

**SPRING 2024
ISSUE**

University of Connecticut Neag School of Education

UConn | NEAG SCHOOL OF EDUCATION **Journal**

UConn
NEAG SCHOOL OF EDUCATION

TABLE OF CONTENTS

MISSION STATEMENT	3
LETTER FROM THE BOARD	4
2023-2024 BOARD MEMBERS	6
NEUROSCIENCE CONCEPTS SUPPORTING TEACHERS' ADAPTIVE EXPERTISE	7
KRISTIN SIMMERS, SAMEERA MASSEY	
FROM FIXED TO FLEXIBLE: NEEDED CONDITIONS TO PROMOTE ELEMENTARY TEACHERS' EQUITABLE USE OF WITHIN-CLASS ABILITY GROUPING	19
SARAH CAROLEO	
DOES THIS MEAN WE GET AN A? CAUSAL IMPLICATIONS OF CHANGES IN SCHOOL ACCOUNTABILITY	42
SAMUEL KAMIN	
PERCEPTIONS OF HIGHER EDUCATION PROFESSIONALS ON THE UTILITY OF THE ACTIVITIES, PROGRAMS, OR POLICIES TOOL TO PROMOTE SELF-DETERMINATION FOR COLLEGE STUDENTS WITH DISABILITIES	71
ASHLEY TACONET, EMILY TARCONISH, JOSEPH W. MADDAUS, NICHOLAS GELBAR, LYMAN DUKES, III, MICHAEL FAGGELLA-LUBY, DON MILLS	
APPENDICES	90
FOUNDING BOARD MEMBERS	92

MISSION STATEMENT

The Neag School of Education Journal is an editor-reviewed, open-access, annual journal founded and run by graduate students and published through the Neag School of Education at the University of Connecticut. Its primary purpose is to offer a platform for graduate students to share their research and knowledge with academic communities, to broaden and deepen the literature of education, as written and experienced by graduate and doctoral students, as well as early-career scholars.

The Neag School of Education Journal highlights the strongest, most robust student and early-career work from a broad range of disciplines such as educational psychology, curriculum and instruction, teacher education, as well as educational leadership.

Of particular interest are pieces providing reflection on student experience with their research processes and manuscripts showcasing research in the preliminary stages. The journal offers students and early-career scholars the chance to publish work through diverse types of academic writing including, but not limited to, research articles (e.g., qualitative and quantitative research), essays, literature review, as well as personal experience and reflective pieces.

Aligning with the mission of its governing institution, the Neag School of Education Journal places significance on pieces seeking to improve education and social systems in order to facilitate increasingly effective, equitable, and socially just practices for educators and practitioners from a variety of fields, perspectives, and theoretical lenses as they serve their local communities. Reflections are also sought after to foster relations and collaboration amongst graduate students and their colleagues, to pass along wisdom, innovation, and creativity in pursuit of fostering a graduate community rooted in rigorous and intentional research design and practices. The journal's open access ensures it as a source for current and future practitioners.

LETTER FROM THE BOARD

We are overjoyed to share the 2nd issue of the Neag School of Education Journal. Leading from our mission, our journal provides a unique space for graduate and early career scholars to develop and share a broad variety of scholarly work, including research articles, essays, literature reviews, and reflective pieces. We take pride in providing a supportive “testing-ground” for graduate authors to refine their original work in collaboration with our graduate-led editorial board. Fundamentally, the Neag School of Education Journal is committed to the growth and development of emergent educational researchers across fields. After much hard work and dedication from our authors and editorial board, we are thrilled to unveil the culmination of their efforts – four carefully selected pieces that showcase the excellence of our 2024 edition.

Each of this year’s articles exemplify the equity-grounded, methodologically rigorous, and innovative research that this journal endeavors to elevate.

Our first piece - *Neuroscience concepts supporting teachers’ adaptive expertise*. This essay, by Kristin Simmers and Sameera Massey, delves into the integration of neuroscience concepts in teaching, emphasizing the importance of understanding cognitive and neural mechanisms for effective instructional decisions. It argues that adopting the Mind, Brain, and Education (MBE) framework enhances teachers’ adaptive expertise, enabling them to navigate unpredictable teaching scenarios with efficiency, analyze classroom dynamics holistically, and maximize student engagement and learning outcomes. At a time when learning and brain myths are rampant in the teaching profession, this piece puts scientific understanding and research-informed approaches at the bedrock of good teaching and learning.

Second - *From fixed to flexible: Needed conditions to promote elementary teachers’ equitable use of within-class ability grouping*, by Sarah Caroleo. This essay explores the historical context of fixed and flexible within-class student grouping practices and examines their impact on students’ access to equitable learning outcomes. The author advocates for flexible grouping strategies, supports the argument with empirical evidence, details the practical application of flexible grouping, and offers organizational and policy recommendations for its implementation. In this way, this piece pushes those supporting and training teachers as well as policymakers to think more expansively about how we ensure differentiation serves to meet all children’s needs rather than reinforce inequities.

Next - *Does this mean we get an A? Causal implications of changes in school accountability* by Samuel Kamin - utilizes regression discontinuity to draw causal conclusions about the impact of school rating categorization through the New York City Department of Education’s most recent school report card rating system. Notably, this piece provides causal evidence that schools that score just below the cut-score categorization of “Meeting” Student Achievement rating (i.e., a score of “Approaching”) show positive and significant math test score gains the following year. As such, the piece has implications for how we think about both the impact of external accountability systems for schools and whether and to what degree such policies need to include carrots or sticks to provoke change.

Finally - *Perceptions of higher education professionals on the utility of the activities, programs, or policies tool to promote self-determination for college students with disabilities*, by Dr. Ashley Taconet. This qualitative exploratory study investigates the experiences of 30 post-secondary disability resource and student affairs professionals with the Activities, Programs, or Policies (APP) Tool. Derived from work showing the benefits of developing and providing opportunities for self-determination for post-secondary students and specifically those with disabilities, this manuscript offers an overview of practitioners' experience with the tool and recommendations for further development of this instrument. As such, the piece provides analysis of a tool for campus leaders to promote and improve student self-determination.

We look forward to your enjoyment of this issue's work and the outstanding contributions from our graduate and early career authors. We further hope that authors use the feedback they received during the editing process and choose to publish these manuscripts in professional peer reviewed journals in the future. At the Neag School of Education Journal, we focus on the development of student work by employing a high-dose, collaborative review process. Our novel copyright policy is designed to empower students and early career scholars, allowing them to maintain the copyright for future publication.

We have many fantastic and dedicated people to thank for the realization of our 2nd issue. To Dr. Jennie Weiner, our advisor, thank you for your tireless dedication to this journal and to students. You model to us what a human-centered and compassionate review process can be and have taught us enduring lessons as reviewers and researchers.

Another thank you to Dr. Jason Irizarry, our dean, who has enthusiastically supported the journal from its inception and made it clear that our work and voices matter.

Thank you to Shawn Kornegay and the design team at UConn who helped ensure a second issue as beautifully apportioned as the first. We look forward to continuing to uplift graduate students' work in years to come.

Thank you to the members of the journal whose hard work and enthusiasm made this issue possible. We are excited to continue advancing this work with your assistance going forward.

Finally, thank you to the authors of the pieces featured in this issue and all who submitted work. It goes without saying that this would be impossible without your contributions. We are immensely proud to feature your work in this issue.

To learn more about our team and mission, please visit us at <https://education.uconn.edu/neag-journal/>

2023-2024 BOARD MEMBERS

Caitlin Blacksmith

Research Methods, Measurement, and Evaluation Ph.D. Student

Jessica Bourget

Special Education Ph.D. Student

Luis Orione Ferreria

Giftedness, Creativity, and Talent Development Ph.D. Student

Brenna Fitzmaurice

School Psychology Ph.D. Student

Talbot Hook

Giftedness, Creativity, and Talent Development Ph.D. Student

Emily Lisy

Curriculum and Instruction (Science Education) Ph.D. Student

Gina Norman

School Psychology Ph.D. Student

Taylor Strickland

Learning, Leadership, and Educational Policy Ph.D. Student

Ani Terterian

Higher Education and Student Affairs Master's Student

Charles Wentzell

Learning, Leadership, and Educational Policy Ph.D. Student

Faculty Advisor: Dr. Jennie Weiner

NEUROSCIENCE CONCEPTS SUPPORTING TEACHERS' ADAPTIVE EXPERTISE

KRISTIN SIMMERS

PH.D. STUDENT

NEAG SCHOOL OF EDUCATION, UNIVERSITY OF CONNECTICUT

ORCID 0009-0009-3369-2959

SAMEERA MASSEY

PH.D. STUDENT

TEXAS A&M UNIVERSITY-CORPUS CHRISTI

DEPARTMENT OF CURRICULUM, INSTRUCTION, AND LEARNING SCIENCES

ORCID 0000-0001-8838-6005

ABSTRACT

Effective, responsive teaching benefits from a foundational understanding of the cognitive and neural mechanisms that underlie student learning, which can inform teachers' instructional decisions. This article explores the potential influence of neuroscience concepts on teachers' adaptive expertise, which can empower educators to navigate unpredictable teaching scenarios with efficient flexibility. The interdisciplinary framework of Mind, Brain, and Education (MBE) is proposed as a means to enhance this adaptability. Two core MBE concepts are explored: neuroplasticity, underscoring the brain's adaptability through learning experiences, and the intertwined nature of emotion and cognition in shaping the learning process. Integrating MBE insights offers educators a holistic lens with which to better analyze and respond to classroom dynamics using research-informed approaches, thus maximizing student engagement and learning outcomes.

Keywords: Educational neuroscience, teacher education, adaptive expertise, mind brain and education.

The demands of modern education require that teachers have a deeper understanding of how students learn and the underlying brain processes of learning to best inform their decisions (Council of Chief State School Officers, 2013). Teachers are faced with the challenge of raising all students' achievement, in addition to supporting students' holistic development. Achievement can be measured in numerous ways, including nationally normed tests, teacher-created assessments, and quantified observations. While there are several factors impacting student achievement, teacher quality has been identified as a critical factor (Dudley-Marling et al., 2006; Kunter et al., 2013; Stronge et al., 2007) and some say the greatest controllable factor (Hattie, 2003). Highly effective teachers, defined as those producing at least a year's worth of growth in their students, are distinguishable from less effective teachers by their beliefs about learning, critical and adaptive thinking, and judgements (Hattie, 2023). Mastering these process skills has greater influence on teacher effectiveness than mastery of content knowledge and specific instructional methods (Hattie, 2023; Männikkö & Husu, 2019).

Teachers make a high volume of planned and in-the-moment decisions in response to the varied emotional, social, developmental, and instructional information they gather about their students (Kennedy, 2019). Responding effectively to unusual or new situations requires the ability to identify and analyze novel situations, engage in flexible problem solving, and generate innovative solutions. This flexible application of knowledge and skills in the context of new situations is at the core of adaptive expertise (Ng et al., 2022). Adaptive expertise is the ability to use deep knowledge to flexibly address new and unforeseen challenges (Carbonell et al., 2014; Hatano & Inagaki, 1986; Novick & Holyoak, 1991) and is contrasted with routine expertise, which also involves a depth of knowledge but without the ability to adapt in response to unexpected or novel situations (Carbonell et al., 2014; Hatano & Inagaki, 1986). For example, a teacher with routine expertise might be successful if everything in a lesson goes according to plan, but struggle if unexpected situations arise (as they often do in teaching!). Novel situations in teaching that require adaptive expertise could be anything that adds a layer of complexity to the teaching and learning dynamic, such as a shortened class period, a technology glitch, students' emotional states, or a teacher realizing that the students are not learning as expected. In response to these dynamics the teacher has to make decisions to adjust and adapt. Teachers' adaptive expertise is influenced by the ability to interpret events in the classroom and by their knowledge and skills (Fairbanks et al., 2010; Männikkö & Husu, 2019), and we propose this may be of particular importance if the teacher has limited practical experience to draw upon. Therefore, we propose that a broad repertoire of research-based knowledge about teaching and learning could be advantageous for adaptive decision making and support adaptive expertise, as it equips a teacher with more lenses through which to adeptly interpret classroom events and choose an effective response (Blake & Gardner, 2007).

Adaptive expertise requires pulling from a broad knowledge foundation as opposed to deep knowledge in one area (Grotzer et al., 2021) and a dynamic interplay between practical and theoretical knowledge (Männikkö & Husu, 2019). This allows an adaptive expert, such as a teacher, to access prior knowledge from across domains and respond flexibly rather than prescriptively, allowing for more dynamic problem-solving (Grotzer et al., 2021). A teacher's ability to adjust their approach in response to multiple variables can be thought of as adaptive instruction or adaptive teaching (Parsons et al., 2017), and

researchers have identified this ability as an essential component to effective teaching, beyond basic content knowledge (Fairbanks et al., 2010). Furthermore, choosing an effective adaptive strategy for a particular situation requires analysis of cognitive as well as emotional conditions (Tokuhamma-Espinosa et al., 2023), making a holistic lens based on multiple disciplines ideal for constructing a more complete assessment of the situation.

With an incomplete or inaccurate knowledge base, adaptive expertise becomes more challenging. It benefits teachers to have a more complete conceptual framework for making research-informed judgments about why, when, and how to apply specific strategies, responses, and tools (Hohnen & Murphy, 2016; Howard-Jones et al., 2020). Researchers studying thoughtful, adaptive teachers identified these common traits: teachers know when to apply “what” and “how” knowledge, and when not to; they know why certain knowledge would be appropriate in one situation but not another; and they proactively look for multiple perspectives and pursue multiple possibilities because they recognize and respond to the complex needs of their students (Fairbanks et al., 2010).

Providing teachers with a foundational understanding of neuroscience concepts related to teaching and learning provides an opportunity to both refine and expand teachers’ knowledge base, enhancing their adaptive expertise from the start of their teaching careers. While there are many terms for the area where neuroscience and education overlap (such as educational neuroscience, neuroeducation, cognitive neuroscience, and even neurocognitive pedagogy), we will generally use the term Mind, Brain, and Education (MBE), as it is most clearly aligned with our integrative view of the field. MBE is an interdisciplinary field that explores the connections between neuroscience, cognitive psychology, and education, giving all fields equal weight, with the goal of creating a scientific groundwork to support teaching and learning (Fischer, 2009). Early evidence suggests that providing teachers with knowledge from MBE can influence their planning decisions (Schwartz et al., 2019), but more evidence is needed to understand this more fully. We propose that core concepts from MBE research have the potential to enhance teachers’ adaptive expertise by providing a research-informed transdisciplinary lens through which to better understand complex teaching and learning dynamics, so that teachers are best supported in making informed decisions in response to their learners.

KEY MIND, BRAIN, & EDUCATION CONCEPTS SUPPORTING ADAPTIVE EXPERTISE

Identifying foundational “key concepts” of MBE can be complex because there are numerous robust resources to choose from, each providing slightly different perspectives, all well-supported by research. However, despite this diversity, several common MBE principles consistently emerge across multiple studies, including those highlighted in a 2020 international survey conducted by Tokuhamma-Espinosa and Nouri (2020), which identified 18 MBE concepts with which teachers should be familiar. We found the majority of the core concepts identified by Tokuhamma-Espinosa and Nouri (2020) to be related to teachers’ adaptive expertise, such as “The brain is plastic and can change as a result of learning experience” and “Affective and cognitive processes are inextricably linked” (Tokuhamma-Espinosa & Nouri, 2020, pp. 67-68). We interpret most of these core concepts as falling under two domains: the concept of neuroplasticity and the link between emotion and cognition. Here we explore them further.

NEUROPLASTICITY

Neuroplasticity is arguably one of the most impactful concepts for shaping teachers' decisions. Stated simply, "Plasticity is key to education" (Ansari et al., 2017, p. 200). Neuroplasticity is the idea that the brain is constantly changing and adapting in response to stimuli (Sousa, 2011). The term encompasses the formation and pruning of connections between neurons and how the brain continuously changes in response to experience and the environment (Ansari et al., 2017). Given its centrality to the learning process, the concept of neuroplasticity tends to be prominent in MBE research. We propose that understanding neuroplasticity can empower teachers with the knowledge that learning can shape the brain's development and thus inform adaptive decisions that foster diverse and effective teaching strategies.

In considering which aspects of MBE are most important for teachers to understand, a Delphi panel identified eighteen key concepts, many of which relate to neuroplasticity, either noting physical changes in the brain, such as "The human brain undergoes enormous development across the lifespan" or referencing the brain's malleability, such as "Intelligence is a malleable biopsychological process..." (Tokuhamma-Espinosa & Nouri, 2020, pp. 67-68). This concept is also recognized in two of the six Principles of MBE established through the Delphi panel: "Neuroplasticity" and "Constant Changes in the Brain" (Tokuhamma-Espinosa & Nouri, 2020, p. 65). It is also found among the "Concepts of MBE Teacher Literacy," which states "The brain is neuroplastic and can change as a result of learning experiences" (Tokuhamma-Espinosa & Nouri, 2023, p.172). Several other researchers include neuroplasticity as a key MBE concept for teachers (Brick et al., 2021; Carrasco et al., 2015; Chang et al., 2021; Dubinsky et al., 2013, 2019; Schwartz et al., 2019), and some note that teachers who participate in neuroscience related professional development or interventions then report changes in their instructional decision making, such as using more student-centered practices (Brick et al., 2021; Schwartz et al., 2019), or allowing for repetition to enhance plasticity (Chang et al., 2021).

When teachers appreciate that neuroplasticity is present throughout the lifespan but is highest in early childhood, it underscores the importance of early childhood education and encourages providing children with access to a rich learning environment during this key stage of cognitive development. Understanding neuroplasticity builds teachers' knowledge base and provides support for principles such as growth mindset and for adopting a more positive perspective of student potential (Carrasco et al., 2015). Ultimately this challenges perspectives that hold intelligence as a fixed trait, and rather presents intelligence as a "malleable biopsychological potential to process information and problem solve" (Tokuhamma-Espinosa & Nouri, 2020, p. 67). This recognition that intelligence is not fixed is important in education as it highlights the crucial roles that education and the environment play in helping learners to reach their full potential (Sousa, 2011) and is also identified as a core concept by the Society for Neuroscience (2008). Increased understanding of neuroplasticity can support teachers' adaptive expertise by shifting perspectives about student potential.

EMOTION AND COGNITION

The idea that emotion and cognition are inextricably linked is another prominent core MBE concept, and several researchers examine this connection in the context of education.

We maintain that recognizing the interconnectedness of emotion and cognition can enhance a teacher's adaptive expertise by informing their use of strategies that address both emotional well-being and cognitive development, thereby fostering a more holistic, responsive, and effective learning environment. Researchers have found relationships between emotion and several other aspects of learning, such as memory, attention, motivation, associative learning, and interpersonal factors in the classroom. Hamann (2001) and LaBar and Cabeza (2006) explored the role emotion plays in encoding, showing that valence (positive or negative emotion) and arousal (the intensity of the emotion) impact activity in the hippocampus and amygdala. They found that arousal, regardless of valence, can enhance memory encoding (LaBar & Cabeza, 2006). Talmi (2013) explained the neurocognitive role emotion plays in attention and long-term memory, which is essential for learning, exploring the neural mechanisms that support emotional memories being remembered more vividly and accurately. Several other researchers have explored the connection between the neuroscience of emotion and classroom practices, both directly and indirectly (Hammond, 2014; Immordino-Yang & Damasio, 2007; Li et al., 2020; Whiting et al., 2021). In what has become a seminal work in the field, Immordino-Yang and Damasio (2007) suggested that emotion is the first form of cognition and cannot be divorced from learning. They suggest that emotion steers reasoning and decision making and is key to the ability to apply learning in novel contexts. Immordino-Yang and Damasio (2007) introduced two key concepts from their findings: emotion is critical for applying learning in real-world and social situations, and culture shapes our cognition through emotion. We propose that knowledge from MBE, such as how emotion and cognition are interlinked, can support teachers in their instructional decision making, thereby increasing their adaptive expertise.

USING KEY MBE CONCEPTS IN ADAPTIVE EXPERTISE

Teachers' knowledge of MBE principles, including neuroplasticity and the interconnectedness of emotions and cognition can have a positive impact on their decision-making processes (Hohnen & Murphy, 2016; Schwartz et al., 2019). MBE's interdisciplinary approach aims to ensure teaching practices are based on robust scientific research, which should support teaching efficiency, or teachers "working smarter, not harder" (Sousa, 2011). To achieve this, teachers should have a foundational understanding of how the brain learns for adapting their instructional and interpersonal decisions to best support student learning.

For example, we can consider how the concept of neuroplasticity could help support teachers' adaptive expertise. Knowledge of neuroplasticity allows teachers to recognize that the brain is malleable and constantly changing, which can influence teachers' pedagogical practices (Gholami et al., 2022). When teachers are aware that specific brain networks (such as those responsible for planning abilities) continue to develop during adolescence and are influenced by experiences, they can adapt their instruction to provide the necessary guidance and stimulation for students' cognitive growth (Dekker & Jolles, 2015).

Studies have demonstrated that training teachers in educational neuroscience concepts, including neuroplasticity, has tangible effects on their pedagogical practices. For instance, after participating in a course that included neuroplasticity concepts, teachers' lesson

plans became more student-centered, emphasizing approaches that promote individual growth and adaptability (Schwartz et al., 2019). Additionally, instruction in educational neuroscience has been found to significantly increase teachers' mindset beliefs, further supporting their adoption of a growth mindset and the integration of neuroscientific principles into their decision-making processes (Gutshall, 2020).

Furthermore, a recent study suggests that teachers who possess knowledge about neuroplasticity tend to also have a more sophisticated epistemological belief system and are less likely to hold a fixed mindset (Gholami et al., 2022), meaning they are more likely to view intelligence and abilities as traits that can be developed through effort and practice. Teachers' mindset beliefs can also impact their decision-making processes and actions towards students (Gutshall, 2020). Researchers suggest that incorporating the concept of neuroplasticity into teacher professional development could support teachers in developing a holistic and growth-oriented approach (Gholami et al., 2022), which could positively influence their instructional strategies and interpersonal interactions.

The interconnectedness of emotion and cognition also supports adaptive expertise. Talmi (2013) suggests that emotional events, regardless of valence, result in increased involuntary attention as compared to neutral emotions when paired together. This means students pay attention to, and encode, emotionally charged information better than neutral information when presented as a set. A practical way teachers could use this is in deciding which texts, examples, or activities to use in the classroom. Those experiences and examples with positive or negative connotations would be encoded with less effort than neutral connotations. The adaptive teacher may apply this knowledge about the impact of emotional events in planning instruction ahead of time or as an adjustment after observing students. Furthermore, research indicates there is a delay in memory encoding when information is presented during these emotionally charged moments (Talmi, 2013). Again, this is useful for teachers' instructional planning; it suggests utilizing a delay when asking students to retrieve the information may yield better results than immediate retrieval tasks, an instructional practice known as "spaced retrieval" (Karpicke & Bauernschmidt, 2011, p.1250). The expert teacher may use this knowledge by first analyzing students' affect and making an adaptive decision about when to ask students to retrieve the information.

Other researchers have found that positive and negative emotions affect memory, attention, and higher-order thinking (Hardiman, 2012; LaBar & Cabeza, 2006; Li et al., 2020; Zadina, 2014) - a concept that can inform instructional planning. Li et al. (2020) provided a framework for understanding how positive emotions impact learning. Their study positioned social interactions at the center of developing positive emotions to strengthen learning. Mentalizing and mirroring systems in the brain help students process cues from social interactions which become essential in productive cooperation (Li et al., 2020). In essence, the evidence of mirrored and synchronized neural activity when participants are engaged with each other provides scientific credence to cooperative practices, and provides insight into nuances about what effective collaboration looks like (Dikker et al., 2017). In the classroom, knowledge of these processes translates to designing learning opportunities and interactions that improve communication and collaboration. Incorporation of dialogue and communication opportunities in the classroom that cultivate positive interactions is a key feature of utilizing this framework (Sousa, 2011). A teacher with adaptive expertise could use this information to adapt their

decisions based on their observations and analysis of student behaviors during instruction and collaboration.

Further, Li et al. (2020) pointed out the contagious nature of emotion in social situations. Negative emotions in one person spread to others, and the converse is likewise true. Because of mentalizing and mirroring functions in the brain, teachers' awareness of this phenomenon and how it functions can be useful in responding to students experiencing negative emotions or crises as well as provide the rationale for creating intentional opportunities to cultivate positive emotions. Stress and anxiety can positively or negatively impact student performance depending on the degree of arousal (Zadina, 2014), which relates to a common phenomenon among students today: test anxiety (Von der Embse et al., 2018). High levels of anxiety can negatively affect higher-level cognitive functions, including critical thinking and metacognition – skills a student needs to utilize in many educational tasks (Zadina, 2014), whereas milder arousal, such as perceiving a difficult task as a challenge rather than a threat, can improve student performance (Travis et al., 2020).

Yerkes and Dodson (1908) recognized this relationship between arousal and performance over 100 years ago in what became known as the “Yerkes-Dodson Law,” which states moderate arousal is generally better for performance, whereas arousal levels that are too high or too low generally impede performance. This finding relates to testing anxiety because it suggests students who experience high anxiety during testing situations may be likely to have lower achievement on those tests because they are less able to perform optimally in these contexts (Richardson, Abraham & Bond, 2012). The observant and adaptive teacher may engage in more nuanced and deliberate observation of students prior to and during testing and be willing to find flexible solutions that allow students to do their best, such as providing more frequent, low-stakes testing that invokes a modest stress level in these students.

Teachers' understanding of the interconnectedness of emotion and cognition can have a profound impact on their decision-making processes, as emotions play a crucial role in learning and cognitive processes (Immordino-Yang & Damasio, 2007). When teachers recognize the interconnectedness of emotion and cognition, they can use this knowledge to improve learning opportunities for their students. Research emphasizes the importance of creating a positive emotional climate in the classroom as emotions shape attention, memory, and motivation (Hammond, 2014; Hardiman, 2012; Immordino-Yang & Damasio, 2007; Li et al., 2020; Whiting et al., 2021). Teachers who are aware of this connection can intentionally promote a sense of belonging among their students, foster meaningful connections between the content and students' personal experiences, and respond more flexibly to students' needs. By doing so, teachers can create an environment that supports cognitive engagement and deeper learning.

CONCLUSION

MBE principles can deepen teachers' understanding of how students learn and expand teachers' knowledge bases, providing greater potential to respond adaptively to the unique students and circumstances in their classroom. MBE concepts equip teachers with a foundational understanding of the cognitive and neural mechanisms that support learning, which can enhance teachers' adaptive expertise and enable them to make

research-informed decisions more flexibly and efficiently to optimize student learning outcomes. Key concepts in this area include neuroplasticity and the interconnectedness of emotion and cognition. Understanding MBE concepts more fully can inform teachers' pedagogical practices and promote adaptive expertise. The authors agree with previous recommendations that MBE principles be integrated into teacher training curricula. We propose this integration would support teachers with a holistic approach to education and empower teachers to more effectively meet the diverse learning needs of their students through broadened and research-informed adaptive expertise.

REFERENCES

- Ansari, D., Konig, J., Leask, M., & Tokuhamu-Espinosa, T. (2017). Developmental cognitive neuroscience: Implications for teachers' pedagogical knowledge. In S. Guerriero (Ed.), *Pedagogical knowledge and the changing nature of the teaching profession* (pp. 195-222). OECD Publishing.
- Blake, P. R., & Gardner, H. (2007). A first course in mind, brain, and education. *Mind, Brain, and Education*, 1 (2), 61-65. <https://doi.org/10.1111/j.1751-228X.2007.00007.x>
- Brick, K., Cooper, J. L., Mason, L., Faeflen, S., Monmia, J., & Dubinsky, J. M. (2021). Tiered neuroscience and mental health professional development in Liberia improves teacher self-efficacy, self-responsibility, and motivation. *Frontiers in Human Neuroscience*, 15. <https://doi.org/10.3389/fnhum.2021.664730>
- Carbonell, K. B., Stalmeijer, R. E., Könings, K. D., Segers, M., & van Merriënboer, J. J. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational research review*, 12, 14-29. <https://doi.org/10.1016/j.edurev.2014.03.001>
- Carrasco, J. G., Serrano, M., & García, A. G. (2015). Plasticity as a framing concept enabling transdisciplinary understanding and research in neuroscience and education. *Learning, Media and Technology*, 40(2), 152-167. <https://doi.org/10.1080/17439884.2014.908907>
- Chang, Z., Schwartz, M., Hinesley, V., & Dubinsky, J. M. (2021). Neuroscience concepts changed teachers' views of pedagogy and students. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.685856>
- Council of Chief State School Officers. (2013, April). *Interstate Teacher Assessment and Support Consortium INTASC model core teaching standards and learning progressions for Teachers 1.0: A resource for ongoing teacher development*.
- Dekker, S., & Jolles, J. (2015). Teaching about "Brain and Learning" in high school biology classes: Effects on teachers' knowledge and students' theory of intelligence. *Frontiers in Psychology*, 6, 1848-1848. <https://doi.org/10.3389/fpsyg.2015.01848>
- Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J. B., Rowland, J., Michalareas, G., Saxe, R., Ding, M., & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9), 1375-1380. <https://doi.org/10.1016/j.cub.2017.04.002>
- Dubinsky, J. M., Guzey, S. S., Schwartz, M., Roehrig, G. H., MacNabb, C., Schmied, A., Hinesley, V., Hoelscher, M., Michlin, M., Schmitt, L., Ellingson, C., Chang, Z., & Cooper, J. L. (2019). Contributions of neuroscience knowledge to teachers and their practice. *The Neuroscientist*, 25(5), 394-407. <https://doi.org/10.1177/10738584198354>
- Dubinsky, J. M., Roehrig, G. H., & Varma, S. (2013). Infusing neuroscience into teacher professional development. *Educational Researcher*, 42(6), 317-329. <https://doi.org/10.3102/0013189x13499403>

- Dudley-Marling, C., Abt-Perkins, D., & Sato, K. (2006). Teacher quality: The perspectives of NCTE members. *English Education*, 38(3), 167-193.
- Fairbanks, C. M., Duffy, G. G., Faircloth, B. S., He, Y., Levin, B. C., Rohr, J., & Stein, C. M. (2010). Beyond knowledge: Exploring why some teachers are more thoughtfully adaptive than others. *Journal of Teacher Education*, 61(1-2), 161-171. <https://doi.org/10.1177/0022487109347874>
- Fischer, K. W. (2009). Mind, brain, and education: Building a scientific groundwork for learning and teaching. *Mind, Brain, and Education*, 3(1), 3-16. <https://doi.org/10.1111/j.1751-228X.2008.01048.x>
- Gholami, K., Alikhani, M., & Tirri, K. (2022). Empirical model of teachers' neuroplasticity knowledge, mindset, and epistemological belief system. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1042891>
- Grotzer, T.A., Forshaw, T., & Gonzalez, E. (2021). Developing adaptive expertise for navigating new terrain: An essential element of success in learning and the workplace. *The next level lab at the Harvard Graduate Harvard graduate school of education. President and Fellows of Harvard College: Cambridge, MA.*
- Gutshall, C. A. (2020). When teachers become students: Impacts of neuroscience learning on elementary teachers' mindset beliefs, approach to learning, teaching efficacy, and grit. *European Journal of Psychology and Educational Research*, 3(1), 39-48. <https://doi.org/10.12973/ejper.3.1.39>
- Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, 5(9), 394-400. [https://doi.org/10.1016/s1364-6613\(00\)01707-1](https://doi.org/10.1016/s1364-6613(00)01707-1)
- Hammond, Z. L. (2014). *Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students* (1st ed.) [E-book]. Corwin.
- Hardiman, M. (2012). Informing pedagogy through the Brain-Targeted teaching model. *Journal of Microbiology & Biology Education*, 13(1), 11-16. <https://doi.org/10.1128/jmbe.v13i1.354>
- Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Developmental Psychology*, 20, 6, 27-36.
- Hattie, J. (2003, October). *Teachers make a difference: What works?* 乳幼児発達臨床センター年報 [Paper presentation]. Building Teacher Quality: What does the research tell us? ACER Research Conference, Melbourne, Australia. http://research.acer.edu.au/research_conference_2003/4/
- Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Taylor & Francis.
- Hohnen, B., & Murphy, T. (2016). The optimum context for learning; drawing on neuroscience to inform best practice in the classroom. *Educational and Child Psychology*, 33(1), 75-90. <https://doi.org/10.53841/bpsecp.2016.33.1.75>
- Howard-Jones, P. A., Jay, T., & Galeano, L. (2020). Professional development on the

- science of learning and teachers' performative thinking—A pilot study. *Mind, Brain, and Education*, 14(3), 267–278. <https://doi.org/10.1111/mbe.12254>
- Immordino-Yang, M. H., & Damasio, A. R. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, 1(1), 3–10. <https://doi.org/10.1111/j.1751-228x.2007.00004.x>
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1250. <https://doi.org/10.1037/a0023436>
- Kennedy, M. B. (2019). How we learn about teacher learning. *Review of Research in Education*, 43(1), 138–162. <https://doi.org/10.3102/0091732x19838970>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805. <https://doi.org/10.1037/a0032583>
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1), 54–64. <https://doi.org/10.1038/nrn1825>
- Li, L., Gow, A., & Zhou, J. (2020). The role of positive emotions in education: A neuroscience perspective. *Mind, Brain, and Education*, 14(3), 220–234. <https://doi.org/10.1111/mbe.12244>
- Männikkö, I., & Husu, J. (2019). Examining teachers' adaptive expertise through personal practical theories. *Teaching and Teacher Education*, 77, 126–137. <https://doi.org/10.1016/j.tate.2018.09.016>
- Ng, S., Forsey, J., Boyd, V., Friesen, F., Langlois, S., LaDonna, K. A., Mylopoulos, M., & Steenhof, N. (2022). Combining adaptive expertise and (critically) reflective practice to support the development of knowledge, skill, and society. *Advances in Health Sciences Education*, 27(5), 1265–1281. <https://doi.org/10.1007/s10459-022-10178-8>
- Novick, L. & Holyoak, K. (1991). Mathematical problem solving by analogy. *Journal of experimental psychology: Learning, memory, and cognition*, 17(3), 398. <https://doi.org/10.1037/0278-7393.17.3.398>
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. D., Pierczynski, M., & Allen, M. H. (2017). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242. <https://doi.org/10.3102/0034654317743198>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353. <https://doi.org/10.1037/a0026838>
- Schwartz, M., Hinesley, V., Chang, Z., & Dubinsky, J. M. (2019). Neuroscience knowledge enriches pedagogical choices. *Teaching and Teacher Education*, 83, 87–98. <https://doi.org/10.1016/j.tate.2019.04.002>

- Society for Neuroscience. (2008). *Neuroscience core concepts: The essential principles of neuroscience*. The University of Arizona. Retrieved June 8, 2023, from https://amygdala.psychdept.arizona.edu/SfnLocal/misc/core_concepts.pdf
- Sousa, D. C. (2011). Commentary: Mind, brain, and education: The impact of educational neuroscience on the science of teaching. *LEARNing Landscapes*, 5(1), 37-43. <https://doi.org/10.36510/learnland.v5i1.529>
- Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, 20, 165-184. <https://doi.org/10.1007/s11092-008-9053-z>
- Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science*, 22(6), 430-436. <https://doi.org/10.1177/0963721413498893>
- Tokuhamma-Espinosa, T., & Nouri, A. (2020). Evaluating what mind, brain and education has taught us. *Access: Contemporary Issues in Education*, 40, 63-71.
- Tokuhamma-Espinosa, T., & Nouri, A. (2023). Teachers' mind, brain, and education literacy: A survey of scientists' views. *Mind, Brain, and Education*. <https://doi.org/10.1111/mbe.12377>
- Tokuhamma-Espinosa, T., Simmers, K., Batchelor, D., Nelson, A. D., & Borja, C. (2023). A Theory of Mental Frameworks: Contribution to the special issue in Frontiers Psychology on enhanced learning and teaching via neuroscience. *Frontiers in Psychology*, 14, 1220664. <https://doi.org/10.3389/fpsyg.2023.1220664>
- Travis, J., Kaszycki, A., Geden, M., & Bunde, J. (2020). Some stress is good stress: The challenge-hindrance framework, academic self-efficacy, and academic outcomes. *Journal of Educational Psychology*, 112(8), 1632. <https://doi.org/10.1037/edu0000478>
- Von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of affective disorders*, 227, 483-493. <https://doi.org/10.1016/j.jad.2017.11.048>
- Whiting, S. B., Wass, S. V., Green, S., & Thomas, M. S. C. (2021). Stress and learning in pupils: Neuroscience evidence and its relevance for teachers. *Mind, Brain, and Education*, 15(2), 177-188. <https://doi.org/10.1111/mbe.12282>
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Punishment: Issues and experiments*, 27-41.
- Zadina, J. (2014). *Multiple pathways to the student brain: Energizing and enhancing instruction*. John Wiley & Sons.

FROM FIXED TO FLEXIBLE: NEEDED CONDITIONS TO PROMOTE ELEMENTARY TEACHERS' EQUITABLE USE OF WITHIN-CLASS ABILITY GROUPING

SARAH CAROLEO

PH.D. STUDENT

SCHOOL OF EDUCATION, JOHNS HOPKINS UNIVERSITY

ORCID 0000-0002-0041-9616

ABSTRACT

Within-class elementary grouping is a staple of modern elementary instruction, as it ideally provides a structure in which classroom teachers can better manage academic diversity. However, it is often implemented ineffectively and/or inequitably due to various structural, cultural, and political features of school systems and teacher training programs. In this essay, I seek to delineate solutions via flexible grouping that combat historical inequities associated with student ability grouping, ultimately to equip teachers to both manage academic diversity and ensure that all students receive appropriately challenging instruction each day. The arguments put forth are informed by my ten years of work as an educator, instructional coach, and researcher, in which I have witnessed a strong, practical need for elementary small group instruction but have also grappled with how ability grouping often inequitably sorts and fixes students into groups that fuel de facto tracking. This work will benefit school and district leaders and teacher preparation programs, as they seek to address systemic issues related to teachers' ability grouping practices. Most importantly, it will provide tangible strategies and descriptions that equip elementary educators to leverage more flexible, equitable grouping practices in their classrooms.

Key Words: Ability grouping; elementary teaching; differentiation; flexible grouping; small group instruction.

Author Note

I would like to extend my sincere thanks to Dr. Julia Burdick-Will, who provided initial guidance on the structure and framing of this piece, and Dr. Jonathan Plucker, who gave helpful feedback for revisions and edits. I would also like to express my gratitude to the editorial board of The Neag School of Education Journal, who supplied constructive and supportive feedback that expanded and strengthened the essay.

INTRODUCTION

Elementary teachers are typically tasked with teaching all subjects to a body of students displaying highly varied levels of readiness for grade-level content, prompting many to utilize within-class small group instruction to level core reading and math instruction (Esposito, 1973; Sørensen, 1970). Research outlines the potential benefits of within-class ability grouping; it helps teachers provide differentiated instruction to academically diverse students (Adelson & Carpenter, 2011; Adodo & Agbayewa, 2011; Castle et al., 2005; Slavin, 1987) and become more familiarized with students' unique personalities and learning dispositions (Altintas & Ozdemir, 2015). Within-class ability grouping can also increase students' engagement since they interact more with the teacher and peers in small groups (Altintas & Ozdemir, 2015; Castle et al., 2005), as well as increase students' self-regulated behaviors (e.g., monitoring personal progress and talking about thinking) since learning is often more active in small groups (Stright & Supplee, 2002). In his best-evidence synthesis, Slavin (1987) found that within-class ability group instruction can have positive effects on student achievement when three criteria are met: (1) the grouping is based upon targeted skill differences of focus across students; (2) teachers flexibly move students based on current levels of understanding; and (3) teachers alter the pace and level of group-level instruction to correspond to students' readiness and rate of learning. A more recent second-order meta-analysis (Steenbergen-Hu et al., 2016) drew promising findings, as well; across 13 meta-analyses about ability grouping, within-class ability grouping had positive and significant effects (with effect sizes ranging from 0.19 to 0.30) on all students' subsequent academic achievement, regardless of initial ability level.

A notable concern often arises in this deeply normed model, however; teachers frequently do not receive sufficient—if any—pre-service or in-service training and resources that equip them to meet any of Slavin's (1987) three criteria (Fitzgerald et al., 2021; Harris, 2010; Tomlinson et al., 1994). To effectively and equitably group students for small group instruction, teachers need to acquire pedagogical/psychological knowledge and hold a range of skills that extend from that knowledge (Voss et al., 2011): how to evaluate student strengths, needs, and misconceptions related to a unit of study; how to match optimal peer groups for each topic; how to meaningfully differentiate instruction around the content standard; how to know when to move a student to a different group; how to appropriately challenge all learners in every group; and more. Structural, cultural, and political dynamics of schools often leave teachers undertrained in this knowledge and these skills, under-resourced to implement differentiated instruction, and pressured to group students in fixed ways that contribute to de facto segregation patterns (Buttaro et al., 2010).

For example, without needed training and resources, many elementary teachers utilizing within-class ability grouping tend to disproportionately assign students from lower social classes, students with perceived behavioral challenges, students with disabilities, and students of color into lower ability groups and provide decontextualized instruction at a slower pace with conveyed low expectations in those groups (Becton, 2018; Calarco, 2014; Eder, 1981; Esposito, 1973; Oakes, 1992). These inequitable grouping patterns fuel early learning gaps and poor academic self-concept that negatively affect later achievement of students fixed into "low" ability groups from a young age (Plucker & Peters, 2016; Tyson, 2011).

In this essay, I provide historical context for how ability grouping has evolved in American K-12 classrooms, articulate why flexible grouping practices are more effective and equitable than fixed grouping practices, and synthesize how teacher training programs and school leaders can make structural changes to better support teachers' facilitation of flexible grouping. Until intentional, systemic action is taken to move elementary teachers away from the use of fixed ability grouping, American schools will continue to foster opportunity and performance gaps that harm historically marginalized groups of students. Thus, it is imperative to support and guide elementary educators in the successful use and implementation of flexible ability grouping.

THE HISTORY OF STUDENT GROUPING IN THE UNITED STATES

To avoid and repair the problematic aspects of within-class ability grouping in modern elementary settings, it is critical to first understand how grouping practices have manifested in American classrooms over time—and what they have meant for the outcomes of historically marginalized students. Prior to the 1820s, children of American colonizers largely learned from family and community members either at home or in one-room schoolhouses (Tyack, 1974). However, in the 1820s, a mass of immigrants entered the country, making it increasingly harder to serve a large range of children across all ages in one-room settings. Around the 1840s, formal public schools were ideated and formed across the country; by the 1860s, age-based grades were established in most schools to foster more homogeneity of developmental needs in classrooms. This structure proved helpful in accounting for the range of academic diversity among students. From 1900 to 1920, the United States experienced another boom in immigration, meaning that class and school sizes grew quickly in number again (Ellison & Hallinan, 2004; Worthy, 2010). Simultaneously, the eugenics movement, a classist, ableist, and White supremacist movement committed to elevating “genetically superior” individuals and families in society and separating “genetically inferior” from public spaces, was on the rise and bled into schooling policies and practices (Brookwood, 2021).

In the 1910s, school systems began employing homogenous grouping strategies beyond age-based grade levels in schools (Goldberg et al., 1966), likely motivated by both the practical need to address the sharp influx in student enrollment and the eugenicist philosophy that certain people, typically those who were Western European, needed to be equipped for certain roles in society—thus, grouping students according to “ability” was a natural means to prepare children for their “deserved” roles. In elementary schools, within-class ability groups emerged, in which teachers used small group instruction with relatively homogeneous groups of students, particularly for reading (Barr & Dreeben, 1983; Ireson & Hallam, 1999). In high schools in the 1930s, tracking, a form of between-class grouping, arose as a dominant grouping strategy (Ellison & Hallinan, 2004; Oakes, 1985). Tracking was a practice in which students were assigned to a certain level—vocational, general, or academic—based on their past school achievement and/or their Intelligence Quotient (IQ) score, which derived from tests developed by leading eugenicists who endorsed separating “superior” children from “inferior” peers in schools (Brookwood, 2021; Steenbergen-Hu et al., 2016). Students then primarily completed coursework associated with their distinct track and remained fixed in their tracks for the duration of their schooling experience (Oakes, 1985).

Initial research around these forms of ability grouping suggested that they increased student achievement. However, Goldberg et al. (1966) published an article that highlighted how most prior studies had not accounted for several confounding factors, such as class size, number of students involved, teaching methods, and more. Subsequently, researchers began accounting for these confounders in their statistical models and often found either null or negative effects of tracking and ability grouping for students placed within “average” and “low” ability groups (Barr & Dreeben, 1983; Eder, 1981; Esposito, 1973; Gamoran, 1986; Oakes, 1985), meaning that performance gaps widened even further as “high” achieving students were the only group who benefitted academically from tracking and within-class grouping. On average, students from low-income backgrounds, students of color, and students with disabilities disproportionately comprised these “low” ability groups (Esposito, 1973; Oakes, 1985)—fulfilling the eugenicist aim of separating students by class, race, and “ability.”

In elementary settings, several studies found that instruction for the designated “low” ability groups was often facilitated at a slower pace and focused more time on decontextualized skills (Allington, 1983; Eder, 1981; Gambrell et al., 1981). Teachers were more likely to hold and convey lower expectations and negative feelings toward students in “low” groups (Eder, 1981; Good & Brophy, 1972), showing more concern for managing their behavior than providing appropriately challenging instruction (Eder, 1981)—practices likely driven by implicit and/or explicit teacher bias against students from historically marginalized backgrounds. Students remained relatively fixed in their assigned ability groups; Barr and Dreeben (1983) found that 70% of first graders remained in the same reading group throughout the duration of their school year. In high schools, researchers found cumulative effects of students’ previous ability group assignments and their differential access to quality instruction on their achievement; studies suggested that their placements reinforced and exacerbated their initial perceived ability differences (Alexander & Cook, 1982; Ellison & Hallinan, 2004; Gamoran, 1989; Ireson & Hallam, 1999). Scholars also voiced concerns for the negative effects on qualitative aspects of students’ lives, such as motivation, academic self-concept, and confidence, that might arise when certain students are primarily fixed into lower groups or tracks (Esposito, 1973; Oakes, 1992; Oakes et al., 1997).

Although several studies indicated that students’ group placements were most strongly associated with their initial achievement scores, students’ socioeconomic status was often significantly associated with their group placements (Esposito, 1973; Steenbergen-Hu et al., 2016; Worthy, 2010), suggesting that this structure may have served as a capital-reproducing mechanism for students who entered the K-12 system with privilege (Bourdieu, 1973). Some noted how the organizational structure of tracking seemed to perpetuate de facto segregation of students both racially and socioeconomically (Buttaro et al., 2010; Eder, 1981; Oakes, 1985,1992), launching a movement that urged schools to detrack their organizational structures and point resources toward providing all children with high-quality instruction. By the mid-1990s, most schools had minimized or eliminated tracking and within-class grouping (Steenbergen-Hu et al., 2016), and research surrounding the effects of ability grouping faded (Worthy, 2010).

ABILITY GROUPING IN MODERN ELEMENTARY CLASSROOMS

At the turn of the 21st century, with a stronger push for accountability via the passage of the No Child Left Behind Act and increased standardized testing, the use of within-class ability grouping gained new saliency in elementary settings (Ireson & Hallam, 1999; Steenbergen-Hu, 2016; Tomlinson, 2000). In 1998, 29% of elementary students participated in ability groups as part of reading instruction; by 2009, that number jumped to 71% (Loveless, 2009). Likewise, by 2011, 61% of elementary students participated in ability groups for mathematics instruction (Steenbergen-Hu et al., 2016). It makes sense that this structure became useful in a testing-driven context, as homogeneous grouping can allow teachers to better adjust instruction to match students' current needs in relation to grade level standards and thus ameliorate external pressure for students performing just below grade level (i.e., "bubble kids") to meet grade level standards by testing time. Nationally representative data suggested that some overall progress was made in elementary grades during this time of increased accountability and within-class ability grouping, with the National Assessment of Educational Progress (NAEP) average 4th grade reading scale scores significantly moving from 213 to 221 (National Center for Education Statistics, 2012); however, those same growth patterns did not occur for students with diagnosed learning disabilities (Becton, 2018). The concept of differentiated instruction (DI) grew in popularity in both research and school settings with this new boom of within-class ability grouping (Altintas & Ozdemir, 2015; Pozas et al., 2020; Santangelo & Tomlinson, 2012; Tomlinson, 2000), bringing much-needed discussions about how to innovate instruction and better match instruction to students' current Zones of Proximal Development (Vygotsky, 1978).

Some equity issues persist with this modern form of within-class ability grouping, though. For example, using ECLS-K data, Buttaro et al. (2010) found that within-class ability grouping in kindergarten was more frequently used in schools with higher levels of racial and socioeconomic diversity and proportions of students of color; conversely, majority-White schools were the least likely to use within-class grouping. This finding suggests that this structure may still function as a mechanism that inequitably separates students within schools or distills instruction for students wrongly assumed to be incapable of handling grade-level content. Students from low-income backgrounds and students of color continue to be disproportionately assigned to "low" ability groups (Adodo & Agbayewa, 2011; Buttaro et al., 2010; Condrón, 2007), and students placed in "high" ability groups still enjoy better academic and social gains from participating in small group instruction than those placed in "low" groups (Bradbury, 2018; Buttaro et al., 2010; Castle et al., 2005; Marks, 2013; Plucker & Peters, 2016). Some research suggests that within-class ability grouping has yet to provide positive effects for students with disabilities (Becton, 2018), likely because they are over-placed and fixed into low ability groups, and teachers may hold or convey differential, biased expectations of them.

When considering both the pitfalls and benefits of elementary within-class grouping over time, a few themes emerge. First, fixed placements into ability groups or tracks typically lead to inequitable distribution of instructional time and resources, benefitting those who already hold privilege prior to entering kindergarten (Bourdieu, 1973; Oakes et al., 1997; Tomlinson, 2000); thus, any form of within-class grouping that employs fixed grouping practices should be eliminated as much as possible from modern elementary education settings.

Next, labelling practices—where educators label a student as “low”, “average”, or “high”, either implicitly or verbally to other adults—can reinforce notions of fixed ability, which may lead educators to make assumptions about where students belong and of how much they can achieve (Barr & Dreeben, 1983; Eder, 1981; Tyson, 2011). Even if unintentional, these assumptions can greatly impact students’ academic self-concepts and access to prerequisite content needed for secondary and postsecondary success (Buttaro et al., 2010; Corbett Burris et al., 2008; Oakes, 1992; Tyson, 2011). Therefore, educators should be mindful of the labels they assign to children in relation to ability; question their own assumptions about ability, its origins, and its development; and reflect upon how those assumptions affect their teaching of their students.

Finally, when grouping is implemented in a flexible way that builds homogeneity around the target skills of focus, as Slavin (1987) originally found, within-class ability grouping can support teachers in accounting for wide ranges of academic diversity in their classroom—a challenge that often arises in elementary settings. However, the vision of all American teachers using this model flexibly has not yet been realized and requires intentional action to progress.

COMPARING FIXED AND FLEXIBLE GROUPING PRACTICES

As discussed above, within-class ability grouping as an instructional model pervades modern elementary classrooms (Steenbergen-Hu et al., 2016), and its use can vary significantly based on teachers’ training, teaching philosophies, management practices, and personal capacities (Adelson & Carpenter, 2011; De Neve et al., 2015; Chandra Handa, 2020). To increase the efficacy and equity of within-class ability grouping, teacher preparation programs, school instructional leaders, and teachers must become familiarized with the differences between fixed and flexible grouping and reflect on to what extent their classroom and school norms are fixed or flexible. Table 1 serves as a quick guide for comparison, and elaborations are provided below.

Table 1
Characteristics of Fixed Grouping versus Flexible Grouping

<u>Fixed Grouping</u>	<u>Flexible Grouping</u>
1. Groups are formed with the use of standardized or benchmark test results. (Fitzgerald et al., 2021)	1. Groups are formed with the use of frequent diagnostic and formative assessment data. (Borland et al., 2002; Castle et al., 2005; Tomlinson et al., 2003)
2. Students are considered and labelled high-, average-, or low-performing. (Fitzgerald et al., 2021)	2. Students’ current strengths and next steps for growth are routinely considered. (Tomlinson et al., 2003)
3. Little to no movement across groups occurs. (Missett et al., 2014; Tomlinson et al., 1997)	3. Students are regularly moved to a group with their shared goal for the day or week. (Tomlinson et al., 2003)

4. Student behaviors inform their group placement. (Kim et al., 2020; Legette et al., 2021; Tomlinson et al., 2003)	4. Academic needs and growth are valued over behavioral prevention. (Tomlinson et al., 2003)
5. Students with similar demographic backgrounds often comprise the same group. (Borman & Dowling, 2010; Condrón, 2007; Van Houtte et al., 2013)	5. Students routinely work with academically, racially, linguistically, and socioeconomically diverse peers. (Tomlinson et al., 1997; Tomlinson et al., 2003)
6. The teacher assumes that students' performance will be stable across all subjects. (Tomlinson et al., 2003)	6. The teacher expects that students' performance will vary across subjects and topics. (Missett et al., 2014; Tomlinson, 2000; Tomlinson et al., 2003).

FIXED GROUPING CHARACTERISTICS

Certain characteristics arise in classrooms that primarily utilize fixed ability grouping. Fixed ability groups are often formed using results from school-mandated, standardized assessments (Haller & Davis, 1981; Tomlinson, 2000). Teachers may consider students' assessment scores from the previous school year or beginning-of-year or end-of-semester diagnostic scores and categorize them as high-achieving, average, or low-achieving based on their results. Fitzgerald et al. (2021) found that many teachers relied on standardized tests to inform grouping decisions, and results from those tests frequently led them to hold fixed conceptions of students' abilities. Thus, once teachers categorize a child as low achieving, per their test score, they may be more inclined to perceive that child as "low ability" and keep them in a lower-level group. Relatedly, fixed groups experience little to no movement throughout the year (Missett et al., 2014; Tomlinson et al., 1997). This means that students designated as cognitively gifted or high achieving may remain in the advanced group for most of the year, while students with learning disabilities, learning challenges, perceived behavioral difficulties, or other reasons (e.g., discriminatory ideas about who is worthy of receiving academic challenge) remain mostly in lower-classified groups (Eder, 1981; Kalogrides & Loeb, 2013; Tyson, 2011). Classifying children in the same level of ability group across multiple subjects (i.e., average for both math and reading) can also be associated with fixed grouping practices if the classification is based more on teachers' personal perceptions and biases than data that supports the grouping (Tomlinson et al., 2003).

Some fixed ability groups are more contingent on students' behavior—and teachers' racialized ideas about behavior—than exhibited ability (Legette et al., 2021; Tomlinson et al., 2003). Since the groups remain relatively consistent throughout the year, the teacher may prioritize having a perceived "manageable" blend of students within each group. Eder (1981) discovered that teachers were more likely to group students based on their behavior, placing children with higher levels of distractibility and interruption into lower groups, regardless of their current readiness levels. This resulted in those groups covering much less content within the allotted instructional time (than the more behaved groups), which perpetuated a self-fulfilling prophecy of "low-ability" groups learning and achieving

less. Kim et al. (2020) also noted that elementary teachers most prioritized behavior in their grouping decisions, specifically aiming to prevent combinations of classmates who could generate behavioral issues when paired—even if they displayed similar levels of readiness for the content. Furthermore, in fixed settings, it is more likely that students with shared demographic characteristics (i.e., racial, linguistic, socioeconomic) will be placed in similar groups (Borman & Dowling, 2010; Condrón, 2007; Van Houtte et al., 2013). Teachers often hold deficit mindsets, perceiving children from lower socioeconomic, single-family, racial minority, and/or immigrant homes as having parents less invested in education or having less exposure to early learning experiences; this leads them to more frequently place and keep these children in lower-ability groups (Calarco, 2014; Gordon & Nocon, 2008; Van Houtte et al., 2013). Put together, fixed grouping typically perpetuates inequitable outcomes for historically marginalized students.

FLEXIBLE GROUPING CHARACTERISTICS

Instead of a reliance on school-mandated, standardized tests, flexible grouping typically relies upon frequent and formative assessments conducted by the teacher (Borland et al., 2002). The teacher might conduct pre-assessments aligned with the curriculum prior to new units, so they can determine students' readiness for that particular subject and topic (Tomlinson et al., 2003). Exit tickets, which are brief quizzes reviewing the major concepts of the lesson, also inform the teacher of students' daily mastery, so they can evaluate if the child needs additional or different support for the following lesson. As they evaluate students' pre-assessment and exit tickets responses, the teacher considers students' relative strengths and weaknesses and flexibly moves them to the group with which they share the most commonalities at that time. In flexible grouping, teachers can quickly respond to students' changes in needs, achievement levels, and motivations by moving them to the group that will best serve their present status.

For example, if a teacher is working with a grade-level group on a fractions unit and notices that one student continues to speed through learning tasks, answering all questions correctly, they may move the student into the advanced-level group the following day to explore deeper fractional concepts with open-ended application opportunities. For the next unit, the teacher would provide another pre-assessment related to that unit's content to determine the best-fit group for that student again; after all, excelling in fractions does not necessarily guarantee advanced readiness for the following unit's concepts. Such a model requires deliberate teacher planning, not only in structuring the learning time and managing student movement, but also in having frequent, formative assessments readily available (Castle et al., 2005; Rubenstein et al., 2015). A strategic and respectful means of communicating the assignment of groups is required; teachers might use something like a digital/written three-by-three table chart with group names, students' names, and a rotation schedule attached that can easily be updated to accommodate the flexibility of student movement and differentiated tasks as need (Tomlinson et al., 2003).

Missett et al. (2014) described how students sorted into seemingly homogeneous groups (e.g., students with learning disabilities, students at grade level, and students identified as gifted or high-achieving) remain inherently heterogeneous even from each other; two identified gifted children (who are often assumed to be similar based on the label) will not hold the same strengths as each other and may exhibit different learning

and motivational struggles (Tomlinson et al., 2003). When educators lock students into one ability level across subjects, they discount the potential strengths and needs those children hold and carry into learning (Tomlinson, 2000; Tomlinson et al., 2003). Thus, flexible grouping assumes that children will perform differently across subjects and topics. It also discourages sorting students based on perceived behavior and, instead, encourages (a) considering students' strengths, interests, and learning motivations related to the concept/skill of focus and (b) pairing them in groups with peers who might complement those well (Kim et al., 2020). It is a higher priority for academic needs to be met in the moment than it is to prevent behaviors from occurring (Tomlinson et al., 2003). This means that if two students chat frequently or routinely disagree but share similar learning patterns for the present unit, the teacher will still place them in the same group—and use alternative management strategies to keep them focused on learning. Since students can move flexibly across groups, it is more common to see peers across diverse demographics and educational backgrounds interacting with and learning from each other (Tomlinson et al., 1997; Tomlinson et al., 2003). In sum, flexible grouping practices aim to ensure that all children learn in a best-fitting environment throughout the scope of the school year, and the teacher utilizing the practices assumes that those environments will regularly change across time, subject, and topic.

EFFECTS OF FLEXIBLE GROUPING AND FIXED GROUPING

In most classrooms, teachers likely adopt a variety of both kinds of grouping practices, based on their current beliefs, capacities, and resources. There are several reasons why researchers endorse the use of flexible grouping over fixed grouping, though. Fixed grouping perpetuates the long-standing conception that ability is inherent and stable over time—someone either is or is not “high-achieving.” This norm has been confronted by research regarding the malleable, flexible nature of intelligence and the impact of environment on one's development (Barab & Plucker, 2002). However, the societal messaging around stratifications (e.g., those who belong in “low” or “high” strata deserve to be placed in them based on their displayed effort; people are born with a certain level of ability and cannot change) reinforces fixed ability thinking as truth (Gamoran, 1989; von Hippel et al., 2018).

Students in relatively fixed ability groups may then be susceptible to fixed ability thinking, where they internalize societal messaging that they are either inherently good or bad at a given domain, based on how they are repeatedly sorted (Fitzgerald et al., 2021; Hargreaves, 2019; Marks, 2013; Tyson, 2011). They may also experience stereotype threat, which describes how individuals' performance suffers from awareness that the identity group(s) to which they belong are not expected to do well (Hartley & Sutton, 2013). Fixed ability thinking acts as a barrier for students in acquiring the motivation and efficacy needed to persist when they face learning challenges (Marks, 2013). Stereotype threat can activate students' anxiety and ultimate disengagement from learning to protect themselves from feared future failures (Steele & Aronson, 1995). Although most teachers do not explicitly articulate each groups' level or designation, students may be able to infer it, based on their personal grades or comparisons to other peer groups, as witnessed in Hargreaves's (2019) study of elementary students who compared their fixed ability groups and test scores with others and subsequently altered their expectations of their personal

abilities. Since students are moved more frequently with flexible grouping and can also be sorted based on interests and strengths, it is less likely that students will attach themselves and their academic self-concept to one level of ability (Marks, 2013).

Perhaps the greatest criticism of fixed grouping practices is that when used repeatedly and primarily as the mode of sorting students, they can perpetuate historical cycles of inequality in schools (Borland et al., 2002; Buttaro et al., 2010; Condrón, 2007; Eder, 1981; Plucker & Peters, 2016). Students of color, students from lower socioeconomic groups, students with disabilities, males, and students from single-parent families are frequently overrepresented in low-ability groups, while White, female, and middle- or upper-middle class students predominantly comprise high-ability groups (Adodo & Agbayewa, 2011; Becton, 2018; Condrón, 2007; Ford, 2011). This can result in historically marginalized groups receiving less access to rigorous instruction and less exposure to classmates who may enrich their thinking (Adodo & Agbayewa, 2011; Kalogrides & Loeb, 2013).

Flexible grouping, on the other hand, has been suggested to provide benefits to both students and teachers (Castle et al., 2005; Plucker & Peters, 2016; Rubenstein et al., 2015; Slavin, 1987). By measuring significant differences in achievement score percent changes over time, Castle et al. (2005) found that in a high-needs school employing flexible grouping over five years, increases in the percentages of students scoring at mastery increased from 10% to 57%. In a study of third grade teachers who were given a pre-differentiated curriculum with included pre-assessments and tiered learning activities (to ease the planning load for teachers), Rubenstein et al. (2015) confirmed that students enjoyed their flexible groups, and their teachers reported increased student engagement with both academic content and peers. This makes sense because flexible movement grants them frequent exposure to different kinds of learners and thinking, which may also support expansion of social networks and a greater sense of community in the classroom. Furthermore, Carol Tomlinson, perhaps one of the most prolific scholars of DI, consistently touts flexible grouping as the most appropriate and respectful means to meet diverse students' learning needs and ensure that all students, regardless of entry point, learn in the classroom (Tomlinson, 2000; Tomlinson et al., 1997; Tomlinson et al., 2003).

Although flexible grouping can address several equity concerns associated with ability grouping, it poses a few practical challenges. Borland et al. (2002) reasonably argued that it is easier to advocate for flexible grouping than it is to implement it because it (a) presents scheduling challenges for administrators, (b) constitutes a change in school culture to adopt curricular differentiation practices, and (c) requires more instructional resources and time for teachers to implement it well. Furthermore, teachers can believe they use a model of flexible grouping, when, in reality, they do not differentiate tasks, materials, or content to the standard of Tomlinson's (2000) DI framework (Chandra Handa, 2020; Maker & Schiever, 2005; Missett et al., 2014; Pozas et al., 2020). When DI is not accomplished, the positive learning effects associated with ability grouping do not occur (Slavin, 1987). Similarly, some teachers do not use formative assessments correctly and may group students based on their perceptions (which increases the potential for bias) instead of what the assessments reveal; this is the primary risk for continued inequity via flexible grouping (Missett et al., 2014). For example, if a fourth-grade emergent bilingual student scores highly on a division pre-assessment, thus constituting the need for advanced differentiation, but the teacher assumes the child may not be able to access the content due to linguistic barriers, then the teacher may place the child in a lower group,

regardless of their pre-assessment results. This would prevent that child from receiving learning opportunities fit to their level of readiness, based on the teacher's assumptions or bias.

Each of these criticisms are valid, proving that flexible grouping practices are not a fix-all on their own; the model relies upon teachers garnering pedagogical-psychological knowledge around how to facilitate groups, how to reflect upon their own biases, and how to identify appropriate times to move children across groups (Heyder et al., 2017; Santangelo & Tomlinson, 2012). Unfortunately, several studies show that this is not a prioritized knowledge base in teacher preservice programs or in-service teacher professional development (Brigandi et al., 2019; De Neve et al., 2015; Evans & Waring, 2008; Fitzgerald et al., 2021; Rubenstein et al., 2015; Santangelo & Tomlinson, 2012). Therefore, certain conditions at the training and K-12 school levels are needed to shift teachers' grouping practices from more fixed in nature to more flexible.

ORGANIZATIONAL INFLUENCES ON ABILITY GROUPING

In understanding systemic ability grouping patterns in American schools, it is helpful to consider teacher training programs and school organizations acting as influential forces in ability grouping, rather than just examining individual teachers' practices (Buttaro et al., 2010). Tomlinson et al. (1994) attribute fixed grouping practices to teacher training programs that insufficiently cover grouping pedagogy with teacher candidates. She argues that when the programs do not provide sufficient modeling—either in their own instruction of teachers or in practicum teaching opportunities—teachers do not have a comprehensive model to which they can refer for their own grouping practices. Once they are placed in a classroom, expected to manage complex responsibilities with little preparation for grouping or the needed capacity for decision-making required of it, it makes sense that they rely upon more traditional, fixed notions of ability, as likely witnessed in their own schooling experiences and as modeled by their school leadership.

Buttaro et al. (2010) attribute observed de facto segregation via ability grouping to three organizational features of schools: structural, cultural, and political. The structural aspect of school organizations relates to the school's characteristics, such as their resources, school size, class size, leadership, and school body demographics. Looking at kindergarten ability grouping data in a large, nationally representative dataset, the authors found that schools with higher proportions of students of color were more likely to utilize within-class ability grouping; they used a composite variable to determine percent minority population versus percent White, so they did not determine any further racial differences. They also discovered that teachers with larger class sizes more frequently used within-class ability grouping to support the large spread of student needs; their data did not differentiate whether grouping was fixed or flexible.

Furthermore, school organizations often do not supply teachers with the curricular materials needed for effective flexible grouping (Harris, 2010). As mentioned earlier, flexible grouping relies upon frequent, formative assessment, which means that teachers readily require unit pre-assessments and lesson exit tickets (Tomlinson et al., 2003). Considering that many elementary teachers must plan daily lessons for all subjects, and then add small group lessons that require different types of instruction for some of those subjects, they have much to plan, often with insufficient planning time provided by school

leaders (Wu, 2013). If the school's mandated curriculum does not supply pre-made pre-assessments and formative assessments, teachers must create them on their own, which is unlikely to occur if they have not been trained in it or do not understand the value behind it. It is more common for district curriculums to be standardized in nature, centered around grade-level content, with little resources for assessing academically diverse students or differentiating content for them (Tomlinson, 2000). In prioritizing planning of instruction, it is understandable that teachers may not have time to create such formative assessments and therefore rely upon standardized assessments' designated ability levels to group students and standardized curriculum to teach groups.

The cultural feature of school organizations represents the norms and ideas most valued by stakeholders within the school (Buttaro et al., 2010). If school administration highly values certain behaviors or exudes fixed ability thinking about students or teachers, it is likely that those values become normed within the school. Teachers' personal backgrounds and beliefs or the school's specific values may inform how they define ability, which leaves any child not in alignment with what is considered "able" vulnerable to inappropriate group placement (Bradbury, 2018). For example, if a teacher values students who complete and submit homework, they may consider a child who consistently puts forth effort on homework as being better suited to a high-ability group than a disorganized student (Van Houtte et al., 2013). Often, school leaders and teachers are more likely to value students whose academic, behavioral, and social skills align with their own background, so their cultural norms shape how they perceive students' ability (Condrón, 2007). Condrón noted how around 80% of the teaching workforce is comprised of White women, so they may be more likely to unevenly distribute rewards and higher placements towards White students, which could explain the racial disparities often seen in ability groups. Borman and Dowling (2010) similarly found that teachers were more favorably biased towards middle-class students in their grouping decisions, sorting them into groups provided with more enriching tasks.

Finally, political aspects of school organizations may place pressure upon teachers to more frequently utilize fixed practices. For instance, school organizations and site leadership highly concerned with standardized test scores may encourage teachers and staff to practice educational triage through small group instructional time (Bradbury, 2018). Educational triage occurs when students are sorted into three ability groups: safe (meaning they will likely pass the state's standardized test); borderline (meaning they may be close to passing); or hopeless (meaning they are deemed too behind to pass grade-level questions within the year) (Booher-Jennings, 2005; Bradbury, 2018; Tomlinson, 2000). Once students are grouped, these schools ration and allot resources (such as length of instructional small group time, amount of instructional coaching and intervention time, and tutoring opportunities) mostly towards students identified as borderline to increase their likelihood of passing. This rationing—distributed to children in relatively fixed groups—blocks access to instructional resources for students within the two other groups. It is inherently unequal, yet the political pressure imposed upon teachers to improve test scores informs their grouping practices (Bradbury, 2018). Furthermore, if certain parents hold political clout within the school organization (e.g., involvement in Parent-Teacher-Association, frequent volunteering, frequent fundraising donations), they may be more able to assert pressure on school leaders and teachers, who in turn might reward their child with higher placements in groups, access to enrichment opportunities, or

designations of giftedness (Gordon & Nocon, 2008; Harris, 2010; Oakes et al., 1997). When parents frequently communicate with the teacher, asking for help and extension activities to conduct at home, the teacher might conflate effort with ability, thus perceiving the child to have higher abilities than actually exhibited (Calarco, 2014; Gordon & Nocon, 2008). All these structural, cultural, and political features work in tandem and can drive the extent of fixedness or flexibility in teachers' practices.

NEEDED CONDITIONS FOR SYSTEMIC IMPROVEMENT

Teacher training programs and school organizations should consider solutions targeted towards the outlined organizational features and their related issues that drive inequitable grouping in elementary classrooms. Suggestions for consideration are presented below.

PRESERVICE CONDITIONS

Concentrated efforts should be implemented in preservice training programs to better equip teacher candidates to understand and manage a diverse body of learners. De Neve et al. (2015) called for thorough modeling and subsequent explanations of decision-making by course professors, so preservice teachers understand what drives grouping decision-making and can visualize how it is organized and implemented in the classroom. They also recommended that preservice teachers conduct lessons in a lab classroom with each other, where the course professor and classmates can provide subsequent feedback and strategies for future consideration. Because teachers' biases can shape how they view students' abilities or potential, preservice programs could also incorporate critical reflections on personal biases and how they can manifest in teachers' mindsets around ability and grouping of students (Evans & Waring, 2008). Relatedly, Ford (2011) urged programs to train teachers in multicultural giftedness, so they become aware of how potential and talent manifest differently across cultures and environmental contexts and can appreciate talent in children with backgrounds different than their own—an important notion, considering the demographics of the largely White and female teaching workforce.

For flexible grouping to be effective, per Slavin's (1987) findings, teachers should also understand the pedagogy of DI. As mentioned earlier, teachers often believe they differentiate their instruction per small group, but several studies have shown a lack of understanding in what qualifies as true DI of content, process, and product (Maker & Schiever, 2005; Missett et al., 2014; Pozas et al., 2020; Santangelo & Tomlinson, 2012; Tomlinson et al., 1994). Therefore, researchers recommend preservice programs dedicate at least one course to DI, specifically highlighting how to use pre-assessments and other formative assessments to inform equitable and effective grouping decisions (Fitzgerald et al., 2021; Haller & Davis, 1981; Tomlinson et al., 1994). In such classes, leaders could train teacher candidates to recognize academic readiness and not conflate it with compliance; teachers should understand that perceived positive behavior does not translate to qualification for a certain level of ability group (Tomlinson et al., 1994).

Since DI can be difficult to facilitate, teachers should understand how to manage a flexibly grouped, DI classroom and practice leading flexible ability groups during student teaching experiences (Tomlinson et al., 2003). For example, programs can teach candidates how to respectfully communicate and change group assignments, as well as

how to establish behavioral and work expectations for independent or partner work time. Teachers also should understand how to develop appropriate, engaging work materials for these times, as well as how to differentiate tasks and materials (via scaffolding or enrichment) for their small group instruction. If these management items are addressed prior to entry in the field, teachers might be better equipped to flexibly group students and enjoy higher amounts of efficacy in DI (Dixon et al., 2014), which in turn should encourage continued use of flexible grouping practices (Poulou et al., 2018). Even with such training, however, it can take several years to acquire and seamlessly employ these skills (Santangelo & Tomlinson, 2012; Tomlinson et al., 1994).

ORGANIZATIONAL CONDITIONS

In addition to teacher training, in-service conditions must also be satisfied for flexible grouping to develop, especially for most veteran teachers who have not been exposed to the recommended pre-service training. Brigandi et al. (2019), in following teachers' longitudinal development of DI and grouping practices, observed how professional development (PD) increased teachers' grouping knowledge and DI toolkits, but veteran teachers' beliefs and attitudes about grouping were difficult to transform. Therefore, they urged school administrators to avoid one-day PD seminars that might not inspire immediate change and instead supply meaningful, ongoing supports with trusted colleagues that will foster experimentation and risk-taking with flexible grouping and DI. For example, they described how sustained coaching with an instructional coach who models flexible grouping and DI practices can motivate a teacher to try it in their classroom. Relatedly, if teachers are to attempt challenging new practices, school administrators must grant them autonomy to experiment with different methods and provide grace when they fail (De Neve et al., 2015). Professional learning communities (PLCs), which are comprised of teachers committed to studying, sharing, and practicing strategies around a shared topic, are another motivating tool for teacher transformation because they are teacher-led and practical in nature. De Neve et al. (2015) and Tomlinson et al. (2003) highlighted how PLCs around flexible grouping and DI can build a stronger sense of self-efficacy for such complex tasks, and they allow teachers to glean new strategies practiced and approved by colleagues.

Since flexible grouping requires additional instructional and assessment materials, district organizations could structurally ease the planning burden by purchasing high-quality pre-differentiated curricula that includes all needed formative assessments for grouping purposes and tiered activities for all groups' levels of readiness (Azano et al., 2011; Callahan et al., 2015; McCoach et al., 2014; Plucker & Peters, 2016; Rubenstein et al., 2015). High-quality, pre-differentiated curricula can raise teachers' awareness of diverse student needs and support them in adopting flexible grouping practices and providing appropriate lessons for each group (Rubenstein et al., 2015). Thus, if organizations met this structural need, they could reduce the planning and instructional load that flexible grouping can place upon teachers.

Cultural shifts within school organizations also must occur for flexible grouping practices to become more commonly used than fixed ability grouping practices (Harris, 2010). De Neve et al. (2015) suggested that such cultural shifts require support among colleagues and a shared sense of responsibility for students across the building, so administrators

can facilitate staff conversations where such values are communicated and considered. Administrators seeking to increase flexibility in teachers' grouping practices might also consider how their personal practices have fueled fixed ability thinking in their schools and then consider what steps must be taken in staff meetings, PD sessions, informal conversations, and teacher evaluation meetings to alter the present school culture.

To address political pressure from influential parents seeking high grouping placement for their children, administrators can communicate to parents the school's commitment to flexible grouping at the beginning of the school year, providing empirical justification that outline its benefits, so expectations are established about how small group instruction will function. Then, if a parent places pressure on a teacher about their child's "level" or placement in a group, the administrator can provide support to the teacher and offer to sit in on the parent meeting, if needed. Further, Tomlinson et al. (2003) insisted that schools shift their rigid concerns around communal testing into more concern for students' individual strengths and needs; this would remove the political stress from teachers to isolate students into groups and grant more freedom to use formative assessments and experiment flexibly in their classrooms.

POTENTIAL CHALLENGES TO CONSIDER IN THE PURSUIT OF EQUITABLE GROUPING

Structural racism and classism are embedded into school funding formulas based on local tax revenue and thus result in schools with higher proportions of students of color and students from low-income backgrounds receiving less dollars per pupil (Baker et al., 2020; Sosina & Weathers, 2019; Weathers & Sosina, 2022). This reality may make some of the above-mentioned recommendations challenging to implement in historically marginalized communities. For example, schools with less funds may not be able to afford structural changes like purchasing a pre-differentiated curriculum or providing ongoing coaching dedicated to grouping and DI. In these cases, school leaders might consider shifting mandatory staff meetings or PD time into time in which grade level teams can collaboratively plan and design pre-assessments and formative assessments together. Collaborative planning eases the practical and cognitive load on teachers (Thousand et al., 2006), so if they are granted time to design assessments aligned to their current curriculum, that might support their use of flexible grouping. Furthermore, instead of providing one-off PDs for the year, as is traditionally done in public school settings, the principal might instead encourage teachers to conduct individual or collaborative action research related to their grouping and DI practices for the school year; typical time allotted for PD could be granted to teachers to reflect on their goals, use their collected data to monitor their progress, and revise goals and plans for action as they go (Mitchell et al., 2009). Innovation may be required to better equip teachers to flexibly group when funds are limited, but it is possible and imperative to prioritize flexible grouping development in divested communities.

Furthermore, schools with higher proportions of students of color and low-income students often have more difficulty recruiting and retaining high-quality teachers, with structural racism and classism similarly driving these effects (Cherng et al., 2022; Lane et al., 2018). This could result in the students who most require equitable within-class small group instruction not accessing teachers with the pedagogical-psychological knowledge

to facilitate it. If administrators in schools with higher proportions of historically marginalized students notice that their teaching body does not hold pedagogical-psychological knowledge for equitable grouping, then it is critical that they prioritize training efforts, particularly citing the evidence of how grouping practices have historically harmed students and how flexible grouping can prevent those harms from repeating.

CONCLUSION

Elementary teachers undoubtedly carry a great load as they seek to meet students' increasingly heterogeneous academic, social, and emotional needs (Tomlinson et al., 1994). To simplify their work, they often group students into more homogeneous ability groups, so they can teach students according to their needs (Slavin, 1987; Sørensen, 1970). Such ability groups can take on various combinations of fixed and flexible characteristics, meaning that in different settings, students can sometimes become locked into one ability group or can move freely across groups fit to current levels of readiness. While there are multiple reasons to explain why certain teachers sometimes utilize more fixed ability grouping practices, it is important to consider how such practices impact students, especially those who have been historically marginalized through sorting and grouping patterns in schools (Borland et al., 2002; Buttaro et al., 2010; Ford, 2011). Teachers' grouping practices are shaped by school organizational features and their own personal factors (such as how they perceive ability), which often works together to perpetuate longstanding, inequitable grouping of certain groups (Tomlinson et al., 1994). Unless teacher training programs and school organizations employ targeted efforts to shift their organizations' values and resources towards flexible ability grouping practices, teachers will likely remain dependent on more fixed practices since they are professionally familiar and easier to implement. Given the generational cycles of inequity hosted in American schools, that is something we can no longer accept.

REFERENCES

- Adelson, J. L., & Carpenter, B. D. (2011). Grouping for achievement gains: For whom does achievement grouping increase kindergarten reading growth? *Gifted Child Quarterly*, 55(4), 265–278. <https://doi.org/10.1177/0016986211417306>
- Adodo, S. O., & Agbayewa, J. O. (2011). Effect of homogeneous and heterogeneous ability grouping class teaching on a student's interest, attitude, and achievement in integrated science. *International Journal of Psychology and Counseling*, 3(3), 48–54. <https://doi.org/10.5897/IJPC.9000013>
- Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, 47(5), 626–640. <https://doi.org/10.2307/2095163>
- Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal*, 83, 548–559. <https://doi.org/10.1086/461333>
- Altintas, E., & Ozdemir, A. S. (2015). Evaluating a newly developed differentiation approach in terms of student achievement and teachers' opinions. *Educational Sciences: Theory and Practice*, 15(4), 1103–1118. <https://doi.org/10.12738/estp.2015.4.2540>
- Azano, A., Missett, T. C., Callahan, C. M., Oh, S., Brunner, M., Foster, L. H., & Moon, T. R. (2011). Exploring the relationship between fidelity of implementation and academic achievement in a third-grade gifted curriculum: A mixed-methods study. *Journal of Advanced Academics*, 22(5), 693–719. <https://doi.org/10.1177/1932202X11424878>
- Baker, B. D., Srikanth, A., Cotto, R., & Green, P. C. (2020). School funding disparities and the plight of Latinx children. *Education Policy Analysis Archives*, 28(135). <https://doi.org/10.14507/epaa.28.5282>
- Barab, S. A., & Plucker, J. A. (2002). Smart people or smart contexts? Talent development in an age of situated approaches to learning and thinking. *Educational Psychologist*, 37, 165–182. https://doi.org/10.1207/S15326985EP3703_3
- Barr, R., & Dreeben, R. (1983). *How Schools Work*. University of Chicago Press.
- Becton, V. (2018). Re-examining the resurgence of ability grouping in public schools: Does it really impact academic achievement and the social and emotional development of students with disabilities? *International Journal of Arts and Sciences*, 11(1), 253–268.
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268. <https://doi.org/10.3102/00028312042002231>
- Borland, J. H., Horton, D., Subotnik, R. F., Shiang-Jiun, C., Miran, C., Freeman, C., & Yu, J. (2002). Ability-grouping and acceleration of gifted students: Articles from the Roeper Review. *Roeper Review*, 24, 100–102. <https://doi.org/10.1080/02783190209554138>

- Borman, G. D., & Dowling, M. (2010). Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data. *Teachers College Record*, 112(5), 1201-1246. <https://doi.org/10.1177/016146811011200507>
- Bourdieu, P. (1973). Cultural reproduction and social reproduction. In R. K. Brown (Ed.), *Knowledge, education, and cultural change: Papers in the sociology of education* (pp. 71-112). Tavistock Publications.
- Bradbury, A. (2018). The impact of the phonics screening check on grouping by ability: A 'necessary evil' amid the policy storm. *British Educational Research Journal*, 44(4), 539-556. <https://doi.org/10.1002/berj.3449>
- Brigandi, C. B., Gilson, C. M., & Miller, M. (2019). Professional development and differentiated instruction in an elementary school pullout program: A gifted education case study. *Journal for the Education of the Gifted*, 42(4), 362-395. <https://doi.org/10.1177/0162353219874418>
- Brookwood, M. (2021). *The orphans of Davenport: Eugenics, the Great Depression, and the war over children's intelligence*. Liveright.
- Buttaro A., Jr., Catsambis, S., Mulkey, L., & Steelman, L. C. (2010). An organizational perspective on the origins of instructional segregation: School composition and use of within-class ability grouping in American kindergartens. *Teachers College Record*, 112(5), 1300-1336. <https://doi.org/10.1177/016146811011200503>
- Calarco, J. M. (2014). Coached for the classroom: Parents' cultural transmission and children's reproduction of educational inequalities. *American Sociological Review*, 79, 1015-1037. <https://doi.org/10.1177/0003122414546931>
- Callahan, C. M., Moon, T. R., Oh, S., Azano, A. P., & Hailey, E. P. (2015). What works in gifted education: Documenting the effects of an integrated curricular/instructional model for gifted students. *American Educational Research Journal*, 52(1), 137-167. <https://doi.org/10.3102/0002831214549448>
- Castle, S., Deniz, C. B., & Tortora, M. (2005). Flexible grouping and student learning in a high-needs school. *Education and Urban Society*, 37(2), 139-150. <https://doi.org/10.1177/0013124504270787>
- Chandra Handa, M. (2020). Examining students' and teachers' perceptions of differentiated practices, student engagement, and teacher qualities. *Journal of Advanced Academics*, 31(4), 530-568. <https://doi.org/10.1177/1932202X20931457>
- Cherng, H. S., Halpin, P. F., & Rodriguez, L. A. (2022). Teaching bias? Relations between teaching quality and classroom demographic composition. *American Journal of Education*, 128(2), 171-201. <https://doi.org/10.1086/717676>
- Condon, D. J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems*, 54(1), 139-160. <https://doi.org/10.1525/sp.2007.54.1.139>

- Corbett Burris, C., Welner, K., & Murphy, J. (2008). Accountability, rigor, and detracking: Achievement effects of embracing a challenging curriculum as a universal good for all students. *Teachers College Record*, 110(3), 571-607. <https://doi.org/10.1177/016146810811000301>
- De Neve, D., Devos, G., & Tuytens, M. (2015). The importance of job resources and self-efficacy for beginning teachers' professional learning in differentiated instruction. *Teaching and Teacher Education*, 47, 30-41. <https://doi.org/10.1016/j.tate.2014.12.003>
- Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, 37(2), 111-127. <https://doi.org/10.1177/0162353214529042>
- Eder, D. (1981). Ability grouping as a self-fulfilling prophecy. *Sociology of Education*, 54, 151-162. <https://doi.org/10.2307/2112327>
- Ellison, B. J., & Hallinan, M. T. (2004). Ability grouping in Catholic and public schools. *Catholic Education: A Journey of Inquiry and Practice*, 8(1), 107-129. <https://ejournals.bc.edu/index.php/cej/article/view/571>
- Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research*, 43(2), 163-180. <https://doi.org/10.3102/00346543043002163>
- Evans, C., & Waring, M. (2008). Trainee teachers' cognitive styles and notions of differentiation. *Education and Training*, 50(2), 140-154. <https://doi.org/10.1108/00400910810862128>
- Fitzgerald, L., Hunter, J., & Hunter, R. (2021). Shifting teacher practices in relation to grouping: Gap gazing or strengths focused approaches. *Mathematics Teacher Education & Development*, 23(3), 97-110. <https://files.eric.ed.gov/fulltext/EJ1321057.pdf>
- Ford, D. Y. (2011). *Multicultural gifted education (2nd ed.)*. Prufrock Press.
- Gambrell, L., Wilson, R. M., & Walter, N. (1981). Classroom observations of task-attending behaviors of good and poor readers. *Journal of Educational Research*, 74, 400-404.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, 59, 185-199. <https://doi.org/10.2307/2112346>
- Gamoran, A. (1989). Rank, performance, and mobility in elementary school grouping. *The Sociological Quarterly*, 30(1), 109-123. <https://doi.org/10.1111/j.1533-8525.1989.tb01514.x>
- Goldberg, M. L., Passow, A. H., & Justman, J. (1966). *The effects of ability grouping*. Teachers College Press.
- Good, T. L., & Brophy, J. E. (1972). Behavioral expression of teacher attitudes. *Journal of Educational Psychology*, 63, 617-624. <https://doi.org/10.1037/h0034069>
- Gordon, V., & Nocon, H. (2008). Reproducing segregation: Parent involvement, diversity, and school governance. *Journal of Latinos and Education*, 7(4), 320-339. <https://doi.org/10.1080/15348430802143634>

- Haller, E. J., & Davis, S. A. (1981). Teacher perceptions, parental social status and grouping for reading instruction. *Sociology of Education*, 54, 162–174. <https://doi.org/10.2307/2112328>
- Hargreaves, E. (2019). Feeling less than other people: Attainment scores as symbols of children's worth. *Forum for Comprehensive Education*, 16(1), 53–66. <http://doi.org/10.15730/forum.2019.61.1.53>
- Harris, D. M. (2010). Curriculum differentiation and comprehensive school reform: Challenges in providing educational opportunity. *Educational Policy*, 25(5), 844–884. <https://doi.org/10.1177/0895904810386600>
- Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development*, 84(5), 1716–1733. <https://doi.org/10.1111/cdev.12079>
- Heyder, A., Bergold, S., & Steinmayr, R. (2017). Teachers' knowledge about intellectual giftedness: A first look at levels and correlates. *Psychology and Learning*, 17(1), 27–44. <https://doi.org/10.1177/1475725717725493>
- Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education*, 25(3), 343–358. <https://doi.org/10.1080/030549899104026>
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304–316. <https://doi.org/10.3102/0013189X13495087>
- Kim, S., Lin, T. J., Chen, J., Logan, J., Purtell, K. M., & Justice, L. M. (2020). Influence of teachers' grouping strategies on children's peer social experiences in early elementary classrooms. *Frontiers in Psychology*, 11, 3583. <https://doi.org/10.3389/fpsyg.2020.587170>
- Lane, E., Linden, R., & Stange, K. (2018). *Socioeconomic disparities in school resources: New evidence from within-districts*. University of Michigan. <http://www-personal.umich.edu/~kstange/papers/LaneLindenStangeOct2018.pdf>
- Legette, K. B., Halberstadt, A. J., & Majors, A. T. (2021). Teachers' understanding of racial inequity predicts their perceptions of students' behaviors. *Contemporary Educational Psychology*, 67, 102014. <https://doi.org/10.1016/j.cedpsych.2021.102014>
- Loveless, T. (2009). *Tracking and detracking: High achievers in Massachusetts middle schools*. Thomas B. Fordham Institute.
- Maker, C. J., & Schiever, S. W. (2005). *Teaching models in education of the gifted* (3rd ed.). Pro-Ed.
- Marks, R. (2013). "The blue table means you don't have a clue": The persistence of fixed-ability thinking and practices in primary mathematics in English schools." *Forum: For Promoting 3-19 Comprehensive Education*, 55(1), 31–44. <https://doi.org/10.2304/forum.2013.55.1.31>
- McCoach, D. B., Gubbins, E. J., Foreman, J., Rubenstein, L. D., & Rambo-Hernandez, K.

- E. (2014). Evaluating the efficacy using predifferentiated and enriched mathematics curricula for grade 3 students: A multisite cluster-randomized trial. *Gifted Child Quarterly*, 58(4), 272–286. <https://doi.org/10.1177/0016986214547631>
- Missett, T. C., Brunner, M. M., Callahan, C. M., Moon, T. R., & Azano, A. P. (2014). Exploring teacher beliefs and use of acceleration, ability grouping, and formative assessment. *Journal for the Education of the Gifted*, 37(3), 245–268. <https://doi.org/10.1177/0162353214541326>
- Mitchell, S. N., Reilly, R. C., & Logue, M. E. (2009). Benefits of collaborative action research for the beginning teacher. *Teaching and Teacher Education*, 25(2), 344–349. <https://doi.org/10.1016/j.tate.2008.06.008>
- National Center for Education Statistics (2012). *The nation's report card: 2011*. The Nation's Report Card. <https://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf>
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press.
- Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher*, 21(4), 4–47. <https://doi.org/10.3102/0013189X021004012>
- Oakes, J., Stuart Wells, A., Jones, M., & Datnow, A. (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform. *Teachers College Record*, 98(3), 483–510. <https://doi.org/10.1177/016146819709800305>
- Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.
- Pozas, M., Letzel, V., & Schneider, C. (2020). Teachers and differentiated instruction: Exploring differentiation practices to address student diversity. *Journal of Research in Special Educational Needs*, 20(3), 217–230. <https://doi.org/10.1111/1471-3802.12481>
- Poulou, M. S., Reddy, L. A., & Dudek, C. M. (2018). Relation of teacher self-efficacy and classroom practices: A preliminary investigation. *School Psychology International*, 40(1), 25–48. <https://doi.org/10.1177/0143034318798045>
- Rubenstein, L. D., Gilson, C. M., Bruce-Davis, M. N., & Gubbins, E. J. (2015). Teachers' reactions to pre-differentiated and enriched mathematics curricula. *Journal for the Education of the Gifted*, 38(2), 141–168. <https://doi.org/10.1177/0162353215578280>
- Santangelo, T., & Tomlinson, C. A. (2012). Teacher educators' perceptions and use of differentiated instruction practices: An exploratory investigation. *Action in Teacher Education*, 34, 309–327. <https://doi.org/10.1080/01626620.2012.717032>
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of educational research*, 57(3), 293–336. <https://doi.org/10.3102/00346543057003293>
- Sørensen, A. B. (1970). Organizational differentiation of students and educational opportunity. *Sociology of Education*, 43(4), 355–376.

- Sosina, V. E., & Weathers, E. S. (2019). Pathways to inequality: Between-district segregation and racial disparities in school district expenditures. *AERA Open*, 5(3). <https://doi.org/10.1177/2332858419872445>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, 69, 797-811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K-12 students' academic achievement: Findings on two second-order meta-analyses. *Review of Educational Research*, 86(4), 849-899. <https://doi.org/10.3102/0034654316675417>
- Stright, A. D., & Supplee, L. H. (2002). Children's self-regulatory behaviors during teacher-directed, seat-work, and small-group instructional contexts. *The Journal of Educational Research*, 95(4), 235-244. <https://doi.org/10.1080/00220670209596596>
- Thousand, J. S., Villa, R. A., & Nevin, A. I. (2006). The many faces of collaborative planning and teaching. *Theory Into Practice*, 45(3), 239-248. https://doi.org/10.1207/s15430421tip4503_6
- Tomlinson, C. A., Tomchin, E. M., & Callahan, C. M. (1994). Practices of preservice teachers related to gifted and other academically diverse learners. *Gifted Child Quarterly*, 38(3), 106-114. <https://doi.org/10.1177/001698629403800303>
- Tomlinson, C. A., Callahan, C. M., Tomchin, E. M., Eiss, N., Imbeau, M., & Landrum, M. (1997). Becoming architects of communities of learning: Addressing academic diversity in contemporary classrooms. *Exceptional Children*, 63, 269-282. <https://doi.org/10.1177/001440299706300210>
- Tomlinson, C. A. (2000). Reconcilable differences? Standards-based teaching and differentiation. *Educational Leadership*, 58(1), 6-11. <https://www.ascd.org/el/articles/reconcilable-differences-standards-based-teaching-and-differentiation>
- Tomlinson, C. A., Brighton, C., & Hertberg, H. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, 27(2/3), 119-145. <https://doi.org/10.1177/016235320302700203>
- Tyack, D. (1974). *The one best system: A history of American urban education*. Harvard University Press.
- Tyson, K. (2011). *Integration interrupted: Tracking, black students, and acting white after Brown*. Oxford University Press.
- Van Houtte, M., Demanet, J., & Stevens, P. A. J. (2013). Curriculum tracking and teacher evaluations of individual students: Selection, adjustment, or labeling? *Social Psychology of Education: An International Journal*, 16(3), 329-352. <https://doi.org/10.1007/s1123-18-013-9216-8>
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general

- pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>
- von Hippel, P. T., Workman, J., & Downey, D. B. (2018). Inequality in reading and math skills forms mainly before Kindergarten: A replication, and partial correction, of “Are schools the great equalizer?” *Sociology of Education*, 91(4), 323–357. <https://doi.org/10.1177/0038040718801760>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Weathers, E. S., & Sosina, V. E. (2022). Separate remains unequal: Contemporary segregation and racial disparities in school district revenue. *American Educational Research Journal*, 59(5), 905–938. <https://doi.org/10.3102/00028312221079297>
- Worthy, J. (2010). Only the names have been changed: Ability grouping revisited. *The Urban Review*, 42(4), 271–295. <https://doi.org/10.1007/s11256-009-0134-1>
- Wu, E. H. (2013). An interview with Carol Tomlinson. *Journal of Advanced Academics*, 24(2), 125–133. <https://doi.org/10.1177/1932202X13483472>

DOES THIS MEAN WE GET AN A? CAUSAL IMPLICATIONS OF CHANGES IN SCHOOL ACCOUNTABILITY

SAMUEL KAMIN, PH.D.

POSTDOCTORAL RESEARCH ASSOCIATE

NEAG SCHOOL OF EDUCATION, UNIVERSITY OF CONNECTICUT

ORCID 0009-0001-6049-3420

ABSTRACT

Many states and districts in the United States use school report cards to share accountability data in which K-12 schools are rated on a variety of metrics, including test scores, which create a categorical grade or rating. These report cards are shared with the public as a mechanism of school accountability and in the process of school choice. This paper explores the causal impact of a school report card used by the New York City Department of Education which was not attached to specific rewards and/or sanctions. I use a regression discontinuity approach to analyze the impact of receiving a lower rating. I find that just receiving a low rating leads to an increase in Math score growth in comparison to similar schools just beyond the cut point, although no such effect is found in English score growth. I also explore implications in the context of school/district policy and leadership.

Keywords: School accountability, testing, regression discontinuity, bounded rationality

The introduction of the No Child Left Behind (NCLB) Act ushered in a host of changes to U.S. public schooling, including new content standards, the introduction of Annual Yearly Progress requirements for all students, and a substantive increase in testing requirements (Linn et al., 2002). While schools in the past were accountable to municipalities and states to varying degrees, NCLB formally required all schools to collect annual testing data to verify their progress towards the goal of all students achieving proficiency in math and reading by 2013–2014 (Dee & Jacob, 2011).

Another key element of NCLB was a new focus on the public sharing of these aforementioned school-level data (Dee & Jacob, 2011), with most states and many districts beginning to publish reports on individual school quality. States and districts often formatted these reports as report cards, sometimes even mimicking classic A to F grades. For these report cards, schools are rated on a collection of measures of varying scales, but a final grade is determined through some scaling mechanism. School report cards have been widely examined and researchers have found far-reaching consequences of their implementation, including changing parents' choice of schools away from low scoring schools (Friesen et al., 2012), and impacting housing markets as high scoring schools drive up property prices (Figlio & Lucas, 2004). There is also evidence that school report cards shift behavior within schools; Chakrabarti (2007) found that schools receiving low scores on school report cards focus on students at or near minimum criteria cutoffs for proficiency.

A salient question, then, is whether school report cards are working efficiently and as intended: to communicate school quality to parents, as well as share data with district and school employees to effect change. A second related question is whether schools substantively change practices based on the information provided to increase student achievement, rather than limited and particular effects. For example, if test scores are increasing, they may only be increasing for specific subgroups within a given school, suggesting only certain students are receiving increased attention because of a new focus on the rating. Last, it is possible that schools and their leaders may respond to the report card rating itself as an inherent signal, as opposed to some particular reward or sanction that may come attached to a particular rating.

New York City provides a particularly interesting opportunity for investigating the impact of school report card systems. In 2015, the New York City Department of Education (NYCDOE) transitioned from an A to F report card system with attached consequences and rewards to a goal-based system with less specific grade metrics. In this paper, I contribute to the causal literature on mechanisms of school accountability by examining the impact of this post-2015 report card system in New York City. Specifically, this paper addresses the following research questions:

1. What is the causal impact of just receiving a lower school report card rating on exam scores?
2. What is the causal impact of just receiving a lower school report card rating on relevant achievement-oriented subgroups of students?

In this paper, I leverage the fact that NYCDOE-defined categorical Student Achievement ratings are sharply determined from a continuous score and use a regression discontinuity approach to examine the causal impact of just receiving particular low Student Achievement ratings in comparison to schools just receiving the higher score. This quasi-

experimental approach yields causal estimates of the impact of just receiving the lower rating.

While prior papers have examined the NYCDOE's A to F report card system and found positive impacts on learning outcomes (Rockoff & Turner, 2010; Winters & Cowen, 2012), this paper examines the impacts of a newer, more holistic report card system which is substantively different in design and intention (discussed more in the following pages). Because the new, post-2015 system was entirely separate from sanctions and rewards, as opposed to the A-F system of the past, any measurable impacts on learning outcomes from this new system can be directly attributed to the rating itself and not any potential consequences. To measure the potential impact on learning outcomes, I developed multiple test score growth metrics from the years since the shift in policy to examine the impact of just receiving a given rating, namely score growth across grades and movement of specific student subgroups.

In summary, I find that being just assigned a particular low rating ("Approaching Target") has three notable impacts: first, there is a positive impact on math score growth; second, there is a negative impact on the proportion of students in the lowest, Level 1 math achievement category (i.e., there are proportionally fewer students in the lowest performance category the following year); third, there is a positive impact on the proportion of students in the proficient categories. There is not statistically significant evidence of similar trends in English test scores, however. Additionally, I do not find any significant impact of just receiving a rating of "Meeting Target" in comparison to similar schools receiving a rating of "Exceeding Target."

In the following sections, I discuss prior research on accountability systems and provide context regarding the specific report card policy in New York City. I then discuss my analytic approach and report my findings. Finally, I discuss these findings in the context of policy and note potential future areas of research on the subject.

BACKGROUND

PRIOR RESEARCH ON ACCOUNTABILITY SYSTEMS

Researchers have examined the impact of the strict accountability imposed and inspired by NCLB. Indeed, extant literature found evidence that the use of strict accountability scores had some notable positive impacts on student achievement and school practice. Chiang (2009) examined the threat of low accountability scores using a regression discontinuity design and found evidence that the pressure of a low score and the sanctions that are threatened therein increase math scores, shift pedagogy, and lead to the introduction of new curricula. The evidence that accountability scores lead to real, substantive shifts is largely replicated by Rouse et al. (2013). Relatedly, Carnoy and Loeb (2002) examined the relative strength of accountability systems across states, defining strength as the amount of pressure placed on schools to improve test scores based on state mandates; they found low strength accountability systems have little to no state-level accountability to increase student test scores, while high strength accountability systems place specific demands (including rewards or sanctions) on schools that meet or fail to meet testing thresholds. They also found that the strength is positively related to NAEP math scores.

Prior research has also examined temporal differences within states as they shift from simply providing public reports of achievement to threatening sanctions and offering rewards for low or increasing performance (respectively). Hanushek and Raymond (2005) defined two distinct categories of accountability systems. First, they described a “system” as a mechanism in which states “[publish] outcome information on standardized tests for each school along with providing a way to aggregate and interpret the school performance” (p. 12). They differentiated, however, between “report card” states and “consequential” states; the former simply report out the data, while the latter attach specific consequences. Throughout the 1990s and early 2000s, many states transitioned from no data-based accountability system to a low (or no) consequence “report card” system to a “consequential” system with rewards and/or punishments; Hanushek and Raymond (2005) leveraged that shift to find that the introduction of consequential systems increased state-level NAEP scores, although not for all subgroups of students.

There is no consensus that strict accountability mechanisms are a panacea for issues of low student achievement, however. Jacob (2005) demonstrated that while there were increases in test scores in Chicago Public Schools after the introduction of a strict accountability system; those gains were driven by positive shifts in test-specific skills and student effort on exams. These gains may not be entirely productive or efficient if the long-term goal is raising student achievement, given the mechanisms identified are limited and test-specific. Deming et al. (2016) found that while the risk of receiving a low school rating may have positive impacts on schools receiving a high score had little impact. Further, low-scoring students in schools pressured to receive a higher rating may have actually experienced negative impacts on exam scores, as well as an increased likelihood of being classified into special education. Last, Deming and Figlio (2016) demonstrated that high-stakes testing (and its related accountability measures) led to increased and disproportionate attention being paid to “bubble” students (students on the threshold of achieving proficiency on a particular exam).

SCHOOL PROGRESS REPORTS AS AN ACCOUNTABILITY SYSTEM IN NEW YORK CITY

In 2007, the NYCDOE created a new “School Progress Report” protocol to assess its schools. Using a combination of student achievement data including test scores and credit accumulation, parent survey data, and other observational data gathered during superintendent review, numeric scores were calculated on a 1-100 scale, which were then collapsed by predefined bands into A-F grades. The grades were also intended to be linked to rewards and consequences, including bonus pay for teachers for successful schools and potential school closure for those with lower grades (Gootman & Medina, 2007). Further, schools with a D or F rating were required to implement formal plans of school improvement, students in F schools were eligible for a special transfer process, and schools that met high grade thresholds were eligible for school-based budget bonuses, as well as principals earning personal bonuses (Rockoff & Turner, 2010). Low scores are also used as justification by district administration for staffing and administrative changes in schools in which they were received (Winters & Cowen, 2012). The fact that the scores were directly linked placed this policy squarely in the “consequential” bucket, as defined by Hanushek and Raymond (2005, p. 306).

These reports were also widely available for public consumption, and designed to be interpretable by parents, educators, and others. The NYCDOE created carefully presented digital and print versions of these reports which prominently featured schools' assigned letter grade, as well as selected other information. These reports were circulated at schools, in school selection publications developed by the NYCDOE, and made available both at each school's official website and that of the NYCDOE, including past years' reports (Corcoran & Pai, 2013).

There were novel elements to these report cards, beyond simply their accessibility to the public, that attempted to correct prior issues in school accountability policies. Specifically, they used school peer groups, used to compare schools within more similar groupings, as opposed to comparing against the entire city school population. There were numerous ways the use of peer groups is important. First, strict and universal school accountability policies are often influenced by out-of-school factors (Hamilton & Koretz, 2002). Knowing this, the report cards were supposed to allow schools to be compared against peer groups with similar out-of-school circumstances (e.g., number of students in poverty, entering student academic preparation, etc.), and provide a report card grade that was contextualized in the reality that different schools serve different students. Second, these report cards included some growth measures instead of city-wide normed achievement metrics, again presenting the opportunity for equity in the consideration of schools with differing circumstances out of their control, in this case prior student achievement. Simply, schools would not be punished with a low accountability grade for serving students who entered with lower prior test scores than other schools in the city. Yet it is still unclear whether the use of these peer groups had the intended balancing impact on schools' grades; according to Corcoran and Pai (2013), the Peer Index (the collapsed measure developed by NYCDOE which was used to group schools) did not have a notable impact on schools' overall grades due to the diversity within the peer groups. This suggests the use of peer groups may not have actually adjusted the scores towards the end of providing balance across differing out-of-school circumstances.

The causal impact of the NYCDOE version of school report cards on student achievement has been investigated in two papers: Rockoff and Turner (2010) and Winters and Cowen (2012). Moreover, both studies used regression discontinuity approach to examine the impact of receiving a particular grade on exam scores. Rockoff and Turner (2010) examine grade 3–8 test scores and find significant positive impacts of receiving an F relative to a D, or a D grade relative to a C in both math and reading scores. Winters and Cowen (2012) add specificity to a similar analysis by adding student-level characteristics and identifiers, providing the ability to follow students from school to school across years. They find positive impacts on student test scores of receiving an F relative to a D, particularly in English scores, and those gains were persistent across multiple years. Together, the pieces suggest that NYCDOE's school report cards do have a positive impact on test scores for schools at or near the cut points.

A NEW PROGRESS REPORT: THE “SCHOOL QUALITY REPORT”

Despite the positive impacts of the prior report card system, NYCDOE made substantive changes to the School Progress Report in 2015. There were a number of elements to this shift in policy. First, as noted by Corcoran and Pai (2013), the peer groups that were

designed to balance the prior report grades by comparing schools against similarly situated “peer” schools actually had little impact on the overall scores in the initial iteration of the report cards. Not only did the peer groups not work as designed, but the NYCDOE also believed these peer groupings created an unfair competitive attitude between the schools being compared (NYCDOE, 2018). Thus, the new progress reports removed the use of the comparison group in score calculation, although interestingly the NYCDOE did choose to include some reference to an unpublished comparison group on the reports themselves, merely suggesting the relevance of the comparison group and not actually using the group to calculate scores and/or ratings.

The most substantive change, though, was a shift from the aforementioned A to F categorical grade scale to a new four-level categorical scale, which labeled schools as Excellent, Good, Fair, or Poor in reports designed for parents. In reports designed for teachers and administrators, the same four-level categorical scale was labeled as Exceeding Target, Meeting Target, Approaching Target, and Not Meeting Target (the latter of these labeling schemes will be referred to for the rest of the paper). These new metrics were described as “gentler” (Wall, 2014, title), and described by then-chancellor of the NYCDOE Carmen Fariña as “looking beyond test scores and focusing on making sure... each school has what it needs for sustained and continuous growth” and a “transformed... approach [to] school accountability” (Darville, 2014, quoted speech).

Still, the most prominently placed measure was for “Student Achievement” which combined student test scores and credit accumulation. As before, these four-tier categorical ratings are assigned by collapsing a continuously calculated numeric score. While the numeric score was and remains publicly available, it is published in a format perhaps too complex for the general population and not formatted, designed, or documented for those without some knowledge of statistics. This suggests any decisions by parents, students, or teachers may be made, not from the continuous numeric score, but rather from the EGFP label. In addition, the scores were no longer criterion referenced — rather, they were built on pre-set targets determined by the NYCDOE, although these targets were not consistent across years.

Last, the new School Quality Reports were also no longer tied to accountability measures or bonuses; instead, these reports were designed for schools and leaders to inform their planning and allow families to learn more about their school (NYCDOE, 2015). This marks a distinct shift away from the “consequential” school accountability mechanism as described by Hanushek and Raymond (2005) towards one in which school performance is still aggregated and publicized, but without the same predetermined rewards or sanctions.

CONCEPTUAL FRAMEWORK

This paper examines the causal impact of receiving a particular categorical Student Achievement rating on a school report card beyond the impact of the numeric Student Achievement score. Policymakers and reformers in NYC adopted report cards to “set expectations for schools and promote school improvement” (NYCDOE, 2018, p.1). However, if there are measurable impacts of the categorical rating beyond that of the numeric rating, it is not obvious why the categorical Student Achievement rating specifically would have any impact on the activities of a school or leader. Indeed, all the information (e.g., test scores and survey responses) used to build the numerical achievement score, which

then determines the rating, is known to school leaders ahead of the release of the grade. Further, the information contained in the newer, more holistic report card was specifically designed to help schools “identify and address specific strengths and weaknesses” (NYCDOE, 2015, Overview section). Last, this policy shift represents a move away from a consequential accountability system (as defined by Hanushek & Raymond, 2005) to one without specifically pre-known consequences. Why, then, might the categorical rating itself have any impact above the variety of known information that informs the rating?

From a purely rational approach, schools and their leaders should work to maximize student achievement outcomes and thus improve all progress report numerical scores, regardless of the cut points and letter grades with which those numeric scores are associated. Simply, if you raise test scores, you raise your achievement score. Unless there are stated punishments or rewards for entering/exiting certain categorical ratings, there is no obvious reason why a rating category would cause any change above and beyond the impact of the numeric score. Further, given the information for specific schools within each school’s report card, the most efficient or rational behavior may be to specifically target areas of weakness in the report card.

However, rational choice theory (Simon, 1956) explains that not all behavior is as rational as expected. Actors may not search for the best option; rather, a good move might be chosen as it is safer. When actors respond in these ways, they are “satisficing” (Simon, 1956, p. 9). Understanding why schools and their leaders may behave by satisficing is further explained by Simon’s (1955) theory of bounded rationality. Simon (1956) argues that actors can rarely take advantage of all the information provided to them, and instead make choices about how much and which information of which to take advantage. Considering these concepts, school leaders may have an overwhelming amount of information at their disposal, to the point where they may not be able act on all of it. Thus, leaders may only use some of that information in deciding which proverbial levers to pull to impact student learning. In this paper, I examine the possibility that the comparatively limited information in the Student Achievement categorical rating demonstrably causes some schools and their leaders to make changes that lead to positive academic outcomes in the form of test score growth.

While the actual behaviors of school leaders are not observed in this study, there are numerous ways extant literature has established schools’ responses to accountability reforms. For example, Shipps and White (2009) examine the differences in school principal behavior before and after increased accountability policies in New York City. They found that principals paid closer attention to bureaucratic expectations and market-style accountability, each of which are directly connected to the New York City progress reports.

Bureaucratic expectations are inherently part of school progress reports in that they are developed and shared by the Department of Education; they are treated as reviews of schools’ performance for parents and school staff alike. Further, all forms of standardized progress reports inherently align with market ideology (Engel, 2000) in that they suggest intra-district comparisons and competition. In this system, even though schools are scored at least in part against their own achievement goals, each school is still given a label on a consistent and comparable metric against other schools in the city. While not every student can choose their elementary school, and thus elementary schools may

not fall cleanly into the market phenomenon described above, the reports still provide information on how a school is doing in direct comparison to its peers. School leaders may use comparatively lower Student Achievement ratings, then, as signals to change their behaviors in ways that are different from schools assigned higher ratings.

Existing literature also demonstrates ways schools and their leaders respond to accountability pressure in unequal ways across student groups, further suggesting a satisficing approach. For example, Booher-Jennings (2005) uncovered the use of educational triage in response to Texas' accountability system, in which teachers and administrators diverted resources to attend to students close to the threshold of passing (i.e. "bubble kids") and students that were known to count for the school's accountability rating. A similar set of circumstances could be relevant in New York City. Schools are commonly judged by their percentage of students meeting proficiency (NYCDOE, 2019). Similar to the findings in Booher-Jennings (2005), schools in New York City may also be practicing educational triage and addressing some of these subgroups differentially based on their Student Achievement rating.

METHODS

DATA/SAMPLE

To answer research questions on the impact of school report card grades, I used the "Student Achievement" ratings from all Elementary, Middle, and Kindergarten through eighth grade schools ($n = 1091$) in the New York City Department of Education from 2014-15 through 2018-19 school years. These years were selected because these were the first years the new reports were used and include all available years of data at the time of writing (with an exception for 2016-17 described below). As noted before, the Student Achievement rating is on a four-level categorical scale and built from a 1-5 continuous measure known as the Student Achievement Score. This continuous metric is built from a complex formula taking into account student achievement, future credit accumulation, and performance relative to a NYCDOE-assigned target. These data create my assignment (score) and treatment (rating) variables.

For my outcome variables, I construct a variety of grade-level test score growth metrics for grades 3 through 8. In New York State, every 3rd through 8th grade student completes an annual Math and English (ELA) exam each Spring. These data were downloaded from the NYCDOE website in Excel format and merged with the quality report data by a NYCDOE-assigned school ID number and year. Schools without test scores for both years, generally new or closed schools, were excluded. Similarly charter schools, who have different reporting requirements, were also excluded. These test scores are reported as collapsed at the grade by subject by school level, from 2014 through 2019; I then use them to construct year-to-year growth measures for each grade-subject-school. Excluded from the analysis is growth from the 2016-17 to 2017-18 school years, as New York State revised the exam in the 2017-18 school year to reduce the number of days tested and substantively changed the scaling of exam scores. As a result, the scores from 2017-18 are comparable to the following year (2018-19), but not prior years. These metrics are described below.

DESCRIPTIVE STATISTICS

In Tables 1 and 2, I present descriptive data. Table 1 presents the distribution of Student Achievement ratings, grouped both by school-year (one observation for each school-year combination) and grade-subject (one observation for each grade-subject combination within each school-year). Of note is the uneven distribution across the rating categories. There are few ($n=55$) schools that received the “Not Meeting Target” rating label across all years, less than 2%. Further, the majority of schools received a “Meeting Target” rating; just over half. The remaining schools were roughly evenly distributed between “Approaching Target” and “Exceeding Target.” Also of note were the relatively consistent percentages across the school-year and grade-subject breakdowns, which suggests no substantive differences within the grades served between schools with different Student Achievement rating categories.

There are a few key observable differences between schools receiving different Student Achievement ratings; in Table 2, I present key variables that highlight some of those differences. For example, schools that received lower scores tended to have a larger percentage of students of color. Schools that received lower scores tend to have slightly higher percentages of students with disabilities, higher Economic Need Index scores, and more students chronically absent. Schools with higher Student Achievement ratings tend to have more experienced principals, although not more experienced teachers.

Table 1
Student Achievement Ratings

	<u>School-Year</u>		<u>Grade-Subject</u>	
	Count	Percent	Count	Percent
Not Meeting Target	55	1.52	153	1.42
Approaching Target	919	25.47	2725	25.29
Meeting Target	1742	48.28	5328	49.45
Exceeding Target	892	24.72	2569	23.84
<i>Total</i>	3608	100.00	10775	100.00

Notes: All school-year combinations of NYCDOE schools serving students in grades 3–8, from 2014–15, 2015–16, and 2017–18. 2016–17 is excluded due to policy shifts in test timing and scaling which inhibit comparability.

OUTCOME MEASURES

My outcomes of interest are generally described as test-related growth in the year following the assignment of a given Student Achievement rating. Table 3 presents outcome averages for each of the four Student Achievement rating levels, grouped into two panels by subject area. In each panel, the first row represents growth after receiving the specified Student Achievement rating. The subsequent rows represent other test-related outcomes of interest.

While I neither individually examine the behaviors of an individual school and its

Table 2
Summary Statistics by Student Achievement Rating

	Not Meeting Target	Approaching Target	Meeting Target	Exceeding Target
Percent English Language Learners	0.0996 (0.0634)	0.152 (0.122)	0.145 (0.118)	0.131 (0.119)
Percent Students with Disabilities	0.224 (0.0719)	0.233 (0.0662)	0.221 (0.0702)	0.196 (0.0735)
Economic Need Index	0.756 (0.160)	0.777 (0.166)	0.691 (0.217)	0.579 (0.254)
Percent Asian	0.0440 (0.0779)	0.0618 (0.108)	0.121 (0.165)	0.233 (0.244)
Percent Black	0.513 (0.279)	0.400 (0.281)	0.282 (0.275)	0.155 (0.202)
Percent Hispanic	0.344 (0.216)	0.449 (0.265)	0.432 (0.264)	0.357 (0.255)
Years of principal experience	4.494 (4.538)	6.218 (4.854)	6.644 (4.572)	7.729 (5.032)
Percent of teachers with 3+ years of experience	0.742 (0.210)	0.765 (0.159)	0.785 (0.136)	0.779 (0.125)
Percent of Students Chronically Absent	0.316 (0.132)	0.293 (0.111)	0.227 (0.112)	0.149 (0.103)
Teacher Attendance Rate	0.962 (0.0116)	0.961 (0.00989)	0.962 (0.0100)	0.965 (0.00987)

Notes: Mean values of selected variables for all school-year combinations.

leaders, nor track individual students in and out of these levels, school-level measures of proficiency across years can be a good measure for student performance and a proxy for administrator behavior. New York State also converts student exam scores to a 1 to 4 scale to indicate level of proficiency for each student; category 1 is the lowest, category 2 follows, and categories 3 and 4 are each considered proficient. As schools may be interested in improving subsets of student scores, I construct growth metrics for numbers of students in the following categories: 1 (lowest), category 2 (“bubble”) and category 3 / 4 (proficient). The city-defined definitions for each of these categories is presented in Table 4.

The importance of the signs of these metrics is worth discussing specifically as they are not uniformly interpreted across categories; a negative “growth” in the lowest category, for example, means a school had less students in the lowest category (in a given grade-subject) than in the prior year — what most would consider a good thing, despite the negative numeric change. A positive growth in the proficient category, though, means a school has more students achieving proficiency (again in a given grade-subject) than in the prior year — also a good thing.

In Table 5, I present four possibilities for various combinations of signs of three performance category outcome measures at four hypothetical schools. Schools A and B

Table 3
Outcome Measures by Student Achievement Rating

	<u>Student Achievement Rating</u>			
	Not Meeting Target	Approaching Target	Meeting Target	Exceeding Target
<u>Math</u>				
Score Growth (in points)	2.007 (7.375)	1.014 (8.906)	0.0444 (8.630)	-0.0921 (7.889)
Change in lowest score category (in pp)	-5.124 (13.85)	-2.387 (13.61)	-0.123 (11.79)	0.526 (8.892)
Change in “bubble” category (in pp)	0.916 (10.64)	-0.175 (10.70)	-1.163 (9.461)	-0.965 (8.469)
Change in proficiency (in pp)	4.208 (10.32)	2.561 (10.54)	1.285 (11.03)	0.439 (10.74)
<u>English Language Arts (ELA)</u>				
Score Growth (in points; $\mu = 406$)	1.279 (8.630)	2.587 (8.425)	2.428 (7.900)	1.845 (7.359)
Change in lowest score category (in pp)	-1.942 (14.46)	-2.988 (12.65)	-2.144 (10.62)	-0.901 (7.993)
Change in “bubble” category (in pp)	-1.264 (11.27)	-0.691 (10.60)	-1.532 (9.451)	-1.846 (8.591)
Change in proficiency (in pp)	3.206 (11.99)	3.679 (10.49)	3.676 (10.78)	2.747 (10.65)

Notes: The first row in each panel represents raw score growth in test points between consecutive years. Mean score ≈ 300 ; $sd \approx 15$ for both exams, corrected for between year differences in scaling. Rows 2–4 in each panel are measured in percentage point change in number of students in listed categories.

Table 4
Scoring Levels and Distribution for New York City Elementary Exams

	<u>Description</u>	<u>% ELA</u>	<u>% Math</u>
Level 1	Students performing at this level are well below proficient in standards for their grade. They demonstrate knowledge, skills, and practices that are considered insufficient for the expectations at this grade.	24.18%	29.75%
Level 2	Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices that are considered partial but insufficient for the expectations at this grade.	32.31%	28.04%
Level 3	Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices that are considered sufficient for the expectations at this grade.	27.04%	20.51%
Level 4	Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices that are considered more than sufficient for the expectations at this grade.	16.46%	20.19%

Note: Descriptions from NYCDOE (2019). Percentages are weighted averages; grades 3-8, 2014-2019.

present the most obvious interpretations. At school A, the proportion of students in the lowest category increases, while the number of proficient students decreases. This is a school that is not showing improvement, regardless of the difference in level 2 students. At school B, the lowest category decreases while the proficient category increases. This is a school that is clearly improving; whether the students are moving out of the lowest category into level 2 (or “bubble”) or proficient category is certainly important, but with this current data we have no way of knowing if that’s the case. Schools C and D are slightly more complicated; in school C, we see increases in both the lowest and proficient category, suggesting that students are being pulled from the bubble to both extremes, suggesting heterogeneity by prior performance level. Conversely at school D students are leaving both the lowest and proficient category, congregating at the bubble, again suggesting heterogeneity, although with notably different results. As students clump at the bubble category, this could indicate School D is increasing scores of its lowest students while suppressing proficiency.

EMPIRICAL STRATEGY

To examine the causal impact of a particular categorical rating, I echo approaches from previous scholarship on school accountability grades, namely those conducted on New York City data (i.e., Rockoff & Turner, 2010; Winters & Cowen, 2012). Ideally, to replicate a fully controlled trial, I would examine the same school under two different conditions; for example, one in which they receive a label of “Approaching Target,” and one in which they receive a label of “Meeting Target.” For obvious reasons, this is not possible; schools only receive one score/rating each year, and schools have already received these labels. Further, it would be difficult and unethical to randomly assign something so important in

Table 5
Potential Subgroup Differences Across Years

	<u>% Level 1</u> <u>[Low]</u>	<u>% Level 2</u> <u>[Bubble]</u>	<u>% Level 3+4</u> <u>[Proficient]</u>
School A	+	[+ or -]	-
School B	-	[+ or -]	+
School C	+	-	+
School D	-	+	-

Note: + / - refers to an increase / decrease in the percentage of students in the specified category across two school years

Table 6: Student Achievement Rating Characteristics

	Mean	Standard Deviation	Minimum	Maximum
Not Meeting Target	1.80	0.15	1.33	1.99
Approaching Target	2.62	0.26	2.00	2.99
Meeting Target	3.49	0.28	3.00	3.99
Exceeding Target	4.34	0.24	4.00	4.99
Total	3.45	0.69	1.33	4.99

assignment inherent in a carefully controlled experiment. However, there are ways, with some assumptions, to create (almost) as-good-as randomization.

In a regression discontinuity (“RD”) approach, the underlying notion is that observations close to the left and right of any given cutoff are essentially statistically identical based on their close proximity on the assignment variable which determines their categorical label, which Lee and Lemieux (2010) described as the “Local Randomization” assumption (p. 295). While in a controlled trial, treatment is assigned based on strict randomization, here treatment is assigned to those close to the cut point in what is assumed to be a near-random way.

This approach leverages the discontinuous treatment assignment mechanism built into the School Quality Reports. Table 6 and Figure 1 demonstrate this assignment mechanism clearly. The Student Achievement Score (referred to from here on as “score”) is generated on a 1 to 5 continuous scale, and depending on this score, schools are assigned one of four Student Achievement Ratings (referred to from here on as “rating”). Note the lack of overlap between the rating categories; the maximum for each category is exactly .01 below the whole number threshold for the next category. Plot point sizes in Figure 1 highlight the cluster in the middle two ratings, and the few schools gathered on the extremes.

In this case, the continuous student achievement score concretely determines the categorical rating, but schools close to the predefined cut point (for example, 2.99 vs. 3.00) are so close that they are essentially randomly distributed on either side of the cut, meaning the difference between receiving a rating of “Approaching Target” and “Meeting Target” is essentially random. Thus, creating localized regression models around the cut point can estimate the causal impact of treatment; in this case, treatment is defined by receiving a particular rating relative to another.

The models implemented are of the following form:

$$E_{(y+1)sgc} - E_{ysgc} = \beta_0 + \gamma T_{ys} + \beta_i (S_{ys}) + \beta_j T_{ys} (S_{ys}) + \beta_k X_{ys} + \mu_y + \varepsilon_{ys} \quad (1)$$

where $E_{(y+1)sgc}$ represents an outcome metric for year $y+1$, or one year following the assignment of a Student Achievement rating, in school s , grade g , and content area c (either Math or English), while E_{ysgc} represents the same metric for the year the rating was assigned. Together, the left side of equation (1) represents growth in a specified outcome. T_{ys} represents a dummy variable for receiving a lower rating at a given school in a given year for a specified cut point between two ratings; for example, $T_{ys}=1$ if a school received an “Approaching” rating and $T_{ys}=0$ if the school received a “Meeting,” if examining the “Approaching vs. Meeting” cut point.¹ S_{ys} is a vector which represents the continuous Achievement Score for a given year/school and its quadratic term, $T_{ys}(S_{ys})$ represents a

Table 6: Student Achievement Rating Characteristics

	Mean	Standard Deviation	Minimum	Maximum
Not Meeting Target	1.80	0.15	1.33	1.99
Approaching Target	2.62	0.26	2.00	2.99
Meeting Target	3.49	0.28	3.00	3.99
Exceeding Target	4.34	0.24	4.00	4.99
<i>Total</i>	<i>3.45</i>	<i>0.69</i>	<i>1.33</i>	<i>4.99</i>

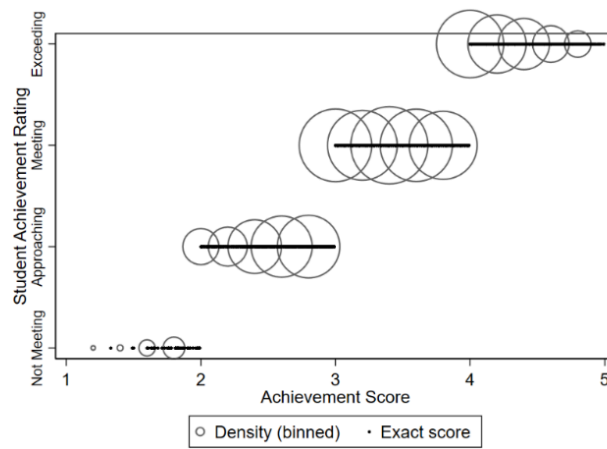


Figure 1: Achievement Score and Achievement Rating

vector of interactions between treatment and Achievement Score, allowing for differing coefficients on either side of the cut point. X_{ys} represents a vector of school-year covariates and μ_y is a year fixed-effect; these terms are added in later models. Lastly, ϵ is an idiosyncratic error term. Finally, γ is the parameter of interest, and given the assumptions of the regression discontinuity design, represents the causal impact of being just assigned a particular label relative to one Student Achievement rating higher.

ESTABLISHING THE VALIDITY OF THE RD IDENTIFICATION STRATEGY

An important preliminary check for internal validity is to assess the possibility of manipulation at the cut point (Lee & Lemieux, 2010). Because the assignment variable — achievement score — is assumed to be continuous, there should be little evidence of significant jumps anywhere along the spectrum, but specifically not at the cut points. If there were to be a jump at the cut point, it might signify an unobservable manipulation to the assignment variable at the cut point, violating a core assumption of the regression discontinuity approach, and thus rendering our analysis inaccurate. To see potential evidence visually, a histogram is the most appropriate choice, and presented in Figure 2. There are three distinct areas, defined by the two cut points of the achievement score. Schools that were designated “Not Meeting” or “Approaching” the target have been classified as “Below Target”, while the “Meeting Target” and “Exceeding Target” labels are directly from the achievement rating. The cut point lines are presented in red for convenience as well.

Upon simple visual analysis, while the cut at 4 seems to not be an issue, there does appear to be a small jump from “Below Target” to “Meeting Target” where Achievement Score equals 3. This is potentially statistically problematic; if there is manipulation happening to move scores from immediately below the cut point to immediately above, this would violate a core assumption of RD and render any inference based on the RD inaccurate. However, considering the nature of the School Quality reports and their underlying statistics, it would be difficult for any real manipulation to take place. First, it is impossible to predict and manipulate the wide range of scores that will eventually be used to calculate an achievement score and therefore rating. While administrators may have

¹While it may seem more intuitive to code this in the reverse, I chose to code treatment in this way as to more intuitively interpret the effect of the lower score relative to the higher score, given the theory that schools may be motivated in particular by a lower score.

had access to their scores before they were finalized, it would still have been difficult to manipulate scores after tests were concluded and shift scores in one direction or the other. Finally, as I address more fully in my limitations section, any manipulation of this sort may actually lead to underestimating the effects at that cut point, given the results presented below.

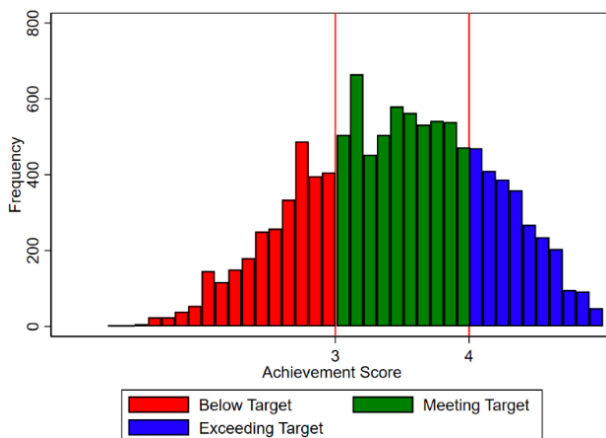


Figure 2: Histogram of achievement ratings

As a final check of the regression discontinuity assumptions, I present a parallel analysis using a covariate as an outcome measure as suggested by Lee and Lemieux (2010) in Table 7. If there was notable manipulation or some other difference in the groups on either side of the cut points, this difference may be revealed by differences in covariates, which, based on the assumptions of regression discontinuity, should be similar across both sides of the cut point. Table 5 presents results from RDRobust (Calonico et al., 2017) for all school-level covariates used later in analysis at both cut points: percent of English language learners, percent of students in special education, economic need index (calculated by NYCDOE to represent schoolwide economic need), a variety of race percentages, principal and teacher experience, and student and teacher attendance. Columns 1–3 indicate balance at the Approaching vs. Meeting cut point, while columns 4–6 indicate balance at the Meeting vs. Exceeding cut point. The first columns (1 and 4) utilize data-driven bandwidth selections, while the remaining use a predefined smaller and larger bandwidth. There should not be any significant results in these tests; if there were, it would signify a discontinuity in one of our covariates, violating the local randomization assumption, and would suggest that there was a statistical difference between the two groups close to the cut point. As suspected, the estimates are small in magnitude, indicating little difference, and only a small handful are significant, and only at the $p = .05$ level. Indeed, given the large number of statistical test results being presented in this table (60), it is not surprising that some may appear significant. This helps reinforce (yet not necessarily fully confirm) the original assumption of local randomization around the two cut points.

RESULTS

In the following section, I discuss the findings of the regression discontinuity design. First, I explain visual differences at the cut point using binned plots with local linear specifications mapped on for ease of interpretation, finding that being just assigned (i.e., assignment based on being just past the cut point) a rating of “Approaching” has a positive impact on some math-related outcomes. Those results do not appear to be

Table 7*Estimates of Differences on Either Side of Cut points for Selected Covariates*

	<i>Approaching vs. Meeting</i>			<i>Meeting vs. Exceeding</i>		
	(1) MSE Opt.	(2) .2	(3) .4	(4) MSE Opt.	(5) .2	(6) .4
Bandwidth:						
ELL	-0.0186 (0.0195)	-0.0197 (0.0235)	0.000406 (0.0136)	-0.00206 (0.0167)	0.00109 (0.0199)	0.00402 (0.0116)
<i>N</i>	3110	1885	6755	2582	1828	6611
Sp. Educ.	-0.0133 (0.00864)	-0.0237 (0.0122)	-0.00796 (0.00692)	-0.0128 (0.0151)	-0.00795 (0.0141)	0.000383 (0.00790)
<i>N</i>	4196	1885	6755	1669	1828	6611
Econ. Index	-0.0690* (0.0286)	-0.0819* (0.0347)	-0.0268 (0.0200)	0.0475 (0.0346)	0.0627 (0.0422)	0.0161 (0.0253)
<i>N</i>	2948	1885	6755	2837	1828	6611
Asian	0.0269 (0.0150)	0.0450* (0.0212)	0.0234 (0.0123)	-0.0106 (0.0385)	-0.000887 (0.0426)	-0.0166 (0.0231)
<i>N</i>	4547	1885	6755	2182	1828	6611
Black	-0.0176 (0.0374)	-0.0488 (0.0533)	-0.0306 (0.0284)	0.0278 (0.0348)	0.0363 (0.0407)	0.0103 (0.0230)
<i>N</i>	3938	1885	6755	2376	1828	6611
Hispanic	-0.0610 (0.0417)	-0.0746 (0.0534)	-0.0179 (0.0286)	0.0690 (0.0412)	0.0702 (0.0484)	0.0365 (0.0268)
<i>N</i>	3206	1885	6755	2582	1828	6611
Principal Exp.	1.128 (0.652)	0.910 (0.927)	1.009 (0.523)	-0.511 (0.893)	-0.521 (0.938)	-0.375 (0.512)
<i>N</i>	4106	1882	6743	2039	1828	6602
Teacher Exp.	-0.0171 (0.0236)	-0.0289 (0.0321)	-0.0102 (0.0183)	0.00370 (0.0284)	-0.00234 (0.0319)	0.00619 (0.0170)
<i>N</i>	4031	1885	6755	2376	1828	6611
Student Attend.	-0.0310 (0.0175)	-0.0480* (0.0221)	-0.0105 (0.0119)	0.00481 (0.0167)	-0.00526 (0.0197)	0.000474 (0.0110)
<i>N</i>	3020	1885	6755	2582	1828	6611
Teacher Attend.	0.00163 (0.00132)	0.00188 (0.00197)	0.00106 (0.00101)	-0.00150 (0.00160)	-0.00210 (0.00179)	-0.00113 (0.000967)
<i>N</i>	3938	1843	6599	2325	1789	6459

Notes: Heteroskedasticity robust standard errors clustered by school in parentheses. Each coefficient is the reduced form estimate of the relationship between Student Achievement Rating and the listed covariate. Coefficients are generated by *RDRobust* command, implementing local polynomial (quadratic) regressions with a triangular kernel. The first columns (1 and 4) in each panel utilize a MSE-optimized bandwidth (Calonico et al., 2017), while subsequent columns use a prespecified bandwidth. All models include year fixed effects.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

present for ELA outcomes, nor for being just assigned the “Meeting” rating. Then, I present numeric estimates that bolster the conclusions evident in the graphical approach.

VISUAL RESULTS

Figure 3 demonstrates the four math outcomes of interest at the “Approaching vs. Meeting” cut point. There are apparent discontinuities in all four of the graphs presented; each graph suggests that just being assigned a rating of “Approaching” improves math-related outcomes. First, in graph A, it appears that there is increased score growth below the cut point, suggesting being assigned a lower rating causes a positive impact on math score growth for schools just below the cut point when compared to their similar peers just above the cut point. Similarly, graph D suggests that when schools just below the

cut point are assigned a lower rating, they have a greater increase in their percentage of students in the overall category of proficiency in comparison to their peers just above the cut point.

Graph B (figure 3) shows an opposite visual pattern, as in being just at the lower rating suggests a negative impact relative to schools just above the cut point. However, this may be consistent evidence of improvement regardless of differing signs. Graph B suggests that schools just below the cut point have a larger decrease in the number of students in the lowest category, which would generally be interpreted as overall improvement. Similar to graph A, the positive difference in graphs C and D suggests that schools just below the cut point both increased the number of students in the bubble category (level 2) as well as students in the Proficient category, when compared to schools just above the cut point. However, the evidence in graph C is the least conclusive visually.

Figure 4 demonstrates the four ELA outcomes of interest at the “Approaching vs. Meeting” cut point. These figures appear to be less conclusive than their counterparts in Figure 3 in that the discontinuities are less pronounced. Still, in graphs A and B there appears to be evidence of improvement for schools in both score growth and movement of students out of the lowest performance category just below the cut point receiving a rating of Approaching in comparison to similar schools just above the cut point.

Figure 5 presents the same math-related outcomes as Figure 3 but shifts the perspective to the Meeting vs. Exceeding cut point. In comparison to Figure 3, there do not appear

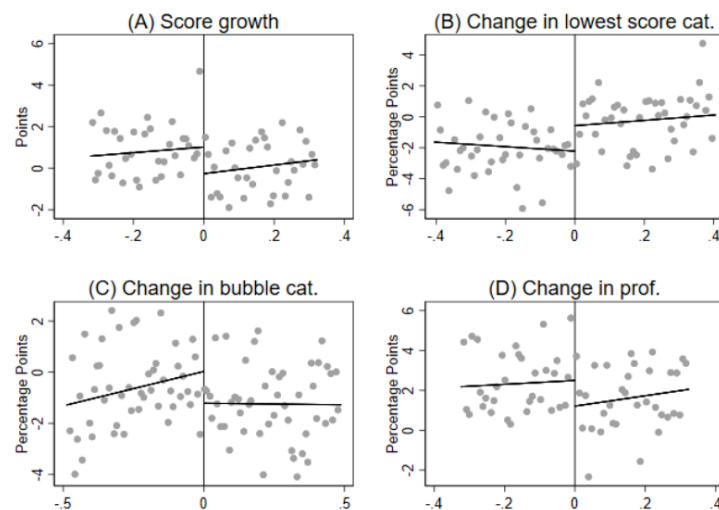


Figure 3: Reduced form impact of Approaching vs. Meeting on math outcomes

to be as notable discontinuities here. A weak argument may be made that graphs A and B present discontinuities; interestingly, the signs of these discontinuities are opposite of the prior evidence, perhaps suggesting that schools just above the higher cut point improved as a function of the higher rating. Still, that claim would need to be confirmed by additional evidence presented below. Figure 6 presents the same ELA-related outcomes as Figure 4, but at the Meeting vs. Exceeding cut point. Similar to math-related results in Figure 5, there may be a weak argument for differences present in graphs A and B, again with different signs than at the prior cut points, but that evidence should be interpreted cautiously and only if verified by additional analyses that follow.

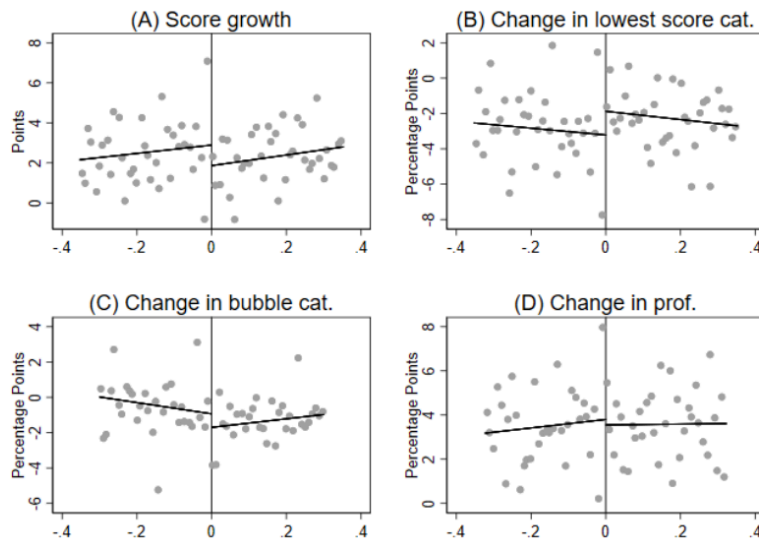


Figure 4: Reduced form impact of Approaching vs. Meeting on ELA outcomes

SENSITIVITY TO BANDWIDTH AND FUNCTIONAL FORM

The results from regression discontinuity designs can be sensitive to the choice of bandwidth; in other words, depending on how one defines the range of scores for which the schools are essentially similar, the analysis may be biased or imprecise. Indeed, the choice of bandwidth is a limitation to the regression discontinuity approach; choosing to widen the bandwidth to improve precision (by including more data points) inherently adds

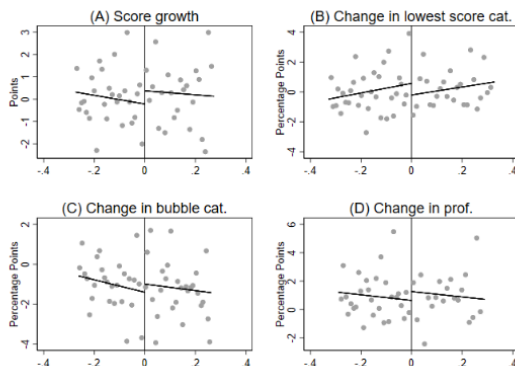


Figure 5: Reduced form impact of Meeting vs. Exceeding on math outcomes

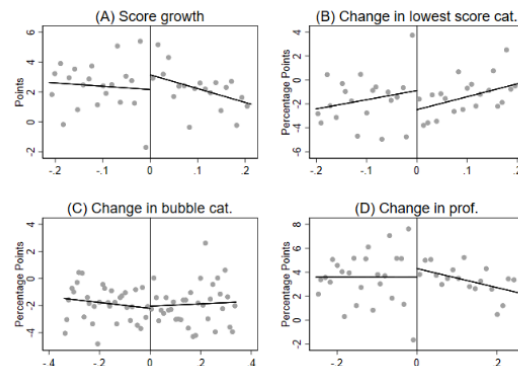


Figure 6: Reduced form impact of Meeting vs. Exceeding on ELA outcomes

bias to point estimates (Lee & Lemieux, 2010). To address this, Lee and Lemieux (2010) suggest exploring the sensitivity of results with a variety of bandwidths and a variety of orders of polynomial, although recent evidence (Gelman & Imbens, 2019) suggest limiting higher order polynomials to second order (quadratic). Ideally, these differing specifications should provide similar estimates of the treatment in both magnitude and sign, revealing a rough approximation of the “true” causal effect.

I present estimates for a variety of bandwidths and polynomial specifications in Tables 8 and 9. Presented in each cell is , the coefficient on the “treatment,” which is defined as being either just below or above the cut point; in Table 6, treatment is just receiving an “Approaching” rating, while in Table 7, treatment is just receiving a “Meeting” rating.

The estimates presented, then, are the causal impact of being just below the cut point, in comparison to otherwise similar schools, for each of the given outcome measures. While there is no expectation that the estimates should be exactly the same across bandwidths and polynomial orders, similarities between the variety of specifications, as well as the visual evidence presented in the earlier figures, may provide a preponderance of evidence of a true causal impact.

The estimates in table 8 are split by math and ELA results in the top and bottom panels, respectively. Rows 1, 2, and 4 each suggest similar results to the visual evidence provided in figure 3; while the magnitudes are not identical, the fact that multiple bandwidths and specifications lead to statistically significant increases in score growth should be taken together as a preponderance of evidence. There is ample evidence, then, that being just below the cut point and receiving an Approaching rating causes increases in math score growth, decreases in the percentage of students in the lowest proficiency category, and increases in the percentage of students scoring proficient, relative to similar schools just beyond the cut point receiving a Meeting rating. There does not seem to be substantial evidence that there is an impact on the “bubble” student category. Conversely, in examining the bottom panel for ELA estimates, there does not appear to be statistically significant evidence of differences at the Approaching vs. Meeting rating, the exception being suggestive evidence of differences in score growth. This is not surprising considering the less substantial visual evidence presented in figure 4.

In Table 9, I present a similar set of estimates for the higher Meeting vs. Exceeding cut point; the top panel is Math outcomes and the bottom panel for ELA. These results confirm the visual evidence in figures 5 and 6; there does not appear to be much evidence of an impact of just receiving the Meeting rating, with the exception being row 2 in the lower panel for ELA. These results weakly suggest that just being rated Meeting may cause an increase in the percentage of students in the lowest category of ELA performance relative to schools just above the cut point receiving a rating of Exceeding.

ROBUSTNESS CHECKS AND ADDITIONAL LIMITATIONS

One mechanism to potentially increase precision is the addition of covariates. The addition of covariates should strictly not shift the magnitude or direction of the results, only the precision, and if the addition of covariates does in fact shift the results the implication is that there was either a manipulation issue or a specification issue (Lee & Lemieux, 2010). In Table 10, I present results for both Math and English scores at the Approaching vs. Meeting cut point with a collection of school-level covariates added (see Table 2 for the comprehensive list of covariates). I choose to omit further results for Meeting vs. Exceeding as the preferred specification was not statistically significant. The results in Table 10 resemble the results in Table 8, as they should, including the relatively weak, suggestive evidence that there may be an impact on ELA score growth.

There are some additional considerations and limitations that must be addressed. The first is regarding the strength and significance of the conclusions; while there is ample evidence that receiving a lower label increased future growth in test scores, particularly for math exams, the results are by no means wholly conclusive. Because there are only three sets of paired years data, with more data the conclusions would be more robust and perhaps more precise.

Table 8
Estimated Effects of Just Receiving a “Approaching” Rating

	Polynomial order = 1			Polynomial order = 2		
	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth	MSE Opt.	.2	.8	MSE Opt.	.2	.8
<i>Math</i>						
Score Growth	1.994** (0.656)	2.449** (0.798)	0.997* (0.438)	2.003** (0.711)	2.185+ (1.117)	1.775** (0.624)
<i>N</i>	2981	1863	6631	5213	1863	6631
Lowest cat.	-2.376** (0.858)	-3.340** (1.157)	-1.518* (0.610)	-3.649** (1.222)	-3.006+ (1.768)	-2.461** (0.878)
<i>N</i>	3472	1863	6631	3733	1863	6631
Bubble cat.	0.656 (0.595)	0.603 (0.965)	0.662 (0.479)	0.733 (1.049)	0.0243 (1.361)	0.628 (0.729)
<i>N</i>	4707	1863	6631	3472	1863	6631
Proficient	2.090** (0.758)	2.737** (0.943)	0.856+ (0.501)	2.832** (0.962)	2.982* (1.322)	1.833* (0.731)
<i>N</i>	3070	1863	6631	3970	1863	6631
<i>ELA</i>						
Score Growth	0.855+ (0.455)	0.954 (0.689)	0.695* (0.344)	1.085+ (0.651)	0.213 (0.961)	0.984+ (0.511)
<i>N</i>	4169	1876	6702	4437	1876	6702
Lowest cat.	-1.012 (0.752)	-1.023 (1.047)	-0.912+ (0.516)	-0.967 (1.017)	-0.466 (1.469)	-1.229 (0.768)
<i>N</i>	3564	1876	6702	4262	1876	6702
Bubble cat.	0.127 (0.876)	0.104 (1.086)	0.695 (0.503)	-0.0592 (1.054)	-0.160 (1.626)	0.490 (0.781)
<i>N</i>	2933	1876	6702	4262	1876	6702
Proficient	0.805 (0.774)	0.919 (0.995)	0.217 (0.493)	0.813 (0.820)	0.626 (1.448)	0.738 (0.735)
<i>N</i>	3094	1876	6702	5711	1876	6702

Notes: Heteroskedasticity robust standard errors clustered by school in parentheses. Each coefficient is the reduced form estimate of the relationship between Student Achievement Rating and the listed outcome. Coefficients are generated by *RDRobust* command, implementing local linear (Cols 1–3) and quadratic (Cols 4–6) regressions with a triangular kernel. The first columns (1 and 4) in each panel utilize a MSE-optimized bandwidth (Calonico et al., 2017), while subsequent columns use a prespecified bandwidth. All models include year fixed effects.
+ p<.10, * p<0.05, ** p<0.01, *** p<0.001

The second and more important consideration is the issue of potential manipulation at the Approaching vs. Meeting cut. Visually, the histogram appears as if there may be a “jump” at the cut point, which ideally should not be the case; there will of course be some idiosyncratic lumpiness throughout the distribution but seeing a particular “jump” at the cut point suggests there may be schools manipulating their scores right at the cut point to move from just below to just above. Further, there is a policy-related chance that manipulation was happening. Because schools had access to the data used to calculate the Student Achievement score ahead of the report’s publication, they could have theoretically calculated their (future) Student Achievement scores relative to the cut point prior to their official label assignment. Knowing their label assignment, they could have attempted to interfere with students’ testing or attempt to manually edit their data to account for the potential lower rating.

Table 9
Estimated Effects of Just Receiving a “Meeting” Rating

	<u>Polynomial order = 1</u>			<u>Polynomial order = 2</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth	MSE Opt.	.2	.8	MSE Opt.	.2	.8
<i>Math</i>						
Score Growth	-0.792 (0.709)	-1.208 (0.805)	-0.372 (0.434)	-1.278 (0.903)	-2.465* (1.099)	-0.761 (0.638)
<i>N</i>	2622	1792	6439	3049	1792	6439
Lowest cat.	0.659 (0.808)	0.948 (1.039)	0.305 (0.508)	0.939 (1.165)	2.434 (1.485)	0.700 (0.795)
<i>N</i>	3209	1792	6439	3209	1792	6439
Bubble cat.	-0.201 (0.694)	-0.338 (0.808)	0.224 (0.403)	-0.408 (0.870)	-0.825 (1.196)	-0.0937 (0.604)
<i>N</i>	2622	1792	6439	3482	1792	6439
Proficient	-0.459 (0.824)	-0.609 (0.959)	-0.529 (0.532)	-0.522 (1.057)	-1.609 (1.336)	-0.606 (0.760)
<i>N</i>	2772	1792	6439	3379	1792	6439
<i>ELA</i>						
Score Growth	-0.470 (0.520)	-0.566 (0.600)	-0.0762 (0.328)	-0.659 (0.632)	-1.202 (0.866)	-0.417 (0.469)
<i>N</i>	2567	1816	6559	3541	1816	6559
Lowest cat.	1.080 (0.681)	1.420+ (0.766)	0.518 (0.410)	2.080* (0.991)	2.281+ (1.185)	0.862 (0.582)
<i>N</i>	2362	1816	6559	2437	1816	6559
Bubble cat.	-0.624 (0.690)	-1.161 (0.790)	-0.133 (0.402)	-1.862+ (1.039)	-1.756 (1.214)	-0.499 (0.581)
<i>N</i>	2437	1816	6559	2362	1816	6559
Proficient	-0.317 (0.660)	-0.259 (0.847)	-0.385 (0.478)	-0.372 (0.873)	-0.525 (1.227)	-0.363 (0.680)
<i>N</i>	3436	1816	6559	3876	1816	6559

Notes: Heteroskedasticity robust standard errors clustered by school in parentheses. Each coefficient is the reduced form estimate of the relationship between Student Achievement Rating and the listed outcome. Coefficients are generated by *RDRobust* command, implementing local linear (Cols 1–3) and quadratic (Cols 4–6) regressions with a triangular kernel. The first columns (1 and 4) in each panel utilize a MSE-optimized bandwidth (Calonico et al., 2017), while subsequent columns use a prespecified bandwidth. All models include year fixed effects.

+ $p < .10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Manipulating test scores to move from one side of the cut point to another seems like an unlikely and unwieldy task for an administrator, however. The administrator would need both a significant amount of time to develop the calculations, as well as plausible rationale to manually adjust scores. Further, a savvy administrator would likely manipulate their grade to be higher. Assuming a savvy administrator would also lead to higher test scores, this manipulation would in fact bias results in the opposite direction; those schools that were manipulated to be just beyond the cut point should see more growth in comparison their otherwise similar schools just below the cut point. Any potential manipulation, then, would suggest the results are actually larger than presented here. Adding further analysis to those specific cases immediately past the cut certainly would be beneficial in the future.

An additional potential issue to be considered is that of schools sliding back and forth across the cut point. Because test scores are included in subsequent years' Student

Achievement scores, there is a possibility that the gains described above of just receiving a rating of Approaching will push a given school into the Meeting category the following year, making the school a control school. While the year fixed effect in the model addresses any between-year dependencies by limiting comparisons to within-year, it does not wholly address the issue from an interpretive standpoint. Because the gains are local to the cut point, if schools are simply sliding back and forth across the cut point, the results are far less meaningful, suggesting any impacts are both short-term and immediately reversed.

Table 10
Estimated Effects of Just Receiving an “Approaching” Rating With Covariates Added

	<u>Polynomial order = 1</u>			<u>Polynomial order = 2</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth	MSE Opt.	.2	.8	MSE Opt.	.2	.8
<i>Math</i>						
Score Growth	1.857** (0.656)	2.355** (0.776)	0.826+ (0.436)	1.991** (0.722)	2.329* (1.095)	1.788** (0.620)
<i>N</i>	2756	1820	6472	4750	1820	6472
Lowest cat.	-2.679** (0.894)	-3.521** (1.143)	-1.364* (0.606)	-3.590** (1.173)	-3.568* (1.753)	-2.705** (0.874)
<i>N</i>	3090	1820	6472	3746	1820	6472
Bubble cat.	0.585 (0.620)	0.564 (0.939)	0.536 (0.474)	0.590 (1.045)	0.0432 (1.332)	0.740 (0.716)
<i>N</i>	4291	1820	6472	3270	1820	6472
Proficient	2.295** (0.792)	2.958** (0.934)	0.829+ (0.500)	2.979** (0.976)	3.525** (1.302)	1.965** (0.730)
<i>N</i>	2756	1820	6472	3746	1820	6472
<i>ELA</i>						
Score Growth	0.884+ (0.470)	0.917 (0.675)	0.689* (0.338)	1.042 (0.658)	0.106 (0.949)	1.064* (0.509)
<i>N</i>	3776	1832	6535	4242	1832	6535
Lowest cat.	-1.245+ (0.742)	-1.285 (1.049)	-1.053* (0.516)	-1.229 (1.035)	-0.465 (1.464)	-1.571* (0.779)
<i>N</i>	3675	1832	6535	4158	1832	6535
Bubble cat.	0.514 (0.840)	0.419 (1.089)	0.851+ (0.505)	0.358 (1.058)	-0.208 (1.606)	0.758 (0.789)
<i>N</i>	3112	1832	6535	4242	1832	6535
Proficient	0.686 (0.769)	0.867 (0.977)	0.202 (0.487)	0.826 (0.836)	0.673 (1.441)	0.813 (0.725)
<i>N</i>	3026	1832	6535	5297	1832	6535

Notes: Heteroskedasticity robust standard errors clustered by school in parentheses. Each coefficient is the reduced form estimate of the relationship between Student Achievement Rating and the listed outcome. Coefficients are generated by *RDRobust* command, implementing local linear (Cols 1–3) and quadratic (Cols 4–6) regressions with a triangular kernel. The first columns (1 and 4) in each panel utilize a MSE-optimized bandwidth (Calonico et al., 2017), while subsequent columns use a prespecified bandwidth. All models include year fixed effects and school-level covariates (listed in Table 5).

+ $p < .10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 7 provides a descriptive picture of this potential issue by mapping the treatment/control status of all schools within the preferred bandwidth of year 1 (2015's report card) at the Approaching/Meeting cut point. In the leftmost column are size-weighted markers for schools just below and above the Approaching/Meeting cut point on the 2015 report card. The middle column filters those schools by their rating, if within bandwidth, the following year. Because some schools move out of the bandwidth altogether in the following year, the markers in 2016 do not sum to their respective markers in 2015. Finally, a similar split is demonstrated between 2016 and 2018 (the next analytic year). Of the 157 treatment schools within the analytic bandwidth labeled Approaching in 2015, only 11 moved up to Meeting in 2016 and back again to Approaching in 2018. Similarly, of the 220 "control" schools, only 6 move back to Approaching and subsequently up to Meeting once again, suggesting any problematic sliding back and forth across the cut point is limited.

While the descriptive picture above suggests only a limited impact of "sliding" back and forth across the cut point, there are legitimate policy reasons why the impact might be limited as well. While the measured outcome, test scores, are a part of future Student Achievement scores and ratings, it is not the only measure; there is significant noise in the assignment variable as it is constructed from a variety of metrics (including test scores, attendance, school surveys, etc.).

Last, there may be a concern that prior treatment, including prior year ratings from this system or the prior system, may present an identification issue. For that to be the case,

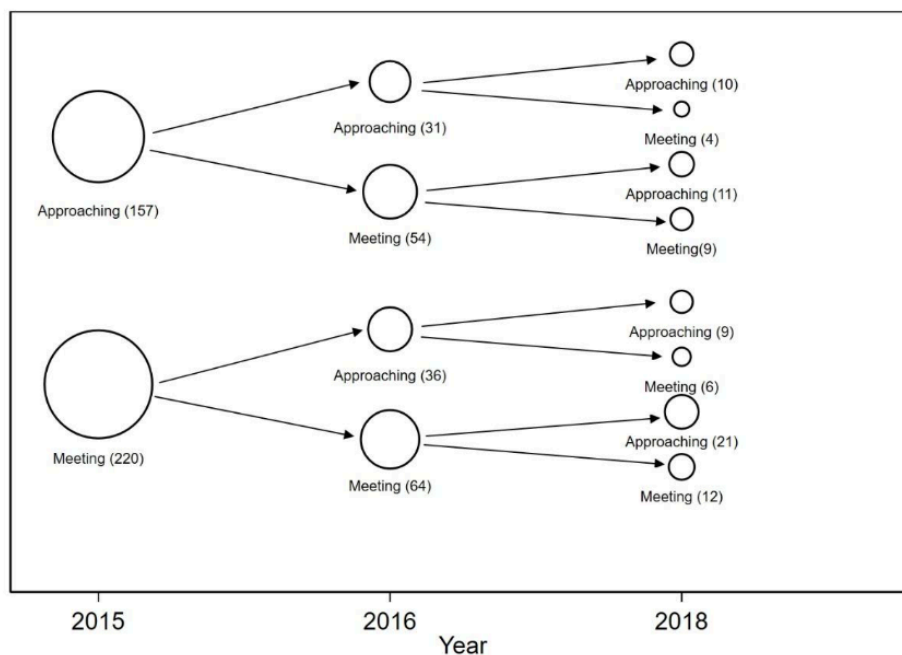


Figure 7: Within-bandwidth distribution of 2015 sample schools

prior years' treatment would need to be endogenously associated with the treatment above and beyond the forcing variable, which is unlikely to be the case. That is to say, even with prior intervention based on prior treatment, which may move schools up (or down) on the continuous forcing variable, there's nothing to suggest that movement would be different immediately surrounding the cut point. While prior treatment may impact some schools, those schools are likely to be distributed across the cut point in a given year, and year fixed-effects ensures only within-year comparisons are estimated.

DISCUSSION

The results suggest two key causal implications of receiving an “Approaching” Student Achievement rating for schools near the cut point in comparison to schools near the cut point receiving a rating of Meeting. First, schools just receiving an Approaching rating causes greater growth in the following year’s math scores, while the evidence is not nearly as strong for ELA. This is consistent with prior evidence of that strict accountability implementation increases in math more dramatically than reading (Hansen et al., 2018).

Second, schools just below the cut point appear to be more effective than otherwise similar schools at moving students out of the lowest category of math score, as well as more effective than otherwise similar schools at moving students into the proficient category of math score. While there is little evidence for the bubble students, this very well may be because schools just below the cut point are moving students both in and out of the bubble, masking any real impact or difference between the two groups of schools despite the progress being made. In conjunction, the evidence I present above suggests that, at least for math, schools and their leaders are responding specifically to the categorical Student Achievement rating above and beyond any information presented by the continuous numeric achievement score.

While the evidence above may indicate a lower categorical Student Achievement rating causes increased math test score growth in the following year for schools near the Approaching/Meeting cut point, it does not address why this happens. There are a variety of potential explanations with policy implications. For example, it may be that for schools just below the cut point there is a differential motivating factor that leads to different tactics leading to increased achievement and/or focus on test scores in the following year relative to their peers, a form of satisficing (Simon, 1956). Alternatively, schools who just barely reach the threshold for a higher score may see this achievement as sufficient relative to their peers and place less emphasis on test scores the following year (another form of satisficing). Again, this is important from a policy perspective because it suggests that schools and their personnel react positively to negative information about their institution, even if that information is comparatively marginal (i.e., puts them just below the cut point). Further, it suggests that despite the availability of the fully continuous Student Achievement score, schools are reacting to the categorical ratings and not numeric score, otherwise there would likely be no discontinuous result.

IMPLICATIONS

FOR POLICY AND PRACTICE

I first consider these findings in the context of the broader education accountability movement. There was a marked increase in available information from No Child Left Behind and other national policies (e.g., Every Student Succeeds Act, Elementary and Secondary Education Act, etc.) that require additional information, namely testing, to be collected by states and municipalities. These findings indicate that while incredibly detailed, individual level data are collected from numerous standardized examinations and/or other data systems, the broad, school-level categorical data causes schools to make some sort of change leading to differences in future test scores. Whether this is because schools are making meaningful changes or not is not determined here; rather, I

show evidence above that at least something different is happening as a function of the categorical rating that may have been difficult to calculate without NCLB's systematic collection of data. Still, while the increase in information ushered in by new accountability policies might be celebrated, it is unclear whether schools have the capacity to use that information effectively; that is, taking advantage of its full level of detail. Therefore, while collapsing data down to a more digestible chunk may have its benefits, in these results there is evidence that it perhaps causes limited response when more comprehensive response may be more beneficial.

While it is dangerous to make broad policy recommendations stemming from a single study or perspective, it is worth considering the impact providing clear and distinct information has on schools, again considering if that information (the Student Achievement rating) is a rough approximation of a more subtle yet just as easily available metric (the Student Achievement score). That is, the evidence above suggests that schools are less likely to respond to a continuous measure and more likely to respond to a categorical one. Perhaps, then, district accountability offices might see more efficiently distributed impacts if they create and distribute more simple, categorical measures of school quality to induce positive changes, especially for schools that are close to given cut points. These could include or expand upon the six sub-areas currently in the NYCDOE School Quality Snapshot (NYCDOE, 2018).

A second implication is that there is a need for district-level support for schools and their data teams. While the New York City Department of Education intended to create a more detailed, nuanced look at school quality, the evidence above suggests the response was similar to the old, "one-dimensional" (NYCDOE, 2015) report cards. If schools at the cut points are responding only to the ratings and not the more detailed information contained in the continuous Student Achievement score, that may be because of a lack of knowledge or resources for doing so. School leaders and teachers would perhaps benefit, then, from additional training carried out by district-provided experts in data analysis. If district leaders could develop processes for schools and their leaders to use the more complex and detailed information, perhaps there could be positive impacts across the spectrum and not just at the cut points.

FOR FUTURE RESEARCH

To advance understanding of school report cards and school accountability systems writ large, researchers should extend these analyses to other outcome variables. For instance, they should examine less traditional outcomes besides test scores, some of which are accessible via publicly shared data on the NYCDOE website. First might be survey-related outcomes, including results from parents and teachers; NYCDOE conducts an annual school climate survey (NYCDOE, 2019) that asks parents, teachers, and older (high school) students questions regarding the functioning of their school including evaluating leadership, school culture, and safety. There are a variety of potential outcome variables of interest embedded in the survey, including shifts in trust for principals based on prior rating, or different parental perspectives on a school based on prior rating.

Also, this analysis could be extended to high schools, which would inherently lead to another compelling application. Because New York City has a system of "universal choice" (see Abdulkadiroğlu et al., 2005 for a summary of the system), students have access to

these scores prior to making their school application decisions. Perhaps, then, there are causal impacts on not only what happens at a given school, but who chooses to attend; does a lower label cause different students to apply to a given school in comparison to otherwise similar schools with a higher (or lower) label? Each of these potential extensions could be explored in the future.

CONCLUSIONS

The RD analyses suggests that receiving a lower categorical Student Achievement rating on a school accountability report may causally increase test score growth on Math exams for those schools who are close to the cut point, as well as decrease the number of students in the lowest proficiency category while increasing the number of students in scoring proficient. While these results are similar to prior work, there are a few key differences; while prior work (Rockoff & Turner, 2010; Winters & Cowen, 2012) demonstrated a causal impact of low grades, they did so at a time when specific sanctions including financial considerations and choice implications were associated with the report cards. Further, the former report cards were scaled only partially, and to a large group of peer schools. Thus, the prior measured impacts may or may not necessarily have been directly attributable to the report card itself; rather, the sanctions, choice threats, or relative performance to peers may have been motivating factors. Indeed, Hanushek and Raymond (2005) note that shifts from simple public-facing accountability to a system involving consequences had positive impacts on student achievement.

In comparison, the results presented here do not necessarily come attached to a consequential system; there were no such threats associated with a low score at the time. In fact, the reports themselves were designed to be more holistic and inspire a more diverse set of changes (NYCDOE, 2015). This strengthens the argument that the rating itself is causing the shift in score growth. While there are a variety of potential explanations for such a phenomenon, the fact that the rating appears to have a causal impact in and of itself provides important information for those working in and around school accountability: even if detailed information is available, the act of labeling a school with a particular rating can have an impact on its own.

REFERENCES

- Abdulkadiroğlu, A., Pathak, P. A., & Roth, A. E. (2005). The new york city high school match. *American Economic Review*, 95(2), 364–367.
- Booher-Jennings, J. (2005). Below the Bubble: “Educational Triage” and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231–268. <https://doi.org/10.3102/00028312042002231>
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression discontinuity designs. *The Stata Journal*, 17(2).
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305–331.
- Chakrabarti, R. (2007). *Vouchers, public school response, and the role of incentives* (Staff Report No. 306). Federal Reserve Bank of New York. <http://eric.ed.gov/?id=ED517702>
- Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93(9), 1045–1057.
- Corcoran, S., & Pai, G. (2013). *Unlocking New York City’s High School Progress Report*. New Visions for Public Schools. https://research.steinhardt.nyu.edu/scmsAdmin/media/users/ggg5/Unlocking_NYCs_High_School_Progress_Report_Corcoran_Pai.pdf
- Darville, S. (2014, October 1). Read Chancellor Fariña’s speech outlining the city’s new school rating system, but no other changes. *Chalkbeat*. <https://chalkbeat.org/posts/ny/2014/10/01/read-chancellor-farinas-speech-outlining-the-citys-new-school-rating-system/>
- Dee, T. S., & Jacob, B. (2011). The impact of no Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446. <https://doi.org/10.1002/pam.20586>
- Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, 98(5), 848–862.
- Deming, D. J., & Figlio, D. (2016). Accountability in US education: Applying lessons from k-12 experience to higher education. *Journal of Economic Perspectives*, 30(3), 33–56. <https://doi.org/10.1257/jep.30.3.33>
- Engel, M. (2000). *The Struggle for Control of Public Education: Market Ideology Vs. Democratic Values*. Temple University Press.
- Figlio, D. N., & Lucas, M. E. (2004). What’s in a grade? School report cards and the housing market. *American Economic Review*, 94(3), 591–604. <https://doi.org/10.1257/0002828041464489>

- Friesen, J., Javdani, M., Smith, J., & Woodcock, S. (2012). How do school 'report cards' affect school choice decisions? *Canadian Journal of Economics/Revue Canadienne d'économique*, 45(2), 784–807. <https://doi.org/10.1111/j.1540-5982.2012.01709.x>
- Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics*, 37(3), 447–456.
- Gootman, E., & Medina, J. (2007, November 6). 50 New York Schools Fail Under Rating System. *The New York Times*. <https://www.nytimes.com/2007/11/06/education/06reportcards.html>
- Hamilton, L. S., & Koretz, D. M. (2002). *Tests and their use in test-based accountability systems* (p. 39). RAND Corporation.
- Hansen, M., Levesque, E., Valant, J., & Quintero, D. (2018). *The 2018 Brown Center report on American education: How well are American students learning?* Washington, DC: The Brookings Institution.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297–327. <https://doi.org/10.1002/pam.20091>
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5–6), 761–796. <https://doi.org/10.1016/j.jpubeco.2004.08.004>
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355. <https://doi.org/10.1257/jel.48.2.281>
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3–16. <https://doi.org/10.3102/0013189X031006003>
- NYCDOE. (2015). *Chancellor Fariña Announces More Students Are Graduating College and Career Ready As Part of First Annual 'School Quality' Reports Release*. <https://www.schools.nyc.gov/about-us/news/announcements/contentdetails/2014/11/10/chancellor-fariña-announces-more-students-are-graduating-college-and-career-ready-as-part-of-first-annual-school-quality-reports-release>
- NYCDOE. (2018). *Educator Guide - New York City DOE School Quality Reports*. Retrieved from <https://infohub.nyced.org/docs/default-source/default-document-library/2018-19-educator-guide-ems---11-13-2019.pdf>
- NYCDOE. (2019). *NYC School Survey*. Retrieved from <https://www.schools.nyc.gov/about-us/reports/school-quality/nyc-school-survey>
- Rockoff, J., & Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal. Economic Policy*, 2(4), 119–147. <http://dx.doi.org.ezproxy.lib.uconn.edu/10.1257/pol.2.4.119>
- 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), 251–281. <http://dx.doi.org.ezproxy.lib.uconn.edu/10.1257/pol.5.2.251>
- Shipp, D., & White, M. (2009). A New Politics of the Principals? Accountability-Driven Change in New York City. *Peabody Journal of Education*, 84(3), 350–373. <https://doi.org/10.1080/01619560902973563>
- Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99. <https://doi.org/10.2307/1884852>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129.
- Wall, P. (2014, October 1). Under gentler rating system, schools will no longer be ranked or graded. *Chalkbeat*. <https://chalkbeat.org/posts/ny/2014/10/01/under-gentler-rating-system-schools-will-no-longer-be-ranked-or-graded/>
- Winters, M. A., & Cowen, J. M. (2012). Grading New York: Accountability and student proficiency in America's largest school district. *Educational Evaluation and Policy Analysis*, 34(3), 313–327. <https://doi.org/10.3102/0162373712440039>

PERCEPTIONS OF HIGHER EDUCATION PROFESSIONALS ON THE UTILITY OF THE ACTIVITIES, PROGRAMS, OR POLICIES TOOL TO PROMOTE SELF-DETERMINATION FOR COLLEGE STUDENTS WITH DISABILITIES

ASHLEY TACONET

POSTDOCTORAL RESEARCHER, CENTER ON DEVELOPMENTAL DISABILITIES
UNIVERSITY OF KANSAS, ORCID 0000-0003-2407-5114

EMILY TARCONISH

TEACHING ASSISTANT PROFESSOR, SPECIAL EDUCATION
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN, ORCID 0000-0001-7434-180X

JOSEPH W. MADAUS

PROFESSOR, EDUCATIONAL PSYCHOLOGY, DIRECTOR, COLLABORATIVE ON POSTSECONDARY
EDUCATION AND DISABILITY
NEAG SCHOOL OF EDUCATION, UNIVERSITY OF CONNECTICUT, ORCID 0000-0003-0248-0629

NICHOLAS GELBAR

ASSOCIATE RESEARCH PROFESSOR
NEAG SCHOOL OF EDUCATION, UNIVERSITY OF CONNECTICUT, ORCID 0000-0002-7734-2207

LYMAN DUKES, III

PROFESSOR OF SPECIAL EDUCATION
COLLEGE OF EDUCATION, UNIVERSITY OF SOUTH FLORIDA, ORCID 0000-0002-9308-6821

MICHAEL FAGGELLA-LUBY

PROFESSOR OF SPECIAL EDUCATION
COLLEGE OF EDUCATION, TEXAS CHRISTIAN UNIVERSITY, ORCID 0000-0002-9735-2293

DON MILLS

DISTINGUISHED PROFESSOR OF HIGHER EDUCATION LEADERSHIP (RETIRED)
COLLEGE OF EDUCATION, TEXAS CHRISTIAN UNIVERSITY, ORCID

ABSTRACT

Self-determination has been associated with academic success for college students with and without disabilities. The APP Tool was designed to allow higher education professionals to examine which campus Activities, Programs, or Policies (APPs) promote, and which might hinder, the development of student self-determination. This study used the qualitative basic interpretive approach (Merriam & Tisdell, 2016) to analyze data from semi-structured interviews of three focus groups of higher education professionals (practitioners) that were conducted to ascertain their impressions of the utility of the tool. Use of the APP Tool led practitioners to reflect on what self-determination included and what campus efforts were currently fostering these skills. Implications of the APP Tool included use as (1) a progress monitoring tool for student self-determination skills and (2) an evaluation tool for current campus programming.

Keywords: self-determination, higher education, student affairs, students with disabilities, disability services, postsecondary education, focus group

Postsecondary education has required students to set both small (e.g., achieve adequate grades on assignments and exams) and large (e.g., complete classes and academic degrees) goals, as well as independently self-regulate their behavior to enable this progress. As such self-determination, a concept that embodies these skills, is noted as critical for all postsecondary students. Field et al. (1998) proposed the following description of self-determination:

Self-determination is a combination of skills, knowledge and beliefs that enable a person to engage in goal-directed, self-regulated, autonomous behavior. An understanding of one's strengths and limitations together with a belief in oneself as capable and effective are essential to self-determination. When acting on the basis of these skills and attitudes, individuals have greater ability to take control of their lives and assume the role of successful adults in our society (p. 2).

Wehmeyer et al. (2007) stated self-determined behaviors encompass "volitional actions" (p. 5) that allow individuals to make choices and direct their behaviors to maintain or alter their lives. The volitional actions that describe self-determination included four essential characteristics: (1) the person acted autonomously; (2) the behavior was self-regulated; (3) the person initiated and responded to the event in a psychologically empowered manner; and (4) the person acted in a self-realizing manner. These characteristics described the function of the behavior that makes it self-determined or not (Wehmeyer et al., 2007). Volitional actions can also be achieved through the development of related attitudes and abilities, or the component elements of self-determination. Definitions of the component elements of self-determination can be found in Appendix A.

Self-determination is noted as an important skill for all students in postsecondary education (Faye & Sharpe, 2008; Graham & Vaughn, 2022; Guiffrida et al., 2013). Specifically, increased levels of self-determination have been associated with higher grade point averages (GPAs) and levels of satisfaction with life in college for all postsecondary students (Graham & Vaughn, 2022), as well as higher rates of student engagement in postsecondary education (Faye & Sharpe, 2008; Guiffrida et al., 2013). For postsecondary students, Guiffrida et al. (2013) also found a relationship between source of motivation, specifically in areas of autonomy and competence, and persistence in postsecondary education as well as GPA.

The benefits of learning and using self-determination skills for college students with disabilities (SWD) have also been well established (Field et al., 2003; Gelbar et al., 2020; Ju et al., 2017; D.R. Parker, 2004; Petcu et al., 2017; Sarver, 2000). Researchers have found positive relationships between GPA in postsecondary education and levels of self-determination for students with learning disabilities (Field et al., 2003; Sarver, 2000) and for students with Attention-Deficit/Hyperactivity Disorder (D.R. Parker, 2004). Ju et al. (2017) found that "teaching self-determination skills or developing self-determined behaviors can enhance overall self-determination leading to academic success" (p. 186), including higher GPA and retention rates for postsecondary SWD.

Self-determination may have been even more crucial for SWD, as accommodation provision is largely reliant on students' proactive behaviors. Postsecondary disability services required SWD to self-disclose their disability or disabilities to their office and professors as needed to receive accommodations (Fleming et al., 2017; Newman & Madaus,

2015; O'Shea & Meyer, 2016). Depending on campus policies students have also needed to request accommodations every semester, deliver or initiate delivery of accommodation letters to their instructors, and follow-up with disability resource personnel or instructors themselves if they required a change in accommodations.

While self-determination is relevant to all postsecondary students, especially those with disabilities, the topic is less researched for students without disabilities. In these settings, self-determination skill development most commonly occurred within disability resource offices (University of California Berkley, n.d.; University of Colorado Boulder, n.d.; University of Illinois at Urbana-Champaign, n.d.); however, this programming only reached students with disabilities, specifically those who chose to disclose and register for services. As only one third of postsecondary students with disabilities registered with disability resource offices (National Center for Education Statistics, 2022), this information reached a relatively small number of students with disabilities. Gelbar et al. (2020) called attention to the fact that not all students register with their disability service office, and students may have invisible disabilities (e.g., learning disabilities, mental health conditions), therefore student affairs professionals may have served students with undisclosed disabilities. Providing opportunities to practice self-determination through both disability resource and student affairs offices could reach a broader population of students with disabilities than through disability resource offices alone. Additionally, self-determination training provided through student affairs offices could reach the general population of college students who may also benefit from these skills.

There has been a need to ensure higher education programming promoted self-determination skills for all students. This included identifying and continuing current programs that effectively cultivate these skills, modifying or eliminating programs that do not effectively develop self-determination, and developing new programs to support skills not being fostered. The Activities, Programs, or Policies (APP) Tool (Mills et al., 2019) provided a systematic means to evaluate whether self-determination skills are supported by campus programming. The APP Tool, which consisted of a three-column form that can be used as a hard-copy or electronic resource, was developed to provide postsecondary professionals, including those within student affairs and disability resource offices, with a guide to foster a campus-wide focus on self-determination skill development. APPs was a broad term used for this tool to signify the various components of higher education, though there is much overlap, and the area or event does not need to fit into one silo. Broadly, activities included activities fairs and new student orientation that are single events or occur less frequently, while programs included first year seminars and intramurals that are structured across a longer period of time. Policies included codes of conduct and attendance rules. The tool did not require users to define an event as an activity, program, or policy but instead identify the whole event, examples provided above, as an APP.

DEVELOPMENT AND USE OF THE APP TOOL

Mills et al. (2019) created the APP Tool for student affairs professionals to facilitate the development of self-determination skills. The APP Tool can be employed in a range of higher education settings and provides professionals a way to “identify self-determination

challenges common to college students and link them to programmatic responses” (Madaus et al., 2020, p. 3). It is completed in a series of reflective steps (described next). Although not necessary, the form can be completed collaboratively, with higher education professionals initially working through each step and then discussing and modifying their responses as a group throughout each stage.

The APP tool (see Appendix B) contained three columns: “Common Challenges,”

“Activities, Policies, Programs” and “Self-Determination Outcomes.” The first step when completing the APP Tool was to list common challenges incoming students, including both first year and transfer students, may face on a separate sheet of paper. This section could also be used to describe common challenges that historically marginalized populations, including, but not limited to, first generation students, SWDs, minority students, and economically challenged students experienced when they enter postsecondary education (Madaus et al., 2020). It should be noted these categories were not discrete, and individual students may identify with more than one challenge.

Next, higher education professionals using the tool would narrow the list to five to seven of the most pressing common challenges for students and record these in the first column of the Tool. If professionals listed more than seven challenges, they should focus on five to seven to ensure enough time is available to think critically about each challenge. They may choose to examine additional challenges with the Tool at a later point in time. These common challenges did not need to be relevant to the entire student population but should encompass issues experienced most frequently by the students currently under consideration when completing the APP Tool. Users then recorded the most impactful five to seven activities, policies, and programs (APPs) offered at their school, which can include the most frequent, most popular, or effective APPs as determined by their institution of higher education. Note that these, can, but do not have to be related to the common challenges listed in the first column.

After APPs have been identified, users can familiarize themselves with a numbered list of 12 specific self-determination outcomes, which are based on the component elements of self-determination (Appendix A), noted at the bottom of the APP Tool form. The self-determination outcomes included the following: (1) choice-making skills, (2) decision-making skills, (3) problem-solving skills, (4) goal setting and attainment skills, (5) independence, risk-taking, and safety skills, (6) self-observation, self-awareness or self-monitoring skills, (7) self-evaluation skills, (8) self-reinforcement skills, (9) self-instruction skills, (10) self-regulation-skills, (11) self-advocacy and leadership skills, and (12) positive attributions of efficacy and outcome expectancy skills. Using this list as a guide, professionals then identified what self-determination outcomes the specified APPs address and indicated the related self-determination outcomes in the third column, as determined by group discussion among the users filling out the APP Tool. The final step of the APP Tool process was to triangulate the data by (a) determining which APPs provide the most support in teaching self-determination skills, (b) determining which APPs provide limited or no support in teaching self-determination skills, and (c) identifying the self-determination skills not being developed by current APPs, which can inform the development of new or modified APPs to foster these skills. It was also possible to identify whether specific APPs hinder the development of student self-determination based on

whether the APP does not address any of the noted challenges or if there are no self-determination outcomes that can be identified for the APP.

The current study examined the reactions of postsecondary student affairs and disability services professionals to using the APP Tool and addressed the following research questions:

1. Does the APP Tool help focus group participants consider self-determination in relation to their work?
2. In what ways could focus group participants see themselves using the APP Tool in their work?
3. What do participant responses tell us about the utility of the APP tool in higher education settings?
4. What recommendations do focus group participants have for improving the APP Tool?

METHODS

The researchers sought to learn what student support services professionals, including student affairs professionals and disability resource office professionals, in higher education thought about the utility, benefits, and drawbacks of using the APP Tool. To do this, we conducted three focus groups with a total of 30 postsecondary education professionals across all groups. Focus groups are established as an accepted practice in a variety of fields, including business, medicine, and the social sciences to evaluate new tools, measure the effects of interventions, and gain perspectives from a variety of users (A. Parker & Tritter, 2006). They provided a vehicle to gain a deeper understanding than a purely quantitative analysis may offer and provide a social setting to evaluate attitudes and opinions towards the topic or item of interest (Breen, 2006). As such, they have been used to support the development, evaluation, and assessment of tools in the field of education (Williams & Katz, 2001).

Conducting focus groups has also been noted as a common approach to collect qualitative data in postsecondary education research (Ortiz & Waterman, 2016), as open-ended questions are used to gain multiple “perspectives from a group that shares one or more characteristics,” (Biddix, 2018, p. 146). This method prompts participants to respond to individual questions, as well as engage in conversations with each other, which can “reveal group dynamics and social processes,” and “check for shared understanding” (Biddix, 2018, p. 146). Focus groups have been previously used to study topics in higher education. Specifically, Murphrey et al. (2014) used focus groups to assess the effectiveness of different teaching platforms used at the college level and Sangster et al. (2016) used this method to evaluate undergraduate student involvement in research.

PARTICIPANTS & SETTING

As noted, three focus groups were held with a total of 30 postsecondary education professionals in multiple regions in the United States. We intentionally sought participants from a combination of disability resource and student affairs professionals in order to represent the range of individuals who work with SWD (Lalor et al., 2020; Madaus et al., 2020), but also to glean if they believed the APP Tool had utility in other functions of student affairs work. The first group was held during a national conference on postsecondary disability services and consisted of ten postsecondary disability services professionals who represented an even mix of two-year colleges, small four-year colleges, and large four-year universities. The second and third groups, conducted at two different large four-year universities in the south and southeast United States, each included ten student affairs professionals, including residential life, advising, and veterans' affairs staff, from the institution at which each group was held. The focus group participants were convenient samples of individuals who either chose to attend a conference session or were university staff at an institution the focus group facilitator worked for, though in a different department, and chose to attend the session. Given that each group had a sample size of ten participants they met the requirement of ten participants per focus group suggested by Krueger (1994). In a review of focus groups, it was found that 90% of themes were identified when there were three to six focus groups, with three being the most common number of focus groups (Guest et al., 2017), therefore using three focus groups in this study was appropriate given current literature. Additionally, when analyzing the data, saturation was reached indicating a sufficient number of participants was sampled (Glaser & Strauss, 1967).

FOCUS GROUP FORMAT

Participants in each group were invited to participate via email and each comprised a convenience sample, which is defined here as participants that were interested in the topic and chose to attend the focus group sessions. The disability service professionals were recruited from a conference session that participants had the option to attend if interested. The student affairs professionals were offered the opportunity to attend the focus groups at their respective institutions. The groups were led by one to two of the APP Tool developers, who served as moderators. Following the approved institutional review board (IRB) protocol, each moderator explained the focus group purpose, informed participants the sessions were recorded, that participation was voluntary, and their identity would remain anonymous. At that time, participants could decide if they wished to participate in the focus group or not. If an individual chose to stay for the focus group, that indicated they provided consent. The four creators of the tool, who also served as the focus group moderators, came together to create the focus group protocol. The moderators all had a copy of the protocol which they followed during the focus groups. The moderators began each session with a 10-to-15-minute overview of self-determination based upon the theoretical framework presented by Field et al. (1998) and other relevant literature on the topic, notably the key components of self-determination and how its development can affect postsecondary students. The moderators next presented the APP Tool and asked participants to collaborate in order to complete it. It was suggested

that participants evaluate the APPs at their own institution when possible. They also had the option to discuss APPs at the institution of other participants if that was preferred. Participants worked in self-chosen groups for approximately 20 minutes. Moderators then lead each group in a discussion and reflection of the utility of the tool. Each focus group lasted between 60-90 minutes in entirety.

Discussions utilized semi-structured interview questions (see Table 1). Questions were generated based upon feedback in previous presentations of the APP Tool at two national student affairs conferences. During these conference sessions, themes emerged from participants’ reactions and feedback, which served as the foundation for the interview questions. The semi-structured interview format allowed the moderators to guide the conversation with pre-prepared general questions and to ask follow-up questions as appropriate. This approach to interviewing can be useful when researchers seek to collect information on a similar topic across settings and enabled participants to express their individualized experience with the topic (Bogdan & Biklen, 2007).

Table 1
Semi-Structured Focus Group Interview Questions

1) How do the components of self-determination fit with students with disabilities? With other diverse populations of students? With students in general?
2) How could these skills be learned by students to help them be successful in college
3) What specific university activities, programs and policies could relate to development of student self-determination and how?
4) What did you learn by using the tool about your university activities, programs and policies?
5) How would this help you in your work? How might it be changed to better help you in your work?

DATA ANALYSIS

The focus group recordings were transcribed, and transcripts were later analyzed using a basic interpretive approach in order to describe, understand, and interpret the participants’ experiences. In this process, data were analyzed by identifying recurring patterns, including initial codes, which are grouped into larger categories, and then interpreted to reveal overall themes (Merriam & Tisdell, 2016). This method was chosen as it is a common method in qualitative research and the authors’ goal was to determine the overall themes across participants. The first two authors conducted data analysis for this

study. These two authors were graduate students studying higher education and disability, previously worked in positions serving students with disabilities in higher education, and both self-identify as individuals with disabilities. The first two authors completed close readings of the focus group transcripts to familiarize themselves with the data. Next, each independently generated initial codes in the form of themes, or words, that captured the units of meaning within participants' accounts (Thomas, 2006). Next, the two researchers met to discuss initial codes and resolve any discrepancies. Discrepancies were determined by comparing the researcher's initial codes and identifying those that have different meanings. Discussion between the two researchers occurred until they came to an agreement on the final list of initial codes. After this meeting, the researchers independently examined the codes and sought to make sense of them through identifying similarities, complements, or patterns. Initial codes were grouped together based on if focus group participants were discussing the same or similar ideas. This process resulted in the development of categories, which were collectively analyzed by the two researchers to identify emerging themes. Themes were described as overarching ideas that encapsulate participants' experiences into a meaningful whole. After establishing themes, the researchers assessed their validity through the examination of initial codes to ensure the themes were representative of the initial data. This process was conducted to ensure that all data present in the initial codes were sufficiently and appropriately reflected in the final themes.

SELF-DETERMINATION AND DISABILITY CRITICAL THEORY

Disability critical theory (Schalk, 2017) guided the coding and analysis process. Specifically, as we completed the analyses for this study, we examined the findings through thinking about how the assessments of the APP tool could help benefit individuals in higher education who do not fit the standard norms, including students with disabilities. To do so we used disability critical theory which defines disability as "socially constructed system of norms which categorizes and values bodyminds based on concepts of ability and disability" (Schalk, 2017, p. 1), with bodyminds indicating the overlap a person's body and mind. Disability critical theory focuses on viewing disability through the various social systems put in place and their impact on individuals with disability, as well as acknowledging other social systems such as race and ethnicity (Schalk, 2017). This theory was the strongest choice for this study given that the goal of the APP tool was to help institutions of higher education foster self-determination and identify areas that may hinder self-determination. Through using disability critical race theory, the participant responses were viewed in reference to the whole campus unless otherwise specified, instead of just being relevant to a disability service office, as people with disabilities should be supported by all campus offices. This was reflected in discussions of cross-campus collaboration and aligned with the view that self-determination was necessary for the full population of college students.

CREDIBILITY MEASURES

The researchers established credibility, or trustworthiness, of the research process in multiple ways. First, researcher triangulation, or the use of multiple investigators comparing findings throughout data analysis (Merriam & Tisdell, 2016) was used. This supported credibility as the data was not being analyzed by a single individual but by multiple individuals with varied life experiences. Second, we recorded an audit trail, or a thorough description of our analysis process to make these steps transparent (Merriam & Tisdell, 2016). Third, this manuscript also included rich descriptions, as well as the participants’ direct quotes, to support the findings of the study.

RESEARCHER POSITIONALITY

The two researchers who performed the interpretive analysis both approached this task from the position of former postsecondary education disability resource professionals, as well as former graduate-level SWDs. They were both also actively involved with student

Table 2
Examples of APPs and Self-Determination Skills

APP	Self-Determination skill
Study abroad, student organization fair, first year-experience, advising office, career planning office/fair	Choice making, decision making, problem solving, leadership skills, self-knowledge, self-awareness
First-year experience, financial literacy class	Independence, risk taking, safety, goal setting and attainment
Student code of conduct/conduct meetings	All self-determination + self-awareness
Mental health services, residence assistant support	Self-regulation
Mental health services, peer mentoring, Wellness Center, residence assistant support	Self-efficacy
Military, Veterans Affairs Offices, Multicultural Affairs	Self-advocacy, connecting activities

advocacy groups for undergraduate SWDs. As these experiences had the potential to influence their analysis of the data, both were intentional to acknowledge and recognize these positionalities and put in place three checks, explained above, among researchers throughout the research process to ensure the trustworthiness of findings.

RESULTS

Several themes were generated from the focus group data surrounding participants' use of the APP Tool during the focus group sessions. Themes included that the tool (a) helped to identify APPs that both fostered and hindered development of self-determination, (b) guided professionals and students to focus on self-determination skill development, (c) can have multiple uses and implications, (d) facilitated collaboration among diverse campus offices, and (e) suggestions for revisions to the tool. Participants in two of the groups reflected on their experiences before discussing their impressions of the tool itself. The participants claimed the tool helped them reflect on which APPs fostered self-determination. Participants debated how the APPs may foster the self-determination concepts, as well as how various programs can target the same skills, thus reinforcing their development and value. Table 2 provides a summary of self-determination components and related campus APPs as discussed by participants. For example, participants noted the APPs of mental health services and residence assistant support to promote self-regulation.

RELEVANCE TO PARTICIPANT WORK

Focus group participants discussed the relevance of the APP Tool to their work including fostering self-determination and hindering self-determination. These topics are discussed subsequently.

FOSTERING SELF-DETERMINATION

Focus group participants noted value in the opportunity to examine current campus APPs, specifically whether they are meeting their intended goals, and whether those goals also reflect the development and use of self-determination skills. One participant shared, "What I found as helpful is you can look at which outcomes we may not be hitting." Moreover, participants discussed how APPs that are intended to foster self-determination are developed constantly in higher education settings; however, strategic ways to assess if goals are achieved may not exist. One participant envisioned using the APP Tool to assess whether programs met intended objectives:

Okay, this is what we hoped for, but what actually is this program hitting and what actually is this program not hitting? And kind of comparing the [program at the beginning of the year to the program at the end of the year] to figure out how we can improve something or get rid of something, honestly, if it's not doing what we want it to do.

Others felt the APP Tool can be used to ensure students receive opportunities for skill development. One participant stated:

If we're complaining about how we see students aren't leaving our institution with a certain thing, where is this missing from our programs or what programs is it in that our students aren't taking part in and how can we get them wrapped in?

Another participant added that program goals also needed to match student expectations; if students are expected to graduate with certain self-determination skills, it is imperative these skills are explicitly taught. She noted:

It's like if we identify the outcomes, if we're seeing like a gap missing in whatever we're doing—say it's problem-solving skills ... That's the outcome that we're trying to really focus in on, and then we're going to structure it. What kind of program do we want? How are we assessing that or even reaching that outcome? And then how can we structure in that manner?

Most participants agreed using the APP Tool would reveal where programs were lacking self-determination components. One participant said, "I like the tool just because it offers you an aerial view of what it is that you're missing." Participants discussed how well-intentioned APPs, whether newly developed or carried on from previous years, may not clearly articulate the self-determination skills addressed. They felt the APP Tool helped them to deliberately consider what concepts needed to be honed in APPs.

HINDERING SELF-DETERMINATION

Members of all three groups shared the APP Tool helped them identify APPs that hindered and highlighted self-determination skills not yet specifically addressed by resources on their campuses. Importantly, group discussants reflected that APPs limiting student choice may work against the development of self-determination. For instance, "Progression Policies" encourage students to complete a certain number of courses in a specified amount of time or limit a student's ability to change his/her/their major; participants discussed how these policies directly impede students' ability to make choices about their academic careers. Participants also shared how well-intentioned APPs may impede development of self-determination skills by "solving problems for students," and not explicitly teaching them how to problem solve independently. Describing a program targeted at building community, one participant expressed "(the program) orientates you to an institution so you get a sense of belonging. Helpful, helpful, helpful. But because we're providing you a packaged sense of belonging, you're not utilizing your skills for self-determination to find your way, right?" Student Codes of Conduct were another policy discussed that may confirm what is expected of students but may not address the reason for behavioral expectations, thus limiting what students can learn from them. Discussants indicated the APP Tool not only emphasized programs that promote or hinder self-determination but additionally promoted thought about APPs that can be revised to include a focus on such skills.

APP TOOL USES

Focus group participants discussed ways that they could use the APP Tool with the themes of implications and fostering collaboration discussed subsequently.

IMPLICATIONS

Participants expressed they could use the APP Tool to not only help themselves think about self-determination, but also to assist students, their parents, and other higher education professionals. They discussed working collaboratively with students to fill in the “challenge” section, which could lead students to articulate self-determination concepts. Next, they indicated the APP Tool could be utilized as a roadmap to match students with programs that foster such skills. In this way, the APP Tool may lead students to think concretely about skill development. One participant said, “We address an issue without ever addressing the skill that’s lacking— [this tool] could provide the opportunity to find out what the lacked skill is and if the student wants help to develop that skill.” Several other participants added the APP Tool may be used in similar conversations with parents to guide them to understand the skills students are developing.

FOSTERING COLLABORATION

The fourth theme addressed how the APP Tool may facilitate collaboration among different campus offices as they collectively work to foster student self-determination.

Discussants noted the APP Tool could be used to provide an overview of the skills students should ideally develop, and departments could collaborate to determine which APPs addressed the same or different skills. One participant proposed:

I can see it being used at different levels ... this tool I think can be used more at a higher level thinking overall about everything being offered on campus and then at an individual office level could use the tool to think about. How can we impact these in each of the programs?

Another participant believed the APP Tool could enable a universal process of addressing student decision making and problem-solving needs, stating, “I think if a tool like this were tied in, it would be easy to follow up the steps. What is the problem? What is the office that it should go to? Did the student follow up?” Participants also suggested providing the tool to faculty so they might consider the incorporation of self-determination goals into their courses, and to gather faculty feedback regarding potential student self-determination needs. One participant mentioned collaborating with faculty would “allow you to develop an institutional perspective on how you better build in strategies that help students become more self-sufficient and who are able to advocate for themselves.” Overall, participants indicated utilizing the APP Tool institution-wide would enable a focus on self-determination and provide its users with a common language to describe potential self-determination goals for all students.

Participants also saw themselves using the APP Tool collaboratively with students,

especially to inform the “student challenges” section. They discussed collecting this information from students using campus-wide surveys or using the APP Tool to drive conversations with individual students. One participant proposed,

I think this could be a helpful activity to do with the student and say like, ‘What do you think the five common challenges are?’ ... instead of just one person looking at this, maybe you need more like a dynamic activity with the student so that they can share what they’re nervous about and we can point them in the right direction rather than us assuming what they don’t know or the challenges are.

Whether used by a variety of professionals or students, most participants felt the APP Tool would encourage a pro-active approach to addressing student needs. “It’s an early alert platform,” one participant expressed. She went on to describe how using the tool had the potential to streamline communication about where students could go to develop specific skills, stating “I think if a tool like this were tied in, it would be easy to follow up the steps. What is the problem? What is the office that it should go to?” Many participants shared the problem of how many students currently “wait until they need services” to seek them out. One participant noted the APP Tool might address this challenge:

It becomes more of a reactive than a proactive approach ... well, one, I may not seek out services just because I don’t know how to do that or I’m not comfortable, but even if I’m going to, I’m not going to do it until I’m hitting that point.

The group discussed how the APP Tool could help professionals anticipate student needs by highlighting which self-determination skills are not addressed by current programming.

Equipped with this foresight, professionals can develop strategic programs to enhance these skills and connect students with necessary supports before problems occur.

SUGGESTED REVISIONS

Each focus group was also asked potential ways the APP Tool could be improved. Several suggestions emerged for improving the tool, which were (a) making changes to definitions and formatting and (b) providing different versions of the tool for different populations. Several participants communicated concerns about potential users not understanding the self-determination definitions, especially if these concepts were not a common component of their discipline. To alleviate this potential issue, several participants stressed the need to ensure clarity of concepts and possibly provide examples of each (see Appendix A for examples).

Proposed changes to formatting included creating both paper and digital versions of the APP Tool and reconsidering the order of columns. Several participants advised arranging the “student challenges” and “self-determination outcome” columns adjacent to each other to emphasize their connection; this suggestion was promoted in two of the three focus groups, whereas members of the third group did not indicate this was necessary.

Another participant suggested a future iteration of the APP Tool could include descriptors, or “characteristics associated with” each component of self-determination, to make these concepts tangible for students. He shared this potential addition:

[It] would then help students identify ‘Oh, this is what I have. This is what I think I have. This is where I may be lacking,’ ... If you’re going to set it as a goal to improve an area, you can have observable, measurable characteristics to say you’ve acquired those over time and then you help them see their own growth.

The second category of suggested revision involved producing different versions of the form catered to a variety of users. Versions would incorporate either more or less explanation of the self-determination concepts based on the users’ familiarity; additional clarification would be especially relevant to student users, who may not have previously encountered language describing self-determination. A third proposed addition to the APP Tool included adding a column to indicate how APPs are being advertised, which may make it more helpful to students. Participants discussed how professionals must not only create programs to address self-determination, but also must ensure that students are learning about and participating in these opportunities. One participant added, “students might be a little more into self-advocating if they just knew where to go directly.” Participants also saw the tool as being useful to all students, not merely those with disabilities.

Finally, while most participants described the tool as useful and could see its future use on their campuses, there were several areas they felt the APP Tool did not address. They mentioned requiring more guidance regarding factors that undermine development of self-determination development, as well as how to achieve student buy-in.

DISCUSSION AND IMPLICATIONS

This analysis addressed several research questions involving participants’ use of the APP Tool, including how and what aspects of the tool helped them consider self-determination, and how and in what ways they see themselves employing the tool in their future work. Several themes emerged from the interpretive analysis of participants’ feedback. First, the APP Tool prompted participants to consider the intended outcomes of APPs at their institution and compare them to what they accomplish. Thus, utilizing the APP Tool led participants to reflect on what self-determination outcomes were being supported by campus programming, as well as what aspects of self-determination were not reflected in their programming.

Second, participants were also prompted to think about the meaning of self-determination, connecting the concept to student developmental needs, and to think strategically about how programs might be adapted or developed to incorporate this focus. Third, in addition to affecting program development, participants stated the APP Tool could be used to make the concept of self-determination explicit to professionals, students, and even parents. Finally, participants also indicated using the tool would foster collaboration, as various campus offices could consider how their APPs do or do not complement each other. As participants considered collaborative use of the tool, suggestions for improving the tool included making it maximally accessible to different audiences.

IMPLICATIONS FOR PRACTITIONERS

The APP Tool may have a variety of uses and implications as noted by Madaus et al. (2020). Though additional studies should be conducted on the APP Tool, this exploratory study provided promising findings and implications. First, the APP Tool can be used at multiple timepoints throughout the year as a way to track progress and gather ongoing data on APPs. Second, it can be used as an evaluation and assessment tool for institutions. Specifically, the tool can be used by student affairs professionals to identify and evaluate which current programs are most effective in fostering student self-determination. Highlighting the common challenges students faced can guide professionals to determine whether they are being supported properly by the APPs currently in place. To address financial constraints related to program assessment, the tool could be used to determine which APPs are most cost effective based on related self-determination outcomes. Another implication of the APP Tool involved better informing staff on the importance of self-determination and allowing for cross-program collaboration. During professional development, the APP Tool can be featured to teach staff about self-determination and how it relates to the current campus programs. Additionally, as the APP tool involves perceptions of outcomes each APP addresses, not objective assessments of whether the outcomes were achieved, it may be used as part of a comprehensive outcome assessment process that allows for triangulation across different assessment tools or professionals using them. Utilizing the APP Tool can also allow various campus offices, including student affairs and disability resources, to have coordinated planning and enable the creation of targeted support for students. Including user feedback from higher education professionals was beneficial and ensured that future iterations of the tool will be most helpful to those end users.

IMPLICATIONS FOR RESEARCHERS

The information collected from the three focus groups not only provided insights regarding how professionals could use the APP Tool, but also informed its future development. Additional trials in which postsecondary education professionals explore and reflect on the APP Tool may produce further insights. Feedback should also be sought from higher education students to ensure that the tool is meeting their needs. Researchers can continue to explore making revisions to the tool to increase its usability to a wider group of individuals. Beyond direct implications for future iterations of the APP Tool, these findings highlighted the overlap between student affairs and disability resource professionals. Specifically, members from both groups discussed the need to support development of self-determination skills in the students they serve. Given student affairs and disability resource professionals share common goals, future research might examine collaboration between these two groups as a way to reach and support more students. The broader discussion of activities, policies, and programs at institutions of higher education that emerged from this study has implications for school policy by noting the inaccessibility embedded into higher education. Future research could continue to explore barriers to student success at the college or university level to ensure that students from all disadvantaged groups have an opportunity for success in postsecondary education.

LIMITATIONS

Although the study was conducted using rigorous qualitative analysis, some limitations are still present. The focus groups were intentionally conducted by different APP Tool authors; however, using a variety of moderators may have introduced some variability to the focus group procedures and questions. To mitigate procedural differences, a common PowerPoint slideshow was used between moderators during each focus group. Second, participant demographic information beyond type of institution where they were located was not collected, therefore, we could not comment on participants gender, race, ethnicity, age, or other characteristics. Third, this study included a convenience sample, which means participants attending sessions on the topic were offered the opportunity to participate in the focus group. Therefore, given that this was an exploratory study, a more representative sample could not be established at this point in time. Fourth, potential limitations when using focus groups, including the current project, are “the tendency for certain types of socially acceptable opinions to emerge” (Smithson, 2000, p. 116), as well as the possibility of certain participants dominating the conversation and research process. Fifth, the focus groups only included the professionals who would be utilizing this tool, and no student feedback, which should be a focus of future studies, as challenges students faced are the focus of this Tool. For example, the challenges that professionals listed may be different from the challenges experienced by students. Sixth, the data was coded only by two researchers and an additional party was not consulted, which would have provided additional perspectives on the data that the two researchers may not have. Finally, feedback on the APP Tool was provided after participants heard about and examined the tool. They did not have the chance to put the tool into practice, thus limiting some reflection.

CONCLUSION

Self-determination may present a useful framework with which to guide campus programming and foster student-development. The APP Tool was therefore created to support student affairs professionals to connect activities, programs, and policies (APPs) with common self-determination related challenges that college students may face (Madaus et al., 2020). This study contributed to the self-determination and postsecondary education research literature as it gathered and analyzed data regarding the usability of the APP Tool from the perspective of higher education professionals. The findings of this study supported that the APP Tool has the potential to shape campus programming and promote a campus-wide focus on self-determination for all college students including SWD.

REFERENCES

- Biddix, J. P. (2018). *Research methods and applications for student affairs*. Jossey-Bass.
- Bogdan, R., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and method (5th edition)*. Allyn & Bacon.
- Breen, R. L. (2006). A Practical Guide to Focus-Group Research. *Journal of Geography in Higher Education*, 30(3), 463-475. <https://doi.org/10.1080/03098260600927575>
- Faye, C., & Sharpe, D. (2008). Academic motivation in university: The role of basic psychological needs and identity formation. *Canadian Journal of Behavioural Science / Revue Canadienne des sciences du comportement*, 40(4), 189-199 <https://doi.org/10.1037/a0012858>
- Field, S., Sarver, M. D., & Shaw, S. F. (2003). Self-determination: A key to success in postsecondary education for students with learning disabilities. *Remedial and Special Education*, 24(6), 339-349. <https://doi.org/10.1177/07419325030240060501>
- Field, S., Martin, J., Miller, R., Ward, M., & Wehmeyer, M. (1998). Self-determination for persons with disabilities: A position statement of the Division on Career Development and Transition. *Career Development for Exceptional Individuals*, 21(2), 113-128. <https://doi.org/10.1177/088572889802100202>
- Fleming, A. R., Oertle, K. M., & Plotner, A. J. (2017). Student voices: Recommendations for improving postsecondary experiences of students with disabilities. *Journal of Postsecondary Education and Disability*, 30(4), 311-328.
- Gelbar, N. W., Madaus, J. W., Dukes III, L. L., Faggella-Luby, M., Volk, D. T., & Monahan, J. (2020). Self-determination and college students with disabilities: Research trends and construct measurement. *Journal of Student Affairs Research and Practice*, Advance online publication. <https://doi.org/10.1080/19496591.2019.1631835>
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.
- Graham, M., & Vaughn, A. (2022). An adapted self-determination measure and college student first-year achievement. *International Journal of Teaching and Learning in Higher Education*, 33(2), 135-142. <https://files.eric.ed.gov/fulltext/EJ1345545.pdf>
- Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, 29(1), 3-22. <https://doi.org/10.1177/1525822X16639015>
- Guiffrida, D. A., Lynch, M. F., Wall, A. F., & Abel, D. S. (2013). Do reasons for attending college affect academic outcomes? A test of a motivational model from a self-determination theory perspective. *Journal of College Student Development*, 54(2), p. 121-139. <https://doi.org/10.1353/csd.2013.0019>

- Ju, S., Zeng, W., & Landmark, L. J. (2017). Self-determination and academic success of students with disabilities in postsecondary education: A review. *Journal of Disability Policy Studies*, 28(3), 180-189. <https://doi.org/10.1177/1044207317739402>
- Krueger, R. A. (1994). *Focus groups: A practical guide for applied research*. Sage Publications Inc.
- Lalor, A. R., Madaus, J. W., Newman, L. S. (2020). Leveraging campus collaboration to better serve all students with disabilities. *Journal of Postsecondary Education and Disability*, 33(3), 249-255.
- Madaus, J., Gelbar, N., Dukes III, L. L., Faggella-Luby, M., Mills, D., & Taconet, A. (2020). Using the APP Tool to promote student self-determination skills in higher education. *Journal of Postsecondary Education and Disability*, 33(3), 265-273.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey Bass.
- Mills, D., Faggella-Luby, M., Gelbar, N., Madaus, J., & Dukes III, L. L.. (2019). *The APP Tool*. Fort Worth, TX.
- Murphrey, T. P., Rutherford, T. A., Doerfert, D. L., Edgar, L. D., & Edgar, D. W. (2014). An examination of usability of a virtual environment for students enrolled in a college of agriculture. *Journal of Agricultural Education*, 55(4), 38-52. <https://doi.org/10.5032/jae.2014.04038>
- National Center for Education Statistics (NCES). (2022, April 26). A majority of college students with disabilities do not inform school, new NCES data show. https://nces.ed.gov/whatsnew/press_releases/4_26_2022.asp
- Newman, L. A., & Madaus, J. W. (2015). Reported accommodations and supports provided to secondary and postsecondary students with disabilities: National perspective. *Career Development and Transition for Exceptional Individuals*, 30, 173-181. <https://doi.org/10.1177/2165143413518235>
- Ortiz, A. M., & Waterman, S. J. (2016). The changing student population. In G. S. McClellan & J. Stringer (Eds.), *The Handbook of Student Affairs Administration* (4th ed., pp. 267-285). Jossey-Bass.
- O'Shea, A., & Meyer, R. H. (2016). A qualitative investigation of the motivation of college students with nonvisible disabilities to utilize disability services. *Journal of Postsecondary Education and Disability*, 29(1), 5-23.
- Parker, D. R. (2004). *Voices of self-determined college students with ADHD: Undergraduates' perceptions of factors that influence their academic success*. [Doctoral dissertation, University of Connecticut]. Digital Commons @ UConn. <https://digitalcommons.lib.uconn.edu/dissertations/AA13156407>
- Parker, A., & Tritter, J. (2006). Focus group method and methodology: current practice and recent debate. *International Journal of Research & Method in Education*, 29(1), 23-37. <https://doi.org/10.1080/01406720500537304>

- Petcu, S. D., Van Horn, M. L., & Shogren, K. A. (2017). Self-determination and the enrollment in and completion of postsecondary education for students with disabilities. *Career Development and Transition for Exceptional Individuals*, 40(4), 225-234. <https://doi.org/10.1177/2165143416670135>
- Sangster, S. L., Loy, K. L., Mills, S. D., & Lawson, K. L. (2016). Engaging first-year university students in research: Promise, potentials, and pitfalls. *Canadian Journal for the Scholarship of Teaching and Learning*, 7(1). http://ir.lib.uwo.ca/cjsotl_rcacea/vol7/iss1/3
- Sarver, M. D. (2000). *A study of the relationship between personal and environmental factors bearing on self-determination and the academic success of university students with learning disabilities*. [Doctoral dissertation, University of Florida]. ProQuest Dissertations and Theses Global. <https://www.proquest.com/docview/304596688?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses>
- Schalk, S. (2017). Critical Disability Studies as Methodology. *Lateral*, 6(1). <https://doi.org/10.25158/L6.1.13>
- Smithson, J. (2000). Using and analyzing focus groups: Limitations and possibilities. *International Journal of Social Research Methodology*, 3(2), 103-119. <https://doi.org/10.1080/136455700405172>
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237-246. University of California Berkley. (n.d.). Autism spectrum services. https://dsp.berkeley.edu/support_services/autism-spectrum-services
- University of Illinois at Urbana-Champaign. (n.d.). Strategies. <https://www.disability.illinois.edu/strategies>
- University of Colorado Boulder. (n.d.). Academic coaching. <https://www.colorado.edu/disabilityservices/academic-coaching>
- Weiss, M. P. & Faggella-Luby, M. (2020). *Transition services for students with learning disabilities*. In J. P. Bakken and F. E. Obiakor (Eds.). *Transition services for students with disabilities* (Vol. 35, pp. 15-34). Emerald Group Publishing Limited.
- Wehmeyer, M. L., & Agran, M., Hughes, C., Martin, J. E., Mithaug, D. E., & Palmer, S. B. (2007). *What works for special-needs learners. Promoting self-determination in students with developmental disabilities*. Guilford Press.
- Williams, A., & Katz, L. (2001). The use of focus group methodology in education: Some theoretical and practical considerations. *IEJLL: International Electronic Journal for Leadership in Learning*, 5(3), 1-10. <https://journals.library.ualberta.ca/iejll/index.php/iejll/article/view/496/158>

APPENDIX A

COMPONENTS AND EXPLANATIONS OF SELF-DETERMINATION

Self-Determination Component	Brief Explanation
Choice-making skills	The ability to identify and select a preferred activity or item from several options without coercion.
Decision-making skills	The use of a process to determine a preferred solution based on a list of relevant action alternatives and with consideration of overall risk.
Problem-solving skills	A process of identifying a solution to resolve a quandary in which response alternatives are identified, selected and verified often through self-instruction.
Goal-setting and attainment skills	Developing a plan to accomplish a targeted behavior or outcome (distal or proximal) through self-regulated behaviors and with regard to consequences of actions and contingencies of an environment.
Independence, risk-taking, and safety skills	Recognition of the individual, acting within an environment of consequences mitigated by assurances or boundaries.
Self-observation, Self-awareness, or Self-monitoring skills	Involves the individual observing his or her own behavior toward identifying an inconsistency between what occurs and a target behavior. (A prerequisite to self-regulation.)
Self-evaluation skills	The specific identification of an inconsistency between what occurs and a target behavior.
Self-reinforcement skills	Rewarding oneself for matching an observed behavior to a target behavior.
Self-instruction skills	Vocalized performance guidance by oneself to direct action toward a targeted behavior; Viewed as a critical step in problem solving.
Self-regulation skills	The ability of the individual to carry out the vocalized performance guidance to direct action toward a targeted behavior.
Self-advocacy and leadership skills	An individual's ability to effectively communicate or assert a desired outcome, often related to achieving specific goals.
Positive attributions of efficacy and outcome expectancy skills	Noting that an outcome or problem may be adequately addressed by response alternatives directed by the individual.

APPENDIX B

MODEL APP FORM

The APP Tool		
Common Challenges	Activities, Programs, Policies	SD Outcomes
Self-Determination (SD) Components		
1. Choice-making skills 2. Decision-making skills 3. Problem-solving skills 4. Goal-setting & attainment skills 5. Independence, risk-taking, and safety skills 6. Self-observation, self-awareness or self- monitoring skills	7. Self-evaluation skills 8. Self-reinforcement skills 9. Self-instruction skills 10. Self-regulation skills 11. Self- advocacy & leadership skills 12. Positive attributions of efficacy and outcome expectancy skills	

FOUNDING BOARD MEMBERS

Caitlin Blacksmith

Research Methods, Measurement, and Evaluation Ph.D. Student

Luis Orione Ferreria

Giftedness, Creativity, and Talent Development Ph.D. Student

Brenna Fitzmaurice

School Psychology Ph.D. Student

Gina Norman

School Psychology Ph.D. Student

Sandra Sears

Educational Psychology/Special Education Ph.D. Student

Vaishnavi Sivaprasad

Educational Psychology Ph.D. Student

Ani Terterian

Higher Education and Student Affairs Master's Student

Sierra Trudel

MA, SYC, NCSP, LPC: School Psychology Ph.D. Student

Jennie Weiner

Neag School of Education Faculty Advisor

Emily Winter

PhD, NCSP: School Psychology Ph.D. Alumna

UConn

NEAG SCHOOL OF EDUCATION

education.uconn.edu/neag-journal