rom, 1989), lifting prescriptive policy recipes from the federal government may not necessarily lead state government to craft new approaches to meet locally defined needs, especially in the case of school turnaround policies for their most impoverished schools. This is partly accounted for in studies of electoral securities that suggest perceived greater insecurity in reelection and "unpopular" policy to electorates are negatively associated with states' responsiveness to bringing about new thinking (Berry & Berry, 1990, p. 400; Berry & Berry, 1992, p.716). This finding implies that competitive dynamics that trigger policy innovation might be at work in a different manner in the policies for the poorest,

lowest-performing schools in the states: In general, the theory of competitive federalism predicts state politicians are encouraged to tailor-make policies to fit local needs promptly and try out fresh ideas to attract potential recipients. From this perspective, state political leaders could feel more inclined to attract taxpaying households by improving lagging schools where students from low-income families are particularly targeted for assistance (McGinnis & Somin, 2004); this approach would follow the well-established national conditions without investing in more fundamental experimentations with a risk of failure. Even if there are, political leaders are likely to make only marginal, symbolic changes in order to avoid controversies of electoral responsiveness (Lieberman & Shaw, 2000).

Empirical evidence on the link between decentralization and policy innovation

Presumably one of the most relevant policy cases for this study is the NCLB Waiver, where the federal government handed controls and responsibilities back to state education agencies to push for changes for disadvantaged students in high-needs schools (Dougherty & Weiner, 2019). Announced in 2011, more than 80 percent of states were offered discretion to make decisions on two of multiple key broad areas of the accountability system: identifying schools in greatest need of changes, and developing and implementing creative strategies to support them (Wong, 2015). Although the studies on state policy innovation under the NCLB Waiver have been less prevalent in empirical research, recent studies have examined how states navigate and manage their own accountability systems to target resources and to build school capacity compared to the federal statutory requirements (Polikoff et al., 2014; Wrabel et al., 2018). Analyzing 42 approved waiver applications, Polikoff et al. (2014) found the evidence on the effects of decentralization on state-level policy innovation is mixed, showing both substantive changes pursuing their own policy direction and continued compliance with

bureaucratic rules. For example, many states have started to incorporate non-test-based measures such as college/career readiness, attendance, or school climate into their identification strategies of lowest performing schools. These experiments are in sharp contrast to one of the most controversial NCLB provisions that laid considerable emphasis solely on test results (Trujillo & Renee, 2012). Meanwhile, authors observed that few experiments were conducted in their attempts to include additional subjects other than math and ELA, nor to account for the multiple years of performance into the growth measure. The work of Wrabel et al. (2018) provides evidence that, despite the widespread controversy of ill-advised design features of federal rules, most states' tendency to retain already-established federal routines or to embark on new initiatives were related to internal state politics and capacity for reform, which vary across states.

Another line of studies looked at state accountability plans under the ESSA where the federal government stepped back from the responsibilities to design and implement school improvement and state governments got to handle innovative learning opportunities for students. Aldeman et al. (2017), for instance, conducted a review of accountability plans in all 50 states and the District of Columbia. Although mostly having possibilities for revision until federal approval, the analysis of plans for intervening in struggling schools found that most states remained stuck in their established routine or embarked on slight modifications to their support system. For example, not a single state among 34 utilized the additional 3% reservation of their overall Title I funds for direct student services into their ESSA proposal at the time of proposal application. Similarly, McGuinn (2019) introduced critical voices from federal or state leaders regarding states' uncreative approach to schools serving mostly vulnerable students. Aldeman et al. (2017) and McGuinn (2019) both suggested that much of the gap between the ESSA goals and early findings of state practices could be partly attributed to a lack of useful information and

detailed non-regulatory guidance aligned with the requirements of the new law (Rentner et al., 2017).

Challenges of examining the link between decentralization and policy innovation

Collectively, these studies outline the conceptual link of how a more expanded state role in U.S. education may (or may not) affect policy innovation and provide mixed empirical evidence in the context of policies for supporting underperforming schools. As noted earlier, one key challenge in the prior studies is how a varying mix of organizational inertia, capacity, local knowledge, and politics would play out in a multi-layered, ever-changing reality. Another challenge stemmed from how to define and measure the extent of innovativeness of the policy in the first place. I define innovation as an organization's strategic learning process of a knowledge base. Indeed, an organization leverages extant knowledge and structures to create a new set of strategies for experimentation; such new alternatives become established routines, generating further opportunities for modification and refinement. A large proportion of current innovation studies are grounded in several specific actions and strategies diffused from one context to the next (e.g., McLendon et al., 2005), providing insights into the qualitative aspects of unprecedented ideas on its own. This line of research, in turn, calls my attention to different aspects of innovation: a state's orientation for change. The approach is consistent with the perspective of Adams (2020), who analyzed California school districts' tendency for experimentation in response to fiscal decentralization.

This approach has several strengths. One strength is placed on its comprehensiveness in scope. While studies that cover a handful of preselected, often bold policies, are insightful to examine the diffusion of innovative policies, they are constrained in part because completely novel ideas, ironically, can be hardly captured by predetermined policy categories. In contrast, an

organization's tendency to change can be captured by using a broader portfolio of policies including both emerging strategies to the whole population and those unique to their own context, thereby enabling more comprehensive understanding of the changes.

Another strength is that this enables more accurate calibration of patterns in policy innovation because this approach accounts for key characteristics of innovation, that is, innovation is a relative concept (Cohen D. & Ball, 2007; Lavie et al., 2010; Spillane et al., 2002; Vincent-Lancrin et al., 2019; Walker, 1969). As in the long history of tension surrounding regulatory law imposed by a principal, compliance by and discretion granted to the subnational agents under "United States' particular brand of federalism" (McDonnell, 2005, pp. 20-21) implies that while the federal government shapes state and local governments' priorities and implementation of federally prescribed requirements, constant interactions between them also yield state-by-state variations in paths toward meeting those goals (Grissom & Herrington, 2012). In the present policy context, even the identical school turnaround strategy could be conceived as either more 'novel' or more 'conventional' across different states than within the same state. As a result, one may feel less confident about the accuracy of gauging innovativeness through the binary scale based on the adoption and diffusion of particular programs or strategies across multiple states.

The versatile potentials of tracking innovation in the long-term are another benefit of this perspective. Because departures from conventional practices generally accompany trade-offs between appreciable investment and uncertainties to stakeholders (Bowling & Pickerill, 2013), researchers have suggested the importance of longitudinal investigation of the full innovation process. A longitudinal instrument with a focus on a state's innovation orientation enables

measurement of constructs across time, broadening our knowledge base. With this focus in mind, I develop a conceptual framework that guides this study in the next section.

Conceptual framework: Innovation as exploitative learning and explorative learning

This paper presents an analytical framework drawn from March's (1991) organizational learning, which provides a lens to characterize the evolutionary patterns of innovation in policy. March's (1991) model has been highly useful to explain a reiterative process between an organization's learning and their strategic behaviors along with an explicit consideration of the essential roles of cognition ("knowledge, understanding, and beliefs") underlying such process (Crossan et al., 1999, p. 535). Thus, the organizational learning perspective is widely used in both theoretical and empirical studies to describe the mechanism of government innovation (e.g., Berry & Berry, 1999) and to understand leaders' and teachers' behaviors relevant to improvement in school contexts (e.g., Pietsch et al., 2020; Strunk et al., 2016).

March (1991) introduced two different, yet interrelated types of activities, 'exploitation' and 'exploration,' to describe how organizations systemically acquire, interpret, integrate, and disseminate knowledge to adapt to a rapidly changing environment while remaining efficient. Exploitative learning behaviors are related to the use and refinement of a pre-existing knowledge base but are not limited to "choice, production, efficiency, selection, implementation, execution" (March, 1991, p.71). In contrast, explorative learning behaviors are related to the navigation of new possibilities, but are not limited to "variation, risk taking, experimentation, play, flexibility, discovery, innovation" (March, 1991, p.71). It draws attention to the knowledge, information, and resources that have been considered as an important conduit of continuous educational change for student learning (Cohen D. & Ball, 2007; Farrell & Coburn, 2018; Honig, 2008; McCharen et al., 2011; Peurach & Glazer, 2007). Education research has stipulated ample

evidence that organizations, especially in the context of schools or school districts, refine the existing knowledge and routines with known payoffs within their boundaries, making incremental adjustments (e.g., see Fitzgerald, 2000; Triant, 2001). They also search for novel knowledge and non-routines with uncertainties across their boundaries, as well as assimilating significant changes.

Some researchers (e.g., Li et al., 2008) argue that an organization's exploitative innovation and explorative innovation are distinguished according to the distance between the knowledge base in which each type of innovations is grounded. However, questions remain: how do we systemically measure the distance between the conventional knowledge and new knowledge? Studies in the field of strategic management suggest three dimensions of knowledge search for this question: temporal, spatial, and cognitive dimensions. In the present study, a temporal dimension and a cognitive dimension were applied, which considers whether or not changes were made between before and after the decentralization, and, of those changes, whether they are novel or familiar to state or federal government, compared to the existing interventions, respectively. I focus on a temporal dimension (i.e., whether or not changes are made) and a cognitive dimension (e.g., of those changes, whether they are novel or familiar compared to the existing interventions) in a state's new school turnaround interventions to compare those changes established before and after decentralization.

Temporal distance considers the history of a knowledge base as a reference for the creation of new knowledge (Nerkar, 2003). Indeed, some organizations may elicit new knowledge and practices from a temporal locality, that is, by scanning the near, recent experience and knowledge and combining them with the current knowledge (i.e., temporal exploitation). Alternatively, other organizations may challenge this approach, casting light on the

remote, older knowledge inputs and discovering new ways of interpreting them in the current context (i.e., temporal exploration). The notion of temporal dimension has been used in many forms to evaluate innovation in education. For example, an OECD research team defined what amounts to be innovation at the classroom or school levels: the greater shifts in teaching practices and resources in a given domain that students are exposed to between at least two different time points, the greater the levels of innovation are taking place (Vincent-Lancrin et al., 2019).

New state policy approaches warrant neither success nor failure on its own (McLendon et al., 2005). Therefore, assessing policy innovation solely on the temporal dimension may be insufficient to bring insights into whether such changes are meaningful to a state's ongoing effort to potentially benefit schools at risk. Cognitive distance, which refers to the shared perceptional familiarity in terms of the content between new knowledge and extant knowledge (Cramton, 2001), is another key dimension to understand the characteristics of innovation that organizations seek. This view leads us to expect that exploration of promising possibilities may occur more often across organizations with distinct levels of professional knowledge (Rosenkopf & Nerkar, 2001) or institutional knowledge (Zukin & DiMaggio, 1990). In contrast, exploitation of familiar perspectives may occur more often across organizations with homogeneous knowledge base. Such perspective can be further extended to explain the fluidity of knowledge in policies and routines not only across organizations in different professional fields or relations but also between subunits at different levels within a given organization (Farrell & Coburn, 2018; Farrell et al., 2019). Some researchers (e.g., Ahuja & Lampert, 2001) argue that 'new to the world' can be cognitively more exploratory than 'new to the organization,' itself. It should be noted that, however, greater cognitive distance may not necessarily yield greater learning potentials of the

organizations (Fleming & Sorenson, 2001) because the search for new information that facilitates experimentation is built upon on the preexisting knowledge base of each organization which also accounts for their similarities.

While temporal and cognitive distance tells us much about the different aspects of knowledge search, they are not mutually exclusive of each other in this study context. Thus, retaining essential logic, I operationalized the degree of exploitation/exploration along the temporal and cognitive dimensions as follows. If a state's new approach after decentralization maintains the status quo, the temporal distance between the new and preexisting approaches is considered to be extremely proximate. Conversely, if a state's new approach departs from the status quo, this is considered to be temporally distant. To this, I add the cognitive dimension. That is, among those with temporal changes, if a state's new approach uses earlier programs or strategies with slight modifications already familiar to a given state or federal government, respectively, the cognitive aspect of new initiatives is thought to be local. In contrast, if a state's new approach is unknown, compared to the existing strategies of a given state or federal government, respectively, this is mapped onto the distant knowledge search. Given the stickiness of the information (Von Hippel, 1994), the same principle is applied to the cases where the knowledge is transferred across different organizations, professional fields, and subunits at different levels.

Consistent with Lavie et al. (2010), exploitative innovation and explorative innovation in this study are ranged along a single continuum with two bipolar ends, rather than as two independent modes. Such a dualistic perspective is justified due to not only an organization's transitive dependencies between exploitation and exploration but also to the relative nature of these two concepts. Indeed, a group or organization may constantly switch between exploitative

22

learning and explorative learning over time to adapt to environmental turbulence. This shifting approach leverages preexisting knowledge and structures to create a new set of strategies for experimentation; such new alternatives become established routines, generating further opportunities for modification and refinement. These shifting patterns are better characterized and understood by using "in the form of zero-sum game" along the continuum (Gupta et al., 2006, p. 695). Thus, the relative position on this continuum between the exploitation and exploration may indicate the extent to which an organization's design, strategy, and behavior rest on the combination of both present knowledge and new knowledge with varying weights. Figure 2. Conceptual framework on policy innovation



With few exceptions, most of individuals and organizations have sought to find optimal combinations of exploitation and exploration. At the same time, the contradictory nature of these approaches creates a dilemma surrounding the allocation of limited resources (Levinthal & March, 1993). For example, they may strategically tend towards the existing routines with predictable returns because adaptation helps them to find solutions using relatively fewer resources. Early wins and high-efficiency gains through exploitation may lead them to focus more on exploitation, resulting in a "success trap" where proximate success gradually reduce

their capabilities and incentives to explore fresh insights, repeating the same recipes. Conversely, they may tend less towards past routines and knowledge because alternative strategies help them to gain strengths in a rapidly changing environment. While such resource-intensive, futureoriented arrangements may bring in failure rather than a successful outcome in practice, this provides organizations with positive reinforcement for further exploration, resulting in a "failure trap" where failure drives them to look for newer alternatives with a greater risk of failure.

Despite the inherent paradox of exploitation and exploration, an organization, as Murphy and Meyers (2007) suggest, can benefit more from being ambidextrous—manage interwoven tensions by pursuing both of them simultaneously, yet in separate organizational units, domains, or timelines. Strategic domain separation has been observed in the studies of education innovation. For instance, charter schools, often viewed as "laboratories" (Lubienski, 2004, p.72), reported higher engagement in the new practices in the administrative areas of governance and staffing policies such as the adoption of merit pay or no tenure protection (Podgursky & Ballou, 2001; Preston et al., 2011), or strategies to facilitate traditionally underrepresented parents involvement (Smith et al., 2011), whereas less experiment in class-level instructional support (Preston et al., 2011), compared to traditional neighborhood schools. For this reason, I examine the degree of exploitation-exploration in the four subcategories of education function (OECD, 2018), which will be discussed in the following section.

In addition, conflicting interpretation on the cognitive innovation by the federal and a state government provides active support for assessing the policy innovation through two separate lenses. Traditionally, concern over a free ride on innovation by others led a state government to focus more on the strategies that would work within their own context, rather than strategies that would work for both themselves and other states (Callandar & Harstad, 2015;

Strumpf, 2002). Therefore, it is more likely that a state determines the degree of novelty in a new approach within its own context by comparing their own pre- and post-interventions. In contrast, the federal government may prefer significantly different approaches, that is, new to the country, not just new to a few states only. Thus, when the federal government assesses the cognitive aspects of innovation, it may compare a given state's new approach with existing federal mandates as well as already prevalent approaches in other states before the decentralization was introduced. In this regard, it would be reasonable to examine patterns of exploitation and exploration in multiple domains from the perspective of a state and the federal government, respectively.

Data and Methods

The present study arose out of an attempt to address the prior-mentioned literature gap by examining a case of Florida under decentralization. This case study proceeded in three steps. First, as I discussed earlier, I developed a conceptual framework based on March's (1991) organizational learning to guide the analysis of a state's new support system. Second, employing a qualitative content analysis approach, I assessed the temporal and cognitive aspects of each item of the new system, both across and within four education functions, from the perspective of the federal and state government, respectively. Finally, the estimates were mapped on a fourquadrant and then characterized in light of the exploitation-exploration framework.

Analytic sample: Florida's differentiated accountability pilot program

This study drew on a case of a waiver program, the differentiated accountability pilot, that was implemented in Florida. The program was launched by Margaret Spellings, who served as President Bush's secretary of Department of Education (USED) from 2005 to 2009. Amid the controversy surrounding the prescriptive school turnaround approaches under NCLB, the Secretary promised additional flexibility to encourage states and local leaders to develop and experiment with more nuanced accountability systems; these systems were for identifying schools and districts in most urgent need of improvement and promoting creative interventions to help students achieve success (Doan, 2008). The Secretary emphasized policy innovation as one of the goals of this plan, which was well stated in her letter to Chief State School Officers (USED, 2008a), "This plan, Building on Results, is designed to improve accountability by providing additional flexibility for innovation while working toward grade-level proficiency for all students in reading and math by 2014." To fulfill this promise, the Bush administration initiated a differentiated accountability pilot program in 2008.

Florida's case provides a unique opportunity toward extending our understanding of the impacts of decentralization on policy innovation. The flexibility pilot neither promised additional funding nor compromised their continued focus on equity and accountability in return for the flexibility (USED, 2008a)—NCLB's 100% proficiency goals and annual assessment requirements. While the pilot program clarified a few conditions for the authorization process, it did not attach significant regulatory strings such as the adoption of teacher evaluation and common core standards in exchange for the flexibility unlike other federal policy initiatives imposed (e.g., NCLB waiver, Race to the Top, and School Improvement Grant). To retain existing arrangements except decentralization enables better identification of its consequences as a policy instrument. Altogether, Florida's case with a history of a strong state accountability system can be used for developing typologies of state policy innovation under decentralization.

Data collection

Data for this study consist of publicly available policy documents on school turnaround programs and relevant research issued from the pre-decentralization period (i.e., 2005-2007;
except for NCLB provisions of 2001) and post-decentralization period (e.g., 2008-2010). The study periods for pre-decentralization and post-decentralization were determined based on two considerations. First, the interstate variances in the timing of designing restructured practices may have affected the agent's perception on the novelty of the programs. Some states (i.e., first-generation accountability systems) launched corrective actions and restructuring plans for low-performing schools earlier (i.e., 2005) than other states. These other states initiated the staged school improvement interventions after they were mandated (Mintrop & Trujillo, 2005). Therefore, school turnaround approaches of the pioneer group may have been more familiar to the federal government when it transferred greater control back to SEAs. Second, two years of post-decentralization documents were reviewed for analysis because policy agents frequently modify their initial reform plans or change implementation timeline (McGuinn, 2012).

Of these, forty-two policy documents and studies were identified as fulfilling the predefined inclusion criteria, where the temporal and cognitive dimensions of the innovation were taken into account. For example, key documents pertaining to Florida's school turnarounds issued from 2005-2010 and school improvement provisions of NCLB of 2001 were included to examine whether or not there were substantial changes in Florida's new school turnaround system (post-decentralization) compared to their old system as well as federal mandates (pre-decentralization). Further, policy documents and studies on the school turnaround plans in California, Georgia, Maryland, Michigan, and Ohio issued from 2005 to 2007 were included using the identical selection criteria to assess the cognitive dimension of innovation viewed from the standpoint of the federal government (e.g., whether or not Florida's new school turnaround system are new to the local or to the country). This group of states was chosen not only because they are geographically representative samples, but also because they were already implementing

school turnaround plans within their own contexts before Florida launched new plans (Scott, 2008). This suggests that the federal government was more likely to perceive these states' strategies as less cognitively innovative than those of others. Florida's guidance for implementation serves as a primary source of outcomes of post-decentralization.

The types of documents reviewed include legislation, technical assistance reports, guidelines, and proposals published by federal or state governments as well as research institutions. These were sourced from the websites of USED and each state. Specific search terms entered in each website include: "differentiated accountability," "school improvement AND low performing schools", "school turnaround", "restructuring", "corrective action", "school improvement", and "Supplemental Instructional Services." In addition, relevant research briefs and technical reports published by research institutions were also reviewed to pursue a more comprehensive collection of data and to get a better understanding of the context in which the support systems were situated. Among these, the relevant passages within the documents were identified and coded through a detailed review of all documents at the preparation stage. A sample list of the analyzed documents is presented in Table 1.

Data Analysis

A qualitative content analysis based on a predetermined conceptual framework (Hall & Steiner, 2019; Hsieh & Shannon, 2005) was employed to assess the effects of decentralization on policy innovation for school turnaround. Content analysis based on systematic coding and categorization through the appraisal of texts (Bowen, 2009) is a powerful analytical approach to "examin[e] the patterns and trends in documents" (Stemler, 2000, p. 1s). This approach also helps to examine the alignment between the objectives of the programs or interventions and the exercised practices (Stemler, 2000). The results of content analysis were visualized using

Table 1. Examples of document list for analysis

Florida	Florida's Differentiated Accountability Model –Guidance for Implementation 2008-09 School Year— (2008)					
	Florida's Differentiated Accountability Pilot Proposal (2008a)					
	2005-2010 Florida Statutes K-20 Education Code					
	- Chapter 1001, 1003, 1004, 1006, 1008, 1012					
	- 2005-2010 Florida Administrative Rules 6A. State Board of Education					
	Chapter 6A-1, 6A-3, 6A-4, 6A-5, 6A-8, 6A-9, 6A-11, 6A-15,					
	Technical Assistance Paper NCLB school restructuring (2007)					
	Technical Assistance Paper Providing Title I Services in Schoolwide and					
	Targeted Assistance Programs (2007)					
	Florida's Differentiated Accountability Plan -Frequently Asked Questions-					
	(2008)					
	The Florida Problem Solving/Response to Intervention Project Year 1 Evaluation					
	Report (2007)					
Federal law	Public Law print of PL 107-110, the No Child Left Behind Act of 2001					
	- Title I (Part A, B, D, E, F), Title II (Part A, B, D), Title IX (Part A)					
Michigan,	Managing more than a thousand remodeling projects -school restructuring in					
Ohio,	California (Center on Education Policy, 2008)					
Georgia,	Making mid-course corrections -school restructuring in Maryland (Center on					
California,	Education Policy, 2007)					
and	Hope but No Miracle Cures: Michigan's Early Restructuring Lessons (Center on					
Maryland	Education Policy, 2005) and many more.					

quadrant mapping, a tool to represent the extent of innovation and characteristics of the new turnaround efforts induced from decentralization in light of the conceptual framework. The fourquadrant is represented by a function of a point with the ordered pair where the horizontal axis indicates a percentage of innovation in the temporal dimension and the vertical axis indicates one in the cognitive dimension. A greater percentage in each of temporal and cognitive dimensions denotes a higher exploration orientation of a given state (i.e., lower exploitation), whereas a smaller percentage denotes a higher exploitation orientation of a given state (i.e., lower exploration).

Four subcategories on education function were drawn from the domains of decisionmaking defined by OECD (2018). These four areas in Table 2 were selected because they cover broad areas of decision-making in education and allow for capturing the innovativeness across different level of government. I further customized these areas to reflect the evolving state roles more comprehensively in education reform: 1) teaching and learning; 2) personnel management; 3) planning and monitoring; and 4) resource management. Teaching and learning encompass interventions on the improvement of instructional systems such as professional development, curriculum alignment with state standards, and the use of continuous improvement model focused on the formative or summative assessments and Response to Intervention (RtI). The domain of personnel management concerns "hiring and dismissal of staff, duties and conditions for staff, and fixing salary levels" (OECD, 2018, p. 413). Planning and monitoring are closely linked to the traditional roles of SEAs such as the development of improvement plans and monitoring of the implementation in terms of input (i.e., the use of resources) and output (i.e., achievement). For example, to fund transportation costs for those who participate in the school choice and to expand the supplemental educational service fall into the domain of resource management.

It should be noted that not all the new approaches should necessarily indicate improved strategies (McLendon et al., 2005; Vincent-Lancrin et al., 2019). However, nearly all the approaches were relevant to the ones found in the studies of effective schools to a certain degree. At this stage, parts of text passages that were not directly related to these functions or were redundant were removed. The unit of analysis is a text passage, and each was sorted into one of four subset categories. It is assumed that a state may strategically respond to a given policy flexibility window by undertaking more exploration in certain domains of education function at one point in the process, while enacting more exploitation in the other domains at another point.

As noted earlier, two primary coding schemes were constructed a priori based on temporal and cognitive dimensions of March's (1991) organizational learning framework. The

Education function	Description			
Teaching and	Interventions directly related to teachers' pedagogical approach and			
learning	student's learning such as curriculum alignment with state standards, and			
	the use of the continuous improvement model focused on the formative			
	or summative assessments and Response to Intervention (RtI)			
Personnel	Interventions relevant to hiring and dismissal of staff, duties and			
management	conditions for staff, and fixing salary levels			
Resource	Interventions relevant to allocating and managing supplementary			
management	resources for educators and students (e.g., professional development,			
	transportation)			
Planning and	The development of improvement plans and monitoring of the			
monitoring	implementation in terms of input (i.e., the use of resources) and output			
-	(i.e., achievement)			

Table 2. Four domains of education function

Note. These categories were drawn from the domains of decision-making defined by OECD (2018) and customized by the author to more comprehensively reflect the evolving state roles in education reform.

temporal distance between pre- and post-decentralization policy was initially conceptualized as whether or not changes were made in school turnaround systems compared to those changes established before decentralization. The cognitive distance was operationalized as the extent to which the new intervention is cognitively familiar or distant with existing interventions (Li, Vanhaverbeke, & Schoenmakers, 2008; Lubienski, 2004). An example of codebook and description is shown in Table 3.

After the codebook was refined, the analysis of text passages was conducted in four-stage operations. First, each provision listed in the new Florida's new school turnaround strategies was evaluated along the temporal and cognitive dimensions. Florida's differentiated accountability guideline provides an important resource to look into how Florida as a policy implementer conceptualizes the innovation because it presents a list of specific interventions that Florida ED view as changed or existing interventions and further clarified the sources of the statute from which the existing intervention was drawn. In a review of Florida's depiction of their new system, a given post-decentralization provision was coded as either status quo or transformation, depending on the extent to which it is (or not) a continuation of the existing strategies.

Category	Code label	Description	Example
Temporal		Whether or not changes were	A greater proportion of
dimension		made in the state systems of	interventions in 'No change'
(Nerkar, 2003)		support after decentralization	indicates 'temporal
		compared to those established	exploitation;' a greater
		before decentralization	proportion of interventions in
			'transformation' indicates
			'temporal exploration.'
	Transformati	Whether a full change is made	"School is organized into
	on	with the discovery of ideas	professional learning
		after the decentralization.	communities (PLCs) aligned
			with district focused delivery
			model (Florida DOE, 2008c,
			p. 72). this action item was
			any other interventions of the
			five states
	Status quo	Whether an existing approach	Choice with transportation
	1	is maintained without change	and Supplemental
		after the decentralization.	Educational Services (ESEA
			Sec. 1116(b)(9-10), PL 107-
			110).
Cognitive		The extent to which the new	A greater proportion of
dimension		approach is (un)familiar with	'Familiar' interventions
(Li,		the existing systems of support	indicates 'cognitive
Vanhaverbeke,		from the perspective of an	exploitation;' a greater
&		entity of analysis.	proportion of 'Novel'
Schoenmakers,			interventions indicates
2008; Lubionalsi	Noval	Whather a new anneach is	"Togehow in torgeted
2004	Novel	whether a new approach is	subgroups are organized into
2004)		compared to existing	Lesson Study groups "
		approaches. Whether a new	(Florida DOF 2008c p 72):
		approach is observed in extant	this action item is unknown
		local/state/federal approaches	to the federal government.
	Familiar	Whether a new approach	this is familiar to Florida FD
		extends existing	because this has been
		state/local/federal	included in the Florida
		mandates/approaches in other	Response to Intervention
		states.	Model and Bureau of School
			Improvement has
			implemented this since 2007.

Table 3. Codebook to Guide Data Coding

Further, the cognitive aspect was assessed to understand the patterns in which different sources are layered to establish state-level policies. Among the list of interventions reported as changed, a given strategy is coded as either novel if unknown, or as familiar if known compared to existing Florida's school improvement strategies (cognitive aspect).

Second, once all the provisions in Florida's guidelines were coded, the percentage of provisions that the Florida perceived as innovation was calculated and visualized in a fourquadrant. The estimated proportion was tabulated to capture the heterogeneous patterns of innovation within each educational function and was aggregated to indicate an overall Florida's perception of innovation. Each pair of parameters were mapped on the quadrant where X value represents the extent of innovations along the temporal dimension and Y value represents the extent of Florida ED (see Figure 2).

Third, all the priori school turnaround statements of Florida state/local, five states, or federal government were re-read and categorized for further analysis if they were relevant to changes in Florida's school turnaround interventions. Then, the interventions that Florida perceived as being changed were compared against these statements line by line and coded to understand the extent to which decentralization resulted in creative changes compliant with the expectancy of the federal government. If a pre-decentralization statement from five states and the federal government is distinctive from any of Florida's post-decentralization interventions, its cognitive aspect is assessed as 'novel,' implying a non-conventional idea to the federal government. In contrast, if a given statement is identified as marginal changes of existing interventions of each source, it was coded as 'familiar.'

In the final step of the analysis, the percentage of innovative provisions evaluated from the federal government perspective was calculated by each education function and mean percentages were plotted in the other four-quadrant as described in the second step. Comparing the location of plots enables me to intuitively understand the extent of innovation that the decentralization gives (or not) rise to from a policymaker's angle (i.e., federal government) and a policy implementer's angle (i.e., a given state government) and their emphasis in achieving policy goals.

Results

Before discussing the findings for the first research question, I briefly summarize Florida's new system to support low-performing schools under decentralization.

The most distinctive feature of Florida's new model was "a consolidation of federal and state accountability systems" to identify and support schools with poor performance (Florida ED, 2008a, p.1). For example, Florida defined at-risk schools based on the combination of four different measures with varying levels of thresholds: 1) the percentage of AYP criteria met, 2) the number of years identified as Schools In Need of Improvement (SINI) status, 3) Title I receipt status, and 4) A-F school grades. Of those, the first three indicators including AYP had been mandated by the federal NCLB act and the last indicator, an A-F school grade system which largely relied on the achievement score in Florida Comprehensive Assessment Test (FCAT), had been issued by the state before the devolution of federal oversight. Both measures heavily relied on proficiency rates, rather than on growth. While AYP was carried over to the new state system because the federal government mandated the state to retain the measure to ensure all students as well as subgroups of students meet state academic content standards, the other three indicators were chosen at the state's discretion. Based on the above proficiency

measures, Florida groups schools into three categories: 1) No intervention schools, 2) Targeted support and intervene group, and 3) Comprehensive support and intervene group. Comprehensive support and intervene group was our main interest and thus it was further categorized into three subgroups: Prevent schools consist of schools that met less than 80% of AYP criteria and/or D or F grades for more than two consecutive years; among those, schools that missed AYP for more than 5 consecutive years were classified as Correct schools; and schools with repeating F grades with 65% or more of non-proficient students in reading or math were identified as Intervene schools in need of most intensive interventions.

While support and requirements for low-performing schools contained some common components, schools in each classification were provided with an escalating intensity level of interventions as their performance declined. In addition, the roles and responsibilities of SEA and local school districts were increased in more challenging schools. Some of those services such as data collection and monitoring, professional development, and direct support/established network were delivered through the regional support system where regional leaders, school improvement facilitators, and connecting partners coordinate reform. The federal government, however, did neither promise additional funding nor compromise their continued focus on equity and accountability in return for the flexibility—NCLB's 100% proficiency goals and annual assessment requirements.

As shown in Table 4, 36.84% of all interventions in Florida's new school turnaround plan were centered around personnel management, followed by planning and monitoring (26.32%), resource management (22.81%), and teaching and learning (14.04%). Interventions in personnel management mostly focused on staffing struggling schools with high-quality principals, leadership teams, teachers, and staff members with a demonstrated record of success in a similar

	Examples of interventions	Frequency
Teaching and learning	 Implementation of evidence-based curriculum programs aligned to the Next Generation Sunshine State Standards. Use of diagnostic assessments for students not demonstrating progress in reading intervention. Implementation of Response to Intervention (RtI) with an integrated data collection/assessment system to inform decisions at each tier of service delivery. Implementation of (student-level) progress monitoring in tested core content areas twice per year. 	8 (14.04)
Personnel management	 Assignment of a principal who is experienced with a clear record of increasing student achievement and overall school performance, in a similar school setting. All teachers must be highly qualified and certified in-field. Assignment of a professional partner (to each principal) with experience in improving student achievement in the subgroups or a similar school setting. (In)voluntary transfer of high-performing teachers from high-performing schools to the low-performing school when any student subgroup that has continued to fail to make AYP for the past three years has a decreasing performance. Inclusion of student achievement goals specifically targeting subgroup(s) not making AYP in principal's appraisal. Assignment of Reading, Math/Science coaches to the school. 	21 (36.84)
Planning and monitoring	 Incorporation of a systematic problem-solving process in the School Improvement Plans. Assignment of a Community Assessment Team (CAT) to review school performance data and determine the cause for low performance. Plan for equitable access to technology resources and technology integration professional development for school personnel in the School Improvement Plan. 	15 (26.32)
Resource management	 All protocol standards and sufficient human resources for delivery, follow-up, and evaluation of PD. Direct real-time access to student achievement data through a computer-based analytical system. Inclusion of professional development (PD) targeting the subgroup(s) not making AYP in the Principal's individual leadership development plan. Organization of Lesson Study groups for teachers in targeted subgroups. 	13 (22.81)
T (1		57(100)

Table 4. Examples of interventions for low-performing schools by education function

Total Note. percentage in parenthesis. setting. The planning and monitoring section outlined a set of key required components along with priority goals, which addressed: 1) a four-step problem-solving model; 2) specific action items targeting subgroups not making AYP; 3) monitoring of progress on school improvement goals through student achievement data. Interventions relevant to the resource management consisted of two types of support: direct support for students through the expansion of supplementary educational services and indirect support for students through building the capacity of leaders and teachers (e.g., job-embedded professional development). Interventions with a teaching and learning focus included a broad range of strategies such as screening, diagnostic, and progress assessment at the student-level, and the establishment of an integrated data system that enables educators to track and assess students' progress over time. *The temporal aspect of innovation in Florida's school turnaround policies after the decentralization*

Evaluated on the temporal dimension, overall, Florida has explored a considerable number of alternative school turnaround strategies in the wake of decentralization. While 35.09% of Florida's school turnaround interventions remained unchanged, 64.91% of them, be it a small amendment of the existing strategies or an original approach, contained components that differed from the previous effort. This seems to hold from the vantage point of Florida or the federal government (see Table 5).

Figures 3-1 and 3-2 show the percentage of temporal changes made in Florida's turnaround plan by four education functions. The first value of each ordered pair corresponds to percentage changes in those before and after the decentralization. A majority of personnel management arrangements (85.71%) were redesigned, which include: 1) the principal is assigned to a professional partner with experience in increasing student achievement; 2) school is

	Cognitive dimension					
Temporal dimension	Florida	USED	Florida	USED	Florida	
	ED	USED	ED	USED	ED	USED
	No	No	Similar	Similar	Novel	Novel
	change	change				
Status quo	20 (35.09)	20 (35.09)	-	-	-	-
Transformation	-	-	28 (49.12)	21 (36.84)	-	-
	-	-	-	-	9 (15.79)	16 (28.07)
Total	20 (35.09)		37 (64.91)			

Table 5. Frequency of Innovation in Florida's Interventions for Low-performing Schools Based on the Florida State Education Agency (SEA)'s and the Federal Government's Perspective

Note. percentage in parenthesis.

fully staffed on the first day of school; 3) school does not have a higher percentage of out-offield nor first-year teachers than the district average or Model Title I school average. Changes (61.54%) were also made in the domain of resource management. Supplemental educational services, for instance, were expanded to schools in need of improvement in year 1. Resource management was planned for building a computer-based analytical system that enabled educators to have direct access to students' academic progress in real-time. Similarly, changes were underway in over half of the interventions (53.33%) in the domain of planning, monitoring, and structures. For example, Florida ED proposed more frequent analysis of progress monitoring in tested core-content areas at the school level, and intervention schools were requested to generate more concrete plans to address equitable access to technology resources. While relatively small compared to temporal changes in previously mentioned domains, 37.15% of interventions in teaching and learning were accompanied by changes. The implementation of evidence-based curriculum programs aligned to the state standards is one of such temporal changes.

Altogether, considering a large proportion of temporal changes (64.91%) and relative position of them mostly mapped on the right hand of the horizontal axis (Figure 3-1 and 3-2), temporal patterns of innovation in Florida's policy reforms for low-performing schools is skewed

more toward the exploration, rather than exploitation, except for teaching and learning. In other words, decentralization drove the state more toward temporal exploration in the administrative support spaces.

Figure 3. Quadrant mapping of innovation in Florida's school turnaround interventions based on the perspective of Florida ED and federal government



Note. The quadrant is represented by a function of a point with the ordered pair where the horizontal axis indicates a percentage of innovation in the temporal dimension and the vertical axis indicates one in the cognitive dimension. For example, 'Overall (65, 24)' in Figure 3.1 indicates overall, changes were found in 65% of interventions of Florida's differentiated accountability plans and 35% (=100-65) of interventions were carried over from Florida's existing school turnaround approaches. Among these, 24% of them were unknown, novel interventions to Florida, whereas 76% (=100-24) of interventions with changes were similar interventions with preexisting school turnaround interventions of Florida that were implemented before decentralization.

The cognitive aspect of innovation in Florida's school turnaround policies after the

decentralization

The cognitive aspect of these changes revealed both common and conflicting patterns of innovation between Florida ED and the federal government. Of 37 temporal policy changes for high-needs schools, 54.05% (n=20) of them were identified as essentially similar to predecessor strategies, whether looked at from a state or federal angle. Florida did bring some universally

new and different thinking (24.32%, n=9), emphasizing, for instance, the statewide systematic performance appraisal of principals and other leadership team members based on student achievement goals, staffing school with qualified student support service personnel with demonstrated success, and transfer of high-performing teachers from high-performing schools to persistently low-performing schools. However, such unprecedented solutions for both state and federal governments were pronounced more in the narrow arena of human resource management, less in other domains.

Some of the contrasting features of innovation on the novelty between the state and federal government were echoed visually by the comparison of the positions of each point on the quadrant, particularly on the vertical axis (Figure 3-1 and 3-2). To illustrate, the largest gap in the extent to which new decisions were considered as truly "novel" laid in how to realign resources for professional development for teachers and extra education services for students: all of which were found to be exploitation of existing strategies and resources in Florida (0%; (62, 0)); half of the temporal changes could be considered to be an exploration of new ventures for the federal government (50%; (62, 50)). These contradictions can be explained by the contextdependent nature of 'innovativeness.' For example, the core elements of Lesson Study and professional learning communities, which refer to job-embedded professional development through collaborative inquiry, have been introduced and set as standards under Florida's professional development system stipulated in Section 1012.98–Florida School Community Professional Development Act since 1997 (Florida H.R. Schools & Learning Council, 2008). The evaluation team of the professional development system found the increasing need of helping districts and schools to adopt these alternative forms of school-based professional development (Bergquist, 2006) and to build a system in place through providing common

planning time and job-embedded professional development within the master in-service plan (Bergquist, 2006; Bergquist, 2011). When given the flexibility to craft its own turnaround system, Florida ED revisited the ideas of lesson study and professional learning communities to address instructional challenges facing the struggling schools and translate them into concrete action plans. Florida was a pioneer to implement the lesson study as a statewide initiative (Akiba & Wilkinson, 2016) and it had not been commonplace in the five other states. Thus, from the perspective of the federal government, it might have been thought an of as unorthodox approach. Such conflicting views between the national (50%; (86, 50)) and subnational governments (61.11%; (86, 61)) were also observed in the realm of personnel management. Turning around low-performing schools requires shared effort to identify high-priority problems and accomplish organizational changes. As such, building strong leadership teams to manage improvement efforts has been a commonly stated goal by other states (e.g., Georgia) prior to the decentralization, suggesting the federal government may consider it a similar approach. However, little notion on the launch of a leadership team was found in Florida's school past turnaround engagement.

Moreover, the adoption of the problem-solving model as a school improvement system is another point of potential opposition between the state and the federal government over the cognitive feature of the changes taken in the realm of planning and monitoring. Initially, the principles of the problem-solving model outlined in the Individuals with Disabilities Education Act (IDEA) of 2004 were sought to ensure students with learning disabilities could be challenged to learn high academic standards within general education classrooms (Prasse, 2006). However, Florida's problem-solving model goes beyond the early identification and remediation in special education. Florida's problem-solving model can be understood as a multi-tiered, systematic school improvement process by which the entire school is involved in setting goals, identifying problems, developing and implementing research-based interventions to address defined problems, and evaluating the effectiveness of the interventions. The model has been in place since the spring of 2006 in conjunction with other statewide initiatives such as the Response to Intervention (RtI) project and continuous improvement model. But, Florida's engagement in building infrastructure and capacity to support high level learning dates back to 1991 (Batsche et al., 2007). These findings suggest that the structured problem-solving model should be seen as a 'novel' approach from the viewpoint of the federal government, whereas the state government should consider it to be an evolution of long-time effort in which intensive resource allocation for building infrastructure and local capacity for scale-up implementation, and communication of all stakeholders has resulted. Teaching and learning was the only domain that no variation was found in the level of novelty within- and across-state; however, they all are essentially modifications or recombinations of prior knowledge and changes established under NCLB such as evidence-based curriculum programs and progress monitoring.

In summary, 43.24% (65, 43) of changes made after the decentralization stand on the ground with relatively untapped knowledge and resources, evaluated from the perspective of the federal government. In other words, the remaining 56.76% of such changes contained a blend of new ideas and NCLB mandates or preexisting strategies found in school turnaround approaches of the five states. Viewed from the vantage point of Florida ED, 24.32 % (65, 24) of interventions with temporal changes were composed of relatively novel experimentations that were not included in their older school turnaround plans and the remaining 75.68% were carried-over strategies with a slight modification of either federal mandates under NCLB or state extant practices.

Discussion

The policy community has emphasized that thinking and acting anew are essential to school improvement. To this end, arguably, low-performing schools serving a large population of disadvantaged students could be the ones in the greatest need of new, or even disruptive changes to break its perennial cycle of underperformance. While the federal government has taken great initiatives to improve these schools, some policymakers claim that subnational units, including state governments, may be better positioned to press innovative approaches. Building on prior literature that advocated for the diffusion of innovative approaches, this paper presents new empirical evidence on two types of state's innovation orientation—exploitative learning and explorative learning—for such schools in response to the decentralization.

Florida's orientation for exploitative learning broadly echoes some earlier findings of states' tendency to rely on established knowledge and routines (Aldeman et al., 2017; McGuinn, 2019; Polikoff et al., 2014; Wrabel et al., 2018). At a first glance, decentralization seems to have led Florida to broadly explore paths different from the status quo: such exploration occurred more extensively in personnel management, yet less in the interventions directly relevant to teaching and learning. However, a closer look at the school turnaround initiatives enacted after decentralization by Florida, five other states, and the federal government unmasked different stories. A vast array of these changes in Florida's school turnaround plan included approaches already in use under NCLB or state accountability system with slight corrections or additions. Such a pattern was preponderated especially in the areas directly related to teaching and learning—standards-based curriculum development, progress monitoring, and assessment of students' learning.

Given the growing backlash from states against the 'recipe book' type of NCLB reform models (Cohen D. & Moffitt, 2010; Goertz, 2005; Sunderman & Kim, 2007), it is surprising that Florida used its hard-earned flexibility to adhere to or recombine old school improvement strategies available under the pre-decentralization era. Further, these findings raise a more troubling question: Why did Florida favor explorative learning in the management of school principals, teachers, and staff? Why did they choose exploitative learning in the instructional efforts such as curriculum development, progress monitoring, and assessment?

While I can only conjecture, one plausible explanation may involve the chronic capacity challenge of most SEAs as previous studies have suggested (e.g., Cohen D. & Moffit, 2010; Duke, 2012; Harris, 2012; Jochim & Murphy, 2013; Manna, 2010; McGuinn, 2012; Peurach & Neumerski, 2015; Shelly, 2008; Tanenbaum et al., 2015; VanGronigen & Meyers, 2019). Certainly, improving schools that perennially fall behind is, by its nature, challenging work that requires "intensive, continuing assistance from capable agents that work closely with schools for years" across curriculum planning, leadership development, teacher education, and assessment (Cohen D. & Moffit, 2010, p. 177). While setting and communicating high expectations, and the management of instructional approaches is generally within school leaders and teachers' control, a design of supportive learning environment by allocating financial, human, and technical resources is within the control of external actors, such as states, districts, and third-party service providers (Mintrop, 2008). As such, SEA's insufficient organizational capacity—"the interplay of resources and knowledge within SEAs and the strategies applied to improving low-achieving schools" (Childs & Russell, 2017, p. 244)-may impede SEAs to invest in new and meaningful ways to identify schools with the greatest needs of support, provide districts and schools with tailored technical assistance, and monitor the delivery of reform efforts. Indeed, many SEAs

have touted chronic capacity constraints in the arena of "infrastructure," "professional resources," and "political resources" even prior to the decentralization, in the present context, under NCLB (Le Floch et al., 2008, pp. 3-5), which has resulted in the limited success in accurately carrying out mandated educational reform agenda of NCLB (Dahill-Brown & Lavery, 2012; Goertz, 2005; Rhim et al., 2007; Sunderman & Kim, 2007). Without a parallel commitment to strengthening the capacity of the SEAs, affording them legal authorities to craft their own path alone hardly lead SEAs to make sharp departures from conventional strategies for poorly performing schools.

Another alternative explanation is that a state's educational interests and priorities may not align with the bold educational changes that federal policymakers desired to accomplish through decentralization (Grissom & Herrington, 2012). In the present study context, misalignment between policy intention on the federal policy maker's end and priorities set by the policy implementer's end is observed in the program application process. For example, Secretary of Education, Margaret Spellings, once stated the goal of the decentralization pilot as "tak[ing] dramatic action to improve our lowest-performing schools" (USED, 2008c) and set priority on the proposals with "innovative models of differentiation and interventions." (USED, 2008d) Meanwhile, as Commissioner of Florida ED, Eric Smith wrote in the proposal that Florida appeared to consider this pilot as a legitimate opportunity to "harmonize differences in the state and federal requirements." (Florida ED, 2008b) Similar movements were observed in some of the other states' proposals such as Georgia and Maryland where separate statewide accountability systems had been established before NCLB and retained after NCLB. States with layered accountability systems have suffered from the confusion because dualistic systems send mixed signals to the education community about districts and schools' performance and needed

improvements (Linn, 2005; Sunderman & Kim, 2004). Hence, Florida ED may have been more motivated to leverage the flexibility afforded by decentralization to slowly streamline the existing approaches for better delivery of services to vulnerable groups of students, maintaining coherent reform efforts, rather than seeking a fundamentally novel approach. That being said, it can be argued that Florida's attempts to develop a way to merge existing dual accountability systems into a single one are bold experiments to a certain extent. However, I argue that simply consolidating components of state and federal systems while keeping them running separately may be insufficient to be regarded as 'exploration' because it is unlikely to prompt substantial changes that build on the strengths of each system.

Above all, a state's orientation toward refining and extending school turnaround strategies that exist may be the results of interaction of a group of factors, many of which are often found to be in tension in practice. These factors include the political dynamics surrounding the nature, rationales, and goals of the policy instrument, and capacity constraints of SEAs layered upon the pre-existing policy and state context. For instance, consider the literature that suggests that accountability to the people is an important trigger for new policy experimentation. In the context of school turnaround, that mechanism might not be as straightforward as it seems. Rather, accountability to the people may be the same force that encourages the state officials to search for new approaches that eventually lead them to stay in the arena of established routines. Florida has been under ongoing accountability pressure for achieving an early boost in student performance since the administrative decentralization did not exempt them from the NCLB's ambitious goal of all the students reaching proficiency level in reading and math by 2014. Despite increasing demands and expectations on quick wins in a decentralized world, in exchange for more freedom and discretion, states could no longer use powerful 'political cover'

from the federal government making harsh choices that targeted schools serving high numbers of poor and minority students (Hess & Petrilli, 2005; Loveless, 2007). Moreover, turning around most-in-need schools has been the area wherein states have seldom reported success at providing them with appropriate technical aid due to limited knowledge and skills about how to improve these schools (Cohen D. & Moffitt, 2010; Meyers & Smylie, 2017; Peck & Reitzug, 2014). Thus, states could have been more motivated to free-ride the benefits of innovations provided by other states rather than risking costly experiments and disrupting a status quo that spawned the underperformance by themselves (Callandar & Harstad, 2015; Strumpf, 2002). The combination of such tensions and trade-offs may have posed new challenges, leading state officials to strategically and readily respond to the demanding accountability pressure on the performance by refining existing systems for greater efficiency within a short time rather than of involvement in experiments with unknown payoffs.

Likewise, the interaction of multiple factors may partially account for why exploitative learning is more dominant in some educational functions than others. The proximate, foreseeable returns from the exploitative learning with lower risk may appear to be a more attractive option for state officials under increasing accountability pressure on students' performance. Yet, there still exists accountability pressure on the input and process that may remind them of the continuous push for discovering new opportunities for these schools. Elected officials including superintendents and local school boards could be also conscious of the significance of continuous signaling to electoral constituencies that they are making changes (Hess, 1999). However, the quality of such services is neither easily monitored by nor communicated to citizens even in a decentralized system (Keefer & Khemani, 2005). Therefore, it is a common observation that officials selectively translate these responsibilities into policy, presumably by sending signals of exploration in the arenas immediately apparent to the people, whereas making only marginal differences to the current system in the arenas with less visibility. Amid the weighty accountability for the process with fundamental constraints on the capacity, Florida state officials might have been motivated to send out signals about candidate quality and commitment to public services by showing more sweeping changes in highly visible ways such as identifying, replacing, and retaining some school leaders and staff. On the other hand, neither changes to the standards-based curriculum, progress monitoring, use of assessment data, and classroom teaching nor returns for such changes may be easily understood and measurable except in the long term. This may partially explain why we could see only a few feedback actions in teaching and learning over the other areas. Indeed, the political power of both state and federal officeholders is as likely to be used to undermine the explorative potentials of federalism as to facilitate them. Nevertheless, many of these speculations need to be examined further in a future study.

Perhaps the most notable finding of my study is the evidence that can indicate the gaps in the characteristics of states' orientation for changes as a function of decentralization depending on whether viewed from the perspective of federal or state government. These incongruences could be attributed in part to the differential cognitive process in which each agent has been involved throughout the decentralization. The cognitive process in policy enactment and implementation entails the "perception, interpretation, processing, storage, and retrieval of information as a basis for action" (Slembeck, p. 230) which are shaped by their own "prior knowledge, beliefs, and experiences" and "context." (Spillane et al., 2002, pp. 388-389) Innovation as a policy goal is an abstract term compared to many other specific policy issues in education (Spillane et al., 2002) including a transition to charter schools and an offer of extra services for tutoring, thereby allowing a myriad of interpretations that leave room for an agent's cognition to speak for what can be defined as innovation.

Using different labels on a given school turnaround approach could originate from at least two sources of the cognitive process. First, each policy agent perceives the novelty (or similarity) of incoming changes relative to their own reference (Spillane et al., 2002; Walker, 1969). The U.S. federal education system, for instance, defines an innovative solution according to whether a school turnaround strategy is cognitively "new to the nation" and not just "new to a given state or district." This may be unsurprising because decentralization in the school turnaround context was fueled to renew the emphasis on local control as much as to find better ways for all students across the country to equally have a chance to learn. Secretary Spellings also clearly held values about the informative benefits of decentralization for the country without risk to the rest of the country, saying "I hope to gain valuable information about SES from these pilot programs—information that can be shared with other states and districts to help them improve the quality of these activities" (cited in Doan, 2008, p. 220) In contrast, a state government's view on innovation is more tightly linked to its own jurisdiction boundary rather than to other neighboring jurisdictions or the country as a whole (Berry & Berry, 1999). In other words, whether the approach has been adopted by a majority of states for quite some time may matter more to the federal government than the state; to the extent that the approach is dissimilar to a given state and local context, it will be seen as a "novel discovery" with potentials to help schools.

In addition to a different point of reference in intergovernmental relations, there could be another source of the gap in judgments between the principal-agent relationship. An asymmetry of information between the federal and state governments can be the very condition that opens spaces for differential characterization of a given new school turnaround approach. As the literature on the intergovernmental relationship suggests (i.e., Fuhrman & Elmore, 1990), state and local governments as agents are generally in a better position to understand their school conditions and their needs on the ground. Moreover, they may have a superior understanding of what has been tried, what worked, and what did not in supporting their schools than the federal government as a principal. Hence, unless explicit attempts are made to fill the asymmetrical information gap, the federal government as a principal is subject to make decisions based on information that lacks breadth and depth (Laffont & Martimort, 2002). Further, it could also experience systematic constraints in its ability to design policy instruments that drive changes in a state's approach to help low-performing schools and monitor the implementation of such efforts, yielding inconsistent evaluation of the effectiveness of the instruments.

One example of information asymmetry includes the intervention that supports the professional development of staff through active learning communities. After decentralization, these team-based, job-embedded professional development arrangements have evolved into a coordinated statewide initiative to connect the curriculum and instructional delivery centered on identified needs in all poorly performing schools in Florida. While they were launched before the decentralization, records show poor implementation at the school level (Bergquist, 2006; Bergquist, 2011; Florida H.R. Schools & Learning Council, 2008). Florida ED used its newfound latitude to change to identify practices in high need of improvement in light of their own historical and indigenous knowledge of practices and current policy goals and to integrate them with assistance from regional school support teams for successfully embedding learning opportunities within the structure and culture of struggling schools. Nevertheless, it would be

insufficient to argue that this is a substantive shift away from existing practices in the eyes of Florida ED.

On the national level, however, it is unlikely that the federal government has ready access to details about Florida's past initiatives designed to establish job-embedded, collaborative learning structures, key actors in this process and their new roles, and challenges of implementing and sustaining them in the low performing schools. The federal government also documented in its guidance on the peer review process on the state proposal that the most current information from the previous school year was considered for the federal government's decision making for the pilot to assess the feasibility of the proposed programs (USED, 2008b). The use of incomplete data may result in, as in the case of the above, the federal government views the adoption of the professional learning community as a niche approach that it expects to achieve through decentralization. This is probably because not all the local interventions would not be found in their temporally recent available database—conventional macro-level school turnaround approaches under NCLB and five other states, whereas it comes, in fact, closer to leveraging discretion to reorganize established programs with potentials of significant growth in the fidelity of implementation.

In sum, while the information asymmetry between federal and state governments provides rationales for the delegation of authorities to states in urgent need of more innovative statewide support systems, this could, in turn, generate cognitive friction to USED's understanding of the state's practices underway. This could further undermine USED's efforts to seek creative solutions to provide all students with the education they deserve in combination with the lack of incentives and capacities for state governments to explore, experiment, and disclose new ideas with high cost and uncertainties, on the one hand, the paucity of effective tools and incentives for USED to collect and utilize detailed information for the oversight of turnaround initiatives, on the other hand.

In line with Spillane et al. (2002), I do not claim that the policy intention of the federal government should be fulfilled by each state government in a normative sense. This study is far from suggesting a state's explorative learning should be the sole ideal outcome of decentralization. Neither do I argue that explorative learning is more effective than exploitative learning in reducing educational inequality. Florida's school turnaround case does, however, generate a discussion on the unduly optimistic premise that a fundamentally reoriented school turnaround approach will be sparked solely by the withdrawal of national prescriptive regulation. In a decentralized education world, the prevalence of exploitative learning in policies pursuing equality—albeit temporally diverged, yet largely a coherent extension of its predecessors—may be attributed to much more complex dynamics, not all related to the mismatch among capacity, preferences, politics, incentives, and policy goals. Florida's case offers an additional perspective that the cognitive tension embedded in intergovernmental relations brings challenges to reach those goals, which has been overlooked in the literature and policy conversation. A lack of consideration to build a common knowledge base about the 'innovation' and the resulting misalignment in measurement may widen the gaps in the perspective on actual policy learning for educational equality, perpetuating the broken link between the rhetoric of decentralization and reality.

Moreover, Florida's uneven path to support low-performing schools raises the red flag for scholars and practitioners. Florida ED embarked on new attempts in personnel management while staying on established routines in areas directly related to the teaching and learning of under-served students. While it is argued that this type of reform may fuel the process of breaking the inertia of poor performance, it is unclear whether this will lead to improvement that sustains in the absence of temporary intensive support (Meyers & Smylie, 2017). Given that the switch between exploitation and exploration across different domains does not occur in a vacuum (Levinthal & March, 1993), it is worth tracking whether such initial efforts are translated into experimentation in teaching and learning in the long term.

Limitations

In any case, my study is only a first step and has some limitations. First, while all the changes are given equal weight in the current estimation, the magnitude of salience or 'boldness' of each change may well vary in practice. For example, a replacement of a principal and staff in low-performing schools alone can be a highly disruptive change for other states or local school districts, even if all other approaches remained unchanged. This may not pose a fatal problem to my analysis because Florida's school restructuring options under NCLB have been largely aligned with federal strategies in terms of principal and teacher replacement (Florida ED, 2007) and retained with more granular conditions about the replacement qualification after decentralization. That said, the analysis may be improved by combining quality measures of each approach and qualitative investigations. Another limitation is that while my study was focused more on identifying characteristics of changes in the emerging school turnaround approaches, policies that made the complete disappearance after decentralization were not covered in the calculation (Adams, 2020). Finally, the study is limited in its generalizability of the findings based on a case study.

At the same time, the enhancement of conceptual understanding concerning policy innovation opens avenues for further research. For instance, while I examined the short-term policy changes right after decentralization as if the learning process were stable, it may be interesting to consider the evolutionary transition of state' policy orientation between exploitation and exploration to improve schools with low-performing students by using longitudinal data. A longitudinal study would be especially meaningful, as the pursuit of exploitation and exploration involve inherent trade-offs among efficiency, uncertainties, and future gains, and self-reinforcing (Levinthal & March, 1993), and both the federal and state government should always make strategic choices for successfully turning around schools with limited resources. In that sense, identifying if and how the transition is driven by different types of policy instruments and factors will be a topic for other subsequent research to achieve educational equality, along with its impact on student performance. Building on an in-depth case study, it will be also important to develop a typology of all 50 state-level policy innovations specific to those support schools that need the most help.

Conclusion

The complicated dynamics surrounding decentralization and policy innovation on school turnaround highlight the value of the conceptual framework for systematically navigating the transition. In this work, I developed an integrative conceptual framework that draws from the scholarship of organizational learning perspectives and policy innovation. Although the framework is relatively crude, the education community can benefit from it in three ways.

First, the framework broadens our conceptual understanding of policy innovation that lays the groundwork for more advanced research. Within the framework, I conceptualize the innovation as changes in policies along the temporal and cognitive dimensions. The changes are further characterized as explorative learning—the sharp departure from the pre-existing knowledge base—and exploitative learning—the refinement of already-known ideas and routines (March, 1991). Although I do not argue this framework should be a single universal conceptual scheme, building a shared understanding of the definition and nature of policy innovation may be an essential stepping stone before larger policy discussions consider whether states take advantage of decentralization for policy innovation. Once constructed, it may help us to gain insights into the patterns of policy learning in each educational function, considered separately, and across them, considered together, as well as the trajectories in the short-term and long-term across multi-level actors. It takes on more importance, given that, despite the voluminous study on the scale-up of innovation at the school or local district level (Cohen D. & Ball, 2007) and increasing roles of SEA (Childs & Russell, 2017; McDonnell, 2005), there has been only a small body of work that looks at the patterns of innovation in education at the state level.

Second, the conceptual framework contributes to extending empirical literature regarding the relationship between decentralization and policy innovation for educational equality. Indeed, the empirical evidence suggests that, despite the universal calls for change to the school improvement requirements in NCLB, a state's overall policy learning after decentralization follows more strongly oriented toward the evolutionary path, building on its current improvement strategies or ones mandated under NCLB. While further reflecting on each education function, it became clear that a more explorative learning pattern was pervasive in the personnel management arena, and a more exploitative learning pattern was found in the arenas directly relevant to teaching and learning. Findings in this study furthermore demonstrate that the policy learning process warrants reconsideration within a new cognitive frame in addition to the conventional analytical frame focused on the cross-sectional and macro-level approach.

The third contribution is the most practical—a generic 'toolbox' that helps the policy community to understand and monitor a portfolio of state turnaround efforts at the system level. As the policy instrument to achieve equality has shifted from the sanction-based regulation to

negotiation-based deregulation (Duff & Wohlstetter, 2019), so must how federal and state governments communicate with each other. The conceptual framework can be a useful tool for both governmental officials as explorative and exploitative learning in multiple areas are quantified and visualized intuitively on the four-quadrant. While the diffusion of specific innovative interventions across states is beyond the scope of this study, it may enhance the visibility of policy coherence by adding an extra axis to indicate the diffusion patterns of specific innovative ideas. Altogether, the framework may provide policymakers at all levels with collaborative learning opportunities to monitor and build common understandings about the cumulative work for at-risk schools, facilitating the planning of evidence-based resource allocation for experimentations within their own contexts.

References

- Adams, B. E. (2020). Decentralization and Policy Experimentation in Education: the Consequences of Enhancing Local Autonomy in California. *Publius: The Journal of Federalism*, 50(1), 30-54.
- Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, 22(6-7), 521-543.
- Akiba, M., & Wilkinson, B. (2016). Adopting an international innovation for teacher professional development: State and district approaches to lesson study in Florida. *Journal* of *Teacher Education*, 67(1), 74-93.
- Aldeman, C., Hyslop, A., Marchitello, M., Schiess, J. O., & Pennington, K. (2017). An independent review of ESSA state plans. *Sudbury, MA: Bellwether Education Partners*.
- Bardhan, P. (2002). Decentralization of governance and development. *Journal of Economic perspectives*, *16*(4), 185-205.
- Batsche, G. M., Curtis, M. J., Dorman, C., Castillo, J. M., & Porter, L. J. (2007). The Florida problem-solving/response to intervention model: Implementing a statewide initiative. In *Handbook of response to intervention* (pp. 378-395). Springer, Boston, MA.
- Bergquist, C. C. (2006). *Encouraging systemic changes in professional development: a short summary of Florida's evaluation protocol system*. Unpublished report prepared for Florida Department of Education. <u>http://www.fldoe.org/core/fileparse.php/5636/urlt/0072420-</u> <u>firstcycle-short.pdf</u>
- Bergquist, C. C. (2011). Professional development system evaluation protocol: second cycle technical report school years 2006-2009. Unpublished report prepared for Florida Department of Education. http://www.fldoe.org/core/fileparse.php/5636/urlt/0072417crossdistrict2ndcyclefullreport.pdf
- Berry, F. S., & Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review*, 395-415.
- Berry, F. S., & Berry, W. D. (1992). Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science*, 715-742.
- Berry, F. S., & Berry, W. D. (1999). Innovation and diffusion models in policy research. *Theories of the policy process*, 169, 5.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2), 27.
- Bowling, C. J., & Pickerill, J. M. (2013). Fragmented federalism: The state of American federalism 2012–13. *Publius: The Journal of Federalism*, 43(3), 315-346.
- Callander, S., & Harstad, B. (2015). Experimentation in federal systems. *The Quarterly Journal* of Economics, 130(2), 951-1002.
- Childs, J., & Russell, J. L. (2017). Improving low-achieving schools: Building state capacity to support school improvement through Race to the Top. *Urban Education*, *52*(2), 236-266.

- Christensen, C. M. (1997). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.
- Cohen, D. K., & Ball, D. L. (2007). Educational innovation and the problem of scale. *Scale up in education: Ideas in principle*, *1*, 19-36.
- Cohen, D. K., & Moffitt, S. L. (2010). *The ordeal of equality: Did federal regulation fix the schools?*. Harvard University Press.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.
- Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science*, *12*(3), 346-371.
- Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. *Academy of management review*, 24(3), 522-537.
- Dahill-Brown, S. E., & Lavery, L. (2012). Implementing federal policy: Confronting state capacity and will. *Politics & Policy*, 40(4), 557-592.
- Doan, K. P. (2008). No Child Left Behind waivers: A lesson in federal flexibility or regulatory failure. *Admin. L. Rev.*, 60, 211.
- Dougherty, S. M., & Weiner, J. M. (2019). The Rhode to turnaround: The impact of waivers to No Child Left Behind on school performance. *Educational Policy*, *33*(4), 555-586.
- Duff, M., & Wohlstetter, P. (2019). Negotiating intergovernmental relations under ESSA. *Educational Researcher*, 48(5), 296-308.
- Duke, D. L. (2006). Keys to Sustaining Successful School Turnarounds. *ERS Spectrum*, 24(4), 21-35.
- Duke, D. L. (2012). Tinkering and turnarounds: Understanding the contemporary campaign to improve low-performing schools. *Journal of Education for Students Placed at Risk* (*JESPAR*), 17(1-2), 9-24.
- Duke, D. L., Carr, M., & Sterrett, W. (2012). *The school improvement planning handbook: Getting focused for turnaround and transition*. R&L Education.
- Faguet, J. P. (2004). Does decentralization increase government responsiveness to local needs?: Evidence from Bolivia. *Journal of public economics*, 88(3-4), 867-893.
- Farrell, C. C., & Coburn, C. E. (2017). Absorptive capacity: A conceptual framework for understanding district central office learning. *Journal of Educational Change*, 18(2), 135-159.
- Farrell, C. C., Coburn, C. E., & Chong, S. (2019). Under what conditions do school districts learn from external partners? The role of absorptive capacity. *American Educational Research Journal*, 56(3), 955-994.
- Finn, C. E., & Hess, F. M. (2004). On leaving no child behind. Public Interest, 35-56.
- Fitzgerald, J. (2000). 1998–99 Colorado charter schools evaluation study: The characteristics, status and performance record of Colorado charter schools. Denver, CO: Colorado Department of Education.

- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, *30*(7), 1019-1039.
- Florida Department of Education. (2007, June 5). *Technical Assistance Paper: NCLB School Restructuring*. Retrieved from https://info.fldoe.org/docushare/dsweb/Get/Document-4454/k12-07-81memo.pdf
- Florida Department of Education. (2008a, May 2). *Differentiated Accountability Pilot Program: Florida's Proposal.* Retrieved from https://www2.ed.gov/admins/lead/account/differentiatedaccountability/flplan.doc
- Florida Department of Education. (2008b, May 2). Cover letter of Differentiated Accountability Pilot Program: Florida's Proposal. Retrieved from https://www2.ed.gov/admins/lead/account/differentiatedaccountability/fldap.pdf
- Florida Department of Education. (2008c). *Florida's differentiated accountability model: guidance for implementation 2008-09 school year.*
- Florida H. R. Schools & Learning Council. (2008). Teacher professional development programs in Florida interim project report. http://www.fldoe.org/core/fileparse.php/5636/urlt/0072410-professdevreport08.pdf
- Forte, E. (2010). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist*, 45(2), 76-88.
- Fuhrman, S. H., & Elmore, R. F. (1990). Understanding local control in the wake of state education reform. *Educational evaluation and policy analysis*, *12*(1), 82-96.
- Gilbert, C. G. (2005). Unbundling the structure of inertia: Resource versus routine rigidity. *Academy of management journal*, 48(5), 741-763.
- Goertz, M. E. (2005). Implementing the no child left behind act: Challenges for the states. *Peabody Journal of Education*, 80(2), 73-89.
- Grissom, J. A., & Herrington, C. D. (2012). Struggling for coherence and control: The new politics of intergovernmental relations in education. *Educational Policy*, 26(1), 3-14.
- Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of management journal*, *49*(4), 693-706.
- Hall, D. M., & Steiner, R. (2019). Insect pollinator conservation policy innovations at subnational levels: Lessons for lawmakers. *Environmental science & policy*, 93, 118-128.
- Harris, D. M. (2012). Postscript: Urban schools, accountability, and equity: Insights regarding NCLB and reform. *Education and Urban Society*, 44(2), 203-210.
- Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). Turning Around Chronically Low-Performing Schools. IES Practice Guide. NCEE 2008-4020. National Center for Education Evaluation and Regional Assistance.
- Hess, F. M. (1999). *Spinning Wheels: The Politics of Urban School Reform*. Washington, D.C., The Brookings Institution.
- Hess, F. M., & Eden, M. (Eds.). (2017). *The Every Student Succeeds Act (ESSA): What it means for schools, systems, and states.* Harvard Education Press.

- Hess, F. M., & Petrilli, M. J. (2005). The Politics of No Child Left Behind: Will the Coalition Hold?. *Journal of Education*, 185(3), 13-25.
- Honig, M. I., & Rainey, L. R. (2012). Autonomy and school improvement: What do we know and where do we go from here?. *Educational Policy*, *26*(3), 465-495.
- Honig, M. I. (2008). District central offices as learning organizations: How sociocultural and organizational learning theories elaborate district central office administrators' participation in teaching and learning improvement efforts. *American Journal of Education*, 114(4), 627-664.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.
- Jochim, A., & Murphy, P. (2013). The capacity challenge: What it takes for state education agencies to support school improvement. *Center on Reinventing Public Education*.
- Keefer, P., & Khemani, S. (2005). Democracy, public expenditures, and the poor: understanding political incentives for providing public services. *The World Bank Research Observer*, 20(1), 1-27.
- King, E., & Ozler, B. (1998). What's decentralization got to do with learning? The case of Nicaragua's school autonomy reform. *Development Economics Research Group, Working Paper Series on Impact Evaluation of Education Reforms*, 9, 278200-1099079877269.
- Kober, N., & Rentner, D. S. (2011). Common Core State Standards: Progress and Challenges in School Districts' Implementation. *Center on Education Policy*.
- Laffont, J. J., & Martimort, D. (2002). *The theory of incentives: The principal-agent model*. Princeton, NJ: Princeton University Press.
- Lake, R. J. (2008). In the eye of the beholder: Charter schools and innovation. *Journal of School Choice*, 2(2), 115-127.
- Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *Academy of Management annals*, *4*(1), 109-155.
- Lee, C. (2019, March 21-23). *The Effects of State-driven Comprehensive Interventions on Student Achievement in Florida* [Poster presentation]. Annual Meeting of the Associations for Education Finance and Policy, March, Kansas City, MO, United States.
- Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209-231.
- Le Floch, K. C., Boyle, A., & Therriault, S. B. (2008). Help Wanted: State Capacity for School Improvement. AIR [R] Research Brief. *American Institutes for Research*.
- Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic management journal*, *14*(S2), 95-112.
- Lieberman, R. C., & Shaw, G. M. (2000). Looking inward, looking outward: The politics of state welfare innovation under devolution. *Political Research Quarterly*, *53*(2), 215-240.
- Linn, R. L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *education policy analysis archives*, *13*, 33.

- Li, Y., Vanhaverbeke, W., & Schoenmakers, W. (2008). Exploration and exploitation in innovation: Reframing the interpretation. *Creativity and innovation management*, 17(2), 107-126.
- Loeb, S., & Hough, H. (2016). Making Use of Waivers under ESSA. Evidence Speaks Reports, Vol 1(24). *Center on Children and Families at Brookings*.
- Loveless, T. (2007). The peculiar politics of No Child Left Behind. In A. Gamoran (Ed.), *Standards-Based Reform and the Poverty Gap: Lessons for "No Child Left Behind*". Washington, DC: Brookings Institution Press
- Lubienski, C. (2003). Innovation in education markets: Theory and evidence on the impact of competition and choice in charter schools. *American educational research journal*, 40(2), 395-443.
- Lubienski, C. (2004). Charter school innovation in theory and practice: Autonomy, R&D, and curricular conformity. *Taking account of charter schools: What's happened and what's next*, 72-90.
- Manna, P. (2010). *Collision course: Federal education policy meets state and local realities*. CQ Press.
- Manna, P., & Ryan, L. L. (2011). Competitive grants and educational federalism: President Obama's race to the top program in theory and practice. *Publius: The Journal of Federalism*, 41(3), 522-546.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, *2*(1), 71-87.
- McCharen, B., Song, J., & Martens, J. (2011). School innovation: The mutual impacts of organizational learning and creativity. *Educational Management Administration & Leadership*, 39(6), 676-694.
- McDonnell, L. M. (2005). No Child Left Behind and the federal role in education: Evolution or revolution?. *Peabody Journal of Education*, 80(2), 19-38.
- McDonnell, L. M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational evaluation and policy analysis*, 9(2), 133-152.
- McGinnis, J. O., & Somin, I. (2004). Federalism vs. States' Rights: A Defense of Judicial Review in a Federal System. *Nw. UL Rev.*, *99*, 89.
- McGuinn, P. (2012). *Stimulating reform: Race to the Top, competitive grants and the Obama education agenda. Educational Policy, 26*(1), 136-159.
- McGuinn, P. (2019). Assessing state ESSA plans: Innovation or retreat?. *Phi Delta Kappan*, *101*(2), 8-13.
- McLendon, M. K., Heller, D. E., & Young, S. P. (2005). State postsecondary policy innovation: Politics, competition, and the interstate migration of policy ideas. *The Journal of Higher Education*, 76(4), 363-400.
- Meyers, C. V., & Smylie, M. A. (2017). *Five myths of school turnaround policy and practice. Leadership and Policy in Schools, 16*(3), 502-523.

- Mintrom, M. (1997). Policy entrepreneurs and the diffusion of innovation. *American journal of political science*, 738-770.
- Mintrop, H. (2008). Low-Performing schools' programs and state capacity requirements. *Holding NCLB accountable: Achieving accountability, equity, & school reform,* 137-151.
- Mintrop, H., & Trujillo, T. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 13, 1-27.
- Murphy, J., & Meyers, C. V. (2007). *Turning around failing schools: Leadership lessons from the organizational sciences*. Corwin Press.
- Nerkar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. *Management science*, 49(2), 211-229.
- OECD. (2018). Education at a Glance 2018: OECD Indicators. OECD Publishing, Paris. http://dx.doi.org/10.1787/eag-2018-en
- Peck, C., & Reitzug, U. C. (2014). School turnaround fever: The paradoxes of a historical practice promoted as a new reform. *Urban Education*, 49(1), 8-38.
- Peterson, P. E. (1995). State response to welfare reform: A race to the bottom?. Urban Institute Welfare Reform Brief, 8(May).
- Peterson, P. E., & Rom, M. (1989). American federalism, welfare policy, and residential choices. *American Political Science Review*, 83(3), 711-728.
- Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on largescale school improvement. *Journal of Educational Change*, *13*(2), 155-190.
- Peurach, D. J., & Neumerski, C. M. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379-420.
- Pietsch, M., Tulowitzki, P., & Cramer, C. (2020). Principals between exploitation and exploration: Results of a nationwide study on ambidexterity of school leaders. *Educational Management Administration & Leadership*, 50(4), 574-592.
- Podgursky, M., & Ballou, D. (2001). *Personnel policy in charter schools*. Washington, DC: Thomas B. Fordham Foundation.
- Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, *43*(1), 45-54.
- Prasse, D. P. (2006). Legal supports for problem-solving systems. *Remedial and Special Education*, 27(1), 7-15.
- Preston, C., Goldring, E., Berends, M., & Cannata, M. (2012). School innovation in district context: Comparing traditional public schools and charter schools. *Economics of Education Review*, 31(2), 318-330.
- Quinn, J. B. (1980). *Strategies for change: Logical incrementalism*. Irwin Professional Publishing.
- Rentner, D. S., Kober, N., & Frizzell, M. (2017). Planning for Progress: States Reflect on Year One Implementation of ESSA. *Center on Education Policy*.
- Rhim, L. M., Hassel, B., & Redding, S. (2007). State role in supporting school improvement. Handbook on Statewide Systems of Support, Lincoln, Ill.: Academic Development Institute, 21-56.
- Robertson, P. J., Wohlstetter, P., & Mohrman, S. A. (1995). Generating curriculum and instructional innovations through school-based management. *Educational Administration Quarterly*, 31(3), 375-404.
- Romanelli, E., & Tushman, M. L. (1994). Organizational transformation as punctuated equilibrium: An empirical test. *Academy of Management journal*, *37*(5), 1141-1166.
- Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic management journal*, 22(4), 287-306.
- Schoen, L., & Fusarelli, L. D. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st-century schools in an era of accountability. *Educational Policy*, 22(1), 181-203.
- Scott, C. (2008). A Call to Restructure Restructuring: Lessons from the No Child Left Behind Act in Five States. *Center on Education Policy*.
- Shelly, B. (2008). Rebels and their causes: State resistance to no child left behind. *Publius: The Journal of Federalism*, *38*(3), 444-468.
- Shober, A. F. (2017). ESSA and State Capacity. In Hess. F. M. & Eden M. (Eds.) The Every Student Succeeds Act (ESSA): What It Means for Schools, Systems, and States (pp. 107-136). Harvard Education Press.
- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization science*, *2*(1), 125-134.
- Smith, J., Wohlstetter, P., Kuzin, C. A., & De Pedro, K. (2011). Parent involvement in urban charter schools: New strategies for increasing participation. *School Community Journal*, 21(1), 71-94.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of educational research*, 72(3), 387-431.
- Stemler, S. (2000). An overview of content analysis. *Practical assessment, research, and evaluation*, 7(1), 17.
- Strumpf, K. S. (2002). Does government decentralization increase policy innovation?. *Journal of Public Economic Theory*, 4(2), 207-241.
- Strunk, K. O., Marsh, J. A., Hashim, A. K., & Bush-Mecenas, S. (2016). Innovation and a return to the status quo: A mixed-methods study of school reconstitution. *Educational Evaluation and Policy Analysis*, 38(3), 549-577.
- Sunderman, G. L., & Kim, J. S. (2004). Expansion of Federal Power in American Education: Federal-State Relationships Under the No Child Left Behind Act, Year One. Cambridge, MA: The Civil Rights Project at Harvard University.
- Sunderman, G. L., & Kim, J. S. (2007). The expansion of federal power and the politics of implementing the No Child Left Behind Act. *Teachers College Record*, *109*(5), 1057-1085.

- Sunderman, G. L., Kim, J. S., & Orfield, G. (Eds.). (2005). *NCLB meets school realities: Lessons from the field*. Corwin Press.
- Sunderman, G. L., & Orfield, G. (2007). Do states have the capacity to meet the NCLB mandates?. *Phi Delta Kappan*, 89(2), 137-139.
- Tanenbaum, C., Boyle, A., Graczewski, A., JamesBurdumy, C., Dragoset, S., & Kallgren, K. (2015). *State capacity to support school turnaround*. National Center for Education Evaluation and Regional Assistance (NCEE 2015-4012). Washington, DC.
- Triant, B. (2001). Autonomy and innovation: How do Massachusetts charter school principals use their freedom? Washington, DC: Thomas B. Fordham Foundation.
- Trujillo, T., & Renee, M. (2012). Democratic school turnarounds: Pursuing equity and learning from evidence. *National Education Policy Center*.
- U. S. Department of Education. (2008a, March 20). Letter from Secretary of Education to Chief State School Officers regarding the differentiated accountability pilot (first cohort). Washington, DC: Retrieved from https://www2.ed.gov/policy/elsec/guid/secletter/080320.html#Enclosure
- U. S. Department of Education. (2008b, April 3). *Peer Review Guidance for the NCLB Differentiated Accountability Pilot Applications*. Washington, DC: Retrieved from https://www2.ed.gov/policy/elsec/guid/daguidance.doc
- U. S. Department of Education. (2008c, July). Secretary Spellings approves 6 states' differentiated accountability proposals. Washington, DC: Retrieved from https://www2.ed.gov/nclb/accountability/differentiated/factsheet03.pdf
- U. S. Department of Education. (2008d, August 1). Letter from Secretary of Education to Chief State School Officers regarding the differentiated accountability pilot (second cohort). Washington, DC: Retrieved from https://www2.ed.gov/policy/elsec/guid/secletter/080801.html
- VanGronigen, B. A., & Meyers, C. V. (2019). How state education agencies are administering school turnaround efforts: 15 years after No Child Left Behind. *Educational Policy*, 33(3), 423-452.
- Vincent-Lancrin, S., G. Jacotin, J. Urgel, S. Kar., González-Sancho, C. (2019). Measuring Innovation in Education 2019: What Has Changed in the Classroom? Educational Research and Innovation, OECD Publishing, Paris, Available from: <u>https://doi.org/10.1787/9789264311671-en</u>.
- Von Hippel, E. (1994). "Sticky information" and the locus of problem solving: implications for innovation. Management science, 40(4), 429-439.
- Walker, J. L. (1969). The diffusion of innovations among the American states. *American political science review*, 63(3), 880-899.
- Weiss, J., & McGuinn, P. (2016). States as change agents under ESSA. *Phi Delta Kappan*, 97(8), 28-33.

- West, M. (2017). The case for ESSA: A proper balance. In Hess. F. M. & Eden M. (Eds.) The Every Student Succeeds Act (ESSA): What It Means for Schools, Systems, and States (pp. 91-106). Harvard Education Press.
- Wöbmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003. OECD Education Working Papers, No. 13, OECD Publishing.
- Wohlstetter, P., Wenning, R., & Briggs, K. L. (1995). Charter schools in the United States: The question of autonomy. *Educational Policy*, *9*(4), 331-358.
- Wong, K. K. (2015). Federal ESEA waivers as reform leverage: Politics and variation in state implementation. *Publius: The Journal of Federalism*, 45(3), 405-426.
- Wrabel, S. L., Saultz, A., Polikoff, M. S., McEachin, A., & Duque, M. (2018). The politics of elementary and secondary education act waivers. *Educational Policy*, *32*(1), 117-140.
- Zhou, K. Z., & Wu, F. (2010). Technological capability, strategic flexibility, and product innovation. *Strategic Management Journal*, *31*(5), 547-561.
- Zukin, S., & DiMaggio, P. (1990). *Structures of capital : the social organization of the economy*. Cambridge: Cambridge University Press.

CHAPTER 3. A CASE OF EXPLOITATIVE INNOVATION: DECENTRALIZATION, SCHOOL TURNAROUND, AND STUDENTS' ACADEMIC ACHIEVEMENT Introduction

The purpose of this chapter is to examine the impacts of Florida's school turnaround, a new state-design intervention under decentralization, on students' academic achievement. U.S. education policy agendas have consistently foregrounded the difficult problem of how to improve low-performing schools and student populations. Policies have sought to support underperforming schools through research-based comprehensive school reform (CSR) demonstration programs focused on strengthening all aspects of school operations from their curricula to level of parental involvement (Desimone, 2002). Similarly, the School Improvement Grant (SIG) and the Race to the Top (RTT) programs have attended to the effectiveness of teachers and school leaders (McMurrer, 2012). In addition, the recently reauthorized Every Student Succeeds Act (ESSA) strives to provide comprehensive and targeted support to the lowest-performing schools and those with consistently underperforming subgroups (Aragon et al., 2016).

Despite educators' and policymakers' repeated calls for interventions in low-achieving schools, the empirical literature has produced mixed findings when it comes to assessing if these school turnarounds make a difference in student performance. Whereas some researchers have found little evidence that recent school turnaround interventions significantly improve student achievement (Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Dragoset et al., 2017; Hemelt & Jacob, 2020), other researchers have found that under specific circumstances they do (Bonilla & Dee, 2020; Carlson & Lavertu, 2018; Dee, 2012; Papay, 2015; Strunk et al., 2016; Sun et al., 2017).

The mixed findings can perhaps be explained, at least in part, by the evolving role of the federal government and the State Education Agency (SEA) in turning around underperforming schools. Indeed, there have been enduring debates around the balance between federal, state, and local authority in educational responsibilities to help schools with poorly performing students and whether and how a different combination of authority shifts achieve policy goals (e.g., see West, 2017).

As noted in Chapter 2, a devolved government body over school turnaround efforts has the potential to bring meaningful changes in how students learn through novel strategies that serve local preferences, while guiding the effective allocation of resources (Hanushek et al., 2013; King & Ozler, 1998). However, state governments with greater flexibility in crafting turnaround efforts also carry a risk of making the least contribution to students' academic success, for a lift of regulatory controls alone is insufficient to enhance the everyday operational procedures of an organization in which new or improved strategies come from.

Nevertheless, there is little empirical evidence on the impacts of state-driven school turnaround interventions on students' academic achievement. Most prior studies have focused more on identifying factors for successful school turnaround interventions in general (Duke, 2006), or impacts of specific school turnaround models under increasing federal directives and prescriptions yet with limited state flexibility (e.g., Dragoset et al., 2017), or school-level responsiveness made in reaction to the implementation of autonomous school models such as charter schools or school-based management (e.g., Steinberg, 2014). Looking solely at the evaluation of individual school turnaround cases without an explicit appreciation of states' responsiveness to the emerging contextual turbulence—the restoration of state educational authority—may miss crucial nuances into the presumption of recent school turnaround initiatives

is with an emphasis on increased flexibility over school turnaround may trigger innovative changes and its causal link to the effects. To get a comprehensive understanding of the effects of school turnaround interventions for underperforming schools, we should also consider it within the context of when state education agencies (SEA) assumed a greater leadership role to design and implement the school turnaround interventions, especially after they underwent overly prescriptive federal regulations. At this juncture in the school turnaround literature, I seek to answer following two research questions:

RQ 1. What is the impact of Florida's school turnaround interventions on students' reading and math achievement?

RQ 2. Do the effects vary by intervention types at different intervention stage?

With panel data on student achievement and administration from Florida, I use a fuzzy regression discontinuity approach and a difference-in-difference approach to better understand the new state systems for school turnarounds crafted in response to the afforded flexibility under waiver pilot and evaluate the effect of school turnaround interventions on students' academic performance.

Conceptual framework

Florida's school turnaround interventions were launched based on the idea that the deregulation from bureaucratic mandates and external authorities opens new, creative opportunities that may be better aligned with local conditions. Knowing what works and what doesn't in their own contexts may positively affect students' learning (Hanushek et al., 2013; King & Ozler, 1998), as it enables subnational units to effectively use available local resources (Fuhrman & Elmore, 1990; King & Ozler, 1998; Loeb & Hough 2016; Schoen & Fusarelli, 2008; Steinberg, 2014; Weiss & McGuinn, 2016). Given this logic, it will be of little wonder that

educators and policymakers show positive outlook on the improved learning outcomes of recent policy shifts, including Florida's school turnaround case, that brought State Education Agencies (SEA)' expanded authority to implement comprehensive support to the forefront of the school turnaround initiatives.

The literature regarding enhanced autonomy at the state level has pointed out that not all the SEAs are well-prepared in terms of their capacity—expert knowledge and resources (e.g., see Childs & Russell, 2017; Kober & Rentner, 2011; McGuinn, 2012). One example of the illequipped capacity is their lack of technical assistance and financial support to develop and sustain effective school turnaround efforts for local districts and schools, which suggests the possibility that already-existing variation across states may be enlarged (Tanenbaum et al., 2015).

Moreover, prior research on autonomy in broader contexts has found that an authority separation is often constrained and complicated in practice by the sheer number of factors induced from the interaction of multi-layered players in different areas of decision-making within a narrow time frame (Finnigan, 2007; Honig & Rainey, 2012; Wohlstetter et al., 1995). For example, while the facades of flexibility arrangement may appear to be formalized through varying forms of legislation, such as waivers in principle, this may not be fully realized shortly to support struggling schools with transformational interventions in practice. As devolving increased flexibility to local actors entails considerable "uncertainty" and "ambiguity" when clarifying the scope of the flexibility, especially within the contexts of accountability (Bowling & Pickerill, 2013; Finnigan, 2007, p.521; Watson & Supovitz, 2001), states and local school districts may still remain under significant federal control through the complex ratification

process with attached strings, or in the form of new systematic structures such as tertiary external organizations in a real sense.

It should be noted that recent school turnaround movements do not return the flexibility over improvement strategies to local districts and schools, but to the SEA level (West, 2017). Thus, the effects, scholars (e.g., Baroody, 2011; Manna, 2010) suggest, could be further confounded by challenges stemmed from a lack of guidance and support from the state and organizational culture at the school- and teacher-level, which manifest themselves during the implementation process by preserving their status quo rather than radically changing them. Together, this suggests that while it is assumed that shifting more decision-making power to state level over crafting school turnaround efforts may be a driving force toward innovative approach for turning around failing schools, such an assumption alone in a "real world" context where dynamic political tensions and structural barriers are at interplay is questionable to ensure promoted students' performance.

Effects of school turnarounds under decentralization on students' academic achievement

While the NCLB Waiver is distinguished from Florida's pilot with respect to the balance between flexibility afforded to the states and greater federal control in return, research on the NCLB Waiver provides insights into the individual state's practices and their effects on school performance as well as administrative changes in reaction to the reclaimed authority. One recent research study found significant, positive effects on student's proficiency rates in reading and math after a year of implementation (Bonilla & Dee, 2020). Capitalizing on increased flexibility, Kentucky identified struggling schools based on new strategies in which all of the at-risk students were consolidated into one "super-subgroup." Under NCLB, small groups of vulnerable students were likely to be ignored due to regulation that required a minimum number for a subgroup to determine school performance. Authors suggest that schools in Kentucky were held accountable for the performance of smaller subgroups through the new super subgroup approach. On the whole, this may have led to improved student learning outcomes, combined with the implementation of school improvement plans with high fidelity guided by SEA and teacher professional development.

However, for many states, flexibility-based school turnarounds do not appear to be successful at better supporting students in highest-need schools, as evidenced by poor implementation at the local level (Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2019; Hemelt & Jacob, 2020). In a study from Louisiana where the SEA focused on providing technical assistance according to the needs of each specific school ("Focus school"), Dee and Dizon-Ross (2019) noted that a lack of capacity of districts to provide intensive support could have led to poor quality of design and challenges to carry out customized strategies with fidelity. Similarly, two other studies in which schools labeled as lowest performing schools in Michigan and Rhode Island received supplemental assistance show null impacts on improving students' academic performance or closing gaps (Dougherty & Weiner, 2019; Hemelt & Jacob, 2020). They suggested that the null effects could have been addressed by building capacity of different levels of actors to identify the degree and areas of support they deemed necessary and ensuring they were undertaken fully.

Effects of school turnaround strategies on students' academic achievement

At the same time, a topic of ongoing interests among researchers and policymakers has been the effects of specific turnaround strategies and the potential mechanisms by which they impact student learning. Some researchers found strong causal effects where students in schools implementing a turnaround model with bold actions, such as replacing the principal and instructional staff, improved their performance on tests (Dee, 2012; O'Brien & Dervarics, 2013; Strunk et al., 2016; Sun et al., 2017). In contrast, Heissel and Ladd (2018) noted that simply replacing a principal along with increased professional development time for teachers does not lead to improved student achievement in reading and math, and even resulted in lower test scores. The authors suggested such results might have been driven by teachers' reduced engagement in instruction because the remaining teachers tend to experience a pervasive sense of administrative overload due to sweeping changes in leadership.

Expanding on findings on the turnaround model focused on leadership and personnel change, three high-profile studies had conflicting results on the effects of closing schools on displaced students (Bifulco & Schwegman, 2019; De la Torre & Gwynne, 2009; Engberg et al., 2012). While the short-term effects of a school closure might vary among these studies, the researchers argued it is possible that students might experience unintended consequences of a school closure based on the quality of the new schools. This is problematic, given that high performing students tend to transfer to high performing schools, whereas low achieving students tend to transfer to weak schools.

Another line of inquiry (Strunk et al., 2016; Zimmer et al., 2017) investigates how students' academic outcomes are affected by the "restart" model, in which schools are closed temporarily and then reopened under the management of alternative governance including charter operators. Using a difference-in-differences approach, Zimmer et al. (2017) found that, in schools maintaining the extant governance with a new management structure (i.e., governed by the district with additional flexibility and resources for "recruiting and retaining highly effective teachers" (p. 692)), students performed better in reading, math, and science compared to their peers attending schools taken over by new governance (i.e., the state or contracted, private management firms such as a Contracted Management Operator (CMO)).

Current study

This paper extends existing literature by making a case of how a state may or may not leverage the opportunities of flexibility to support low-performing schools and provides empirical evidence to inform the decision making of policymakers and practitioners who seek answers on how such integrated arrangements may engender student success. Detailed description of Florida's school turnaround policy is presented in results section in Chapter 2.

Data and Methods

Data and sample

This paper analyzes student outcome data and administrative data for multiple cohorts of students in Grades 3 through 10 in Florida between 2007-2008 and 2008-2009. The data were collected from three sources: 1) the Florida Department of Education's (DOE) K–20 Education Data Warehouse (EDW), 2) the Common Core of Data (CCD), and 3) Florida DOE's webpage.

Three criteria were established for determining the sample schools and students included in this analysis. First, since overall school management systems could be important mediating attributes for how interventions play out at the school-level, the school samples had to be from regular public schools. Second, the emphasis on interventions mainly refers to the influence of school-wide reforms on the general pool of public-school students. Thus, school samples had to be schools that serve a majority of general student populations in grades 3 through 10. Thus, charter schools, special education schools, vocational education schools, and alternative schools were excluded from the school sample data sets. Lastly, since student-level test score data were needed, this study excluded students with missing FCAT test scores. The final full sample includes approximately 920,000 students from 1,722 untreated schools and 373 Comprehensive intervention schools. For parametric RD analysis, the focus is on the samples within a bandwidth of 24 SG points.

Student-level outcomes were measured by reading and math test scores on the statewide standardized test, Florida Comprehensive Assessment Test (FCAT) and attendance rates for 2007-2008 and 2008-2009 obtained from the Florida EDW. Test scores in reading and math are standardized by grade level, year, and subject for comparability purposes. Additional student-level outcome variables, HIGH STANDARD and LEARNING GAINS, are binary variables that are indicative of high levels of student academic achievement and the growth students make over time, respectively. HIGH STANDARD is defined based on the current proficiency status, which "1" presents that a student has earned a high achievement level of 3, 4, or 5 in each subject, and "0" otherwise. In another measurement, LEARNING GAINS draw attention to a student's improvement during a given academic year, and has the value of 1 for the following three cases in each subject: 1) a student has moved toward the higher achievement level, 2) a student with a relatively high achievement level of 3, 4, or 5, has maintained at the same achievement level during a given year, 3) a student with a relatively low achievement level of 1 or 2 has obtained sufficient growth during a given year, while remaining at the same achievement level.

Forcing variable used for intervention assignments, *SG points for 2007-2008*, is defined using a school grade point system imposed by Florida DOE. The SG points reflect eight components including the percent of students meeting high proficiency standards in reading, math, and writing, and the percent of students making learning gains. The schools were assigned a grade of A through F based on the SG point, which represents a school's overall performance.

74

Since those schools that received below 435 SG points, i.e., D or F, got comprehensive support, I re-centered the SG points by subtracting the minimum point (i.e., 435) for C-rated school.

To reduce sampling variability (Lee & Lemieux, 2010; Imbens & Lemieux, 2008), this study used a set of student-level demographic information (e.g., race, gender, free or reduced lunch (FRL) eligibility status, and limited English proficiency (LEP) status) as well as test scores during the preintervention period from Florida DOE. Common Core of Data (CCD) and Florida School Indicators Report (FSIR) provided data that were used for constructing the covariates on teacher characteristics (e.g., average teaching experience in years, a percentage of teachers with advanced degrees, and a percentage of courses taught by out-of-field teachers) and school characteristics (e.g., log-transformed enrollment size, school level, pupil-teacher ratio, and Title I status), which are also included in the model.

Table 6 reports the descriptive statistics for the analysis of student and school samples broken down by the intervention status in the pre-treatment year, 2007-2008. The first two columns of numbers display the means and SDs for students, teachers, and schools in each intervention category, respectively, whereas the last column provides descriptive insights into the full samples. Comparisons of column 1 and 2 show that Black, Hispanic, and students with limited English proficiency were more prevalent in comprehensive intervention schools than in non-comprehensive intervention schools. Students in comprehensive intervention schools also appear to be more likely to receive free or reduced lunch and be less academically prepared in reading and math than their peers in non-comprehensive intervention schools. Not surprisingly, students in the comprehensive intervention schools received a lower quality of educational inputs related to the teacher characteristics. For example, even though comprehensive intervention schools showed a slightly smaller pupil-teacher ratio than the other schools, these schools tended

Variable	Non- comprehensive intervention schools	Comprehensive intervention schools	All schools
Category II schools	0	1	0 178 (0 383)
(Intervention schools)	0	1	0.170 (0.505)
% of AYP met (centered)	11.260 (9.089)	-8.048 (6.230)	7.807 (11.381)
School grade point (centered)	125.424	17.776	106.196
	(54.505)	(40.368)	(66.568)
Student characteristics			
Female	0.499 (0.500)	0.498 (0.500)	0.499(0.500)
White	0.542 (0.498)	0.217 (0.412)	0.491 (0.500)
Black	0.165 (0.371)	0.411 (0.492)	0.204 (0.403)
Hispanic	0.224 (0.417)	0.324 (0.468)	0.240 (0.427)
Free or Reduced-priced Lunch	0.395 (0.489)	0.696 (0.460)	0.442 (0.497)
Limited English Proficiency	0.159 (0.365)	0.320 (0.467)	0.184 (0.387)
Math z-score	0.156 (0.944)	-0.308 (0.985)	0.083 (0.965)
Reading z-score	0.171 (0.963)	-0.328 (0.983)	0.093 (0.983)
Attendance rate	0.953 (0.049)	0.943 (0.063)	0.952 (0.051)
Baseline year math z-score	0.153 (0.950)	-0.341 (0.985)	0.076 (0.972)
Baseline year reading z-score	0.151 (0.956)	-0.383 (0.986)	0.069 (0.980)
Baseline year attendance rate	0.958 (0.042)	0.950 (0.053)	0.957 (0.044)
Teacher characteristics			
Average teacher experience in	12 035 (3 327)	10 230 (2 701)	11 713 (3 297)
years	12.035 (3.527)	10.230 (2.701)	11.715 (3.277)
% Teachers with advanced degree	33.376	31 131 (0 501)	32.976
	(10.969)	51.151 (7.501)	(10.755)
School characteristics			
Enrollment (log transformed)	6.753 (0.587)	6.695 (0.579)	6.743 (0.586)
Pupil-teacher ratio	15.588 (2.506)	14.518 (2.422)	15.396 (2.524)
% Classes taught by out-of-field	6 524 (9 151)	11.828	7 469 (10 378)
teacher	0.524 (9.151)	(13.990)	7.407 (10.570)
% Disabled students	14.159 (5.411)	15.676 (5.078)	14.429 (4.217)
Student stability	0.942 (0.048)	0.910 (0.033)	0.936 (0.047)
Elementary School	0.587 (0.493)	0.539 (0.499)	0.578 (0.494)
Middle School	0.218 (0.413)	0.196 (0.397)	0.214 (0.410)
High School	0.164 (0.371)	0.236 (0.425)	0.177 (0.392)
Title I status	0.163 (0.369)	0.810 (0.393)	0.278 (0.448)
	1 700	272	2 005
N (schools)	1,/22	5/5	2,095
N (students)	//6,50/	143,988	920,495

Table 6. Descriptive Statistics

Notes: Sample is limited to K-12, non-charter, regular schools open as of the spring of 2008. Source: Florida Department of Education Data Warehouse

to have a lower fraction of teachers with advanced degrees and less teaching experience, i.e. qualified and experienced teachers. This pattern of small pupil-teacher ratios and more qualified teachers was pronounced in that a greater fraction of classes in comprehensive intervention schools were covered by teachers outside their fields of qualification and expertise. A majority of intervention schools (81%) are Title I schools.

Estimation strategy

In recent years, regression discontinuity design (RDD), an empirical approach first proposed by Thistlethwaite and Campbell (1960) has gained popularity for complementing the selection bias problem in causal inference under a non-random assignment setting (Berk et al., 2010; Cook et al., 2008). As such, this approach has been increasingly used in studies examining the effects of school intervention models on student outcomes. The main idea behind the regression discontinuity design stems from the presumption that the assignment to treatment is determined and modeled by the observed, fixed point of the variable(s), partly or entirely (Thistlethwaite & Campbell, 1960; Imbens & Lemieux, 2008). As long as the RD designs are carefully and properly executed, then it is unlikely that units on either side of the threshold within narrow window would be substantially different in their characteristics as if in randomized experiments. Thereby, a discontinuity in outcomes at the cutoff can be interpreted as the evidence of a causal impact of the treatment (Imbens & Lemieux, 2008; Lee & Lemieux, 2008; Shadish et al., 2001; van der Klaauw, 2002). However, in some situations, the estimates may be biased due to incorrectly specified functional forms on the association between the assignment variable and the treatment variable, and between the assignment variable and the outcome variable in the analytic stage (van der KlaauW, 2002). It is reasonable to be concerned that implementation of the RD designs can result in a disproportionate sorting around the cutoff and manipulation of the assignment variable to avoid or receive treatment likely leading to

biased inferences (Lee & Lemieux, 2008). This potential problem will be explored in greater detail below.

In this study, a fuzzy regression discontinuity based on instrumental variable approach is employed to identify and estimate the effects of CSI on student achievement outcomes. Florida DOE tiered schools were identified as in need of intensive support when their school grade was below D (i.e., D or F) or the percentage of AYP criteria met was less than 80. Because many cases below the thresholds of either of forcing variables received treatment, each forcing variable may partially determine the treatment exposure, and thus created a 'fuzziness' around the boundary (Reardon & Robinson, 2012; Wong et al., 2013). For example, it is possible that only 20 percent of subgroups made the yearly performance target in schools that demonstrated overall acceptable performance (i.e., B). In the RDD setting with multiple forcing variables, an instrumental variable (IV) approach using each forcing variable has been proposed as a strategy to address such fuzziness (Reardon & Robinson, 2012; Wong et al., 2013). Its major advantage is that, not only is IV estimate analogous to the fuzzy RDD estimate conceptually in typical RDD contexts (Hahn et al., 2001), but also an IV approach for each forcing variable evaluates the heterogeneity of treatment effects without loss of statistical power (Wong et al., 2013). Though, it is worth noting that the point estimate for one of two forcing variables, a percentage of AYP criteria met, was unattainable in practice because it consists of a limited number of discrete values, and this challenges the construction of the appropriate functional forms (Lee & Card, 2008). In addition, a random selection to treatment is hardly expected to be satisfied (Catteneo et al., 2015). Therefore, an index of school grade points was used as the main forcing variable to exploit the intervention effects.

For the first stage, I estimate the propensity that a school is to be treated as a function of the assignment variable, the forcing variable, and the school-level characteristics as follows:

$$CSI_{j} = f(SG \ point)_{j} + \beta_{1}I(SG \ point_{j} < 0) + \beta_{s}S_{j} + \varepsilon_{j} - (1)$$

Where CSI_j is a dummy variable indicating "1" that a school *j* participates in the intervention program, "0" otherwise in effect. $f(SG \ point)_j$ represents the functional form for the distance of the school grade points from the cutoff, which is assumed to be flexible by including the interaction terms and the higher order of polynomials. The forcing variable, the school grade points (SG points), here is centered around the cutoff. $I(SG \ point_j < 0)$ indicates whether the school grade point is below the cutoff (1=assigned to the intervention group, 0=assigned to the comparison group), which induces a change in the probability (β_1) of their actual receiving of treatment (CSI). S_j is a vector of school-level control variables and ε_j is the error term for school *j*.

In the second stage, the predicted propensity of the actual CSI treatment receipt from the first stage model, the forcing variable, and covariates are included to estimate the 2 SLS estimate of the effect as follows:

$$Y_{ij} = f(SG \ point)_{ij} + \gamma_1(\widehat{CSI_{ij}}) + \gamma_2 S_{ij} + v_{ij} - (2)$$

 Y_{ij} is the standardized test score in math or reading for student *i* in school *j*. γ_1 is the two-stage estimator of the causal effect of the comprehensive intervention on student academic outcomes. We can interpret this estimate as the local average treatment effect (LATE), which refers to the average effect of treatment on the outcome for the subpopulation who are induced to receive it by the exogeneous instrument (Hahn et al., 2001). A set of control covariates on school-level characteristics and on student-level characteristics, denoted as S_{ij} , are also entered to gain more

precise treatment estimates (Frölich & Huber, 2017). All standard errors were adjusted at the school-level to account for the nested nature of the data.

The coefficient estimates of fuzzy regression discontinuity in this study indicates the local average treatment effects (LATE) and hence, the findings may only be generalizable to sub populations around the threshold, with little consideration of schools away from the threshold.

In the RDD, the size of the discontinuity at the cutoff is estimated based on the extrapolation, which indicates how important it is to correctly specify the functional relationship between the forcing variable and the dependent variable (Schochet et al., 2010). Therefore, I provide evidence obtained from linear and quadratic splines and compare whether they show consistent results. A Nonparametric approach with a rectangular kernel is also considered, which complements the boundary issue by focusing on the estimation by a local linear regression model (Hahn et al, 2001; Lee & Lemieux, 2010). Even though I weigh the regression by the enrollment size, the rectangular kernel is applied to give the identical weight to the observations regarding the cutoff. The choice of bandwidth size for running a local linear regression is determined based on the mean squared error optimal bandwidths approach, as suggested by Calonico et al. (2014).

Another challenge with utilizing the RDD is that the estimates of regression discontinuity are sensitive to the choice of the analytic window around the threshold. For example, using samples within a wider range of bandwidth may lead to a greater likelihood of attaining biased estimates, because, intuitively, samples far from the threshold would differ from those near to the cutoff, and thus, unobservable, omitted variables may come into play more substantially as the widths of window of observations expands (Green et al., 2009). While narrowing a bandwidth might reduce the potential risk of bias, it comes with a trade-off—a loss of efficiency—by decreasing the effective sample size and inflating the standard errors (De la Cuesta & Imai,

2016). To balance the bias and variance, researchers have proposed several methods to estimate the optimal points for the regression discontinuity design (Calonico et al., 2014; Imbens & Lemieux, 2008; Ludwig & Miller, 2007). A set of bandwidths suggested by Imbens & Kalyanaraman (2012) and Calonico et al. (2014) show that bandwidth sizes vary by outcomes, ranging from as little as 6 SG points to as much as 28 SG points. I present results performed at various window sizes to check if the estimates are driven by the choice of analytic windows.

Subgroup analyses. In a set of subsequent analyses, I conducted subgroup analyses to examine whether different turnaround strategies may lead to heterogeneous effects on students' academic achievement. Instead of leveraging a regression discontinuity design, the analyses used a difference-in-differences (DiD) approach that measures the causal impacts by comparing the difference in mean values of outcomes across pre- and post-treatment periods between the treatment and comparison groups. While regression discontinuity design is widely used to evaluate the impacts of an intervention due to its strong internal validity, DiD approach was more appropriate for the subgroup analysis in the present study than RDD for two reasons. The first was that the estimates obtained from the RD for each subgroup may be biased in cases where the samples identified under RD framework were assigned to different treatment groups in a nonrandom manner on their observable characteristics (Carril et al., 2018). The second is that the DiD approach allows for leveraging the larger sample size beyond the narrowly defined bandwidth of RDD whereby otherwise the reduced subsample size within the window may undermine the precision of the estimates (Schochet, 2008).

Key idea behind the DiD approach is that if both the treated and the nontreated groups are similar enough other than their treatment status, unbiased treatment effects can be estimated by comparing the means of both groups after the treatment. Relatedly, the validity of the DiD approach partially rest on the fulfillment of parallel trends assumption, which states that, in the absence of the treatment, the trend in outcomes of a treatment group would have been identical to that of an untreated group. This parallel trend assumption, however, is difficult to test in reality as it is infeasible to observe counterfactual trend in outcomes of the treated group in the absence of treatment during the post-treatment. Albeit not ideal, but an alternative to cope with this problem is to examine whether the difference in outcomes between the two groups during the pre-treatment period stays constant, which will be discussed below.

Comparison groups similar to the intervention groups were formed as students attending schools where they were untreated for a given year, yet within the identical window of school grade points. I further narrowed down school samples to those falling into the identical intervention category in the following year, as suggested in Strunk et al. (2016). The identification strategy was informed from Florida DOE's rubric to classify schools as requiring escalated intensity of interventions. The rubric considers three overlapped dimensions, including 1) school grade, 2) adequate yearly progress, and 3) years of being identified as schools in need of improvement. Even though some schools currently might not be at a given intervention stage yet, such soon-to-be intervention schools might have gone through challenges and underperformance equivalent to a given intervention group other than their exposure to the intervention. This approach helps to deal with one of the most common criticisms of DiD, an establishment of a highly balanced comparison group along both observable and unobservable characteristics. Given the intensity of the intervention and the sample size, the identification strategy left two sets of analytic subsamples: 1) 165,875 students in 359 schools that encompass the untreated groups and the treated groups that were subject to more intensive intervention

efforts (i.e., Correct 2 and Intervene schools); 2) 76,721 students in 140 schools that includes the untreated groups and the treated groups that received the usual interventions.

Separately for each matched subsample, I ran the following specifications:

$$Y_{igst} = \beta_0 + \beta_1 Year_{2005} + \beta_2 Year_{2006} + \beta_3 (T_s \times Year_{2005}) + \beta_4 (T_s \times Year_{2006}) + \beta_5 Year_{2008} + \beta_6 (T_s \times Year_{2008}) + \beta_7 X_{igst} + \varphi_s + \varepsilon_{igst}$$

Where Y_{igst} is an outcome in reading and math for students *i* in grades 4 through 10 in school *s* at year t=2005,...,2008. Year t=2005,...,2008 is a set of dummy variables with t=2007 as the reference year, the last time point prior to the intervention, and t=2008 as the post-treatment period. T_s is a dichotomous variable indicating schools that ever received a given intervention (=1). A vector of covariates that capture students' characteristics are represented in X_{igst} , which includes students' prior year test scores for the corresponding subject, gender, race and ethnicity (Black, Hispanic), eligibility for free or reduced-price lunch, ELL status, IEP status, as well as grade level. I also introduced a school-level fixed effects, φ_s , to all models to control for unmeasured, time-invariant school-specific effects. ε_{igst} is an error term. β_1 and β_2 adjust for non-linear time-specific effects that may vary from year to year, yet common to both the treated and untreated groups before the intervention occurred. In contrast, β_3 and β_4 capture differences in time-specific effects between two groups, testing whether pre-intervention time trends are parallel. For the post-intervention period, β_1 represents an estimate for the differences in mean outcomes common to all groups. β_6 , the coefficient of primary interest, indicates the treatment effect measured immediately after the intervention. Standard errors are clustered at the school level.

Results

Tests of the Validity of the Regression Discontinuity Design

RDD identifies unbiased causal effects under the assumption that the forcing variable is not potentially manipulated to include or exclude samples from the treatment assignment. It could be understandably thought that the states and schools might have behaved strategically, easing the standards benchmark and focusing on "bubble kids" whose proficiency is close enough to the threshold to have a greater impact on the determination of intervention status (Ballou & Springer, 2017; Booher-Jennings, 2005; Carey, 2007; Lauren & Gaddis, 2016). In the literature on school support in high-stakes environments, the general pattern of a school's strategic behavior is to avoid the intervention (Schoen & Fusarelli, 2008). Yet, school-driven manipulation, in contrast with state-level manipulation, may not be possible in Florida. While the state school accountability system is grounded in the A-F grading model where the cutoff values between the grades and the elements for rating are well known and such values may suggest who the target students are, it is perceived to be challenging to point out a new cutoff value for the other primary assignment variable, a percent of AYP criteria met before the first year of implementation. Further, the assignment to a comprehensive intervention group is made as a function of multiple separate variables with varying assignment rules. In fact, a schoolwide comprehensive intervention may also lead to two types of parents' sorting behavior. One is that, for example, some parents may be motivated to move their child to the intervention schools, expecting their child, directly or indirectly, to benefit from the intensive school interventions such as evidence-based curriculum programs. On the contrary, the other is that the intervention may incentivize parents to take out of their child from the target school because the assignment to intervention schools informs parents of their child's school as a "failing school." Such systematic sorting behaviors, if they occur, may accompany a discontinuous jump in their distributions as well as in their observable characteristics before or after the interventions are

proposed, ultimately raising concern over whether samples in either side of the threshold are comparable.

Figure 4 presents visual evidence that the potential for endogenous sorting at the schoollevel might be unlikely in this analysis. Although the density of schools just to the left of the threshold seems to be somewhat higher than just to the right, overall schools are smoothly distributed around the threshold. This finding is further supported by McCrary (2008)'s test of manipulation presented in Figure 5^2 .

Figure 4. Distribution of Students and Schools





Figure 5. McCrary manipulation test



² McCrary's test of manipulation (2008) shows that the log difference in heights is -0.133 (0.477).

In the subsequent analysis, I assessed whether the observable characteristics of student samples lying on either side of the cutoff are equivalent overall to account for the school-level and student-level strategic responses. The balance of preintervention covariates can be tested by plugging them as outcomes in the reduced form equation. The set of covariates used for balance test consists of school-level student demographic information (i.e., the proportion of Black students, Hispanic students, students with limited English proficiency, students eligible for free or reduced-price lunch), the average test scores in reading and math, and the stability rate in the preintervention academic year. Panel A in the seemingly unrelated regression approach to address this possibility. The results in **Error! Not a valid bookmark self-reference.** reveal that the coefficients are not significantly different from zero with respect to a series of covariates for either pre intervention period or post intervention period.

Table 7 shows that there is no evidence that the observable characteristics of students on average change with a jump across the cutoff (student-level). Nevertheless, there exists a chance that a set of covariates may be correlated with each other and make a difference jointly in the case when they are entered in one, whole model. I use the seemingly unrelated regression approach to address this possibility. The results in **Error! Not a valid bookmark self-reference.** reveal that the coefficients are not significantly different from zero with respect to a series of covariates for either pre intervention period or post intervention period.

А	% Black	%Hispanic	% ELL	% FRL	Math	Reading
					Test score	test score
Comprehensive	0.063	-0.119	-0.047	-0.043	-0.002	0.016
intervention effect	(0.321)	(0.073)	(0.041)	(0.059)	(0.059)	(0.056)
(2007-2008)						
Observations	196	196	196	196	196	196
Joint significance test		0.547				
В	% Black	%Hispanic	% ELL	% FRL	Math	Reading

Table 7. Test of mean covariates balance at the cutoff

Comprehensive intervention effect (2008-2009)	0.054 (0.082)	-0.116 (0.072)	-0.039 (0.040)	-0.029 (0.048)	Test score NA	test score NA
(2000 2007)						
Joint significance test		0.459				
17 data 0.01 dat	0.07	2.4				

Note. *** *p*<0.01, ** *p*<0.05, * *p*<0.1

Samples are school-by-year observations within a bandwidth range from -24 to 24 during the 2007-2008 academic year. Linear spline was used. It should be noted that not only pre-treatment sorting behavior, but also post-treatment endogenous selection may threaten the validity of RDD (Dee, 2012; De la Cuesta & Imai, 2016). While this phenomenon is interpreted as an unexpected impact of the intervention, the effects on the outcomes of primary interests may be confounded by the irregular shift in the student compositions after the intervention has been implemented. To address these issues, I examine whether the student features for the post-intervention year are balanced at the cutoff. Panel B on the seemingly unrelated regression approach to address this possibility. The results in Error! Not a valid bookmark self-reference. reveal that the coefficients are not significantly different from zero with respect to a series of covariates for either pre intervention period or post intervention period.

Table 7 shows that there is little indication that students in neighborhood schools of the cutoff are dissimilar in terms of their sociodemographic characteristics such as race, ELL status, and free or reduced lunch eligibility status. When taken together, these findings suggest that the strategic responses are implausible in this context, and, even if there were such responses, they are less likely to be non-random.

Discontinuity in treatment probabilities

Next, I investigated whether the probability of receiving comprehensive intervention changes discontinuously at the cutoff. As seen from The first-stage point estimates derived from the equation (1) also confirm that the

Figure 6, all schools just below the centered SG points cutoff were assigned to receive comprehensive interventions and actually received them. In contrast, even though schools just above the centered SG points cutoff are assigned to the non-intervention group, approximately 80 % of the sampled schools are actually exposed to the interventions at the cutoff. The main cause behind the imperfect compliance for the comparison schools stems from the fact that the multiple variables are simultaneously acting on determining whether a school is eligible for the comprehensive support, rather than from school's strategic responses to the assignment³.

The first-stage point estimates derived from the equation (1) also confirm that the Figure 6. Probability of receiving comprehensive intervention by SG points



probability of receiving comprehensive interventions significantly hinges on the SG points threshold. All the coefficients in the first row of Panel B in Table 8 are positive and statistically significant, indicating that the propensity for receiving comprehensive support increases discontinuously at the cutoff. Despite schools' partial compliance to the assignment rule, a substantial jump in the probability of receiving interventions is still observed at the centered SG points threshold. Although I perform all estimation using a fuzzy regression discontinuity described by Wong et al. (2013), a full compliance is observed on the treated group because the schools were assigned to the intervention in the case when the school did not pass either of the

³ I find, using a 'party' R package, that state's decision rules based on multiple forcing variables accounts for approximately 99 % of school-level intervention assignment.

benchmark thresholds. In the next section, I turn to the estimation of causal effects and check the robustness of the findings.

Comprehensive support and student outcomes

Figure 4 presents a graphical interpretation of ITT effects of the comprehensive school intervention on the mean of student academic outcomes. Aggregated math achievement measures are plotted in the graphs on the left and aggregate reading achievement measures are plotted in the graphs on the right. For the fitted data points displayed in Figure 7, linear splines in the SG points with different slopes on each side of the cutoff can be used to detect a discontinuity in outcomes. If the comprehensive school intervention had successfully improved student achievement, we would expect to observe a discontinuous jump at the threshold (Imbens & Lemieux, 2008). For both math and reading outcomes, students attending schools with a lower SG points cutoff tended to score lower than their peers in comparison to schools with a higher SG points cutoff. For math achievement, there is no significant drop or jump in any of the outcomes at the cut-off points, which denotes that one year of comprehensive school intervention did not make a substantial difference in their standardized test scores, achievement of high standards, or achievement gains. A similar pattern is observed for test scores in reading, with no clear change found at the threshold. Although a greater share of students in comprehensive support schools achieved high standards and had achievement gains in reading as compared to math, such differences are not statistically significant.

Table 8 displays the causal estimates of student academic outcomes on various measures. I first show reduced-form estimates (the effect of the intention-to-treat) by regressing the outcomes on their assignment status in panel A. As a next step, I compare them with 2 SLS estimates (the effect of receipt of the comprehensive intervention, i.e., of treatment-on-thetreated) as reported in panel B. The coefficients in each cell represent the estimates based on a linear function or a quadratic function for three outcomes: *standardized test scores*, *high standards*, and *learning gains*. Column (1), (3), (5), and (7) present the effect estimates without additional covariates. In the rest of the columns, I report the estimates of the models in which a set of covariates, such as pretest scores, student demographic information, and teacher-level and school-level characteristics, are entered to reduce standard errors by accounting for variances that may result from these variables impact on treatment outcomes.



Figure 7. Mean student outcomes by SG points and estimated ITT discontinuity

		М	ath			Rea	ding	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Reduced form (below 0 SG p	oints)							
A.1. Standardized test score	-0.047	0.011	-0.088	0.008	-0.053	-0.003	-0.020	0.031
	(0.080)	(0.025)	(0.125)	(0.036)	(0.074)	(0.021)	(0.117)	(0.027)
A.2. High standards	-0.029	0.002	-0.067	-0.015	0.019	0.014	0.059	0.028*
	(0.041)	(0.013)	(0.063)	(0.017)	(0.033)	(0.012)	(0.053)	(0.015)
A.3. Learning gains	-0.010	0.007	-0.026	-0.002	0.008	-0.001	0.056	0.026
	(0.022)	(0.017)	(0.031)	(0.024)	(0.036)	(0.015)	(0.059)	(0.020)
B. 2 SLS								
1 st stage (below 0 SG point)	0.402***	0.336***	0.446***	0.387***	0.402***	0.336***	0.446***	0.387***
	(0.004)	(0.006)	(0.009)	(0.008)	(0.004)	(0.006)	(0.009)	(0.008)
2 nd stage								
B.1.Standardized test score	-0.118	0.032	-0.198	0.020	-0.131	-0.009	-0.045	0.080
	(0.197)	(0.076)	(0.266)	(0.095)	(0.180)	(0.061)	(0.255)	(0.082)
B.2. High standards	-0.072	0.007	-0.151	-0.040	0.047	0.040	0.133	0.072
	(0.102)	(0.038)	(0.136)	(0.048)	(0.082)	(0.039)	(0.125)	(0.056)
B.3. Learning gains	-0.025	0.021	-0.058	-0.004	0.021	-0.002	0.126	0.067
	(0.054)	(0.050)	(0.070	(0.061)	(0.088)	(0.044)	(0.125)	(0.061)
Linear	Y	Y	Ν	Ν	Y	Y	Ν	Ν
Quadratic	Ν	Ν	Y	Y	Ν	Ν	Y	Y
Controls	Ν	Y	Ν	Y	Ν	Y	Ν	Y
n (School)	196	196	196	196	196	196	196	196
n (Student)	79552	79552	79552	79552	79552	79552	79552	79552

Table 8. Estimates of effect of comprehensive intervention on student achievement (h=24)

Note. *** *p*<0.01, ** *p*<0.05, * *p*<0.1 Standard errors in parenthesis are clustered by school.

As Row A.1. in Table 8 shows, the ITT effect estimates on standardized FCAT test scores are statistically insignificant regardless of subject, type of functional forms, or inclusion of covariates. The coefficients in Row A.2. also indicate that the probability of a student reaching high standards in math and reading does not change significantly between non-intervention schools and intervention schools, with the exception of reading in the case when a quadratic smoother is used with additional covariates. Similarly, there is no evidence that comprehensive intervention leads to variation in the probability of a student having progressed in both reading and math from one year to the next.

The TOT parameter estimates for the same outcomes are presented in Section B of Table 8. The ITT findings remain consistent under the 2 SLS model. For example, the estimates of the effect in Row B.1. are not significantly different from 0, indicating that students in schools which actually received comprehensive assistance do not perform significantly better or worse than their peer students in schools which did not receive the intervention. On the other hand, the signs indicating parameter estimates for *high standards* and *learning gains* in Rows B.2. and B.3. change depending on the functional forms used and covariates added; the local average treatment effects on these outcomes are extremely close to 0 and thus statistically insignificant. Altogether, The study failed to find convincing evidence that comprehensive intervention improves student academic outcomes in reading or math, at least for the short-term.

Robustness check

The next issue at hand is whether the effect estimates derived from the parametric functions used above are robust to the selection of different bandwidths and functional forms. One strategy to address these issues entails comparing the original parametric estimators with non-parametric estimators using various bandwidths. Table *9* presents the 2 SLS coefficients and

standard errors for the effect estimators for new samples within a series of redefined windows based on the following bandwidth values: Imbens-Kalyanaraman (2012) (IK) optimal bandwidth, 20 points, and 30 points. I find that, across each category of student outcomes and each subject, all of the coefficients in Row 2 of Table 9 estimated with samples defined by the IK optimal bandwidth (Imbens & Kalyanaraman, 2012) are positive and insignificant, which is consistent with the findings obtained from the 2 SLS main analyses. For the student samples restricted to the bandwidth suggested by Calonico et al. (2014), the coefficients in Row 3 of Table 9 are again not statistically different from 0. These results are similar when limiting samples to schools which earned between -20 and 20 SG points or expanding samples to schools which earned between -30 and 30 SG points. As I explained for Table 8 in the previous section, even_with the addition of the covariates and the use of the higher polynomial order, there was little evidence that one year of comprehensive support increases student scores on standardized tests, the probability of achieving high standards, or the probability of making gains in their learning in reading and math.

		Math		Reading				
	Standardized	High	Learning	Standardized	High	Learning		
	test score	standards	gains	test score	standards	gains		
Baseline	0.032	0.007	0.021	-0.009	0.040	-0.002		
estimates	(0.076)	(0.038)	(0.050)	(0.061)	(0.039)	(0.044)		
Bandwidth=IK	0.038	0.035	0.028	0.042	0.122	0.133		
	(0.078)	(0.048)	(0.052)	(0.061)	(0.083)	(0.097)		
Bandwidth=20	0.027	-0.013	0.013	0.020	0.043	0.017		
	(0.072)	(0.036)	(0.049)	(0.058)	(0.037)	(0.041)		
Bandwidth=30	0.001	-0.012	0.004	-0.032	0.022	-0.002		
	(0.058)	(0.029)	(0.038)	(0.044)	(0.028)	(0.034)		

 Table 9. Sensitivity analysis on student achievement (2 SLS estimates)

Note. *** p<0.01, ** p<0.05, * p<0.1

Standard errors are clustered by school. Linear spline with covariates was used (2 SLS).

Heterogeneous effect

A particular focus of this study includes whether students exposed to different types of interventions benefit more (or less) from these interventions. Since one of the goals of Florida's differentiated accountability was to create more nuanced support systems for schools with varying needs, Florida DOE grouped schools that need the most intensive intervention to the least depending on their historical subgroup performance. Therefore, even though an overall effect of the differentiated accountability as a whole package is essentially zero, it may have masked heterogeneous effects by each type of interventions. To identify whether effects are heterogeneous across interventions, I estimated Difference-in-Difference (DiD) models as presented in Table *10* and Table *11*.

The credibility of DiD estimators is bolstered when the parallel time trends assumption is met. Accordingly, I compared the flexible time trends in outcomes between the intervention and comparison schools over the three years prior to the implementation of each intervention. For math outcomes, the coefficients of time-fixed effects were generally insignificant across all intervention types except for intervene group with respect to the probability of meeting high standards and making gains. Meanwhile, for reading, I found that the gaps between matched prevent 2 schools or matched correct 2 schools are inconsistent during the pre-intervention periods, especially in terms of the probability of meeting high standards or achieving learning gains. These findings suggest cautions for readers in interpreting the results of subgroup analyses.

Overall, the intervention yielded only little and statistically insignificant effects on the test scores, the percentage of students making high achievement level and gains in math for students in the Prevent 2 schools receiving universal assistance from the district (Table 10). While the effects of the intervention for these schools were generally negligible on reading

outcomes, I found a statistically negative impact on the percentage of students making test score gains in reading (p < 0.05).

For those in the Correct II and Intervene schools where state and district were directly involved in concentrated support that better matched their needs, I found statistically significant and positive evidence of the intervention on nearly all outcomes in reading and math. This indicates that, while the changes in the percentage of students making higher levels of achievement in math or in reading test scores between the Correct II and comparison schools is small and insignificant, the interventions have overall positive impacts on test scores, advancement in their level of achievement, and learning gains in reading and math at both Correct II and Intervene schools. As hypothesized, the DiD estimates reported on the ninth row of Table *11* are substantially larger than the others on the eighth row, suggesting that the Intervene schools with the highest level of support and oversight from state and district, demonstrated greater improvement in all three outcomes across subjects than the Correct II schools in the lower-tiered intervention system as well as similar comparison schools.

Table 10. H	eterogeneous e	ffect by the type	of intervention:	Prevent 2 schools
-------------	----------------	-------------------	------------------	-------------------

			Mat	h		Reading						
	Standardized		High standards		Learning		Standardized		High star	dards	Learning gains	
	test s	core			gains		test score					
2005 (t-2)	.024†	.014	056***	.01	.009	.012	.008	.015	035***	.01	.002	.012
$2005 \times \text{Prevent 2 school}$	017	.015	007	.009	01	.012	.001	.015	.002	.01	.036**	.012
2006 (t-1)	004	.013	047***	.009	026*	.011	.009	.012	019*	.008	015	.01
$2006 \times Prevent 2 $ school	.001	.013	.01	.009	.005	.011	002	.012	001	.007	02†	.011
2008 (t+1)	.007	.023	.028*	.014	.001	.015	.022	.015	.027**	.009	.03*	.014
$2008 \times Prevent 2$ school	.015	.024	.012	.015	0	.016	.006	.016	007	.01	034*	.014
R^2	0.0	56	0.457		0.082		0.627		0.429		0.06	51
Observations (student)	76,7	721	76,72	21	76,721		76,721		76,721		76,721	
Observations (school)	14	0	140		14	0	140		140		140	

Note. *** p < 0.01, ** p < 0.05, * p < 0.1, †p < 0; h=121. Standard errors are clustered by school.

	Math						Reading					
	Standardized High standards		Learning	gains	Standardiz	zed test	High standards		Learning gains			
	test sc	core					scor	e				
2005 (t-2)	.027***	.007	037***	.004	.022***	.005	.013*	.006	026***	.003	.031***	.005
$2005 \times \text{Correct } 2 \text{ school}$.016*	.009	002	.005	003	.007	.033***	.008	.015**	.005	007	.007
$2005 \times$ Intervene school	012	.019	02†	.012	026†	.014	014	.018	018	.012	044**	.016
2006 (t-1)	.006	.006	025***	.003	017***	.004	-0	.005	022***	.003	036***	.004
$2006 \times \text{Correct } 2 \text{ school}$	004	.008	006	.004	.004	.006	.003	.008	.005	.005	.018**	.006
$2006 \times$ Intervene school	013	.023	018*	.008	003	.012	.029	.019	.013	.009	.027*	.013
2008 (t+1)	009	.006	.031***	.004	013*	.005	001	.007	.004	.003	014**	.005
$2008 \times \text{Correct } 2 \text{ school}$.03**	.009	004	.005	.011†	.006	.013	.009	.028***	.005	.042***	.006
$2008 \times$ Intervene school	.149***	.026	.045**	.017	.07**	.024	.058*	.025	.058***	.014	.081**	.023
R^2	0.65	54	0.44	9	0.10	1	0.615		0.416		0.047	
Observations (student)	165,8	375	165,8	75	165,8	75	165,8	75	165,8	75	165,8	75
Observations (school)	359	9	359	1	359)	359)	359	1	359	

Note. *** p < 0.001, ** p < 0.01, *p < 0.05, †p < 0.1; h=75. Standard errors are clustered by school.
Discussion and Conclusion

The effectiveness of turning around low-performing schools cannot be understood apart from the deliberation among the government authorities that have different incentives and capacities to provide technical and strategic support. This chapter investigated the effect of a school turnaround led by Florida SEA—one of nine state educational agencies that re-gained responsibilities to formulate their own policy direction for struggling schools in 2008. The analysis finds no evidence that the Florida-brand school turnaround generated major differences in the students' test scores, probability of meeting high standards, or making gains in math and reading. However, the subgroup analyses showed evidence of heterogeneity in its effect, indicating substantial and positive impacts across all outcomes for students at schools with the most intensive level of support, yet only modest or no effect for the remaining schools.

These findings make contributions to the accumulating literature on the state school turnaround, particularly when the federal oversight is lifted. Some recent studies in similar policy contexts have documented the limited success in reducing the within-school achievement gap in Michigan (Hemelt & Jacob, 2021) as well as little impacts on students' test scores in Rhode Island (Dougherty & Weiner, 2019) or in Louisiana (Dee & Dizon-Ross, 2019), adding more cases that transferring responsibilities and resources to the SEA does not necessarily lead to improvement in students' academic performance. On the contrary, the case of Florida provides strong evidence in support of state activism in turnaround efforts by showing that the state school turnaround built on the newfound flexibility is indeed yielding positive results in students' learning at least in states' neediest schools, a result generally consistent with the case of Kentucky (Bonilla & Dee, 2020).

Although this study is unable to fully disentangle why the Florida-led school turnaround was successful at improving students' academic performance at the state's most struggling schools, this could be partly accounted for by the interplay of contextual or organizational contingencies that occurred along with the withdrawal of federal constraints. Contextual contingencies include an alignment of the state and federal accountability systems, which enabled the SEA to approach school turnaround differently than before in two major ways. The first, and maybe the rudimentary way is related to the question of who gets priority when allocating resources amongst an increasing number of schools being identified as in need of assistance under two separate systems. While Florida's new model did not lead to a "seamlessly" consolidated accountability system like many other models observed in the NCLB waiver, it helped the SEA to better prioritize resources for schools that require immediate and intensive support, sending clear and consistent messages about the performance of schools (Reed et al., 2012).

The second approach may have shaped the way for the Florida SEA to use such resources to support the lowest-performing schools. As discussed in Chapter 2, Florida's approach focused more on the implementation of programs in place with fidelity with slight modifications, rather than experimentation with new intervention designs. In general, resources were reallocated to boost within-school organizational changes through professional development and consultation on the use of data, supplemental and intensive instruction, and leadership development, recruitment of onsite coaches to engage in whole school reform models within their own school contexts, and implementation of research-based instructional materials and strategies (Florida DOE, 2008). A group of school improvement specialists and instructional specialists in five regional support centers also directly worked with districts, Correct II, and Intervene schools to

provide technical assistance in areas noted as localized problems (Florida DOE, 2008). Other funds such as the Florida Partnership Grant and the Southern Regional Education Board were redirected to intensive tier schools to deliver training for school leadership teams or professional development for teachers to improve knowledge of advanced courses.

Finally, in addition to the intensive resource allocation and implementation, accountability pressure could account for why the new Florida model yielded larger effects in Intervene schools, compared to null or weak effects found in other intervention schools. Previous research has found that students at schools faced with strong sanctions of accountability pressure demonstrated large gains in their test scores (Ahn & Vigdor, 2014; Rouse et al., 2013; Strunk et al., 2016). Such sanctions were thematically equivalent to those observed in Florida Intervene schools where they were subject to implement a closure or a restart of the school when they still failed to show significant progress either on the state or federal evaluation metrics based on the state standardized tests. Thus, the threats of sanctions imposed by the state and districts could have motivated educators in the Intervene schools to explore areas of weaknesses and change their leadership and instructional practices with the intensive assistance of relevant specialists, increasing students' academic performance even in the short-term.

The positive effect of Florida-led school turnaround for some of Florida's underperforming schools leads to a dilemma among policy makers and practitioners who attempt to displace the old system in compliance with federal rules and formulate more localized ones, as the new system appears to be a combination of essentially the same support strategies and incentive structure that subnational units aimed to change that they landed eventually when they took greater initiatives. Some advocates of federal activism would may even ask why not roll back legislative authorities over school turnaround efforts to the federal government if SEA uses the flexibility to reiterate the interventions prescribed under greater federal control.

Unfortunately, the data do not allow for the comparison of the extent to which students' learning was affected by two similar turnaround programs that are designed and implemented in different temporal and spatial contexts. Furthermore, I argue that even though Florida's new turnaround is heavily past dependent, reflecting a majority of interventions that were found to be effective in the previous literature or were in place prior to the decentralization, it is more than a static sum of the past approaches. Rather, the turnaround initiative could be closer to the learning outcomes of the Florida SEA that have resulted from an iterative process of incorporating insights from the past challenges and coordinating moving parts within its own context, including the past interventions, multi-level actors, incentive structures, available resources, institutional environment, and temporal turbulence. In that sense, what became clear from this study is that the "Made in Florida" turnaround efforts did generate a significant and positive impacts on students' achievement in reading and math at the most struggling schools—evidence in favor of state-directed initiatives.

Limitations

Although findings from this study may provide empirical evidence on the causal impacts of school turnaround policy on students' achievement, this study suffers from two weaknesses. First, the generalizability of findings is limited to schools close to the school grade threshold, as the research design—regression discontinuity—estimates the local average effects in nature.

Second, this study examines only early phases of school turnaround after decentralization. Research on school reform often noted that even if reform efforts are successful at improving students' academic performance while they are receiving intensive support, it could be more challenging to retain its effectiveness especially after the withdrawal of the assistance and resources. In this study context, there are reasons to believe that intervention schools may follow varying improvement trajectories depending on the type of interventions as the intensive resources and oversight erode over time. Hence, the next step is to examine the longitudinal impact of Florida-directed school turnaround on students' academic performance.

References

- Ahn, T., & Vigdor, J. (2014). The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina (No. w20511). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20511
- Aragon, S., Griffith, M., Wixom, M. A., Woods, J., & Workman, E. (2016). ESSA: Quick Guides on Top Issues. *Education Commission of the States*.
- Baroody, K. (2011). Turning around the Nation's Lowest-Performing Schools: Five Steps Districts Can Take to Improve Their Chances of Success. *Center for American Progress*.
- Berk, R., Barnes, G., Ahlman, L., & Kurtz, E. (2010). When second best is good enough: A comparison between a true experiment and a regression discontinuity quasi-experiment. *Journal of Experimental Criminology*, 6(2), 191-208.
- Bifulco, R., Duncombe, W., & Yinger, J. (2005). Does whole-school reform boost student performance? The case of New York City. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management,* 24(1), 47-72.
- Bonilla, S., & Dee, T. S. (2020). The Effects of School Reform Under NCLB Waivers: Evidence from Focus Schools in Kentucky. *Education Finance and Policy*, *15*(1), 75-103.
- Bowling, C. J., & Pickerill, J. M. (2013). Fragmented federalism: The state of American federalism 2012–13. *Publius: The Journal of Federalism*, 43(3), 315-346.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295-2326.
- Carlson, D., & Lavertu, S. (2018). School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis*, 40(3), 287-315.
- Childs, J., & Russell, J. L. (2017). Improving low-achieving schools: Building state capacity to support school improvement through race to the top. *Urban Education*, 52(2), 236-266.
- Cook, T. D. (2008). "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2), 636-654.
- Dee, T. (2012). *School turnarounds: Evidence from the 2009 stimulus* (No. w17990). National Bureau of Economic Research.
- Dee, T. S., & Dizon-Ross, E. (2019). School performance, accountability, and waiver reforms: Evidence from Louisiana. *Educational Evaluation and Policy Analysis*, 41(3), 316-349.
- De la Cuesta, B., & Imai, K. (2016). Misunderstandings about the regression discontinuity design in the study of close elections. *Annual Review of Political Science*, *19*, 375-396.
- De la Torre, M., & Gwynne, J. (2009). When Schools Close: Effects on Displaced Students in Chicago Public Schools. Research Report. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented?. *Review of educational research*, 72(3), 433-479.

- Dougherty, S. M., & Weiner, J. M. (2019). The Rhode to turnaround: The impact of waivers to No Child Left Behind on school performance. *Educational Policy*, *33*(4), 555-586.
- Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., ... & Giffin, J. (2017). School Improvement Grants: Implementation and Effectiveness. NCEE 2017-4013. *National Center for Education Evaluation and Regional Assistance*.
- Duke, D. L. (2006). What we know and don't know about improving low-performing schools. *Phi Delta Kappan*, 87(10), 729-734.
- Engberg, J., Gill, B., Zamarro, G., & Zimmer, R. (2012). Closing schools in a shrinking district: Do student outcomes depend on which schools are closed?. *Journal of Urban Economics*, 71(2), 189-203.
- Finnigan, K. S. (2007). Charter school autonomy: The mismatch between theory and practice. *Educational Policy*, *21*(3), 503-526.
- Florida Department of Education. (2008). *Florida's differentiated accountability model: guidance for implementation 2008-09 school year.*
- Frölich, M., & Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5), 1645-1666.
- Fuhrman, S. H., & Elmore, R. F. (1990). Understanding local control in the wake of state education reform. *Educational Evaluation and Policy Analysis*, 12(1), 82-96.
- Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis*, 17(4), 400-417.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Hanushek, E. A., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, 104, 212-232.
- Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain?. *Educational Evaluation and Policy Analysis*, 32(2), 273-298.
- Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, 62, 302-320.
- Hemelt, S. W., & Jacob, B. A. (2020). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance?. *Education Finance and Policy*, 15(1), 45-74.
- Honig, M. I., & Rainey, L. R. (2012). Autonomy and school improvement: What do we know and where do we go from here?. *Educational policy*, 26(3), 465-495.
- Huberman, M., Parrish, T., Hannan, S., Arellanes, M., & Shambaugh, L. (2011). Turnaround Schools in California: Who Are They and What Strategies Do They Use?. *California Comprehensive Center at WestEd*.

- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.
- King, E., & Ozler, B. (1998). What's decentralization got to do with learning? The case of Nicaragua's school autonomy reform. *Development Economics Research Group*, *Working Paper Series on Impact Evaluation of Education Reforms*, 9, 278200-1099079877269.
- Kober, N., & Rentner, D. S. (2012). Year Two of Implementing the Common Core State Standards: States' Progress and Challenges. *Center on Education Policy*.
- Ladd, H. F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, 18(1), 1-16.
- Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655-674.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2), 281-355.
- Loeb, S., & Hough, H. (2016). Making Use of Waivers under ESSA. Evidence Speaks Reports, Vol 1,# 24. *Center on Children and Families at Brookings*.
- Manna, P. (2010). Collision course: Federal education policy meets state and local realities. Washing, D.C.: CQ Press.
- McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants and the Obama education agenda. *Educational Policy*, 26(1), 136-159.
- McMurrer, J. (2012). Schools with Federal Improvement Grants Face Challenges in Replacing Principals and Teachers. *Center on Education Policy*.
- Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple ratingscore variables. *Journal of research on Educational Effectiveness*, 5(1), 83-104.
- Reed, E., Scull, J., Slicker, G., & Winkler, A. M. (2012). Defining Strong State Accountability Systems: How Can Better Standards Gain Greater Traction. Washington, DC: Thomas B. Fordham Institute.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, 5(2), 251-81.
- Schochet, P. Z. (2008). Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Education Evaluations. NCEE 2008-4026. National Center for Education Evaluation and Regional Assistance.
- Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for Regression Discontinuity Designs. *What Works Clearinghouse*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc rd.pdf
- Schoen, L., & Fusarelli, L. D. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st-century schools in an era of accountability. *Educational Policy*, 22(1), 181-203.

- Steinberg, M. P. (2014). Does greater autonomy improve school performance? Evidence from a regression discontinuity analysis in Chicago. *Education Finance and Policy*, 9(1), 1-35.
- Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016). The impact of turnaround reform on student outcomes: Evidence and insights from the Los Angeles Unified School District. *Education Finance and Policy*, 11(3), 251-282.
- Sun, M., Penner, E. K., & Loeb, S. (2017). Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants. *American Educational Research Journal*, 54(4), 607–643. https://doi.org/10.3102/0002831217695790
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, *51*(6), 309.
- U. S. Department of Education. (2008). Peer Review Guidance for the NCLB Differentiated Accountability Pilot Applications. Retrieved from https://www2.ed.gov/policy/elsec/guid/daguidance.doc
- U.S. Department of Education. (2012). *Early Implementation of State Differentiated Accountability Plans Under the* No Child Left Behind Act, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Washington, D.C.
- Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review*, 43(4), 1249-1287.
- Watson, S., & Supovitz, J. A. (2001). Autonomy and Accountability in Standards-Based Reform. *education policy analysis archives*, 9(32).
- Weiss, J., & McGuinn, P. (2016). States as change agents under ESSA. *Phi Delta Kappan*, 97(8), 28-33.
- West, M. (2017). The case for ESSA: A proper balance. In Hess. F. M. & Eden M. (Eds.) The Every Student Succeeds Act (ESSA): What It Means for Schools, Systems, and States (pp. 91-106). Harvard Education Press.
- Wohlstetter, P., Wenning, R., & Briggs, K. L. (1995). Charter schools in the United States: The question of autonomy. *Educational Policy*, 9(4), 331-358.
- Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2), 107-141.
- Zimmer, R., & Gill, B. (2007). State and Local Implementation of the "No Child Left Behind Act." Volume I—Title I School Choice, Supplemental Educational Services, and Student Achievement. A Report from the National Longitudinal Study of "No Child Left Behind"(NLS-"NCLB"). US Department of Education.
- Zimmer, R., Henry, G. T., & Kho, A. (2017). The effects of school turnaround in Tennessee's achievement school district and innovation zones. *Educational Evaluation and Policy Analysis*, *39*(4), 670-696.

CHAPTER 4. A CASE OF EXPLORATIVE INNOVATION: A VIRTUAL INTERNSHIP AS AN ALTERNATIVE WORK-BASED LEARNING SPACE TO IMPROVE EQUITY⁴ Introduction

The purpose of this study is to assess the accessibility, quality, and outcomes of a virtual internship, a format that may provide underrepresented college students with rich learning opportunities for a successful future career. Internships can have pervasive impacts on college students' academic performance as well as employment and career outcomes, including final academic achievement, degree classification, career adaptability, job interview opportunity, job attainment, and salary (e.g., Baert et al., 2021; Binder et al., 2015; Di Meglio et al. 2020; Gault et al., 2000; Mansfield, 2011; Nunley et al., 2015; Oswald-Egg & Renold, 2021; Sterling & Fernandez, 2018). Colleges and universities, thus, have integrated internship programs into their curricula, which have come under increasing praises and alarms (Burke & Carton, 2013), partly for their general neglect of structural inequality embedded in the traditional in-person internship landscape. Particular concerns with conventional in-person internships, especially with unpaid internships, include the high cost of relocation, unpaid labor, and uneven socio-cultural resources and support (Allen et al., 2013; Curiale, 2009; Perlin, 2012). These barriers keep students from historically marginalized races, ethnicities, genders, socio-economic backgrounds, and geographic regions out of the learning experience that may bring high value to the job market (Bayerlein & Jeske, 2018; Hora, Wolfgram, Zi, & Lee, 2021; Shade & Jacobson, 2015).

Researchers, practitioners, and policymakers (e.g., Kraft et al., 2019) argue that some of these structural inequalities that are inherent in the conventional internship ecosystem may be

⁴ The research reported here was supported by the Seen@Work. The opinions expressed do not represent views of my supporters.

reduced by a new form of work-based learning— the virtual internship, a digital-mediated workbased learning experience that is done remotely (Hora, Lee, Zi, & Hernandez, 2021). For example, the temporally and spatially flexible work mode may present unprecedented opportunities in making experiential learning opportunities accessible to students from lowincome backgrounds or students living in rural communities who otherwise could not attend inperson internships. If these claims are valid, a strong linkage should be found between college students from marginalized social identities, economic backgrounds, and/or geographical regions and their participation in a virtual internship. While these are laudable goals, there is nearly no empirical evidence to support this argument.

Access to a virtual internship, however, may not be the sole source of challenges for disadvantaged college students. Another fundamental challenge lies in whether a virtual internship offers meaningful learning experiences and support for all students and whether such training results in learning outcomes and satisfaction, compared to in-person internships. The consequence of poor-quality workplace experience such as clerical tasks with no feedback could be critical for both industry and the entire society, as it may lead many young college students to leave the profession after graduation (Callan, 1997; Garcia, 2009; Richardson, 2008). Indeed, prior research has found uneven training and support within the conventional in-person internship environment for interns from disadvantaged backgrounds, including Black interns' experience of micro-aggressive behaviors from colleagues during the internship (Proctor et al., 2016).

While internship accessibility itself may have been expanded to some of the underrepresented students in the traditional in-person internship system, it is possible to observe similar pattern of unequal learning experience in the virtual internship setting. Furthermore, the digitally mediated learning that occurs in a virtual internship may provide unique challenges and chances for online workplace learning, affecting the nature of tasks, supervising style, communication, and networking strategies, relative to in-person internships (Jeske & Axtell, 2018). While the recent global pandemic boosted vigorous interests about the roles of a virtual internship as a new learning sphere, little is known about what and how virtual interns learn, especially those interns who otherwise would not have taken the internship in the conventional setting, and how they are systematically supported with respect to achieving academic or career goals and satisfaction, compared with in-person interns.

This study aims to fill in some of these literature gaps by exploring the experiences of college students, from their entry to the outcomes of the internship, using a large-scale dataset collected from the National College Internship Study Pilot of 2021 (N=11,009). Specifically, I focus on four research questions:

RQ 1. Does college students' entry to virtual internships differ by their socio-economic status, race/ethnicity, gender, and geographic location, compared to an in-person internship or no internship?

RQ2. Do college students' learning profiles during the internship vary by online or inperson internship?

RQ3. To what extent do college students' socio-economic status, race/ethnicity, gender, and geographic location of online intern learning profiles differ from those of in-person intern profiles?

RQ4. Does the value of academic and career development and satisfaction vary across internship learning profiles?

Using multinomial logistic regression, I explore the first question by comparing students' characteristics in three pairs of college student groups: 1) virtual interns vs. non-interns; 2) inperson interns vs. non-interns; and 3) virtual interns vs. in-person interns. These comparisons reveal whether the virtual internship modality may serve as a new pathway to "high-impact" learning for particular groups of students who have been marginalized in conventional internship settings.

For the next research question, a latent profile analysis is employed to detect subgroups with different learning profiles. Based on a review of literature, I focus on internship modality (in-person/online) and three overarching areas and related sub-areas to gain a comprehensive understanding of interns' work-based learning: 1) task-related pedagogical practices (incorporation of soft skills, clarity of learning goals, and academic relatedness); 2) supervisory support and style (site mentorship on job performance and responsibilities, site supervisors' emotional support, faculty guidance, autonomy, and professional networking opportunities); and 3) working environment (work hour, duration, and monetary compensation). If group (s) of virtual interns demonstrate distinctive learning experience patterns, this may suggest that a virtual internship provides a niche learning experience, relative to the conventional in-person internship.

Building on these typologies, the third research question identifies the extent to which such unique virtual intern profile (s) are related to students' marginalized identities race/ethnicity, gender, socio-economic status, and geographic location—compared with inperson intern groups. Based on the results from a multinomial logistic regression, I consider if niche learning profiles of virtual interns are associated with disadvantaged groups, compared with learning profiles of in-person interns. This comparison may imply that a virtual internship tends to create a new work-based learning space for that specific marginalized group (s), although it is unclear whether such space provides learning of higher quality compared with inperson. Finally, I compare psychosocial outcomes of different intern profile groups measured by academic and career developmental value, and satisfaction.

This article represents an initial step to provide empirical evidence on whether the virtual internship modality may equalize or reinforce existing disparities in college students' access, learning quality, and outcomes of "high-impact" learning opportunities.

Literature review

Virtual internship and access of historically marginalized college students

Advocates of the virtual internship modality believe that it may create greater access to historically underrepresented college students to an internship than an in-person internship, for both employers and students. By 'historically underrepresented college students', I refer to groups of college students who have been disadvantaged under the conventional in-person internship modality due to restraints caused by their socio-economic status, race/ethnicity, gender, and/or geography (i.e., rural areas).

The greater temporal and spatial flexibility of the virtual internship environment may make internships more accessible to everyone. Traditional in-person internships, in particular unpaid internships, tend to put low-income students of color at a disadvantage because they cannot afford to pay their rent and tuition while doing (unpaid) professional development training (Hora, Wolfgram, Zi, & Lee, 2021). Students working full-time or with family responsibilities, or those in rural areas without a car have also reported financial challenges due to costly relocation fees (Curiale, 2009; Jacobson & Shade, 2018). Perhaps these obstacles may compel marginalized college students to favor a remote, virtual internship over a traditional onsite internship as it enables student employees to keep a flexible work schedule without moving to another location (often cities) or commuting to a physical workplace (Kraft et al., 2019; Ruggiero & Boehm, 2016). Indeed, Waters and Russell (2016) find that internships at virtual schools were viewed as a more attractive option to pre-service teachers, especially female interns, who had to juggle around family demands, financial challenges, and uncertain transportation availability. Similar preference was pronounced among nursing students looking for online international internships (Graber et al., 2019).

Changes in the recruitment and hiring process of virtual interns may serve as another pathway into internships. One of the common staffing strategies for an in-person internship includes on-campus recruitment events where organizations' recruiters and hiring managers visit their target schools and hire talented students through on-campus interviews. Of course, these employers do not necessarily exclude students' applications from non-target institutions. However, their selection of target institutions is often linked to institutions with high selectivity and academic programs of interest within commuting zone of a given employer (Smith & Green, 2021). Such a non-random recruitment practice is even more troubling, given that some institutions fulfill their diversity recruiting goals by sourcing underrepresented candidates within such institutions (Smith & Green, 2021). Boulton (2015) and Reeves (2017) also point out uneven 'opportunity hoarding' by privileged social groups for highly sought-after internship positions in U.S. advertising firms or law firms.

On the contrary, scholars find that some organizations designing and implementing virtual internship programs take overlapping, yet different recruitment approaches by using virtual tools, which have the potential to reach a more diverse pool of historically disadvantaged students (Cutshall et al., 2021; Jeske & Axtell, 2016). Virtual interviews, a digitized application

process, and advertisements through social media platforms or job searching websites are good examples of such efforts. According to a 2021 student survey by the National Association of Colleges and Employers, roughly two-thirds of Black or Hispanic students tend to report having positive virtual interactions with employers through virtual meetups, virtual open house, and virtual interviews, whereas half of white students report the same. However, Harpen et al., (2020) raise concerns about whether a greater transparency in internship postings on social media indeed is linked to greater hiring of underrepresented college students as not many internship listings on social media platforms are fully equipped to provide rich, precise information that may motivate potential applicants to opt-in for the complex application process. **Virtual internships, learning quality, and outcomes of historically marginalized college students**

Virtual internships have the potential to diversify the labor force through recruitment and remote working environments, yet virtual interns may still face challenges while they are getting hands-on experience. The potentials and risks of a virtual internship explored in this section begin by examining important learning elements in a conventional internship, in general, as well as what the history of the conventional internship has clearly documented as a *double jeopardy* for historically marginalized college interns: inequitable organizational structures and cultural differences in tacit norms during the traditional internship (e.g., Allen et al., 2013; Frenette et al., 2015; Gracia, 2009). In conjunction with the unique challenges of a virtual work environment, such barriers and learning components may create a comparative typology of virtual interns in terms of the integrity of their learning, ultimately affecting net learning outcomes and satisfaction. Core learning elements of traditional internships fall into three categories: 1) what college interns from disadvantaged backgrounds learn from their virtual internship 2) how they

learn from their virtual internship, and 3) how their experiential learning is systematically supported.

Numerous studies on the traditional internship form directed their attention to the nature of tasks, including clarity of goals, academic relatedness of the tasks, and contents that enable self-directed learning. Clearly defined objectives of an internship, for example, (Gault et al., 2000; Henry et al., 2001; Ruggiero & Boehn, 2016) reduce ambiguity in interns' understanding of what they are expected to do. The internship's relatedness to an academic program may contribute to interns drawing connections between academic theory and 'real world' practices. Some researchers (e.g., Pretti et al., 2020) emphasize the incorporation of learning contents designed to facilitate independent, self-directed learning into a remote internship program, which is not often reported in traditional internship programs. This is related to the quality of learning for some female interns whose work is often limited to clerical tasks such as photocopying or supporting tasks in male-dominant accounting firms (Garcia, 2009).

A supervisor's mentoring of an intern's job performance also makes an internship into a meaningful learning experience. Personalized mentoring by reviewing task progress, and providing timely feedback and guidance, whether online or in-person internships, paid or not, positively influence interns to make successful learning growth, engage in reflective learning, develop skills, achieve job satisfaction, get acclimated to professional norms, and stay on track of the career path (Chand & Deshmukh, 2019; Chesler et al., 2015; D'Abate et al., 2009; Jeske & Linehan, 2020; Liu et al., 2011). However, research finds not all mentors are equally helpful. Closely related to the effects of mentoring during a virtual internship could be the question of quality of communication mediated through digital technology between them. Studies of experiential online courses support this claim (e.g., Compton & Davis, 2010; Duncan & Barnett,

2009). In the context of a virtual internship, Teng et al. (2021) found that the quality of communication between mentors and team members is negatively associated with the sole use of impersonal communication tools (e.g., e-mail only), but positively associated with the use of both asynchronous and synchronous communication tools. Benefits also appear to vary by the style and level of communication among supervisor mentors and clients that mediates the quality of interactions between intern mentees and clients (Ruggiero & Boehm, 2016).

In addition, some prior research suggests that one key aspect of learning during an internship is autonomous learning that allows individuals to define problems and deal with solutions in complex situations (Jeske & Axtell, 2014; McHugh, 2017; Teng et al., 2021). While Taylor (1988) showed that autonomous interns tend to demonstrate a higher level of vocational concept, making a smooth transition from college to the workforce, Beenen and Rousseau (2010) cautioned that overemphasis on autonomy without direction and guidance from supervisors, especially in the context of remote environments, may undercut the developmental values of an internship.

Along with what and how interns learn, a closer look at how they are systematically supported can provide us with a better understanding of the quality of learning during an internship. Scholars have found that the presence of emotional, academic, social, and financial support enhance active engagement in experiential learning for all. For instance, several studies found a strong linkage between interns' job satisfaction and supervisors' emotional support through encouragement, care for well-being, and respect, regardless of in-person or remote internship settings (D'Abate et al., 2009; Johari & Bradshaw, 2008; Pretti et al., 2020). The significance of support from academic staff are recognized by both supervisors and students (Henry et al., 2001), as regular advising from academic staff helps student interns to clearly

understand tasks and needs of an intern employer, sometimes communicating with employers (Black & Bachman, 2007; Johari & Bradshaw, 2008; Narayanan et al., 2010).

Perhaps, financial support through monetary compensation is the most controversial issue, among other things, related to college internships. This is partly because uncompensated or poorly paid internships pose systematic, yet legitimate barriers for marginalized students that often make them opt-out from these professional development opportunities (Allen et al., 2013). This occurs even though internship experience is highly predictive of more job offers, higher starting salary, job satisfaction and performance, and intention to pursue a future career with the internship host (Jeske & Axtell, 2014; McHugh, 2017). Recent scholarship on paid/unpaid internships has attempted to explicate the dynamics around the paid internships including who enjoys paid internships, how, and why. Frenette and colleagues (2015), for instance, found that female college students in the arts tend to engage in internships more than their male peers; while male peers tend to take paid positions more than female peers. Allen et al. (2013) observed similar patterns for working-class college students, highlighting that unpaid internships (including none) to be taken, the duration, and the location of internships.

One interesting pattern found in recent studies on remote work is that flexibility with commuting and working hours is viewed as a form of compensation to some job applicants, such that they are willing to take a pay cut if they can maintain a flexible working status (Mas & Pallais, 2017). According to Kurter (2021), employers such as Google are keenly aware of such preference, planning to implement pay cuts for work-from-home employees. Extended to the virtual internship context, it remains to be seen how the scale of virtual internships will affect the

level of payment for virtual interns as well as in-person interns while attracting talented college students.

Social support through professional networking opportunities is another controversial area in terms of interns' learning, especially in a virtual internship. Liu et al. (2011) found that interns who actively network are more apt to have mentoring, which, in turn, creates more opportunities to engage in learning through sharing knowledge and experiences. Allen et al. (2013) advanced this issue from the perspective of cultural capital, arguing social activities naturally expose interns to tacit norms and expectations of a given organization, reinforcing the bonds between members with a similar cultural fit. Interns who experienced both face-to-face internships and computer-mediated interactions expressed concerns, Jeske and Axtell (2014) found, which led the researchers to speculate that social support could be underdeveloped in virtual internships, compared to conventional face-to-face internships.

The lack of social support in virtual internships may also bring counter effects to the learning of interns with different cultural norms or social capital, which may not necessarily lead to poor quality of learning and low satisfaction for them. Garcia (2009) reported an example of male-dominant, accounting firms where many socialization opportunities occur outside the workplace. Female interns felt they were conversely isolated from the professional networking opportunities, "miss[ing] opportunities to become known." (p. 20) Allen et al. (2013) also introduced the challenging learning experience of non-white immigrant interns that occurred due to the cultural distance between them and peers in affluent, white-dominant, creative industries. While these cases do not directly support the argument that female or non-white college interns may benefit from virtual internships with less social support, a dearth of social networking may

indirectly distribute opportunities for socialization for all that otherwise would have favored certain groups.

Finally, an implicit, yet important feature of a virtual internship that affects the quality of learning is that interns can gain work experiences without having to commute. Indeed, Ahmad (2020) observed accounting student interns in Malaysia were excited by not needing to commute to a physical internship location. This finding is clearly relevant to the U.S. context, as Rothman (2003) found lengthy commute time was rated as one of the least favored aspects of an internship by college students in one U.S. business school.

Prior studies have found close, negative relationships between a long commute time and employees' job satisfaction and productivity including absenteeism, late arrivals, and less task engagement due to physical and emotional exhaustion (e.g., van Ommeren & Gutierrez, 2011). These negative side effects may result in the affected employees having a poor reputation in the perspective of supervisors and colleagues. A study by D'Abate et al. (2009), however, found counter evidence of this in the context of traditional internship, suggesting no significant link between interns' job satisfaction and a reasonable commute. The authors reasoned that interns (i.e., in-person interns) could have already accounted for the commute distance by accepting the internship position, and thus those who expect a long commute might have opted-out of the position.

Given that the literature on remote work found comparative advantages of no commute in terms of employee well-being and productivity, the emergence of a virtual internship option may add different angles to the relationship between commute time and learning outcomes and satisfaction for those who engage in learning through internships. This may carry special weight to college students from marginalized social identities, economic backgrounds, and/or geographical regions as these groups of students often travel farther to their internship worksite than their peers.

Taken together, these findings show the potential of a virtual internship for traditionally marginalized students in terms of their access, learning quality, and outcomes of important work-based learning opportunities. However, empirical evidence in the context of equity and a virtual internship is largely limited, suggesting further studies are needed.

Data and Methods

This study's primary data sources are from 1) survey data of college students from 17 institutions, 2) geospatial data from Education Demographic and Geographic Estimates Program (EDGE), and 3) institution administrative data from the Integrated Postsecondary Education Data System (IPEDS).

Data and sample

The survey data were drawn from the 2020-2021 National Survey of College Internship Study Pilot, which was conducted in 17 universities or colleges in the U.S. As part of a largescale mixed methods study, the survey was designed to capture college students' internship participation, experience, barriers, life situations, and academic profiles during the pandemic. Unexpected disruptions and lockdowns due to the pandemic pushed some employers and students to design and engage in remote internship programs (Hammoud et al., 2022). Such rapid expansion underscores the importance of a systematic understanding of how this new type of work-based learning changed the internship ecosystem. Additionally, it enables this study to take advantage of variations from sufficient number of cases, allowing for consistent assessment of potentials of virtual internship in a single study. Convenience sampling was utilized to recruit these institutions; they were contacted via internship list-serve or Project Investigators' personal network. All the undergraduate students in these institutions were invited to take a web survey via an anonymous link that was sent at different timepoints between December 2020 to March 2021, resulting in an overall response rate of 4.53% (N=12,130). Thus, students' survey data reflects their internship-related experiences that occurred up to one year prior to the time of data collection at the earliest, with 96.8% of full samples having taken the survey between January 2021 through March 2021.

Among the respondents, most of the hybrid interns (i.e., interns that experienced a blend of remote and on-site work) reported that their internship programs were converted from one mode to another at a random stage of a program due to an exogenous event (i.e., COVID-19 pandemic) rather than it being the result of employers' design or interns' self-selection. Such variance poses significant challenges to capture the characteristics of a hybrid internship program type, for many employers and interns in hybrid internships may have limited success at developing a systematic mode of operating procedures uniquely suited to the hybrid internship circumstances. In addition, students who attended high school outside the U.S. (50 states and federal district) or returned to home countries at the time of data collection may have undergone different internship dynamics than the U.S. Therefore, both group of college students were excluded from the final analytical sample.

While samples in this study were recruited on a non-random basis, potentially limiting the generalizability of findings, this data set still provides information that can guide our understanding of college students' internship experience under different modalities. In addition, to alleviate the biases in estimation, some observable characteristics that may be related to sorting as independent variables were included.

121

As Table 12 shows, the final analytical sample consists of 11,009 college students from 12 public four-year institutions, three private four-institutions, one public two-year institution, and one private two-year institution. White people constitute the largest race and ethnicity group in our samples (62.88%, n=6,923), followed by Asian (13.02%, n=1,453), Hispanic (10.32%, n=1,136), and other race (8.1%, n=892). By gender, respondents identifying as female compose 69.62% (n=7,664) of the entire sample, while male group account for 27.38% (n=3,012), respectively. Among them, one in five (20.5%, n=2,257) respondents reported having taken at least one internship during the past 12 months. Interns who held an in-person position account for 50.78% (n=1,146) of total interns, which is slightly higher than the rate of those who did a virtual internship (49.22%, n=1,111).

Education Demographic and Geographic Estimates Program. Data on students' geographic and spatial contexts of internship are drawn from *Education Demographic and Geographic Estimates* (EDGE), a database that classifies institutions into one of four locale types: City, Suburban, Town, and Rural. The Zip Code Tabulation Areas (ZCTAs) locale file is useful to characterize college students' accessibility to an internship based on the rurality of the area where students' high school is located.

Integrated Postsecondary Education Data System. Data on the institution characteristics come from the Integrated Postsecondary Education Data System (IPEDS), one of the most comprehensive databases on postsecondary institutions collected by the National Center for Education Statistics (NCES). While I used institution-fixed effects in the model, I presented the characteristics of sample institutions in Table *12* to contextualize our understanding of the samples.

Key variables

Survey items were initially informed by extensive literature reviews (e.g., Beenen &

Rousseau, 2014; Jackson et al., 2022; McHugh, 2017; Varghese et al., 2012), and refined after a series of pre-testing and revisions based on expert reviews and focus group interviews.

Student characteristics	N (%)
Gender	````
Male	3,012 (27.38)
Female	7,664 (69.62)
Race	
Asian	1,453 (13.2)
Black	605 (5.5)
Hispanic	1,136 (10.32)
White	6,923 (62.88)
Other race	892 (8.1)
First-generation status	
First-generation students	2,821 (25.62)
Continuing-generation students	8,185 (74.35)
Enrollment type	
Full-time	9,803 (89.05)
Part-time	1,124 (10.21)
Employment type	
No employment	3,440 (31.25)
Part-time employment	6,439 (58.49)
Full-time employment	1,009 (9.17)
Caregivers' income level	
\$0-\$39.9k	2,817 (25.59)
\$40k-\$79.9k	1,969 (17.89)
\$80k-\$119.9k	1,989 (17.24)
\$120k or more	2,626 (23.85)
Not sure/not applicable	1,699 (15.43)
Major	
Arts & Humanities	734 (6.67)
Biosciences, Agriculture, & Natural Resources	1,231 (11.18)
Physical Sciences, Mathematics, & Computer	518 (4.71)
Social Sciences	944 (8.57)
Business	1,185 (10.76)
Communications, Media, & Public Relations	440 (4)
Education	378 (3.43)
Engineering	525 (4.77)
Health Professions	972 (8.83)
Other majors (not categorized)	3,900 (35.43)
Internship requirement	
Yes	704 (29.67)
No/not sure	1,669 (70.33)
Internship mode	
Virtual	1,111 (49.22)
Unsite	1,146 (50.78)
GPA	3.44 (0.54)

Table 12. Descriptive statistics of the sample

High school location	
Rural location	4 029 (36 6)
Urban suburban and town location	-,029 (30.0) 6 458 (58 66)
Unknown	522 (4 74)
	522 (1.77)
Institution characteristics	
Institution type	
Public, 4-year institution	7,837 (71.19)
Private, non-profit, 4-year institution	2,173 (19.74)
2-year institution	999 (9.07)
Minority-serving institution	
HSI	975 (8.86)
Non-HSI	10,034 (91.14)
Acceptance rate	76.91 (16.77)
Institution location	
Rural location	243 (2.21)
City	9,713 (88.23)
Suburban & Town	1,053 (9.56)
Enrollment	21,044.9 (11,258.28)
Socio-economic diversity	24.95 (13.17)
% Full-time students	86.42 (16.99)
Racial composition	
% Asian	9.04 (5.11)
% Black	6.13 (4.23)
% Hispanic	14.06 (13.88)
% White	58.39 (15.96)
N	11,009

*Note. For students' GPA, mean and standard deviation were presented.

HSI: Hispanic-Serving Institution. Socio-economic diversity: percent of full-time, first-time students receiving an income-based federal Pell Grant intended for low-income students.

Source: Institution information from IPEDS.

Internship participation and modality. Students' internship participation (1=yes, 0=no) was measured by one item on the survey asking: "In the past 12 months, have you participated in any internships?" In order to disambiguate the meaning of an internship, respondents were presented with a definition as follows (Hora, Chen, Parrott, & Her, 2020, p. 240):

"An internship is a position held within an established company or organization while also completing a college degree, certificate, or diploma program. It involves working in a position clearly designated as an "internship" by the host organization and performing tasks similar in nature and skill-level to tasks done by entry-level employees in the organization."

Those who have undergone internships were then asked to choose the modality of their internship engagement from three options: In-person internship, virtual internship (online internship), and other. As with the item on the internship participation, an operational definition of a virtual internship was presented as an internship that allows an intern to work remotely via digital technologies (e.g., laptop) without a physical appearance at the organization.

Internship program features. I assess key dimensions of an internship program pedagogical and contextual aspects—most of which were measured using validated instruments in the prior literature. Variables relevant to the pedagogical dimension of an internship program can be divided into two sub-domains pertaining to task-related pedagogical dimension and supervisory support and style dimension. I draw on interns' perception of task-related pedagogical practices during their internship in the domain of clarity of learning goals, academic relevance, and level of tasks. *Supervisory support and style*, another sub-domain of pedagogical aspects of an internship program, composed of site supervisor's mentoring on job duties and performance, site supervisors' emotional support, the use of autonomous learning strategies, support from academic staff, and professional networking opportunities. Contextual dimension of an internship program centered around the working conditions such as work hour, internship duration, and hourly wages.

Task-related pedagogical practices. Clarity of learning goals measures how clearly a respondent understood 1) what their goals were and 2) what tasks to be completed to achieve such goals (Beenen & Rousseau, 2010; McHugh, 2017). Each item was answered on a 5-point Likert-type scale, ranging from 1 (Not at all clear) to 5 (Extremely clear) and were averaged into an overall score (α =0.8) *Academic relevance* was measured by a single item on the survey that asked respondents to indicate the extent to which respondents' internship was related to their academic program on a 5-point Likert-type scale (1=Not at all related; 5=Extremely related). *Incorporation of soft skills development* was operationalized as whether the internship provided opportunities for developing the communication skills, teamwork, problem-solving skills, or supervising others (4 items; 1=None; 5=A great deal).

Supervisory support and style. Supervisor's mentoring on job duties and performance and use of autonomous learning strategies were measured by survey items that were adapted from McHugh's study (2017). While McHuch (2017) used five items to measure the extent of supervisor mentoring, I focused on four items asking the extent to which their supervisors directly engaged in respondents' task-related learning process (e.g., "How often did your supervisor give you feedback regarding job performance?"). Use of autonomous learning strategies was measured by two survey items (e.g., In this internship, how much freedom did you have to decide how to do your work?) on a 5-point Likert-type scale (1=None; 5=A great deal). Cronbach's alpha coefficient for task-related supervisor mentoring and autonomous learning were 0.83 and 0.81, respectively. The composite scores for each construct were computed by calculating the average score of the corresponding survey items. McHugh (2017) also provided the basis for assessing the extent to which student interns learned in the supportive and respectful working environment with their field supervisors. Respondents were asked to rate four items (e.g., In this internship, how much respect did you feel you received?") on a 5-point Likert type scale (1=None; 5=A great deal), which were averaged into a composite score for *supervisor support* (α =0.89). *Support from academic staff* was operationalized as the frequency of contact with a faculty member or academic advisor who was responsible for overseeing their internship experience during the internship (never, twice a month, less than once per month, monthly, twice a month, weekly, daily). *Professional networking opportunities* was measured based on one binary item asking "Do you think your internship experience expanded your professional network?"

Working condition. Concerning the financially supportive context, I perceived an item asking the amount of hourly compensation paid by the internship on the survey to align with this construct. After I assigned 0 to unpaid interns, I transformed the dollar values in hourly compensation to the log-scale so that the skewness of data is mitigated (Molina & Martin, 2018). Another contextual dimension pertains to the temporal work commitments that allow interns to learn. Interns can be distinguished from temporal employees in their work hours and duration. Respondents were asked to report the work hours while they were interning up to 100 hours per week. They also reported the number of weeks they participated in the internship up to 60 weeks. To improve the consistencies across survey instruments, I rescaled two measures by dividing each by 10.

Academic and career developmental values of an internship. Interns' learning outcomes from an internship were assessed along two separate dimensions: academic development and career development. Utilizing survey items developed by McHugh (2017) and Nghia and Duyen (2019), the research team identified 5 items that asked respondents to rate the extent to which the internship improved understanding of academic knowledge gaps and focus on studying and motivated them to look for more hands-on learning opportunities that promote their academic learning. The team also identified another 5 items that assessed the extent to which the internship enhanced the clarity of their career goals, learning of new career-related skills, and confidence in their ability to pursue future career opportunities. The former group of items was labeled as 'academic developmental value' and the latter as 'career developmental value,' and each group was averaged into two composite scores (α =0.88 for academic developmental value; α =0.89 for career developmental value).

Satisfaction. A respondent's perceived satisfaction with an internship was assessed using a single item asking: "How satisfied were you with your internship experience?" Respondents indicated their satisfaction level in a 5-point Likert type scale, ranging from 1 (=Not at all satisfied) to 5 (=Extremely satisfied).

Student characteristics. Because this study examines variations in students' internship attainment or internship experience by the interaction of college students' race and ethnicity, gender, socio-economic status, geographic profile, and internship modality, I constructed variables that indicate all these demographic characteristics. I identified the four largest *race and ethnicity groups*—Asian, Black, Hispanic, and white—in this study, each of which was dummy coded. The *Gender* variable was also operationalized as a binary construct: 1 indicating a female and 0 a male.

Students' *socio-economic status* was measured using two complementary indicators: *student employment status* (Arber et al., 2009; Aschan et al., 2013) and *parental income level*

(Machin & Vignoles, 2004). In prior literature, these measures often have been used as proxies for one's socio-economic status, but were found to have independent effects in those studies. Based on two survey items asking whether respondents were working for pay at a non-internship job and the number of working hours, student's employment status was distinguished into three categories: college students working at a paid non-internship job for more than 40 hours a week were classified as 'student with full-time employment,' while those working less than 40 hours a week as 'student with part-time employment' and those who do not work as 'student with no employment.' The parental income level variable was constructed from a single survey item asking respondents to choose the total gross annual income of their parent(s) from 12 categories, ranging from '0-\$19,999' to '\$200,000 and above.' Following the income group classification developed by Horowitz et al., (2020), I divided these respondents into four income groups: lowincome group defined as parental income as less than \$39.9k; middle-income group as between \$40k-\$119.9k; high-income group as \$200k or above. For respondents who refused to answer this question, I classified them into the 'unknown household income' category.

Students' *geographic profile* was classified as rural (=1) or non-rural area (=0) according to the location of the respondent's high school and current residence area. Given that college students living in geographically isolated rural communities often have limited accessibility to internship opportunities due to physical proximity, I applied a dummy coding for a rural location (=1) based on the location of respondents' high school.

Respondents' first-generation status was determined by the self-reported survey item asking whether the respondent is the first in their family to attend a four-year college/university and attain a bachelor's degree (1=first-generation student; 0=otherwise).

Analytical strategies

In the first part of this study, a series of multinomial logistic regressions with institution fixed effects were performed to examine whether students' underrepresented group membership is associated with their relative chance of doing: 1) an in-person internship vs. no internship (Model 1); 2) a virtual internship vs. no internship (Model 2); and 3) a virtual internship vs. an in-person internship (Model 3). Multinomial logistic regression approach is well-suited for addressing these questions because it allows for dependent variables with three or more nominal categories, modelling probability of being in a certain category compared to the reference category for each pair.

Students' internship engagement is modeled as a function of individual characteristics and institution specific context. Marginalized groups of students, primary interests in this study, were identified based on their socio-economic status (i.e., parental income and students' fulltime/part-time/no employment status, respectively), race/ethnicity (i.e., Asian, Black or Hispanic, and white), gender (i.e., female and male), and geographical location of their high school (i.e., rural vs. non-rural location). Black and Hispanic groups were combined due to the small sample sizes in each disaggregated category. For the same reason, other race category was created. Literature suggests that, of these categories, the marginalized populations of each have also been marginalized within the traditional in-person internship space yet may benefit from niche internship communities. To adjust for potential confounding factors stemming from individual demographic and background characteristics, the following variables also appear in all models: GPA, first-generation student status, past experience in other work-based learning programs such as coop and apprenticeship, grade level, age, and academic programs. Finally, a vector of dummy variables for all institutions (i.e., institution fixed effects) were introduced to the model to capture unobservable institution-specific factors that could explain an engagement in an internship of students in each institution. Standard errors were clustered at the institution level.

The key assumption of my approach is that virtual internship widens work-based learning opportunities to some marginalized college students if they are more likely than nonmarginalized peers to take a virtual internship relative to an in-person internship. Conversely, I consider the possibilities that virtual internships create additional marginalized identities, or that they further exacerbate disparities in access to work-based learning opportunities if some underrepresented students are less likely than overrepresented peers to engage in a virtual internship in comparison to an in-person internship. Positive coefficients for students' marginalized identities in Model 3 support the former assumption, whereas negative coefficients the latter. Nevertheless, the estimated coefficients in the Model 3 alone do not discern underlying source of comparative benefits (or risks) of each internship against others. For instance, the estimated coefficients for female students in Model 3 could be positive and significant in the case where female students are more likely than male peers to take an internship, whether implemented onsite or remotely.

Two preliminary models were generated to set the basis for this understanding. Model 1 yields evidence of underrepresented groups of students in a conventional in-person internship. Similarly, estimates from Model 2 reveal variations in how virtual internship serve students with different backgrounds. It is important to note that as affordances of virtual mode may bring some types of underrepresented students to new learning opportunities, it is also possible that constraints of virtual internships may strengthen old patterns of inequality or yield newly marginalized groups. An integrative review of these findings from three models contribute to our

understanding of potentials of virtual internship over in-person internship as an alternative route to ensure equitable access to work-based learning opportunities.

The second part focuses on student interns' learning experience during the internship. My approach is to identify, first, distinct profiles among student interns based on the three dimensions of learning experiences during the internship, described earlier in this chapter: 1) task-related pedagogical practices; 2) supervisory support and style; and 3) working condition. Combinations and interactions of different aspects of an internship, D'Abate (2010) and Gamboa et al. (2014) say, are what makes a difference in students' learning experience. Nevertheless, the utility of prevailing variable-centered approach often rests on the examination of relationships between each component in isolation (Weller & Bowen, 2020). A latent profile analysis (LPA), on the other hand, is one of the widely used statistical methods in detecting subgroups with homogeneous behavior patterns or characteristic patterns, as this enables us to consider complex combinations of behaviors or characteristics simultaneously (i.e., students' joint learning experiences in varying dimensions) (Hagenaars & McCutcheon, 2002; Keefer et al., 2012). Accordingly, I use an LPA—a person-centered, probability-based clustering approach—to identify qualitatively different intern groups across multiple dimensions of learning.

Latent profile analysis proceeded in two phases. First, the optimal number of intern profiles is identified through a series of exploratory process that find a model with the best model fit. Following determination process suggested by Magidson and Vermunt (2004) and Nylund et al. (2007), I fitted models with increasing number of latent classes (from 1 to 6 in this study) and evaluated the fit of each model. Three commonly used statistics were compared to the selection of the correct number of classes: Bayesian Information Criteria (BIC), Lo-Mendell-Rubin (LMR) adjusted likelihood test, and the percentage reduction in the likelihood ratio chi-square statistic (L^2) as opposed to the unconditional measurement model. Additionally, I reviewed theoretical interpretability of competing models to identify conceptually meaningful subgroups (B. O. Muthén & Muthén, 2000; Nylund et al., 2007). This stage also contains sorting each respondent into a latent cluster based on their estimated posterior class membership probabilities (Masyn, 2013).

The second phase of the analysis examines how virtual interns are distributed across different profiles of interns by comparing the shares of virtual interns in each group. This allows the researcher to systematically analyze whether virtual interns are prevalent in (a) particular latent intern cluster(s), helping to map out where a virtual internship is placed in the overall internship market. If unique learning profiles are found and they are significantly prevalent among virtual interns compared to in-person counterparts, this warrants further investigation of whether these online-based, new patterns of learning experiences (i.e., profile) are more (or less) likely to serve for historically underserved interns, compared with extant profiles. Even if virtual interns tend to have similar learning experiences as in-person interns, still, an important empirical question is left open: Do interns from historically marginalized backgrounds share similar learning experience and environments as those from privileged backgrounds?

To address this question, next I examine whether interns' marginalized identities predict the profile membership with different combinations of quality features, employing a multinomial logistic regression. As in the first research question, key independent variables related to students' marginalized background include race/ethnicity, gender, geographic location, and socio-economic status. Student or internship characteristics were controlled in all models such as first-generation status, academic programs (i.e., business and health), internship requirement, and
type of internship host organization (i.e., government internship, non-profit internship, and forprofit internship).

Then, I proceed to the last part of the study: examining whether these latent profiles are associated with psychosocial outcomes, including career or academic developmental values, and satisfaction. Ordinary Least Squares regressions were performed to statistically compare the mean values of internship outcomes between groups of different profiles. It must be cautioned that this should not be over-interpreted as a causal relationship. Rather, it should be interpreted as relative potentials and challenges of an emerging internship program—virtual internship—that could be linked to more equalized professional employment and life outcomes.

Results

Access to a virtual internship

Table 13 displays the results of multinomial logistic regression models on the question of whether virtual internships widen access to students with marginalized identities compared to inperson internships. For ease of interpretation, regression results are presented in the form of a relative risk ratio, *or* exponentiated coefficients estimated from the model. In the multinomial logistic regression, a relative risk ratio greater than 1 represents that a one unit change in the independent variable is associated with a greater risk of falling in the corresponding outcome group of interests compared to falling in the reference outcome group. On the contrary, a relative risk ratio less than 1 indicates that a one unit of change in the independent variable is associated with a greater risk of being a member of a reference group relative to that of a comparison group of interests. In this study, the reference groups are those who did not take an internship for models 1 and 2 and those who took an in-person internship for model 3, respectively. Importantly, an insignificant coefficient does not indicate that there is little relationship between the corresponding independent variable and being in the corresponding dependent group.

Aligned with previous literature, estimates in Model 1 reveal unequal access to in-person internship exists for some non-white racial/ethnicity groups and working student groups. For example, all else being equal, Asians are 46% less likely than their white peers to take an inperson internship. Similarly, Black or Hispanic students were 25% less likely to have experience working as an intern at an onsite environment. Not surprisingly, students working while in college all tend to get fewer in-person internship experiences compared to non-working students, with those working full-time 52% less and part-time 35% less. The likelihood of taking an inperson internship, however, did not differ by students' gender, rural high school background, or caregivers' income level.

Turning now to the findings for virtual internship participants versus non-interns in Model 2, it is seen that a virtual internship appeals to different groups of students. Indeed, Asian, Black, or Hispanic students were about 26 to 28% more likely than their white counterparts to take an internship program online, after adjusting for students' background characteristics and institution-fixed effects. On the other hand, the likelihood of taking a virtual internship for students from rural high schools was significantly lower by 18%, compared to their non-rural counterparts. For working students, the relative risks of gaining internship experience in a virtual setting are much lower for those working full-time jobs by 59% and part-time jobs by 30%, respectively, relative to non-working counterparts.

Next, Model 3 in Table *13* shows, net of other individual and institutional influences, how virtual interns differed from in-person interns with respect to the marginality associated with racial/ethnic demography, geographic location, and socio-economic status. One of the

categories that shows the biggest and strongest contrast between the two types of interns was found in interns' racial and ethnic identity. To illustrate, Asians' likelihood of engaging in a virtual internship versus an in-person one was much more than twice (238%) as high as their white peers. Black or Hispanic students were also 69% more likely to participate in a remotebased internship program than an onsite-based one, compared to white counterparts. Geographical location turns out to be another substantial determinant that characterizes virtual versus in-person student interns. Surprisingly, students from rural backgrounds tend to take a virtual internship 31% less often than their peers from non-rural backgrounds. Yet, no statistical difference was found among students with varying levels of socioeconomic status in their internship work settings.

In sum, these findings contribute evidence that a virtual internship carries noticeable advantages for ensuring work-based learning opportunities to Asian, Black, or Hispanic students—historically marginalized racial and ethnic communities in traditional internship settings. At the same time, a virtual internship seems to play no more than a limited role in terms of serving working students. Both virtual and in-person internships remain somewhat unreachable for most of students with full-time or part-time employment relative to non-working students. Furthermore, the new access gap for students from rural and non-rural areas was more pronounced among virtual interns than in-person peers.

Internship profiles, modes, and marginalized identities

To further explore the democratization potentials of a virtual internship from the perspective of quality of learning, I identified the appropriate number of profiles of interns based

VARIABLES	Model 1	Model 2	Model 3
	In-person	Virtual	Virtual vs. in-
	vs.	VS.	person
	no internship	no internship	internship
Race/ethnicity			
Asian	0.54***	1.28**	2.38***
	(0.08)	(0.12)	(0.35)
Black or Hispanic	0.75*	1.26**	1.69***
	(0.09)	(0.09)	(0.22)
Other race	0.97	1.08	1.11
	(0.09)	(0.11)	(0.13)
Gender			
Female	1.06	1	0.94
	(0.06)	(0.05)	(0.07)
Geographic background			
Rural background	1.18	0.82*	0.69***
	(0.11)	(0.07)	(0.07)
Household income			
Middle income	0.9	0.87	0.97
	(0.09)	(0.07)	(0.13)
High income	1.11	1.15	1.04
	(0.11)	(0.13)	(0.06)
Unknown income	0.86	1.02	1.18
	(0.07)	(0.1)	(0.12)
Employment status			
Weekly work hours > 20	0.48***	0.41***	0.85
	(0.08)	(0.03)	(0.12)
Weekly work hours ≤ 20	0.65***	0.7***	1.09
	(0.06)	(0.06)	(0.16)
Other individual characteristics		e destadada	
Advanced academic standing	3.45***	6.45***	1.8/***
	(0.24)	(0.59)	(0.17)
Other work-based learning experiences	2.51***	1.42	0.5/***
•	(0.42)	(0.27)	(0.09)
Age		0.99	0.99
	(0.01)	(0.02)	(0.01)
First-generation status	0.97	0.68^{***}	0./***
	(0.09)	(0.02)	(0.07)
STEMM major	1.16	0.51^{*}	0.44^{**}
	(0.1)	(0.14)	(0.11)
Business major	1.55*	1.5/***	1.01
A set maios	(0.1/)	(0.15)	(0.14)
Art major	0.84	0.92	1.1
	(0.17)	(0.12)	(0.3)

Table 13. Multinomial logistic regression estimates on access to an internship

GPA	1.22	2.06*	1.69**
	(0.19)	(0.59)	(0.28)
Constant	0.01***	0^{***}	0.04***
	(0)	(0)	(0.02)
Observations	10,289	10,289	10,289
Institution fixed effects	Y	Y	Y
Pseudo R2	0.16	0.16	0.16

Note. *** p<0.001, ** p<0.01, * p<0.05

Students in advanced academic standing includes those who are in the second half of their programs (e.g., junior or senior standing), making progress toward graduation. Standard errors are clustered at the institution

Model	BIC	L ²	% reduction in L^2	LMR <i>p</i> -value	Entropy
One cluster	72342.59	36078.63			
Two clusters	69124.42	34419.36	4.6	< 0.001	0.92
Three clusters	65894.96	32754.43	9.21	< 0.001	0.98
Four clusters	64341.95	31927.74	11.52	< 0.001	0.95
Five clusters	63745.28	31579.21	12.47	< 0.001	0.94
Six clusters	63149.34	31231.05	13.44	< 0.001	0.91
Seven clusters	62590	30901.19	14.35	< 0.01	0.92

Table 14. Goodness-of-fit Statistics for Cluster Models of Interns

Note. N=2,257. BIC = Bayesian information criterion; LMR = Lo-Mendell-Rubin likelihood ratio test.

on 11 indicators that represent interns' learning experiences. Table 14 presents the estimated fit indices for each LPA model.

I found that BIC and L^2 continuously decreased as the number of clusters increased, suggesting that the model with seven clusters was superior to other models with respect to the goodness of fit. While models with lower values on the BIC and L^2 are generally selected when identifying the appropriate number of clusters, there was only a minimal additional reduction in L^2 for models with four or more latent profiles. With these statistical and parsimony rationales, I narrowed down the range of potential profiles to models with four through seven profiles and considered the theoretical tenability of each model. Models with four or five latent profiles demonstrated theoretical alignment over the ones with six or more profiles, with the five-profile solution capturing greater nuance in the facet of interns' learning environments. Accordingly, I retained the five-latent profile model as the best solution.

While it is often encouraged to standardize all indicators prior to analysis if they are measured on different scales for ease of interpretation, I reported results derived from original scores—one of the affordances of mixture modeling where the transformation of variables is not required (Pastor et al., 2007). This approach is unlikely to affect the interpretability of findings or might even carry comparative benefits since the face values of indicators in this study were not only self-explanatory but also fell within similar score ranges. This resulted in a rich understanding of the absolute quantitative level of each indicator and qualitative characteristics of each profile through a visualized typography of profiles. Nevertheless, there was one exceptional indicator that was measured in a binary manner, students' professional networking contacts during an internship, which suggests that two separate visualizations are required.

Figure 8 and Figure 9 reveal five immediately noticeable intern groups with different learning qualities, with the key differences in combinations of three areas: 1) task-related pedagogical practices; 2) supervisory support; and 3) working environment (described earlier). The model indicators used for evaluating the quality of internships are listed on the x-axis whereas mean scores for each indicator are plotted on the y-axis. Each profile was labeled with descriptors that summarized different response patterns to the indicators.

Two general patterns stand out from the Figure 8 and Figure 9. The first group consists of profiles represented in solid lines and positioned in the upper part of the y-axis of Figure 9 which are differentiated from the second group in terms of their task characteristics and supervisory support. Despite the challenges imposed by the global pandemic, large shares of internships appear to serve as a promising learning venue for student interns. Such is the case with three

profiles of interns (e.g., Profiles 1-3) who went through highly structured, supportive programs, comprising approximately a total of 81% of sampled interns. Typical interns in these profiles had quality educational experiences in which they could learn through academically relevant tasks combined with clearly stated goals. These internship programs, interns report, seem to provide them with opportunities to hone a set of soft skills—communication skills, teamwork, and problem-solving skills—as they interact with team members on a variety of projects. For the most part, they were also placed in contexts in which they felt supported by site supervisors professionally, socially, and emotionally.

These three profiles, however, varied across interns' response to the support from their faculty members or academic advisors as well as working conditions (e.g., internship duration, working hours, and compensation). For instance, Table 15 shows that interns in Profile 1, the largest cluster, representing 42% of all respondents, worked over 30 hours per week for roughly 13 weeks and got paid \$17.48 (=exp (2.86)) hourly on average. The second-largest intern subgroup (33%)—Profile 2— showed the opposite patterns from those in Profile 1, regarding the working conditions for learning: They tended to do nearly unpaid work for less than 20 hours over 18 weeks. They also indicated they were in contact with internship-related academic staff less than once per month, whereas their peers in Profile 1 were 1-2 times per month. Still, interns in Profile 3 appeared to lie between Profile 1 and Profile 2, especially with respect to the extent of support from faculty supervisors and work hours. While Profile 3 comprises relatively fewer interns (6% of the sample) compared with the two other intern subgroups above, the members of this group not only made nearly a year-long commitment (49 weeks) but also were paid an average compensation of \$13.99 (=exp (2.64)) per hours worked.



Figure 8. Proportion of interns reporting professional networking opportunities by intern profile

Note. Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions.

The remaining two subgroups of interns, on the other hand, fare poorly across all the pedagogical practices in the internship ecosystem. Representing 10% (Profile 4) and 7.3% (Profile 5) of the entire respondents, interns in these subgroups commonly reported that they were expected to complete somewhat unclear goals with little latitude to accomplish them, compared to members in Profile 1-3. They also evaluated if their internship programs offered a moderate chance of developing soft skill, tasks that were weakly tied to their academic program, as well as moderate emotional support from site supervisors. The quality of internship for these groups continued to be somewhat disappointing, distinctively in professional cares such as managing tasks, meeting goals and deadlines, and reflecting the performance of both site supervisors and faculty members, which all interns may need to grow as both a future employee and a lifelong learner. As in Profile 1-3, these two groups contrasted with each other mainly in terms of the working conditions including work hours and hourly wages. At least, interns in Profile 4 appeared to learn in a relatively favorable working conditions, working for 30 hours per



Figure 9. Latent profiles based on the internship quality indicators

Note. Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions.

Work hours in hours/10; Duration in weeks/10; Hourly wages transformed in log.

Latent profile	Ave	erage	Prot	file 1	Pro	file 2	Pro	file 3	Pro	file 4	Prot	file 5
	M	(SD)	M ((SD)								
Task characteristics												
Soft skills incorporation	3.58	(0.62)	3.74	(0.4)	3.74	(0.47)	3.8	(0.39)	2.92	(0.73)	2.67	(0.8)
Goal clarity	3.83	(0.87)	4.06	(0.69)	4.05	(0.7)	3.75	(0.82)	3.02	(0.91)	2.76	(0.98)
Academic relevance	3.77	(1.11)	3.9	(1.05)	3.96	(1.01)	3.76	(1.1)	3.18	(1.14)	2.93	(1.16)
Professional networking building	0.89	(0.32)	0.96	(0.2)	0.95	(0.21)	0.93	(0.25)	0.71	(0.45)	0.37	(0.48)
Supervisory support												
Autonomy	4.13	(0.92)	4.31	(0.76)	4.32	(0.76)	4.31	(0.83)	3.24	(1.07)	3.35	(1.15)
Faculty mentoring	2.27	(1.4)	1.9	(1.31)	2.91	(1.33)	2.39	(1.58)	1.67	(1.08)	2.25	(1.3)
Site supervisor task mentoring	3.46	(0.74	3.68	(0.51)	3.67	(0.57)	3.54	(0.57)	2.56	(0.73)	2.32	(0.67)
Site supervisor emotional support	4.24	(0.85)	4.51	(0.55)	4.49	(0.57)	4.39	(0.66)	3.07	(0.91)	2.95	(0.86)
Working condition												
Work hours	2.54	(1.47)	3.32	(1.27)	1.73	(1.25)	2.14	(1.41)	2.94	(1.32)	1.58	(1.09)
Duration	1.75	(1.31)	1.33	(0.7)	1.8	(1.25)	4.98	(0.99)	1.48	(0.85)	1.34	(0.91)
Hourly wages	1.66	(1.41)	2.86	(0.37)	0.01	(0.13)	2.63	(0.32)	2.75	(0.35)	0.01	(0.08)
N (%)	2257		949	(42)	760	(33)	149	(6)	232	(10)	166	(7)

Table 15. Means and standard deviations of all internship quality indicators by latent profiles

week and getting paid for their labor as high as \$15.66 (=exp (2.75)). However, this was not the case for interns in Profile 5; They generally took an uncompensated, part-time position of 16 hours. Notably, Profile 5 had a considerably small percentage of interns (=37%) indicating limited opportunities to build professional relationships and network at an internship, while 72% to 96% of interns in other profiles did.

I further analyzed these profiles to learn more about how virtual interns are distributed across the five internship segments characterized by a different mix of instructional quality and structural learning environment. The data in Table 16 shows the shares of virtual and in-person interns not only for the full sample but also for each subgroup. Overall, the percentage of virtual interns in Profile 1, 2, and 4 resembled an average pattern for the whole group, with nearly half (49.22%) of respondents having had taken a virtual internship. However, the prevalence of virtual interns in Profile 3 and 5 was in sharp contrast, as the quality of learning programs and working conditions in these profiles rest at the opposite ends of a continuum. While Profile 5 consist of a disproportionately excessive share of virtual interns (63.86%) and relatively small share of inperson interns (36.14%), Profile 3 had 35.57% of virtual interns and 64.43% of in-person interns. The proportions of profiles within each internship mode in Table 16 also confirms this finding: Among all virtual interns, the share of Profile 5 (9.54%) is almost twice that of in-person interns (5.24%) in the same profile. Given that poorly designed instructional practices and harsh working conditions were dominant in Profile 5, a relatively high concentration of virtual interns here suggests that a virtual internship may play limited roles for college students in bridging the gap between learning and finding a career, relative to an in-person internship.

Are there systematic relationships between interns' marginalized identities and their predicted profile membership?

Internship mode	Average	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
		(42.09)	(33.67)	(6.6)	(10.28)	(7.35)
Virtual	49.22	47.68	50.79	35.57	48.71	63.86
		(40.77)	(34.74)	(4.77)	(10.17)	(9.54)
In-person	50.78	52.32	49.21	64.43	51.29	36.14
		(43.37)	(32.64)	(8.38)	(10.38)	(5.24)
Asian		10.84	14.61	10.74	14.66	21.08
		(34.45)	(37.12)	(5.35)	(11.37)	(11.71)
Black or Hispanic		10.11	11.58	14.77	5.6	10.84
		(40.51)	(37.13)	(9.28)	(5.49)	(7.59)
White		72.32	65.92	64.43	70.69	62.05
		(44.29)	(32.3)	(6.19)	(10.57)	(6.64)
Female		62.11	79.87	69.8	56.9	80.72
		(37.65)	(38.74)	(6.64)	(8.42)	(8.55)
Rural background		40.84	33.03	40.27	41.81	32.53
		(40.62)	(33.72)	(7.29)	(10.55)	(7.81)
Middle income		32.84	34.08	37.58	10.55	7.81
		(32)	(45.41)	(4.71)	(7.53)	(10.35)
High income		35.89	24.21	28.86	35.78	21.69
		(49.64)	(26.78)	(6.26)	(12.08)	(5.24)
Weekly work hours	> 20	14.32	25.39	13.42	13.79	26.51
		(32)	(45.41)	(4.71)	(7.53)	(10.35)
Weekly work hours :	≤ 20	40.95	43.68	35.57	45.26	43.98
-		(40.86)	(34.87)	(5.57)	(11.03)	(7.67)

Table 16. Distribution of virtual or in-person interns by profiles

Note. Virtual intern: n = 1,111; In-person intern: n = 1,146. The proportion of profiles per identity is presented in parenthesis.

Having identified the characteristics of internships at different quality levels and a relative position of a virtual internship, I examined who is more likely to take an internship of low quality versus high quality. Table *17* presents the means of the quality indicators for each profile and the results from multinomial logistic regressions in which Profile 5 is served as a reference group. For ease of interpretation, I describe these relationships in relative risk ratios (RRR). I found that, all else being equal, Asian, female, or students working more than 20 hours weekly are still disproportionately concentrated in qualitatively disadvantaged internship programs, Profile 5. Asians, for example, had 53-55% less of a chance than white peers to belong to Profile 1 or 3—an internship where decent learning quality and compensation are commonly

reported, as opposed to Profile 5. They were also 39% less likely to be in the Profile 4 versus profile 5 when compared to white peers. The pattern is the same for female students who were also heavily distributed in Profile 5 than in Profile 1, 3, or 4, with the odds of belonging to Profile 5 to the other three profiles being lower by 45% to 68% relative to their male peers. A high concentration in Profile 5 versus other three profiles is even steeper for students whose working hours exceeded 20 hours per week. When contrasting students with or without heavy work obligations, the difference in the likelihood of falling into Profile 3 versus Profile 5 between them was the largest at 74%, followed by a 63% of a difference in Profile 1 versus Profile 5 and 55% for Profile 4 versus Profile 5. Interns' household income or part-time employment (no more than 20 hours per week) also partially differentiated between the latent profiles of interns. As predicted, students from a high-income family are 89% more likely than their low-income counterparts to belong to Profile 1 versus Profile 5. Compared to students without employment, the likelihood of student part-time workers' being in a Profile 3 intern group versus Profile 5 is 51% lower. Students' geographic background, however, appears to have little control over their internship membership assignment. Finally, although not the focus of the present study, students' advanced academic class standing (i.e., juniors or seniors) or voluntary internship (i.e., non-mandatory internship) also play a role in the likelihood of being in Profile 1, 3, or 4—the type of quality internships that provide either meaningful educational experiences or fair working conditions.

How are virtual internship and internship profiles associated with the internship's developmental value and satisfaction?

Given that learning during the internship may shape the educational outcomes, the uneven distribution of marginalized students across varying internship modes and learning

	Profile 1		Profile 2		Pro	Profile 3		Profile 4	
VARIABLES	Mean	RRR	Mean	RRR	Mean	RRR	Mean	RRR	Mean
Race/ethnicity									
Asian	0.11	0.45**	0.15	0.76	0.11	0.47*	0.15	0.61*	0.21
Black or Hispanic	0.10	1.17	0.12	1.00	0.15	1.46	0.06	0.68	0.11
Other race	0.07	1.16	0.08	1.17	0.10	1.85	0.09	1.58	0.06
Gender									
Female	0.62	0.40***	0.80	0.90	0.7	0.55*	0.57	0.32***	0.81
Geographic background									
Rural background	0.43	1.35	0.35	0.92	0.41	1.29	0.43	1.39	0.35
Household income									
Middle income	0.33	1.18	0.34	0.97	0.38	1.18	0.35	1.17	0.36
High income	0.36	1.89*	0.24	1.19	0.29	1.52	0.36	1.74	0.22
Income unknown	0.14	1.32	0.14	1.05	0.11	0.99	0.12	1.06	0.15
Employment status									
Weekly work hours > 20	0.14	0.37***	0.26	0.91	0.13	0.26***	0.14	0.45*	0.27
Weekly work hours ≤ 20	0.41	0.68	0.44	0.94	0.36	0.49*	0.45	0.90	0.44
Other individual characteristics									
Advanced academic standing	0.87	3.07***	0.79	1.45	0.85	2.75**	0.85	2.84***	0.68
Other WBL experiences	0.12	0.80	0.20	1.20	0.18	1.24	0.13	0.89	0.16
Age	22.15	1.00	22.48	1.02	22.45	1.01	21.77	0.96	21.83
First-generation status	0.15	1.19	0.23	1.33	0.28	2.138	0.14	1.22	0.19
STEMM major	0.18	0.84	0.24	0.92	0.28	1.42	0.19	0.93	0.28
Business major	0.16	2.18*	0.08	1.11	0.14	2.24	0.19	2.64*	0.08
Art major	0.03	0.44*	0.10	1.29	0.07	1.06	0.02	0.33*	0.08
GPA	3.54	0.64	3.56	0.83	3.54	0.66	3.56	0.72	3.60
Mandatory internship	0.22	0.50***	0.43	1.21	0.26	0.54*	0.21	0.51**	0.34
Constant		28.21**		4.67		2.45		13.45	
Observations	950		760		149		232		166

Table 17. Means and relative risk ratio of predictors on latent profiles

Note. *p < .05. **p < .01. ***p < .001. RRR = Relative Risk Rati

environments suggest an investigation of the link between the educational outcomes and a virtual internship. Academic as well as career developmental values of virtual interns were consistently lower compared to those of in-person interns (see Model 1 in Table 18 and Model 4 in Table 19). As seen in the Model 7 in Table 20, the net of students' background control variables, the mean satisfaction of virtual interns was significantly lower than that of in-person interns. While these findings suggest that a virtual internship carry relative educational disadvantages over an in-person internship, these models do not conclusively explain the source of its inferior outcomes. Lower outcomes of virtual interns may be attributable to their high concentration in profiles with poor learning programs captured along task characteristics, supervisory support, and working conditions—design and implementation elements that are likely to be improved through proactive and collaborative efforts of host organizations and academic staff. To exploit a such possibility, I added dummy variables for each profile of interns to the previous models.

Models 2, 5, and 8 in Table 18-20 show that virtual interns still tend to report lower levels of internship developmental values and satisfaction, even though the differences between virtual interns and in-person interns were slightly dampened when intern profiles were taken into account. At the same time, results of hierarchical regression models also reveal that the bundle of task characteristics, supervisory support, and working conditions do contribute to higher internship outcomes. Interns of Profile 4, who are similarly grappling with poor work design but indeed at least paid for their long work hours—reported, albeit small, significantly higher internship developmental values and satisfaction than their counterparts of Profile 5, the reference group. Contrasting Profile 5 and three intern groups who took a highly structured program (i.e., Profiles 1, 2, and 3) reveals all significant and much larger gaps in perceived internship outcomes. In general, the largest gaps across all outcomes occurred with Profile 1, with the exception of the academic developmental value being highest reported by Profile 2. In a separate regression analysis for the covariate adjusted mean outcomes by profiles (Table 21), there was not much difference in perceived academic values among interns in Profiles 1, 2, and 3. However, interns in profile 1 reported a significantly higher career developmental value than their peers in Profile 2 or 3. For internship satisfaction, there exist a significant difference between interns in Profile 1 and 3, but neither between ones in Profiles 1 and 2, nor Profile 2 and 3.

In the Models 3, 6, and 9, I further attempted to isolate the heterogeneity across interns' marginalized identities from the overall association between the internship mode and outcomes. This specification shows that interns' learning outcomes in a virtual workplace setting are clearly racialized. I find while White interns assessed the academic value of a virtual internship is substantially lower than in-person one, the gap in academic outcomes between the modes is significantly smaller for Asian, Black, and Hispanic students. Concerning the career developmental value of a virtual internship, the mean difference between a virtual and an inperson mode is not statistically significant for most of the race and ethnicity groups, with virtual interns of the other race group feeling substantially higher positive career developmental potentials, relative to their onsite peers. Interns' satisfaction followed a similar pattern of overall insignificantly higher satisfaction with their experience in comparison to their peers who took a site-based internship.

Together, these results reveal that the significant gaps between interns of Profile 5 and others in all models, net of learning qualities and student background characteristics, could be driven by within-profile differences, suggesting that the disproportionate sorting of experiential learning programs by internship mode may not be the only source of unequal internship outcomes. More importantly, the results suggest that the developmental potentials or satisfaction of a site-based program seem to be higher than a remote one, but non-white interns find more advantages in remote work than the place-based internships.

	Academic developmental value					
	Model 1:	Model 2:	Model 3:			
	Virtual	Virtual	Virtual + profile			
	+ background	+ profile	+ interaction			
	variables	+ background	terms +			
		variables	background			
			variables			
Virtual internship	-0.25*** (0.04)	-0.2*** (0.04)	-0.3* (0.13)			
Asian	-0.02 (0.07)	0.03 (0.06)	-0.24* (0.1)			
Black or Hispanic	0.15* (0.07)	0.11(0.07)	-0.31 (0.13)			
Other race	0.03 (0.08)	0.03 (0.07)	0.26 (0.15)			
Female	0.16** (0.05)	0.15*** (0.04)	0.2*** (0.00)			
Rural background	0.03 (0.05)	0.05 (0.04)	0.06 (0.45)			
Medium income	-0.01 (0.06)	-0.01 (0.05)	0.07 (0.11)			
High income		-0.06 (0.06)	0.08 (0.11)			
Income unknown		-0.11 (0.07)	-0.19 (0.14)			
Weekly work hours > 20		-0.03 (0.05)	-0.08 (0.07)			
Weekly work hours ≤ 20		-0.01 (0.04)	-0.04 (0.06)			
Profile 1		1.28*** (0.08)	1.27*** (0.08)			
Profile 2		1.29*** (0.08)	1.28*** (0.08)			
Profile 3		1.16*** (0.1)	1.15*** (0.1)			
Profile 4		0.36*** (0.09)	0.35*** (0.09)			
Virtual*Asian			0.43** (0.12)			
Virtual*Black or Hispanic			0.31* (0.13)			
Virtual*Other race			0.26 (0.15)			
Virtual*female			-0.13 (0.08)			
Virtual*rural background			0.06 (0.08)			
Virtual*medium income			0.07 (0.11)			
Virtual*high income			0.08 (0.11)			
Virtual*income unknown			-0.19 (0.13)			
Virtual*weekly work hours > 20			0.11 (0.32)			
Virtual*weekly work hours ≤ 20			0.05 (0.58)			
Constant	3.62*** (0.23)	2.56*** (0.22)	2.61*** (0.23)			
Observations	2,129	2,129	2,129			
Adjusted R ²	0.05	0.23	0.23			

Table 18. Hierarchical regression for internship academic developmental outcome

Note. *** *p*<0.001, ** *p*<0.01, * *p*<0.05. Reference profile = Profile 5

Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions. Student background variables include previous work-based learning experience, age, advanced standing, first-generation status, major (STEMM, business, and art), GPA, and mandatory internship.

	Career developmental value				
	Model 4:	Model 5:	Model 6:		
	Virtual	Virtual	Virtual + profile		
	+ background	+ profile	+ interaction terms		
	variables	+ background	+ background		
		variables	variables		
Virtual internship	-0.2*** (0.04)	-0.13*** (0.04)	-0.18 (0.12)		
Asian	-0.12 (0.06)	-0.06 (0.05)	-0.18* (0.09)		
Black or Hispanic	0.16 (0.07)	0.11 (0.06)	-0.00 (0.09)		
Other race	-0.02 (0.08)	-0.01 (0.07)	0.27 (0.13)		
Female	0.09 (0.04)	0.11** (0.04)	-0.08 (0.07)		
Rural background	0.03 (0.04)	0.03 (0.04)	0.04 (0.07)		
Medium income	-0.01 (0.06)	-0.02 (0.05)	-0.02 (0.1)		
High income	0.04 (0.06)	0.00 (0.05)	-0.03 (0.07)		
Income unknown	-0.03 (0.07)	-0.06 (0.06)	-0.1 (0.12)		
Weekly work hours > 20	-0.11 (0.06)	-0.04 (0.05)	0.08 (0.07)		
Weekly work hours ≤ 20	-0.08 (0.04)	-0.04 (0.04)	0.07 (0.08)		
Profile 1		1.54*** (0.07)	1.54*** (0.07)		
Profile 2		1.39*** (0.07)	1.39*** (0.07)		
Profile 3		1.39*** (0.09)	1.39*** (0.09)		
Profile 4		0.50*** (0.08)	0.49*** (0.08)		
Virtual*Asian			0.21 (0.11)		
Virtual*Black or Hispanic			0.21 (0.12)		
Virtual*Other race			0.27* (0.13)		
Virtual*female			-0.08 (0.07)		
Virtual*rural background			0.04 (0.07)		
Virtual*medium income			-0.02 (0.07)		
Virtual*high income			-0.02 (0.07)		
Virtual*income unknown			-0.1 (0.12)		
Virtual*weekly work hours > 20			-0.08 (0.1)		
Virtual*weekly work hours ≤ 20			-0.07 (0.06)		
Constant	3.90*** (0.22)	2.61*** (0.2)	2.65*** (0.21)		
Observations	2,129	2,129	2,129		
Adjusted R ²	0.03	0.28	0.28		

Table 19. Hierarchical regression for internship career developmental outcome

Note. *** *p*<0.001, ** *p*<0.01, * *p*<0.05. Reference profile = Profile 5.

Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions. Student background variables include previous work-based learning experience, age, advanced standing, first-generation status, major (STEMM, business, and art), GPA, and mandatory internship.

	Satisfaction				
	Model 7	Model 8.	Model 9		
	Virtual	Virtual	Virtual + profile		
	+ background	+ profile	+ interaction terms		
	variables	+ background	+ background		
		variables	variables		
Virtual internship	-0.25*** (0.04)	-0.16*** (0.04)	-0.1 (0.12)		
Asian	-0.21** (0.07)	-0.13* (0.05)	-0.2* (0.09)		
Black or Hispanic	0.09 (0.07)	0.03 (0.06)	-0.09 (0.09)		
Other race	-0.05 (0.08)	-0.04 (0.07)	-0.14 (0.05)		
Female	0.05 (0.05)	0.06 (0.04)	0.12* (0.05)		
Rural background	-0.01 (0.04)	-0.00 (0.04)	-0.04 (0.05)		
Medium income	-0.01 (0.06)	-0.02 (0.05)	0.00 (0.07)		
High income	-0.04 (0.06)	-0.01 (0.05)	-0.04 (0.07)		
Income unknown	0.02 (0.07)	-0.02 (0.06)	0.02 (0.08)		
Weekly work hours > 20	-0.07 (0.06)	0.01 (0.05)	0.07 (0.07)		
Weekly work hours ≤ 20	-0.03 (0.05)	0.02 (0.04)	0.03 (0.06)		
Profile 1		1.82*** (0.07)	1.81*** (0.07)		
Profile 2		1.66*** (0.07)	1.66*** (0.07)		
Profile 3		1.73*** (0.09)	1.73*** (0.1)		
Profile 4		0.59*** (0.08)	0.58*** (0.08)		
Virtual*Asian			0.11 (0.11)		
Virtual*Black or Hispanic			0.23* (0.12)		
Virtual*Other race			0.21 (0.13)		
Virtual*female			-0.13 (0.07)		
Virtual*rural background			0.06 (0.07)		
Virtual*medium income			-0.05 (0.1)		
Virtual*high income			0.06 (0.1)		
Virtual*income unknown			-0.07 (0.12)		
Virtual*weekly work hours > 20			-0.14 (0.1)		
Virtual*weekly work hours ≤ 20			-0.04 (0.08)		
Constant	3.75*** (0.22)	2.61*** (0.2)	2.23*** (0.21)		
Observations	2,129	2,129	2,129		
Adjusted R ²	0.02	0.35	0.35		

Table 20. Hierarchical regression for internship satisfaction

Note. ***p < 0.001, **p < 0.01, *p < 0.05. Reference profile = Profile 5.

Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions. Student background variables include previous work-based learning experience, age, advanced standing, first-generation status, major (STEMM, business, and art), GPA, and mandatory internship.

Latent profiles	Academic developmental value		Ca developm	reer ental value	Satisfaction	
Profile 4 (reference)						
Profile 1	0.91***	(0.06)	1.04***	(0.06)	1.23***	(0.06)
Profile 2	0.92***	(0.07)	0.89***	(0.06)	1.08***	(0.06)
Profile 3	0.79***	(0.09)	0.89***	(0.08)	1.14***	(0.08)
Profile 3 (reference)						
Profile 1	0.12	(0.08)	0.15*	(0.07)	0.09	(0.07)
Profile 2	0.13	(0.08)	-0.01	(0.07)	-0.07	(0.07)
Profile 2 (reference)						
Profile 1	-0.01	(0.05)	0.15***	(0.04)	0.16***	(0.04)

Table 21. Covariate adjusted internship outcomes by profile membership

Note. ***p < 0.001, **p < 0.01, *p < 0.05. Reference profile = Profile 5. Profile 1 (42%) = a highly structured internship with favorable working conditions (short-term); Profile 2 (33%) = a highly structured internship with challenging working conditions; Profile 3 (6%) = a highly structured internship with favorable working conditions (yearlong); Profile 4 (10%) = an in-person centered, self-directed internship with favorable working conditions; Profile 5 (7.3%) = a virtual-centered, self-directed internship with challenging working conditions. Student background variables include previous work-based learning experience, age, advanced standing, first-generation status, major (STEMM, business, and art), GPA, and mandatory internship.

Discussion and Conclusion

Many employers have switched their internship programs to a virtual mode in various formats without understanding how this newfound transition will play out for college students who are underrepresented in the traditional work-based learning ecosystem. While the new format appears to be persisting in the post-pandemic area, gaining greater favor, there has been very little research that assesses its democratic potentials over the conventional in-person internship. The present paper has attempted to fill this gap by examining virtual interns' learning experiences from their *access* to *outcomes*, relative to their in-person counterparts.

Consistent with prior research (Graber, 2019; Mack et al., 2020; Waters & Russell, 2016), this study found a stream of empirical evidence to support that virtual internships may facilitate the social mobilization of some students who have been left out of traditional internships. Indeed, the analysis revealed that students of color are more visible in internships

conducted on digital platforms relative to their white peers. Female students or students from poor families are also active in a digital internship sphere just as much as male students or peers from affluent families. Certainly, widened access to work-based learning opportunities could be a potential pathway to reduce what Frenette et al. (2015) dubbed as the '*internship divide*', the gap in access to opportunities.

Another piece of evidence that reflects the equalizer potential of interning remotely can be found in the racial and ethnic variation in the internship developmental outcomes and satisfaction. These comparative benefits of a virtual internship were derived from regressionadjusted comparisons, extending a recent study by Fletcher et al. (2021), who documented a high degree of satisfaction and high chances of developing soft skills in the context of a virtual internship hosting mostly marginalized high school students. In the present work, white students who took a site-based internship gave higher ratings to the academic or career developmental values of their internship than their non-white peers. However, students of color who carried out their internship remotely gave higher ratings to the developmental values than their white peers. The same pattern continues to be observed with respect to their satisfaction level. Moreover, such gaps were statistically significant in different combinations: Asian, Black, and Hispanic students with respect to the academic developmental value, 'other race' students regarding the career developmental value, and Black and Hispanic students regarding satisfaction. These findings indicate that the white/non-white gap in internship outcomes is not only substantially varied by the type of internship, but also smaller when they undertake an internship online. That is, a virtual internship may, at least in part, reduce the race and ethnic gap in the developmental outcomes or satisfaction level that existed among those who took a site-based internship.

At the same time, this study lends strong support to the reinforcement thesis, uncovering more complicated and lingering issues of inequality around the access and quality of learning in the virtual workplace. Interning online versus in-person appears to be far out of reach for rural college students, creating an additional dimension of inequity concerns–geographic disparity. Even if the access to remote internships is expanded for female and Asian students, they are more likely to get on-the-job experiences in an internship with poorly designed learning segments, in deplorable working conditions, or both.

In addition to the racialized, gendered, and spatialized internship divides, the analysis confirms little mobilization potentials of a remote internship for students of low socioeconomic status, a common concern in an onsite program (Carnevale & Smith, 2018; Hora et al., 2021). Clearly, some students from high-income families were disproportionately more visible in a segment of internships with all the good things—highly structured and supportive, rich in networking and soft skills development opportunities, high-paying, and balanced enough to actually learn at work and remain in good grades—reaping the most of career building benefits in both modes. This work further attends to working students who hit hardest in both settings as well as throughout the entire process, finding that they are more likely to miss chances to gain work experiences. When students with full-time jobs take the rare opportunities, they tend to hold low-wage, part-time internship positions, relative to their non-working peers.

Although the mixed bag of evidence around a remote internship as an equalizer may be a new one, previous literature may partially account for the financial and sociocultural forces behind the reinforcement or mobilization thesis. As for social stratification based on one's socioeconomic status, recall that one of the primary hurdles that hold poor young people back from getting placed-based work experiences was simply they could not afford to pay for moving expenses and opportunistic costs, working for free during an internship, while their affluent peers could (Hora et al., 2021; Shade & Jacobson, 2015). As a result, interning online could be an attractive alternative option for those who are looking to move their career development forward without relocation, resulting in the door opened to low-income students. However, working students who need income to support their families or to pay for college may not only have fewer financial resources, but also have fewer times to allocate to work and academic workload (O'Connor & Bodicoat, 2017; Shade & Jacobson, 2015). Thus, the net results here imply that virtual placement opportunity could still be overwhelming for those who juggle between school-related obligations, paid jobs, and internship tasks, despite its greater temporal and spatial flexibility.

On the other hand, a high concentration of students working full-time in poor working conditions may persist even with greater flexibility. One possible interpretation is that the host organizations that provide unpaid internships favor those who have other paid jobs, such that they can work for free, relative to their decently paying counterparts. If this interpretation is correct, it should be asked why such internship hosts do not show the same interest in non-working students who can afford unpaid internships. A competing explanation is that working students have little choice but to take exploitative remote positions in exchange for flexibility or to provide cheap, part-time labor at the expense of meaningful professional experience regardless of the internship format. Some economic sociologists (e.g., Wood et al., 2018) had already discussed some of these findings with a reference to tensions simmering between increased temporal and spatial flexibility, unfavorable working conditions, and low job quality of low-skilled remote gig workers who handle studying, other paid jobs, and remote work altogether.

internship fare the students working full-time poorly in their entry, plus while interning, should be a key focus to researchers, practitioners, and policymakers that seek to build an inclusive and equitable professional pipeline towards social mobility.

Relatedly, the analysis implies that the potentials of a remote internship to advance equality may need more than temporal and geographical flexibilities. Some scholars (e.g., Siebert & Wilson, 2013; Swan, 2015) provide suggestive ideas that social capitals and networks of inner-circle communities may explain some, but certainly not all, of the premium reiterated over the quality of learning, be it in-person or remote. Most current WBL debates elaborate on the increased access flexibility (e.g., Franks & Oliver, 2012; Palmer et al., 2020; Waters & Russell, 2016), whereas going virtual may leave another barrier unresolved, especially a lack of family network and community support. It is a form of resource that marginalized students in terms of race and ethnicity, or social class are unable to build on their own, but may connect them to "valuable resources such as information, assistance, exposure to adult worlds, support, and encouragement." (Ross et al., 2020, p.6) Historically, those resources helped white students, male students, or students from economically privileged backgrounds to land an internship opportunity by reserving seats exclusively for them or at least passing opportunities through closed networks (Boulton, 2015; Reeves, 2017; Smith & Green, 2021). It was so even when unpaid in-person positions were proliferating. And when they do, it is likely that the effects of social capital persist throughout their learning at work, shaping the quality and intensity of ties with currently dominant groups based on mutual acquaintance at work and begetting another powerful pipeline of networks (Swan, 2015). In this context, while the technology may serve as a tool for all students to gain access to 'real world' work opportunities through increased temporal and spatial flexibility and web-based open recruitment approaches, as long as it is unable to

remove the barriers beyond the geographical, economical, and socio-cultural constraints, highprofile internship opportunities, whether place-based or remote, may continue to skew heavily wealthy, male, or white students.

In addition, this study contributes to the growing body of literature on work-based learning at the higher education level, especially to the research focused on typology with respect to the qualities of programs during an internship. The analysis confirms the notion that interns' learning experience occurs in complicated and multidimensional ways, identifying what five distinct typologies of interns look like across pedagogical contents, supervisory support, and structural working conditions. Each intern profiles largely matches with specific types of intern snapshots portrayed in previous studies (e.g., Curiale, 2009; Rogers et al., 2021). All profiles commonly showed a high correlation between overall task characteristics and onsite supervisory support, whereas the between-group difference generally stemmed from the level of advising from academic staff, wage, and working arrangements.

From these profiles, I find some encouraging evidence in two ways. First, nearly four in five interns (Profile 1, 2, and 3) gained educationally meaningful experiences in highly structured programs. One in two were learning in carefully guided programs with intentional incorporation of professional and academic goals and fair work that appreciated interns' effort financially and enabled the development of relationships, whether intensively working as a full-time (Profile 1) or continuing almost a full year as a part-time (Profile 3). Second, three of five interns undertook paid positions (Profile 1, 3, and 4), which is in between the ones found in the U.S. context of government and private industry and corporations (Gardner, 2011).

The analysis also revealed a group of low-quality internships characterized by selfdirected programs with challenging working conditions, although they take up a relatively small percentage (7%, Profile 5) out of the entire interns. While previous literature noticed such unfair types of programs in the context of an in-person internship (Rogers et al., 2021), this study offers new evidence that some technology-fueled internships do provide distinctive features of learning programs, but by adding disproportionately more cases of exploitative programs to the entire internship ecosystem.

Task design, supervisory support, and working conditions of these profiles, in combination, shape the outcomes of the internship programs. Complementing studies that discussed psychosocial or labor market outcomes of an internship in relation to specific features of an internship such as the clarity of goals (Beenen & Rousseau, 2010), task-related mentoring (Morgeon & Humphrey, 2006), or compensation (McHugh, 2015), the analysis show this relationship with the program quality in the aggregate. What becomes clear from this approach is that the mix of highly structured tasks, greater supervisors' support, and fair working conditions are associated with the greatest benefits for interns in terms of the satisfaction, academic as well as career development potentials. Additionally, large differences in all outcomes between profiles with low- and high-quality pedagogical practices reveal that, overall, the structure of tasks and supervisory support together play *more* important roles than working conditions such as wage and time arrangement in these relationships. These results generally align with a study by D'Abate et al. (2009) who recognized the significant contribution of pedagogical practices including "meaningful activities, which build skills and knowledge, as well as work in a context that offers considerable feedback" (p. 535) to interns' level of satisfaction.

At first glance, these findings may seem to suggest support for running up against the recent call for 'fair work' partly through paying interns for their efforts. Nonetheless, these should not be misinterpreted as evidence to justify the 'free training' argument. Rather, the full

story of this study reminds us of the role of 'securing compensation' as an important step that enables *all* college students to access professional development and future job opportunities. The analysis further elaborates the heterogeneous pattern: When little pedagogical considerations were paid to the program, the combination of support from academic staff, compensation, and work duration arrangements play more powerful roles in shaping internship outcomes than when with greater focus. In this vein, the results here provide a more nuanced understanding of what works for whom and when, extending D'Abate et al.'s work (2009). Thus, programs like Profile 2 and 4, in which interns enjoy only one of the two conditions, both raise a mixture of hope and concern.

All things considered, these findings offer implications to those who consider a technology-fueled internship as an alternative to an onsite format as an equitable work-based learning sphere. Faculty and staff, for example, may consider shifting their roles being primarily rolling out mandatory internships to taking greater initiative in providing specialized assistance with the internship search and advising, especially targeting those who are disadvantaged in remote or in-person internship settings. While career professionals at universities do so, they need to guide students to start with a wide-angle framework that enables them to construct a comprehensive vision of what mode of programs would best fit their needs, plus what would help students to make the most out of it. Such decisions could be better informed by a greater awareness of what learning actually looks like while interning at a specific host organization, whereas such detailed information only becomes available after they complete one. Therefore, in the long-term, collaborative efforts across institutions to build a large-scale internship database may be required to address the information asymmetry at least in part, especially for those who have little social capital to connect them to quality internship programs and networks. College

students who have been involved in an internship may voluntarily provide data on their past internship experience with the detailed information on the host organizations as future references for others and themselves. However, it is institutions' role to provide the diagnostic and evaluation tools like those presented in this study, ones that capture the multiple dimensions of an internship including their accessibility, quality of learning, and outcomes, and how they help students to navigate the career development process in meaningful internship experience. Given that one in five full-time students and one in three part-time students in the U.S. work more than 20 hours per week in 2020 (National Center for Education Statistics, 2022), a side-by-side arrangement to consider includes prioritizing financial resources and flexible course arrangements for those who work full-time while enrolled and with other marginalized identities.

Finally, back to the original question: Is a virtual internship a beneficial career development pipeline for college students towards the equitable and inclusive workforce? The answer would be yes: Students with marginalized identities—in terms of the race and ethnicity, gender, geography, and family's SES—enjoyed expanded benefits from technology-fueled internship at least one stage from their entry throughout the outcomes. Then, the next question is: Is a virtual internship comparable pipeline to an in-person internship for them? The answer would be no: This study finds, like site-based internship, not all remote internships are equally opened to all students due to financial, spatial, temporal, and socio-cultural barriers, nor they engage students in meaningful learning experiences that may lead to future careers. In fact, some remote internships may even cause more harm than good to students' academic performance or to the level of satisfaction, as studies have shown with a reference to a mandatory internship (Prescott et al., 2021) or an unpaid internship (McHugh, 2017). In sum, although technologyfueled internship may not be the most effective WBL venue to address issues around the inclusion and equity, it is a "workable" learning sphere that systematically keep young people moving forward.

Limitations

Before concluding, it is worth noting that this analysis has some limitations and potentials. Data on college students' internship experience was collected during the COVID-19 pandemic, which allowed for the examination of this new form of internship modality. While the coronavirus outbreak precipitated a shift in existing trends in a virtual internship, such unique research opportunity poses a notable limitation in that findings from this study provide only narrow insights into college students' workplace learning under unprecedented times. In future research, scholars can expand the present study to examine the equity potentials of a virtual internship outlast in the context of the post-pandemic era.

This study also highlights the need for further examination of the mechanisms underlying the results. Future research could address the sources of uneven distribution of students with marginalized identities across internship typology, with greater consideration of expanded internship positions at certain types of host organizations or recurring hurdles of unearned privilege. In addition, students' self-selection may account for these patterns partially because some of them might have limited choice options but to go with the remote options due to lockdown and disruptions in personal lives. This calls for caution in the interpretation of the results of this study as well as offers clues for future studies.

In addition, mean differences in internship outcomes in this study are estimates derived from a comparison of cross-sectional data, indicating descriptive patterns in nature. More rigorous research design using longitudinal data will be needed to tease out the causal impacts of varying profiles of internships. This can determine whether a technology-fueled internship is a small progressive step or a retreat for making the college-to-workforce connection equitable with respect to its psychological developmental values.

Lastly, the instruments to assess the quality of internships cover only limited aspects of students' internship experience, suggesting more comprehensive dimension is yet to be fully captured. The choice of these instruments was informed by the extant literature that found them to be critical to shaping what and how they learn in a traditional onsite internship setting. This suggests that the current survey might have overlooked the complexity of work-based learning, including taken-for-granted aspects of an in-person internship (e.g., sharing tacit knowledge) or unique factors that shape the outcomes of a remote internship. These quality measures uncovered in the survey may be a key to understanding why interns working in physical workplaces and virtual spaces show a different level of developmental outcomes and satisfaction even after controlling for variables capturing pedagogical practices, support, and working conditions. As such, additional studies should advance our understanding of what and how college interns experience in the two different modes by using more comprehensive measures.

References

- Ahmad, N. N. (2020). Accounting Students' Internship Satisfaction on the Work from Home (WFH) Experience during Internship. *Global Business & Management Research*, *12*(4).
- Allen, K., Quinn, J., Hollingworth, S., & Rose, A. (2013). Becoming employable students and 'ideal'creative workers: exclusion and inequality in higher education work placements. *British Journal of Sociology of Education*, *34*(3), 431-452.
- Arber, S., Bote, M., & Meadows, R. (2009). Gender and socio-economic patterning of selfreported sleep problems in Britain. *Social science & medicine*, 68(2), 281-289.
- Aschan, L., Goodwin, L., Cross, S., Moran, P., Hotopf, M., & Hatch, S. L. (2013). Suicidal behaviours in South East London: prevalence, risk factors and the role of socio-economic status. *Journal of affective disorders*, 150(2), 441-449.
- Baert, S., Neyt, B., Siedler, T., Tobback, I., & Verhaest, D. (2021). Student internships and employment opportunities after graduation: A field experiment. *Economics of Education Review*, 83, 102141.
- Bayerlein, L., & Jeske, D. (2018). Student learning opportunities in traditional and computermediated internships. *Education* + *Training*, 60(1), 27-38. https://doi.org/10.1108/ET-10-2016-0157
- Beard, F., & Morton, L. (1998). Effects of internship predictors on successful field experience. *Journalism & mass communication educator*, 53(4), 42-53.
- Beenen, G., & Rousseau, D. M. (2010). Getting the most from MBA internships: Promoting intern learning and job acceptance. *Human Resource Management*, 49(1), 3–22.
- Binder, J. F., Baguley, T., Crook, C., & Miller, F. (2015). The academic value of internships: Benefits across disciplines and student backgrounds. *Contemporary Educational Psychology*, 41, 73-82.
- Black, G., & Bachman, V. (2007). Virtual internships in the business accounting curriculum: A feasibility study. *Journal of Business Administration Online*, 6(2), 1-15.
- Boulton, C. (2015). Under the cloak of whiteness: A circuit of culture analysis of opportunity hoarding and colour-blind racism inside US advertising internship programs. tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society, 13(2), 390-403.
- Burke, D. D., & Carton, R. (2013). The pedagogical, legal, and ethical implications of unpaid internships. *Journal of Legal Studies Education*, *30*(1), 99-130. http://dx.doi.org/10.1111/j.1744-1722.2013.01115.x
- Callan, R. J. (1997). Supervised work experience in Europe: a profile of UK undergraduate perceptions. *International Journal of Contemporary Hospitality Management*, 9(1), 35-39.
- Carnevale, A. P., & Smith, N. (2018). Balancing work and learning: Implications for lowincome students. Washington, DC: Center on Education and the Workforce, Georgetown University. Retrieved December 30, 2021, from <u>https://1gyhoq479ufd3yna29x7ubjnwpengine.netdna-ssl.com/wp-content/uploads/Low-Income-Working-Learners-FR.pdf</u>

- Chand, V. S., & Deshmukh, K. S. (2019). Addressing the undergraduate internship challenge in developing countries: A "learning-by-doing" project-based virtual internship model. *Education+ Training*, 61(9), 1064-1077. https://doi.org/10.1108/ET-12-2018-0254.
- Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Williamson Shaffer, D. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of biomechanical engineering*, 137(2), 024701.
- Compton, L., & Davis, N. (2010). The impact of and the key elements for a successful virtual early field experience: Lessons learned from a case study. *Contemporary Issues in Technology and Teacher Education*, *10*(3), 309-337.
- Curiale, J. L. (2009). America's New Glass Ceiling: Unpaid Internships, the Fair Labor Standards Act, and the Urgent Need for Change. *Hastings Law Journal*, 61(6), 1531-1560.
- Cutshall, H., Hattaway, R., Singh, N. P., Rais-Bahrami, S., & McCleskey, B. (2021). The# path2path virtual landscape during the COVID-19 pandemic: preparing for the 2020 pathology residency recruitment season. *Academic Pathology*, 8, 1-4. https://doi.org/10.1177/23742895211002783
- D'Abate, C. P., Youndt, M. A., & Wenzel, K. E. (2009). Making the most of an internship: An empirical study of internship satisfaction. *Academy of Management Learning & Education*, 8(4), 527-539.
- D'Abate, C. (2010). Developmental interactions for business students: Do they make a difference?. *Journal of Leadership & Organizational Studies*, *17*(2), 143-155.
- Di Meglio, G., Barge-Gil, A., Camiña, E., & Moreno, L. (2021). Knocking on employment's door: internships and job attainment. *Higher Education*, 1-25.
- Duncan, H. E., & Barnett, J. (2009). Learning to teach online: What works for pre-service teachers. *Journal of educational computing research*, 40(3), 357-376.
- Fletcher, E. C., Minar, N. J., & Rice, B. A. (2021). The Future Ready Lab: Maintaining Students' Access to Internships during Times of Crisis. *Education and Urban Society*. https://doi.org/10.1177/00131245211065413
- Franks, P. C., & Oliver, G. C. (2012). Experiential learning and international collaboration opportunities: virtual internships. *Library review*. 61(4), 1-30.
- Frenette, A., Dumford, A. D., Miller, A. L., & Tepper, S. J. (2015). The internship divide: The promise and challenges of internships in the arts. Strategic National Arts Alumni Project. Indiana University and Arizona State University.
- Gamboa, V., Paixão, M. P., & de Jesus, S. N. (2014). Vocational profiles and internship quality among Portuguese VET students. *International Journal for Educational and Vocational Guidance*, *14*(2), 221-244.
- Gardner, P. (2011). The Debate over Unpaid College Internships. Intern Bridge. Retrieved from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.1710&rep=rep1&type=pdf

- Gault, J., Redington, J., & Schlager, T. (2000). Undergraduate business internships and career success: are they related?. *Journal of marketing education*, 22(1), 45-53.
- Graber, J. S., Saylor, J., & Johnson, A. N. (2019). Innovative Global Health Care Experiences for Nursing Students: Microsemesters and Internships. *Nursing education perspectives*, 40(5), 306-308.
- Gracia, L. (2009). Employability and higher education: Contextualising female students' workplace experiences to enhance understanding of employability development. *Journal of Education and Work*, 22(4), 301-318.
- Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge University Press.
- Halpern, C., Halpern, B., & Cerchiaro, I. B. (2020). Brazilian Undergraduate Students' Perceptions of the Role of Social Media in Online Recruitment for Internship Positions. *Journal of Human Resource and Sustainability Studies*, 8(3), 294-311.
- Hammoud, R., ElGarah, W., & Bouarich, B. (2022). Internship in the Time of COVID-19. *Communications of the Association for Information Systems*, 50(1), 12.
- Henry, J. S., Rehwaldt, S. S., & Vineyard, G. M. (2001). Congruency between student interns and worksite supervisors regarding critical elements of an internship experience. *Information Technology, Learning, and Performance Journal, 19*(1), 31-41.
- Hora, M., Chen, Z., Parrott, E., & Her, P. (2020). Problematizing college internships: Exploring issues with access, program design and developmental outcomes. *International Journal of Work-Integrated Learning*, 21(3), 235-252.
- Hora, M. T., Lee, C., Chen, Z., & Hernandez, A. (2021). Exploring Online Internships amidst the COVID-19 Pandemic in 2020-2021: Results from a Multi-Site Case Study. WCER Working Paper No. 2021-5. Wisconsin Center for Education Research.
- Hora, M. T., Wolfgram, M., Chen, Z., & Lee, C. (2021). Closing the Doors of Opportunity: A Field Theoretic Analysis of the Prevalence and Nature of Obstacles to College Internships. *Teachers College Record*, 123(12), 180-210.
- Horowitz, J. M., Ingielnik, R., & Kochhar, R. (2020). Most Americans say there is too much economic inequality in the U.S., but fewer than half call it a top priority. The Pew Research Center. Retrieved from November 28, 2021, from <u>https://www.pewsocialtrends.org/2020/01/ 09/most-americans-say-there-is-too-mucheconomic-inequality-in-the-u-s-but-fewerthan-half-call-it-a-top-priority/</u>
- Jackson, D., Shan, H., & Meek, S. (2022). Enhancing graduates' enterprise capabilities through work-integrated learning in co-working spaces. *Higher Education*, 84, 1-20.
- Jeske, D., & Axtell, C. M. (2014). e-Internships: prevalence, characteristics and role of student perspectives. *Internet Research*, 24(4), 457–473.
- Jeske, D., & Axtell, C. M. (2016). Going global in small steps: E-internships in SMEs. *Organizational Dynamics*, 45(1), 55-63.

- Jeske, D., & Axtell, C. M. (2018). The nature of relationships in e-internships: A matter of the psychological contract, communication and relational investment. *Revista de Psicología del Trabajo y de las Organizaciones*, 34(2), 113-121.
- Jeske, D., & Linehan, C. (2020). Mentoring and skill development in e-internships. *Journal of Work-Applied Management*, 12(2). 245-258
- Johari, A., & Bradshaw, A. C. (2008). Project-based learning in an internship program: A qualitative study of related roles and their motivational attributes. *Educational Technology Research and Development*, *56*(3), 329-359.
- Keefer, K. V., Parker, J. D., & Wood, L. M. (2012). Trait emotional intelligence and university graduation outcomes: Using latent profile analysis to identify students at risk for degree noncompletion. *Journal of Psychoeducational Assessment*, 30(4), 402-413.
- Kraft, C., Jeske, D., & Bayerlein, L. (2019). Seeking diversity? Consider virtual internships. *Strategic HR Review*, 18(3), 133-137.
- Kurter, H. L. (2021, August 26). Google Plans To Cut Remote Workers' Salaries By 25%. Here's The Impact. *Forbes*. <u>https://www.forbes.com/sites/heidilynnekurter/2021/08/26/google-plans-to-cut-remote-workers-salaries-by-25-heres-the-impact/?sh=5a8359094fb2</u>
- Liu, Y., Xu, J., & Weitz, B. A. (2011). The role of emotional expression and mentoring in internship learning. *Academy of Management Learning & Education*, 10(1), 94-110.
- Machin, S., & Vignoles, A. (2004). Educational inequality: the widening socio-economic gap. *Fiscal Studies*, 25(2), 107-128.
- Mack, K., Das, M., Jain, D., Bragg, D., Tang, J., Begel, A., ... & Potluri, V. (2021, October). Mixed Abilities and Varied Experiences: a group autoethnography of a virtual summer internship. *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1-13).
- Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 175-198). Newbury Park, CA: Sage.
- Mansfield, R. (2011). The effect of placement experience upon final-year results for surveying degree programmes. *Studies in Higher Education*, *36*(8), 939-952.
- Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12), 3722-59.
- Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *Oxford* handbook of quantitative methods (Vol 2., pp. 551-611). New York, NY: Oxford University Press.
- McHugh, P. P. (2017). The impact of compensation, supervision and work design on internship efficacy: implications for educators, employers and prospective interns. *Journal of Education and Work*, *30*(4), 367-382.
- Molina, I., & Martín, N. (2018). Empirical best prediction under a nested error model with log transformation. *The Annals of Statistics*, 46(5), 1961-1993.

- Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of applied psychology*, *91*(6), 1321.
- Moss-Pech, C. (2021). The career conveyor belt: how internships lead to unequal labor market outcomes among college graduates. *Qualitative Sociology*, 44(1), 77-102.
- Muthén, B. O., Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical & *Experimental Research*, 24(6), 882-891.
- Narayanan, V. K., Olk, P. M., & Fukami, C. V. (2010). Determinants of internship effectiveness: An exploratory model. *Academy of Management Learning & Education*, 9(1), 61-80.
- National Center for Education Statistics. (2022). College Student Employment. Condition of Education. U.S. Department of Education, Institute of Education Sciences. Retrieved May 20, 2022, from <u>https://nces.ed.gov/programs/coe/indicator/ssa</u>
- Nghia, T. L. H., & Nguyen Thi, M. D. (2019). Developing and validating a scale for evaluating internship-related learning outcomes. *Higher Education*, 77(1), 1-18.
- Nunley, J. M., Pugh, A., Romero, N., & Seals Jr, R. A. (2016). College major, internship experience, and employment opportunities: Estimates from a résumé audit. *Labour Economics*, 38, 37-46. <u>http://dx.doi.org/10.1007/s10734-018-0251-4</u>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal, 14*(4), 535-569.
- O'Connor, H., & Bodicoat, M. (2017). Exploitation or opportunity? Student perceptions of internships in enhancing employability skills. *British Journal of Sociology of Education*, 38(4), 435-449.
- Oswald-Egg, M. E., & Renold, U. (2021). No experience, no employment: The effect of vocational education and training work experience on labor market outcomes after higher education. *Economics of education review*, 80, 102065.
- Palmer, H. M., Smart, A., & Brumbaugh, S. (2020, December). A marine science internship, without the ocean: transitioning a community college internship program to a virtual setting. In AGU Fall Meeting Abstracts (Vol. 2020, pp. ED040-04).
- Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary educational psychology*, 32(1), 8-47.
- Perlin, R. (2012). *Intern nation: How to earn nothing and learn little in the brave new economy*. Brooklyn, NY: Verso Books.
- Prescott, P., Gjerde, K. P., & Rice, J. L. (2021). Analyzing mandatory college internships: academic effects and implications for curricular design. *Studies in Higher Education*, 46(11), 2444-2459.
- Pretti, T. J., Etmanski, B., & Durston, A. (2020). Remote Work-Integrated Learning Experiences: Student Perceptions. *International Journal of Work-Integrated Learning*, 21(4), 401-414.
- Proctor, S. L., Kyle, J., Lau, C., Fefer, K., & Fischetti, J. (2016). Racial microaggressions and school psychology students: Who gets targeted and how intern supervisors can facilitate racial justice. *School Psychology Forum: Research in Practice*, *10*(3), 321-336.
- Reeves, R. V. (2018). Dream hoarders: How the American upper middle class is leaving everyone else in the dust, why that is a problem, and what to do about it. Brookings Institution Press.
- Richardson, S. (2008). Undergraduate tourism and hospitality students attitudes toward a career in the industry: A preliminary investigation. *Journal of Teaching in Travel & Tourism*, 8(1), 23-46.
- Rogers, S. E., Miller, C. D., Flinchbaugh, C., Giddarie, M., & Barker, B. (2021). All internships are not created equal: Job design, satisfaction, and vocational development in paid and unpaid internships. *Human Resource Management Review*, 31(1), 100723.
- Ross, M., Kazis, R., Bateman, N., & Stateler, L. (2020). Work-based learning can advance equity and opportunity for America's young people. *Washington, DC: Brookings Institution*.
- Rothman, M. (2003). Internships: Most and least favored aspects among a business school sample. *Psychological reports*, *93*(3), 921-924.
- Ruggiero, D., & Boehm, J. (2016). Design and development of a learning design virtual internship program. *International Review of Research in Open and Distributed Learning*, 17(4), 105-120.
- Shade, L. R., & Jacobson, J. (2015). Hungry for the job: gender, unpaid internships, and the creative industries. *The Sociological Review*, 63(S1), 188-205.
- Siebert, S., & Wilson, F. (2013). All work and no pay: consequences of unpaid work in the creative industries. *Work, employment, and society*, 27(4), 711-721.
- Smith, K. N., & Green, D. K. (2021). Employer internship recruiting on college campuses: 'the right pipeline for our funnel'. *Journal of Education and Work*, *34*(4), 572-589.
- Sterling, A. D., & Fernandez, R. M. (2018). Once in the door: Gender, tryouts, and the initial salaries of managers. *Management Science*, 64(11), 5444-5460.
- Taylor, M. S. (1988). Effects of college internships on individual participants. *Journal of Applied Psychology*, *73*(3), 393.
- Teng, C. W. C., Lim, R. B. T., Chow, D. W. S., Narayanasamy, S., Liow, C. H., & Lee, J. J. M. (2021). Internships before and during COVID-19: experiences and perceptions of undergraduate interns and supervisors. *Higher Education, Skills and Work-Based Learning*. 2042-3896.
- Van Ommeren, J. N., & Gutiérrez-i-Puigarnau, E. (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, 41(1), 1-8.

- Varghese, M. E., Parker, L. C., Adedokun, O., Shively, M., Burgess, W., Childress, A., & Bessenbacher, A. (2012). Experiential internships: understanding the process of student learning in small business internships. *Industry and Higher Education*, 26(5), 357-367.
- Waters, S., & Russell, W. (2016). Virtually ready? Pre-service teachers' perceptions of a virtual internship experience. *Research in Social Sciences and Technology*, 1(1), 1–23.
- Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: a guide to best practice. *Journal of Black Psychology*, 46(4), 287-311.
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: autonomy and algorithmic control in the global gig economy. *Work, Employment and Society,* 33(1), 56-75.

Chapter 5. Conclusion

In this chapter, I conclude with a summary of findings and develop more comprehensive explanations of under which conditions, which types of innovation happen, and how such significant changes may be an equalizer or reinforce the status quo.

In Chapter 2, I created an analytical framework for educational innovation drawing on March's (1991) organizational learning framework, *exploitation* and *exploration*. These organizational behaviors are distinguished in terms of the degree of emphasis on the use and refinement of a pre-existing knowledge base and the navigation of new possibilities based on the knowledge, information, and resources. A long history of educational reform has shown that multi-layered actors in education have made incremental changes, refining routines with known payoff, on the one hand, introducing experimental reform ideas with uncertain consequences, on the other hand. I synthesized these two interrelated concepts and developed a measurement that may visualize the characteristics of organizational learning in the temporal and cognitive dimensions.

With this framework, I explored a theoretical link between a policy instrument, *decentralization*, and innovativeness of the equity-oriented policy in K-12 setting. Employing a qualitative content analysis approach, I assessed the innovativeness of Florida's school turnaround model under decentralization from the vantage point of federal and state government, respectively, and mapped the results on the quadrant by four education functions: teaching and learning; planning and monitoring; resource management; and personnel management. The analysis revealed that the Florida turnaround model is generally positioned in the exploitation zone, suggesting a transfer of greater authorities back to SEAs alone is unlikely to yield significant transformation in policy designs targeting traditionally underserved students. In addition, the quadrant mapping showed conflicting views surrounding the innovativeness of the new model—cognitive tension between the federal and state authority. Seen from the federal lens, Florida SEA appear to take greater initiatives to explore fresh ideas, albeit limited, for underperforming schools when the federal authority granted regulatory exceptions to the state. From Florida's perspective, on the contrary, they continued to pursue a mixture of locally developed approaches as well as strategies established under preexisting federal guidelines, making only marginal changes to teaching and learning practices, but relatively radical changes in personnel management.

In Chapter 3, I shifted the focus from policy design to implementation and examined the causal impacts of Florida school turnaround on students' academic achievement, as an ultimate goal of pursuing innovation lies in achieving more equitable outcomes. The analysis using a regression discontinuity design finds no evidence that the Florida-brand school turnaround generated substantial differences in the students' test scores, probability of meeting high standards, or making gains in math and reading. To investigate the heterogeneous effects by turnaround types, I focused on each subgroup and re-estimated the effect using a difference-in-difference approach with matched group of schools. Results of subgroup analyses exhibit the largest positive impacts across all outcomes for students at schools with the most intensive level of support, yet only little or no effect for those students at schools that received less intensive support.

Chapter 4 zoomed into a case of *explorative* innovation fueled by the development of technology—college students' virtual internship from the equity lens. The global pandemic expedited unprecedented demand for remote learning and work, and advances in technology-enabled college students to gain real-work experience on a digital platform through virtual

internship programs. I studied whether a technology-fueled internship program is associated with the social mobilization of college students who have been left out in traditional internships with respect to their access, learning, and outcomes. By examining virtual interns' learning experiences relative to their in-person counterparts, this study uncovered that a virtual internship is a mixed bag of progress and challenges towards the goal of creating an equitable and inclusive career development pathway for young students. The distribution pattern of students with marginalized identities across the entire internship landscape indicates that students of color, female students, and students from low-income households are similarly or more visible in virtual work-based learning opportunities. Manifestation of its equity potentials is stretched over to interns' learning experience during an internship in which four in five virtual interns engage in comparably well-structured tasks and work side-by-side with supportive supervisors. More importantly, I observe relatively smaller racial and ethnic disparities in virtual interns' academic or career developmental outcomes and satisfaction when compared to the between-race gaps for in-person interns. Together, these comparative advantages seem to lend strong empirical support that a virtual internship may be moderating the racial, gender, and wealth gaps in college students' career preparation trajectory.

A virtual internship, nevertheless, is unlikely to be an alternative route to an early career that works for every young people. In fact, Chapter 4 shows that it may reinforce social stratification either through uneven access to opportunities or low-quality of programs, highlighting a new type of divide. Students from rural high school are one of such cases, as the rural/non-rural gap in students' access to an internship is larger in a virtual sphere whereby onthe-job learning primarily takes place. Even when access to remote internships is expanded for female and Asian students, they were more likely to take low-quality programs compared to their counterparts. I also recognize that working students, especially students working full time, suffer from a 'double jeopardy' in a virtual setting, and thus, require immediate attention from educator communities: they were underrepresented in the entire internship landscape, whether virtual or onsite, and when they took one, they were more concentrated in a low-quality virtual internship.

The preceding analyses revealed when and how exploitation and exploration in educational initiatives occur by looking at two instances that were related to equity. While results are based on somewhat "blunt" measures of innovation and our scope goes beyond the analysis of policy making, findings from these studies are largely aligned with the punctuated equilibrium model. Baumgartner and Jones (1991) attempt to understand the development of public policy from the evolutionary perspective that underscores the incremental adjustments punctuated by bursts of transformational changes. Extended to the context of equity-oriented policymaking in K-12 settings, this study shows that SEA tends to enact exploitative policy changes for the state's most struggling schools when federal regulations and oversights are lifted, presumably due to the interaction of the limited capacity of SEAs and pressure for quick wins. However, it is important to note that such little variation may not be the same as those standing federal approaches, as the exploitative changes result from the state's policy learning through prioritization and modifications of the existing strategies within their own context. While only bounded areas may have been reconsidered during this adjustment process, a more optimal combination of available resources, effective interventions, and accountability pressure derived through exploitative learning within its own context is also encouraging in terms of its effectiveness, as shown in the case of Florida that improved students' academic performance in the state's most struggling schools. That said, this study is not to say that SEA's exploitative interventions are sufficient for ensuring equitable public education in the long term because the effect of the new model is

unlikely to be sustained due to little consideration of fundamental problems of underperformance and lock-in to resource-intensive interventions at the expense of other schools. Indeed, among 11 Intervene schools that made significant progress in 2008-09, 58% of them were classified again as schools that need intensive support in 2011-12⁵. This suggests that SEA's exploitative innovation triggered by decentralization can be one of the available and helpful options, but may not be the most effective and sustainable way to close achievement gaps in public education, suggesting alternative approaches are necessary.

On the other hand, the development of technology and external shock such as the global pandemic enabled the explorative social innovation—a virtual internship. Whilst this novel mode of internship is not necessarily targeted at vulnerable groups of young people, this study provides some promising evidence that a virtual internship may build more inclusive college-to-career pathways for traditionally marginalized young people than the conventional in-person WBL opportunities. These comparative advantages were particularly pronounced when the affordances of interning 'remotely'—temporal or spatial flexibility, web-based open recruitment approaches, or reduced emotional toll of feeling microaggression-could potentially alleviate some of these challenges relevant to the corresponding group of students while pursuing an in-person internship, as detailed in Chapter 4. Where the link between these affordances and onsite internship barriers were vague, not only did marginalized students in the conventional internship ecosystem continue to be underserved in the digital internship sphere, but also it spawned a new type of disadvantages for students from rural background. Following recurring access challenges, this explorative career development segment comes with a hidden cost at the expense of its affordances, especially for working students, in that they tend to be more visible in exhaustive

⁵ Author's calculation based on a list of differentiated accountability schools reported by the Florida DOE.

positions where they can rarely gain meaningful learning experiences in a sound environment, making the broadening of 'bad' internships real. Moreover, the academic or career developmental potential of a virtual internship are significantly lower than those of an onsite internship, differences that cannot be attributed to conventional, observable aspects of the WBL.

These findings altogether reaffirm that exploitation and exploration for equitable education are indeed complex undertakings that emerges from multiple sources with multilayered actors and constantly evolve in particular institutional and temporal contexts. Thus, it is highly unlikely that simple policy changes or the development of technology drive innovations that fully address the complicated problems of equity spread in access, quality of learning, and outcomes. This, in turn, highlights the need for more a systematic understanding of the link between exploitation, exploration, and equity with greater specificity by using tools available such as the one described earlier in this dissertation. Innovation scholars need to ask, what coordination or separation mechanism across education functions and stakeholders drives what types of innovation and how? How does exploitation or exploration affect the implementation process? What types of innovation reduce disparities in education, when, and for whom? Answering these questions may not provide us with definitive solutions for the challenges surrounding educational inequity but they will be able to generate a discussion on the possible tools, strategies, and role of actors to be closer to it.

Finally, I close the discussion with a suggestion for future studies to look beyond the three pillars of educational equity—access, quality of learning, and outcomes—whenever possible. These are key dimensions of equity in education, which nonetheless cover only a portion. Recent scholarship broadened the concept of equity to some of the under-examined aspects such as identity (i.e., "be themselves and better themselves") and power (Gutiérrez, 2012,

References

- Baumgartner, F. R., & Jones, B. D. (1991). Agenda dynamics and policy subsystems. *Journal of Politics*. 53, 1044-1074.
- Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education?. In *Equity in discourse for mathematics education* (pp. 17-33). Springer, Dordrecht.