

NUMERACY PROXIES AND PRACTICES: STUDIES IN
APPROXIMATIONS OF THE “REAL”

By

Samuel Luke Tunstall

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics Education—Doctor of Philosophy

2019

ABSTRACT

NUMERACY PROXIES AND PRACTICES: STUDIES IN APPROXIMATIONS OF THE “REAL”

By

Samuel Luke Tunstall

Whether viewed as an ability or a social practice, numeracy centers on what we as humans do with numbers. Given that we are inundated each day with quantitative information in multiple ways and in various arenas of our lives, numeracy is an important construct to study. In this dissertation, I pursue questions related to a tension raised in recent literature: the tension between numeracy as an ability (i.e., something that one has), and numeracy as a social practice (i.e., something that one does). Over the course of three related studies, I ask the following questions: To what extent are numeracy practices captured through processes of measurement or quantification? How do we talk about the *impact* of numeracy based on measurements of it? Finally, how are numeracy practices present in students’ engagement with public issues, and in relation, how might we attend to numeracy practices in the context of general education mathematics at the postsecondary level? I explore the first question more specifically through a validity analysis of a widely known international assessment of numeracy, the Programme for the International of Adult Competencies (PIAAC).

The key finding from the first study was that while PIAAC numeracy scores may be valid for the use of describing proficiency distributions of specific population subgroups, the construct of interest—real-life numerate behavior—is

not what is measured by the instrument. I explore the second question through a critical discourse analysis of a document describing results stemming from the PIAAC. Findings from the second study reveal that authors of the chosen document used micro-linguistic moves to construct a specific relation between numeracy skills and well-being—a relation that is not justified given the authors' data, yet rendered coherent in that it is situated within broader narratives about the influence of skills on well-being. In the third study, I analyze how college students reasoned with public issues in the context of a focus group, finding that students' responses were related to their beliefs and experiences, and that numeracy events—though present—were primarily centered around students' acknowledgements of the importance of numbers, more so than their active articulation in building on that importance through group conversation. I discuss implications from each of these studies in the context of numeracy assessment and numeracy education.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of numerous individuals over the course of my lifetime. I first wish to thank my Dissertation Committee of Tonya Bartell, Lynn Fendler, Beth Herbel-Eisenmann, and Vincent Melfi, for their constructive engagement with my work. Additionally, I thank numerous other individuals involved in the Program in Mathematics Education, including Freda Cruél, Kelly Fenn, and Lisa Keller, for their consistent support in relation to events, logistics, program requirements, and the presence of a puppy in Kedzie Hall. Next, I want to acknowledge the broader education community at Michigan State University, as well as my program cohort—Younggon Bae, Chris Dubbs, Michael Gundlach, Kate Knowles, Molade Osibodu, and Kevin Voogt—for their unconditional love and support since the day we began in Fall 2015. Finally, I acknowledge my family, including those both alive and no longer with us. To my mother and father, brothers, sisters-in-law, nieces and nephews, grandparents, aunts and uncles, and cousins: thank you. I cannot fathom having completed this journey without you.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
Introduction.....	1
A Milieu for Exploring Numeracy	1
Curiosities and Positionality.....	5
Study One: Validity Analysis of the PIAAC’s Numeracy Component	8
Study Two: Critical Discourse Analysis of Relational Links in <i>Skilled for Life?</i>	10
Study Three: College Students’ Numeracy Events and Discussion of Public Issues in Focus Groups.....	12
REFERENCES	15
Study One: Validity Analysis of the PIAAC’s Numeracy Component.....	20
Introduction: Social Context of this Study	20
Warrants for the Study.....	23
Definition of Concepts	26
Assessing Numeracy	31
Challenges to Numeracy Assessment	33
Protocols for this Study	36
Method	37
Data Sources	37
Analytical Framework	39
Findings	41
Interpreting a Measurement for a Specified Use.....	41
What the PIAAC Numeracy Assessment Measures	41
Uses of the PIAAC Numeracy Assessment.....	47
Interpreting PIAAC Numeracy Scores	49
Supporting Interpretations with Theory and Evidence	53
Assessment Content and Response Processes	55
Relations to Other Variables.....	60
Consequences of the Assessment	62
Discussion and Looking Ahead.....	64
Beyond Valid or Invalid.....	65
Towards Caution and Responsibility	68
APPENDIX.....	71
REFERENCES	74
Study Two: Critical Discourse Analysis of Relational Links in <i>Skilled for Life?</i>	81

Introduction	81
Linking Literacy and Numeracy with Development	86
Method.....	94
Critical Discourse Analysis.....	94
Tools for Studying Discourse and Causation.....	97
Present Study.....	99
Findings: Constructions of Association and Causation in <i>Skilled for Life?</i>	102
Construction at the Thematic and Organizational Levels	105
Thickening Links through Clausal Work.....	109
The Text in Its Broader Milieu.....	115
Talking Back	122
APPENDIX.....	125
REFERENCES	130

Study Three: College Students' Numeracy Events and Discussion of Public Issues in Focus Groups	141
Introduction	141
Numeracy and Public Issues.....	145
Method.....	151
Participants.....	151
Focus Groups	153
Protocol.....	153
Media Artifacts	155
Data Collection	157
Analysis.....	159
Findings	163
Question One: How Students Discussed Public Issues.....	164
Leveraging Prior Experiences.....	164
Asking Questions.....	171
Questions Two and Three: Numeracy Events and Their Characteristics	176
When and Where Numeracy Events Occurred.....	177
Discussion.....	184
Numeracy Events	184
Conclusion.....	188
APPENDICES	192
APPENDIX A: Focus Group Artifacts.....	193
APPENDIX B: Characteristics of Participants' Responses to the Three Artifacts	196
REFERENCES	197

Conclusion	203
Circling Back: A Milieu for Exploring Numeracy	203
In Response to Curiosity	205
Study One: Validity Analysis of the PIAAC's Numeracy Component	205

Study Two: Critical Discourse Analysis of Relational Links in <i>Skilled for Life?</i>	207
Study Three: College Students' Numeracy Events and Discussion of Public Issues in Focus Groups.....	210
Lingering and Emergent Clouds.....	211
REFERENCES	218

LIST OF TABLES

Table 1: PIAAC Numeracy Proficiency Levels.....	51
Table 2: Context Considerations for Sample PIAAC Numeracy Items	57
Table 3: Focus Group Participants.....	153
Table 4: Characteristics of Participants' Responses to the Three Artifacts.....	196

LIST OF FIGURES

Figure 1: Publicly Available Numeracy Item from the PIAAC	45
Figure 2: Example of Callout from <i>Skilled for Life?</i>	107
Figure 3: Example of Formatting to Call out Important Information in <i>Skilled for Life?</i>	108
Figure 4: A Graph that Has the Potential for Construing Causality	113
Figure 5: A Graph that Suggests an Association Between Variables Derived from Skills and from an Outcome of Well-being	114
Figure 6: A Snapshot of Alexa’s Annotation of the <i>Mother Jones</i> Article on Glyphosate in Foods	164
Figure 7: An Excerpt of C’s Annotation of the <i>Mother Jones</i> Article on Glyphosate in Foods	166
Figure 8: A Portion of the Annotations from Layla and Jayla of the Beginning of the <i>Mother Jones</i> Article.....	173
Figure 9: An Excerpt of Pashiel’s Annotation of the <i>Fox News</i> Article.....	175
Figure 10: Snapshot of Savannah’s Annotation of the <i>Mother Jones</i> Article	178
Figure 11: Snapshot of Pashiel’s Annotation of the <i>Mother Jones</i> Article	178
Figure 12: Snapshot of Layla’s Annotation of the <i>Mother Jones</i> Article.....	178
Figure 13: Alexa’s Annotation of the <i>Fox News</i> Article	180
Figure 14: Alexa’s Annotation of the <i>Fox News</i> Article	180

Introduction

A Milieu for Exploring Numeracy

It has been said that “The world of the twenty-first century is a world awash in numbers” (Steen 2001, 1). Or, as described through a similar metaphor in a now-famous article from *The Economist*, that we live under an unending deluge of data (“Leaders: The Data Deluge” 2010). The implications of this inundation are vast, and one of particular interest is how our practices in relation to numbers and data shift as new technologies emerge with increasing frequency (Craig, Mehta, and Howard 2019).

Indeed, though the deluge described in the 2010 *Economist* article began before the new millennium (cf. Crowther 1959, 270), our existence in a world awash in data is only nascent relative to the timespan of human history. In particular, the prevalence of quantitative thinking—as manifest in political debates, advertisements, or daily conversation, among other outlets—co-emerged with the rise of industry and colonialism over the past four centuries, only in the last century to figure so prominently within modern discourse at large (Cohen 1999; Crosby 1997; Porter 1995). In light of the rapid naturalization of quantification in our lives, we have reached a point now where, just as literacy scholars are able to detail the diverse and multifaceted ways in which individuals and communities act on, create, or adopt written and spoken texts in their lives (e.g., Heath 1983; Scribner and Cole 1981), we see scholars examining numeracy practices as a unique set of practices

worthy of characterization on their own (e.g., Craig and Guzmán 2018; Oughton 2018).

In accordance with these changes, curricula in mathematics education at both the secondary and postsecondary levels are shifting to meet the ostensible needs of students who interact with numbers and who use quantification as part of their personal and professional lives (e.g., Franklin et al. 2007; National Council of Teachers of Mathematics 2018; Steen et al. 2001). Though labeled in different ways, the notion of using mathematics or quantification for its utility is often called numeracy, quantitative literacy, or quantitative reasoning (Karaali, Hernandez, and Taylor 2016).¹ And coupled with the emergence of these constructs is a desire to ensure that all have access to them. Accordingly, in acting simultaneously with curriculum stakeholders, separate but related entities such as (for example) the College Board or the Organisation for Economic Co-operation and Development (OECD) aim to measure, sort, and surveil—that is, to “govern softly” (Kanes, Morgan, and Tsatsaroni 2014; Ozga 2012)—nations, teachers, and students, to ensure that they meet the needs of a global society awash in numbers. My dissertation lives within and builds from this social milieu.

In particular, I situate this dissertation at the confluence of three major ideas alluded to above: *numeracy*, *practices*, and *measurement*. Because the population I

¹ Where relevant, I will delineate distinctions among the meanings of those terms.

have worked with the most in the past is college students (discussed in further detail in the next section), I foreground that context as I focus on those three ideas. Note that numeracy has traditionally been defined as facility in mathematical skills for use in everyday life—a functional analogue to literacy (Cockcroft 1982, as cited in Karaali, Hernandez, and Taylor 2016). The term practices comes from literacy studies, and conveys the patterned (or regular) ways in which individuals take up literacy in their lives (Barton and Hamilton 2000). One can mix the words *numeracy*, *practices*, and *measurement* in various combinations to describe the curiosities that guided this dissertation. In their most basic form, though, questions I consider over the course of these three studies are: To what extent are numeracy practices captured through processes of measurement or quantification? How do we talk about the impact of numeracy based on measurements of it? How are numeracy practices present in students’ engagement with public issues, and in relation, how might we attend to numeracy practices in the context of general education mathematics at the postsecondary level? Another way to frame the first two questions is to refer to the measurement or quantification as a proxy, where a proxy for a concept is a substitute, or stand-in, for that concept. The questions become: To what extent are numeracy practices they captured through proxies for numeracy? How do we talk about the impact of numeracy based on numeracy proxies? There are more nuances to the specific questions I ask, and notably, the questions are phrased so broadly as to be unanswerable in a dissertation alone.

Nonetheless, the general curiosities I have are captured well through those questions.

Before describing the three studies that constitute this dissertation, I expand on my positionality in relation to this work. Bias is an inherent part of any research endeavor, regardless of the methods (e.g., quantitative, qualitative, etc.) a researcher or research team employs. Despite its negative connotation, bias need not be considered a signal of inferior research (Fairclough 2003). Insofar as social research is not done in a laboratory where factors may be controlled and conditions scrutinized, Dean (2017) notes there is a general expectation in social research that one acknowledge their “humanness” (1). Among other things, this process entails bringing to the fore one’s positionality so as to acknowledge how their personal experiences inform their practices as a researcher, and ultimately their interpretations and conclusions. Foote and Bartell (2011) argue that within mathematics education research in particular, “a field that often keeps invisible the relational aspects of mathematics teaching, learning, and research,” it is imperative that we foreground our positionalities and their relation to the work that we do (65). Doing so is not a move to reduce bias, but to enrich the experience for readers of our scholarship, as well as to move closer towards scholarship that is just. With those points in mind, I briefly share here how I arrived at this work, and how (to my understanding) I believe my background informs the questions that I ask and the interpretations that I make.

Curiosities and Positionality

A starting point for my connection to this work is that I am a White male from a rural area in North Carolina. I attended Title I schools throughout my K-12 education, and came to excel in mathematics (in my view) in large part from having the same teacher in all four years of high school, as well from being a White male working within in a discipline historically (and to this day) dominated by White males (Martin 2009). This teacher encouraged me to pursue study in the subject and led me to consider education as a career path. After obtaining an undergraduate degree in mathematics, I decided that I wanted to teach in the community college context. Subsequently, as a master's student in mathematics, I had the opportunity to work with college students by teaching College Algebra at Appalachian State University and Mathematical Modeling at Wilkes Community College. The large majority of students in the sections that I taught were taking those courses as their final mathematics course to fulfill a general education mathematics requirement; at those institutions in particular, the requirement was specifically labeled a quantitative literacy requirement. In my perception, nearly all of my students were White; slightly more than half were female (I do not know if that is how they actually identified). I sensed dissonance between the content I taught in those courses and the goal of quantitative literacy as operationalized by Steen and colleagues in *Mathematics and Democracy: The Case for Quantitative Literacy* (Steen 2001). To Steen and colleagues, quantitative literacy was defined as the

ability and disposition to work with numbers as they manifest in various contexts in life. At the time, I felt that students “needed” something different than the skills typically found in College Algebra—they “needed” to be quantitatively literate. I did not realize at the time that I was still viewing students through a deficit lens.

That tension led me to pursue the present degree in mathematics education at Michigan State University (MSU), where I have had the privilege of working with various faculty, staff, and students in developing coursework in quantitative literacy over the last four years. Since coming to MSU, my view of seeing students as “in need” has shifted. In many ways, this shift has occurred because I now recognize the hegemony that mathematics has in compulsory schooling (Greer and Mukhopadhyay 2012), and that there are myriad ways in which one can flourish, not all of which involve being skilled in traditional areas of mathematics such as algebra, geometry, and calculus (Tunstall and Ferkany 2017). Furthermore, in my teaching, I have found that the ways my students think about and approach real-world contexts often differs from how I pose or broach them in formal assignments like labs or quizzes. These observations, coupled with subsequent learning in readings and coursework, have led me to view numeracy as a social practice. Though I recognize why constructs such as numeracy and quantitative literacy have traditionally been operationalized as an ability to use mathematics, and my continued involvement with the Mathematical Association of America allows me to converse with others who have that perspective regularly, I am committed to

viewing numeracy and quantitative literacy through a social practices lens. Among other things, doing so allows me to focus on what students *do* and how to build with them from that starting point. I believe in the brilliance of all students, and will strive to acknowledge and embrace that brilliance in my scholarship and teaching. Where relevant, I expand on the distinction between perspectives of functional and social practices to numeracy and quantitative literacy in subsequent portions of this dissertation.

I end this section by noting that my commitment to viewing numeracy as a social practice has led me to interrogate curricula and policies that position students as in need of quantitative skills. One will see that this penchant for interrogation infuses all three studies that constitute this dissertation. The particular topic that I chose to pursue in this dissertation was the tension between numeracy practices and proxies for numeracy—a tension summarized succinctly (if imperfectly) as that between numeracy as it is practiced and numeracy as it captured in some form of an assessment. The questions that I ask are informed by a wariness of numeracy proxies, which stems from my interactions with students in quantitative literacy classrooms and my anecdotal observations of various assessments of numeracy both in my classroom and in other places (e.g., other instructors' classes, online assessments). At the same time that I intended to use this dissertation as a space for generating critique, though, I was genuinely driven by a curiosity about the relationship between proxies and practices. Understanding more about this

relationship might serve to improve assessments of numeracy, as well as how we design curricula that attend to students' existing numeracy practices. I hope that this curiosity shines through the studies that I report here.

Study One: Validity Analysis of the PIAAC's Numeracy Component

The first study of this dissertation tackles the tension between numeracy practices and numeracy proxies through a validity examination of the numeracy portion of a well-known international assessment: the OECD's Programme for the International Assessment of Adult Competencies (PIAAC). As described earlier, this dissertation work sits at the intersections of numeracy, practices, and measurement. A validity exploration of the PIAAC's numeracy assessment forces one to grapple with all three ideas insofar as the assessment serves as a proxy for the construct of numeracy

Note that the PIAAC is an offshoot of another OECD exam, the Programme for International Student Assessment (PISA), but differs from PISA in that it is primarily aimed at individuals aged sixteen to sixty-five, rather than fifteen year-olds—the sole group taking part in PISA. Both assessments include a portion designed to measure students' numeracy;² all of the assessment items require selected or constructed responses, with the constructed responses being limited to numerical input only. Building from the social practices approach to numeracy, as

² As noted by Gal and Tout (2014), PISA test developers uses the term *mathematical literacy*, but operationalize the construct in the same way that numeracy is operationalized in PIAAC.

well as previous theoretical critiques of PIAAC, I am guided in this study by a curiosity concerning how test developers operationalize the construct of numeracy, and how they account for (or not) what an assessment with those characteristics (e.g., questions with closed inputs) is able to capture. My aim is to explore what is maintained and what is lost when attempting to measure numeracy through the PIAAC's approach. Understanding and calling attention to what is gained and what is lost is important if one is to report on results stemming from the assessment. For example, it would be important to know how numeracy is defined and measured before making claims about an individual or nation's "numeracy."

To complete my validity examination, I draw from the literature on validity in educational measurement, leaning heavily on the field's *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), which delineates the complexity of validity in testing, and provides test developers and users with actionable insights and questions to consider in the development of assessments. The use of their framework for validity led me to the following research questions: (1) What does the PIAAC numeracy assessment claim to measure? (2) What are the intended uses of the assessment? (3) How are we to interpret scores with those uses in mind? (4) To what degree do evidence and theory support interpretations for those uses?

I intend to submit this study to the journal *Numeracy*, as it would be the first paper in the journal to focus specifically on the idea of validity. In light of Vacher's (2014) editorial, "Educational Assessment Is an Enduring Theme of Numeracy," and the foundational importance of validity in educational assessment, I believe the paper will provide a valuable contribution to the field of scholarship pertaining to numeracy.

Study Two: Critical Discourse Analysis of Relational Links in *Skilled for Life*?

As noted above, my guiding curiosity in the first study centered on how test makers of the PIAAC operationalized numeracy, as well as the extent to which their numeracy assessment was valid with respect to their scoring scheme and associated score interpretations. It is important to note that that validity exploration is post hoc; the PIAAC has been administered to various countries in the OECD three times since 2008, with the third round of administration currently underway. Accordingly, the OECD has had ample time to compile and publish reports concerning results from its administration. The second study of this dissertation arises from questions I had in reading the diverse and numerous reports published by the OECD on their report site, the OECD iLibrary.³ In my reading of several of the reports concerning PIAAC (e.g., OECD 2013a, 2013b, 2013c), I sensed that

³ See <https://www.oecd-ilibrary.org>.

authors of the document were positioning numeracy—as measured through the PIAAC's proxy—as causally linked with measures of well-being such as wages and health. What was intriguing to me is that they appeared to make such links without making causal claims directly.

While I recognize that there are indeed material benefits potentially available to those who have certain quantitative skills (e.g., nurses) or quantitatively-demanding degrees (e.g., engineers), the claims I read concerning connections between the constructs of literacy and numeracy with well-being seemed grandiose. Though I had not completed the first study concerning the validity exploration before embarking on the second, I felt that such claims would be unqualified in the context of the validity of the assessment. That is, I hypothesized that those claims might not be justified. Furthermore, they (the claims) were written alongside qualifiers that the nature of the assessment and analysis provided no warrant for causal links. The curiosity I had, then, was the following: How, if at all, did the document position association or causation in relation to numeracy skills and measures of well-being? Pursuing this question in depth involves the study of discourse. Given my positionality in relation to this question being one of critique and interrogation, the study is a critical discourse analysis (Gee 2004). Engaging in this critical discourse analysis is important if one is to call attention to particular ways in which documents attempting to influence decision makers make misleading or misguided claims.

With that in mind, for the second study I use ideas from social semiotics (Morgan 2006) and critical discourse analysis (Fairclough 2003) to engage in the analysis of a specific public OECD document, *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b). I draw heavily from Achugar and Schleppegrell's (2005) analysis of causal construction in history texts to complete the analysis. In addition to examining micro-level features of the text to understand how those features purpose to convey a link between skills and well-being, I also studied the broader socio-political context in which *Skilled for Life* is situated to better understand how the text is rendered coherent.

I intend to submit this study to *Educational Studies in Mathematics*, a journal that publishes research of diverse aims, methods, and topics pertaining to mathematics education, broadly construed. Two studies that I draw from in my review of the literature for this study (Kanes, Morgan and Tsatsaroni 2014; Tsatsaroni and Evans 2014), both appeared in *Educational Studies in Mathematics*, and call for systematic research that interrogates the OECD's testing regime. This second study will continue the conversation they broached while contributing to the broader literature on the construction of association and cause in discourse.

Study Three: College Students' Numeracy Events and Discussion of Public Issues in Focus Groups

Whereas the first two studies of this dissertation center around the OECD's PIAAC as a proxy for numeracy, the final study steps away from numeracy proxies to

consider numeracy practices in relation to how undergraduates think about public issues. In particular, the focus in the third study is on how students reason with public issues in the context of a focus group. At MSU, a team of individuals in the last decade has worked to create two new numeracy-focused courses that help students satisfy the University's mathematics requirement—a requirement intended to "build a foundation for quantitative literacy" (Michigan State University Registrar 2019).

This study builds directly from previous work I have done with colleagues at Michigan State University (MSU), where a team of individuals in the last decade has worked to create two new numeracy-focused courses that help students satisfy the University's mathematics requirement—a requirement intended to "build a foundation for quantitative literacy" (Michigan State University Registrar 2019). As part of that study, we explored in the context of an individual interview (Tunstall, Matz, and Craig 2018). The curiosity that drove that work with colleagues was an interest in how students in these courses already thought about public issues, the idea being that understanding potential characteristics of those processes might better inform our practice as curriculum developers and instructors (see Tunstall et al. 2016). We found that students tended to leverage their background experiences—whether related to ethnicity, age, or religion, among other things—to argue certain standpoints on public issues. Furthermore, only some students, for some issues, used quantitative reasoning as they engaged in the

interviews. Left open was a question of what might happen if the students were in dialogue with others as they engaged with the artifacts, as well as what the nature of numeracy events was (if they occurred). Thus, this study builds from that previous study by having students engage in focus groups (rather than one-on-one with me), by having multiple means of data collection (e.g., written, spoken, and follow-ups by email), and by using the lens of numeracy events and practices in my analysis. The research questions that I ask are: How do the students in my focus groups discuss public issues? Do numeracy events occur as they articulate their reactions? If so, what are the characteristics of these numeracy events? I include eight students across two focus groups in this study.

I intend to submit the results of the third study to *The Journal of General Education*, where colleagues and I published the first study related to this work (Tunstall, Matz, and Craig 2018). Though there are other outlets that could be appropriate for this work, I choose *The Journal of General Education* out of a desire to build upon our prior work in that venue. Reviewers of the first manuscript expressed a desire for more dimensions of analysis, including having students in different courses, and this study will allow for precisely that. As a whole, the contribution of the third study is to further our understanding of numeracy from a practices perspective while contributing to—and even disrupting—ongoing conversations about curriculum and policy for general education mathematics at the postsecondary level.

REFERENCES

REFERENCES

- Achugar, Mariana, and Mary J. Schleppegrell. 2005. "Beyond Connectors: The Construction of Cause in History Textbooks." *Linguistics and Education* 16 (3): 298-318.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Barton, David, and Mary Hamilton. 2000. "Literacy Practices." In *Situated Literacies. Reading and Writing in Context*, edited by Barton, David, Mary Hamilton, and Roz Ivanič, 7-15. London: Routledge.
- Cockcroft, Sir Wilfred H. 1982. *Mathematics Counts. Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr. Wilfred H. Cockcroft*. London: Her Majesty's Stationery Office.
<http://www.educationengland.org.uk/documents/cockcroft/cockcroft1982.html>.
- Cohen, Patricia Cline. 1999. *A Calculating People: The Spread of Numeracy in Early America*. Revised ed. New York: Routledge.
- Craig, Jeffrey, and Lynette Guzmán. 2018. "Six Propositions of a Social Theory of Numeracy: Interpreting an Influential Theory of Literacy." *Numeracy* 11 (2): Article 1.
- Craig, Jeffrey, Rohit Mehta, and James P. Howard III. 2019. "Quantitative Literacy to New Quantitative Literacies." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 15-26. Washington, DC: Mathematical Association of America.
- Crosby, Alfred W. 1997. *The Measure of Reality: Quantification and Western Society, 1250-1600*. Cambridge: Cambridge University Press.
- Crowther, Geoffrey. 1959. *The Crowther Report. A Report of the Central Advisory Council for Education (England)*. London: Her Majesty's Stationery Office.

- Dean, Jon. 2017. *Doing Reflexivity: An Introduction*. Bristol, UK: Bristol University Press.
- Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Foote, Mary Q., and Tonya Gau Bartell. 2011. "Pathways to Equity in Mathematics Education: How Life Experiences Impact Researcher Positionality." *Educational Studies in Mathematics* 78 (1): 45-68.
- Franklin, Christine, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Scheaffer. 2007. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A PreK-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.
- Gal, Iddo, and Dave Tout. 2014. "Comparison of PIAAC and PISA Frameworks for Numeracy and Mathematical Literacy." *OECD Education Working Papers* 102.
- Gee, James Paul. 2004. "Discourse Analysis: What Makes it Critical?" In *An Introduction to Critical Discourse Analysis in Education*, edited by Rebecca Rogers, 19-50. Mahwah, NJ: Lawrence Erlbaum.
- Greer, Brian, and Swapna Mukhopadhyay. 2012. "The Hegemony of Mathematics." In *Opening the Cage*, edited by Greer, Brian, and Ole Skovsmose, 229-48. Rotterdam, The Netherlands: SensePublishers.
- Heath, Shirley Brice. 1983. *Ways with Words: Language, Life, and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Kanes, Clive, Candia Morgan, and Anna Tsatsaroni. 2014. "The PISA Mathematics Regime: Knowledge Structures and Practices of the Self." *Educational Studies in Mathematics* 87 (2): 145-65.
- Karaali, Gizem, Edwin Villafane-Hernandez, and Jeremy Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9 (1): Article 2.
- "Leaders: The Data Deluge." 2010. "The Data Deluge." *The Economist*, February 27, 2010. <https://www.economist.com/leaders/2010/02/25/the-data-deluge>.
- Martin, Danny Bernard. 2009. "Researching Race in Mathematics Education." *Teachers College Record* 111 (2): 295-338.

- Michigan State University Registrar. 2019. "Graduation Requirements for a Bachelor's Degree, Mathematics Requirement (Effective Fall 2019)." <https://reg.msu.edu/AcademicPrograms/Print.aspx?Section=284>.
- Morgan, Candia. 2006. "What Does Social Semiotics Have to Offer Mathematics Education Research?" *Educational Studies in Mathematics* 61 (1-2): 219-245.
- National Council of Teachers of Mathematics. 2018. *Catalyzing Change in High School Mathematics: Initiating Critical Conversations*. Reston, VA: Authors.
- Organisation for Economic Co-operation and Development (OECD). 2013a. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013b. *Skilled for Life?: Key Findings from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013c. *Time for the U.S. to Reskill? What the Survey of Adult Skills Says*. Paris: OECD Publishing.
- Oughton, Helen M. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2.
- Ozga, Jenny. 2012. "Governing Knowledge: Data, Inspection and Education Policy in Europe." *Globalisation, Societies and Education* 10 (4): 439-55.
- Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Steen, Lynn A., ed., and National Council on Education and the Disciplines (NCED). 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: NCED.
- Tunstall, Samuel Luke, and Matthew Ferkany. 2017. "The Role of Mathematics Education in Promoting Flourishing." *For the Learning of Mathematics* 37 (1): 25-28.

Tunstall, Samuel L., Rebecca L. Matz, and Jeffrey C. Craig. 2018. "Quantitative Literacy Courses as a Space for Fusing Literacies." *The Journal of General Education* 65 (3-4): 178-94.

Tunstall, Samuel L., Vincent Melfi, Jeffrey C. Craig, Richard Edwards, Andrew Krause, Bronlyn Wassink, and Victor Piercey. 2016. "Quantitative Literacy at Michigan State University, 3: Designing General Education Mathematics Courses." *Numeracy* 9 (2): Article

Study One: Validity Analysis of the PIAAC's Numeracy Component

Introduction: Social Context of this Study

Proceeding from a catchy title, “U.S. Millennials Post ‘Abysmal’ Scores in Tech Skills Test, Lag behind Foreign Peers,” *Washington Post* columnist Frankel (2015) noted

There was this test. And it was daunting. It was like the SAT or ACT—which many American millennials are no doubt familiar with, as they are on track to be the best educated generation in history—except this test was not about getting into college. This exam, given in 23 countries, assessed the thinking abilities and workplace skills of adults. It focused on literacy, math and technological problem-solving. The goal was to figure out how prepared people are to work in a complex, modern society. And U.S. millennials performed horribly.

Frankel is not the only journalist in popular media to participate in discussions about aggregate results of Americans’ performances on international assessments. Similar headlines, sounding nearly identical alarms about performance, abound in relation to both this exam (e.g., Emanuel 2016; Zinshteyn 2015) and similar ones from the past (e.g., Rice 2009; The National Commission on Excellence in Education 1983). In the particular piece excerpted above, Frankel discusses with an Educational Testing Service (ETS) researcher U.S. millennials’ results from the Programme for the International Assessment of Adult Competencies (PIAAC).

Developed by the Organisation for Economic Co-operation and Development (OECD), the PIAAC is an offshoot of another OECD exam, the Programme for International Student Assessment (PISA). The PIAAC differs from PISA in that

(among other things) it is primarily aimed at individuals aged 16 to 65, rather than 15 year-olds—the sole group taking part in PISA. With data collection completed from 2011-2012, the first administration of PIAAC consisted of a survey of 166,000 adults aged 16 to 65 in twenty OECD member countries (in addition to Cyprus and the Russian Federation); the third administration is currently in progress. Per the OECD, PIAAC “assesses the proficiency of adults from age 16 onwards in literacy, numeracy and problem solving in technology-rich environments,” the motivation being that such proficiencies “are relevant to adults in many social contexts and work situations, and necessary for fully integrating and participating in the labour market, education and training, and social and civic life” (2013b, 5). In addition to testing in literacy, numeracy, and problem-solving in technology-rich environments, respondents also complete a detailed questionnaire, which includes demographic information (e.g., the level of education of one's parents) as well as habits in relation to numeracy, literacy, and one’s general home life.

- The first paragraph of Frankel’s article represents the PIAAC from a particular perspective, one that differs from my own in that in my view, the PIAAC assessment is not necessarily daunting (the test lasts around 60 to 80 minutes, which includes time for the background survey),
- is not readily comparable to the SAT or ACT (the format, the constructs tested, and stakes for test takers are different),

- is taken by few Americans (5,010 people in the 2011-2012 administration), and
- aims to assess the construct of numeracy, rather than that of mathematics (which the test developers distinguish, as I discuss later).

It is not wholly surprising that my view of the PIAAC is different from that of Frankel, and my purpose here is not to admonish or belittle Frankel. Journalists often incorporate influences and perspectives that are different from mathematicians and research scientists when adapting research studies into news products suitable for their respective audiences (Woloshin and Schwartz 2002). Given the task that journalists face in translating complex ideas into bites accessible to a wide audience, it is understandable that these differences in perspective might arise. For example, U.S. readers may not be familiar with the term *numeracy*, but they probably have some familiarity with the term *mathematics*. The substitution in terminology likely does little harm in that context. Indeed, it may be a necessary substitution for the work to be accessible to Frankel's readership. That being said, what I have found surprising, and what partially prompted the study I report on here, is the degree to which interpretations of PIAAC results by PIAAC researchers are valid for proposed uses by the assessment's developers.

That is, though I was not familiar with the concept at the time, I was concerned with the *validity* of the PIAAC numeracy assessment in the context of interpretations such as those from Frankel in the title and body of the article. By

validity, I mean the degree to which interpretations of scores are appropriate for their proposed uses. Are Americans, on the aggregate, actually unprepared to work in a “complex, modern society”?

Warrants for the Study

My rationale for this work stems from two areas: (1) my personal connection to coursework centered around numeracy, and (2) calls for increased interest in assessment. With respect to the first, my personal connection comes from teaching courses centered on quantitative literacy at both two- and four-year institutions. I write about this personal connection, or my positionality (Foote and Bartell 2011), because it inevitably informs the work that I do, regardless of whether I desire for it to. In my teaching, I have found that the ways my students think about and approach real-world contexts often differs from how I pose or broach them in formal assignments like labs or quizzes. A recent example of this disconnect occurred in the 2018 Summer Session at Michigan State University (MSU), when I facilitated a unit on gerrymandering for a course I was teaching, Quantitative Literacy II (see Tunstall et al. 2016 for more information about these courses). A bulk of the unit was on the mathematics of the efficiency gap (Stephanopoulos and McGhee 2015), but that topic—even the YouTube video⁴ associated with it—was not the first thing that arose in students’ beginning-of-class discussions; instead, it

⁴ See <https://www.youtube.com/watch?v=IKtbfVmKM3w> for the video from WNYC.

was voter suppression and proportional representation, the former of which had been a hot topic in the news that month. To subsequently read Frankel's headline not long after those conversations, which suggests that Americans' numeracy scores are abysmal, yielded dissonance for me. I saw promise, not deficit, in students' discussions about voting and representation. Students were engaged with the material, and ready to learn about the efficiency gap. Furthermore, my students were not answering the types of questions sampled in Frankel's article in class, and it was difficult to imagine them answering many of them in any current context—whether in or out of class. This raised a question: millennials performed poorly by what standards?

With respect to my second rationale for this exploration, scholars of numeracy and quantitative literacy have expressed increased interest in assessment in the last decade (Cahoon and Kiliç-Bahi 2019; Vacher 2015). This interest stems from larger movements to assess general education outcomes in higher education (Rhodes 2010), as well as the more specific need to gauge the success of novel programs in numeracy, where success is measured by the extent to which (in this case) college graduates are able to demonstrate behaviors and attitudes aligned with—that is, are valid proxies for—what has been defined as numeracy or quantitative literacy. As scholars, our ability to make claims based on an assessment is contingent upon the validity (i.e., alignment of purposes) of that assessment (American Educational Research Association, American Psychological Association, & National Council

on Measurement in Education 2014). While some in the field have alluded to the importance of validity in developing assessments for numeracy (e.g., Gaze et al. 2014), to date there has been no holistic consideration of the validity of a numeracy assessment—that is, the consideration of more than just one facet of validity (in the example case of Gaze et al. 2014, content validity). Until the mid- to late twentieth century, validity was viewed through multiple lenses, or multiple types of validity. These types included (among others) content validity, criterion validity (consisting of predictive and concurrent validity), and construct validity. Insofar as validity is now viewed from a broader lens than just one of a specific type of validity, and the justifiable use of an assessment is contingent upon a foundation of validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), this paper provides an example of what the validation process might look like as we consider the types of claims we can make from an assessment.

My work is informed by a social practices view of numeracy (Craig and Guzmán 2018; Oughton 2018) demonstrating that a social theory of numeracy need not be in opposition with epistemological expectations for rigor and method expected by many individuals in the educational research community (e.g., Scheaffer 2008; Shulman 1981). Working from these rationales, I embarked on a post hoc validity study of the numeracy portion of the PIAAC, using an argument-based approach to validation (American Educational Research Association,

American Psychological Association, and National Council on Measurement in Education 2014; Kane 2012). In particular, I discuss validity and the validation process from an external standpoint of the PIAAC, raising questions and considerations for developing, implementing, and reporting on their assessments related to numeracy. Further, To those ends, I begin by discussing a definition of validity, and then discuss assessments of numeracy. I then transition to the PIAAC, and a discussion of the validity of PIAAC interpretations in light of the test developers' proposed uses. I end with implications and a call for future work in relation to validation and numeracy assessments.

Definition of Concepts

Prior to exploring validity in relation to the PIAAC numeracy assessment, it is important to have a foundation for what validity is. I begin this section with that grounding discussion. The definition I adopt, and that I will explain in further detail below, is that validity refers to “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of test scores. (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014, 11).

As an adjective, *valid* is a relative term insofar as it raises questions of: Valid to whom? Valid with respect to what? And valid by what standard(s)? For example, the declaration, “She brings up a valid point,” bears little meaning without knowing

more about the conversants, the referent for any claim of validity, or the backdrop of their conversation. Even with that information, the extent to which one might agree with the proposition that someone's point is valid, will vary. For example, one person may regard a point as valid because they agree with it; another person may regard a point as valid because it is factually demonstrable; yet another person may regard a point as valid because it is clear and easy to understand. In each case, the assessment of validity is based on a different set of criteria: opinion, fact-checking, or communicative effectiveness. In other words, we cannot make an objective judgment that a test is valid or invalid; rather, we can only make judgments that a given test is more or less valid for which specific purpose, of which version of a construct, or toward what kinds of effects. For this reason, there is no algorithm or criterion or methodology that can serve as a rubric for assessing validity. Rather, judgments of validity are inferences; validity is judged on the basis of inferences about purposes, constructs, and beliefs about what counts as operationalization of any given concept.

Regardless of one's agreement with such a point, *valid* carries with it connotations of power, as it tends to codify a particular thing as sound, as fact, or as knowledge. In the Foucauldian (1980) sense, it signifies to us that something is True (note the capital T). Albeit some scholars dismiss the pursuit of validity in scientific research (Gergen and Gergen 2000; Lather 1993; Wolcott 1990), the characteristic is widely used in the field of educational measurement, where validity

refers to the alignment between what a test measures and what that test claims to measure.

With roots among psychologists studying intelligence and cognition more broadly (e.g., Terman et al. 1915; Thorndike 1916), the meaning of validity and process of assessment validation has evolved significantly over the past century, from purely statistical validations of assessments (e.g., factorial validity) to checks of differing types of validity (e.g., content validity, predictive validity), among other approaches (Sireci and Sukin 2013). Today, though there is still debate (Newton and Baird 2016), validity largely centers on how well a test measures what it claims to measure (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014; Kane 2012; Newton 2012). That is, rather than breaking validity into constituent parts, validity is a unitary concept that “refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of test” (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014, 11). In this way, validity is not broken into a binary of valid/invalid. This is because, regardless of the construct of interest, once we move from construct definition to its operationalization in an assessment, perfection is not feasible. Validity of a given assessment, then, falls along a spectrum of persuasion.

Authors of the *Standards for Educational and Psychological Testing* (a book hereafter referred to as the *Standards*) synthesize perspectives on what counts as persuasion and provide guidance for individuals seeking to validate an assessment (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014). Per the *Standards*, there are five categories of evidence one might draw from (i.e., infer) for validation. These categories address:

1. Assessment content (the extent to which an assessment aligns with the construct of interest),
2. Response processes (test takers should engage with the assessment in ways test developers and the construct anticipate),
3. Internal structure (if some aspects of the construct are to be distinguished, or if the test is to function differently for different groups, there should be evidence for these patterns),
4. Relations to other variables (if the construct of interest relates to external variables, or if construct performance is to generalize to other contexts, evidence should support those propositions)
5. Consequences of the assessment (benefits of the assessment should outweigh its consequences).

It is jointly incumbent on the test maker and test user to provide combinations of these sources of evidence when validating their assessment.⁵ The authors of the *Standards* establish this imperative early on, stating that “Evidence of the validity of a given interpretation for a specified use is a necessary condition for the justifiable use of the test” (11). Similarly, Kane (2012) notes: “If a lot is being claimed, a heavy ‘burden of proof’ is imposed on those making the claims” (70). That being said, there is no combination of these five sources that produces a valid assessment. The validation process varies based on inferences about the assessment itself, the meaning assigned to its outcomes, and the potential use of such outcomes. For example, the validation process of a university’s mathematics placement exam will be different if exam score interpretations are taken as suggestions versus if they rigidly influence a student’s course options; the validation of the same exam will be different yet if the construct of interest is quantitative literacy versus mathematical literacy. We do not talk about the validity of the assessment itself, but rather the validity of the assessment within the broader milieu in which it is administered.

To summarize, then, the validation process for an assessment is contingent upon a variety of factors, including what the test purports to measure, how scores are interpreted, and what the consequences are of such interpretations. The five

⁵ Note the developer and user may be the same individual or collective.

evidence sources discussed above collectively contribute to the justification of proposed interpretations for proposed uses. Later, I will revisit the five evidence sources above in discussing my external validation of the PIAAC numeracy assessment. Note that I use the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014) as the guiding framework for validation, rather than derivative frameworks like Evidence-Centered Design (Mislevy and Haertel 2006) that specify a means of validation, as the *Standards* are broader in scope.

Assessing Numeracy

In the context of quantitative literacy or related constructs, assessment is not a novel concern (Cahoon and Kiliç-Bahi 2019). As a new skill for the twenty-first century, or a new requirement in postsecondary general education programs, quantitative literacy is a construct that administrators, faculty, and policymakers at multiple levels express increasing interest in surveilling. For example, we see this interest manifest in

- the creation of several VALUE rubrics from the Association of American Colleges and Universities, one of which centers on numeracy (Rhodes 2010);
- the recent creation of the HEIghten® assessment of quantitative literacy for postsecondary institutions from the ETS (Roohr et al. 2017);

- a National Science Foundation grant awarded to multiple institutions for the development of a numeracy assessment instrument (Gaze et al. 2014);
- the numeracy assessment on PIAAC, and even a special year of PISA devoted to numeracy (Gal and Tout 2014; Kosko and Wilkins 2011);
- the inclusion of a numeracy domain in the Collegiate Learning Assessment (CLA) (Klein et al. 2007); and
- a special issue on assessment in *Numeracy* (Vacher 2015).

The projects and scholarship listed above represent only a sample of efforts to assess numeracy; they vary in goal, format, funding (or lack thereof), and conceptualization of numeracy, among other things. Regardless of the flourish associated with these assessments—including multi-million dollar funding, white papers, external publications, and uptake in media sources—the *Standards* suggests that results from these assessments have little substantive meaning without accompanying discussions of validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014). As I will argue below, assessments of numeracy (operationalized through a competency perspective) are especially tenuous, as the setting and assessment itself fundamentally obfuscate the construct of interest. An implication of this proposition is that—as numeracy researchers and scholars—we should be particularly demanding in thinking through the validation process of assessments we develop.

Challenges to Numeracy Assessment

Assessments of any construct are necessarily only proxies for that construct, unless those assessments are practical, real-life, real-time engagements. Scholars developing written assessments involving the construct of numeracy face a special hurdle to the first source of evidence in the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), in that the construct nearly always addresses some notion of the real world that is somehow separate, spatially or temporally, from the writing of the definition of numeracy. Similar issues arise in the assessment of constructs such as critical thinking (Rear 2019) or problem solving (Griffin, Care, and Wilson 2018). In contrast, the assessment of skills, such as the ability to graph a rational function or describe the steps of meiosis, is less tenuous, as no claim is made about when and how these skills might manifest. This is not to imply that the development of a numeracy assessment is impossible, because, as noted earlier, validity is not only about construct validity. However, insofar as assessment content feeds into the development of interpretations and proposed uses of test scores, claims of validity require that the content align with interpretations that use language about that construct

Synthesizing the diverse ways scholars have used terms like numeracy, quantitative literacy, and quantitative reasoning, Karaali, Hernandez, and Taylor (2016) converged on a common “thread,” stating that the terms tend to connote “a

competence in interacting with myriad mathematical and statistical representations of the real world, in the contexts of daily life, work situations, and the civic life” (25). As one might imagine, the inherent grounding of the three terms in the “real” differentiates them from other things one might assess, such as the ability to factor a polynomial, where the assessment setting and construct setting (though ambiguous or not provided at all) are likely to align more closely. A host of scholars (e.g., Grawe 2011; Kosko and Wilkins 2011) have discussed this distinction at length, arguing in essence that numeracy assessments with limited response options (e.g., multiple-choice questions, numerical entry questions) fail to capture the essence of the real in numeracy. These scholars suggest that other mediums, such as essays or portfolios (Grawe, Lutsky, and Tassava 2010; Klein et al. 2007; Rhodes 2010; Shavelson et al. 2019; Zerr 2019), are better suited for capturing what one means by numeracy. Though the aforementioned scholars do not take on a social practices perspective of numeracy explicitly, the issue they tackle—that of capturing the real—is explained well through such a perspective.

To expand on this point, we can interrogate the notion of *competence* included in Karaali, Hernandez, and Taylor’s (2016) statement. The inclusion of competence in their thread suggests a functional or skills-based approach to the terms, meaning that, when evidenced through action, numeracy, quantitative literacy, and quantitative reasoning all hinge in some way on some subset of skills (e.g., the ability to convert from a decimal to a percentage). But if the construct we seek to

understand is what it is that people actually *do* with numbers, the definition itself of that action should not hinge on ability. Drawing from scholars largely in the anthropology and literacy studies communities, Oughton (2018), and later Craig and Guzmán (2018), challenged a functional view of numeracy in favor (or acknowledgement) of a practices-oriented view. A practices approach to numeracy views numeracy through the lenses of *practices* and *events*. Craig and Guzmán define numeracy events as events which are mediated in some way by quantification; such events are observable insofar as they "happen," whether mentally or physically. From this, numeracy practices are those patterned (or repeated) things individuals tend to do in numeracy events, coupled with the significance individuals ascribe to such events. Distinct from a functional approach to numeracy, where numeracy is viewed as a set of skills used in context, "A social practice perspective not only takes into account different practical contexts; it also considers how people's life-histories, goals, values and attitudes will influence the way they carry out numeracy" (Oughton 2018, 6). Oughton's remarks are corroborated by a variety of studies in the context of numeracy (Carraher, Carraher, and Schliemann 1985; Kahan et al. 2017; Lave and Wenger 1989; Tunstall, Matz, and Craig 2018) that suggest that skills alone do not dictate the nature of numeracy events.

Indeed, a central benefit of this perspective is that it acknowledges that our actions in the world outside of formal assessments are complex and ill-defined.

Moreover, it disputes any assumption that ability (as measured by a test score) determines action, given that actions are influenced by more than just ability. Hence, if an assessment of numeracy only addresses ability, it raises fundamental questions of validity, that is, whether the test measures what it claims to measure. Though some scholars will circle back to note that, due to a dearth of resources or a desire for efficiency, we are forced to resort to assessments that may be quickly administered and scored (PIAAC Numeracy Expert Group 2009; Shavelson et al. 2019), the analysis here will contribute to conversations about the validity of such an assessment with respect to the interpretations and uses of the assessment. In short, especially when testing policies prioritize expediency, they often marginalize issues of validity in the process. In the analysis that follows, I adopt a social practices view while recognizing that I cannot change the construct that PIAAC developers intended to measure in the numeracy portion of the assessment. This framework for numeracy will manifest when I discuss or assess claims that link scores with action, as well as when I use the term *practices* or *events* to describe particular PIAAC components.

Protocols for this Study

In following the path set out by the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), questions that formally guided this study were

the following: (1) What does the PIAAC numeracy assessment claim to measure? (2) What are the intended uses of the assessment? (3) How are we to interpret scores with those uses in mind? And (4) to what degree do evidence and theory support interpretations for those uses? Though the first three questions require research, the fourth question is the central research question of this study (and invites analysis more so than summary). Taken together, answers to these four questions allow me to talk about the validity of PIAAC numeracy assessment scores with respect to their intended use.

Method

Data Sources. In addition to several analyses of results, the OECD provides various resources for those interested in understanding how the PIAAC numeracy assessment was conceptualized, designed, and then implemented. These sources are available from the OECD's iLibrary, which hosts thousands of books, working papers, policy documents, and data sets, and serves as "the gateway to OECD's analysis and data."⁶ To find documents reporting the PIAAC numeracy assessment, I used the iLibrary's search engine and the terms PIAAC and numeracy, compiling all documents that reported on the conceptualization, design, or implementation of the numeracy assessment. The initial search using the terms PIAAC and numeracy yielded 1,092 results, many of which were not related to what I was searching for,

⁶ See <https://www.oecd-ilibrary.org/>.

so it was necessary to delimit the search to documents (not datasets alone, for example) written in English (some documents in the database are written in French), and then to cull from those results documents that concerned the conceptualization, design, or implementation of the numeracy assessment. If a document referenced a previous OECD-published document to describe any of those elements, I did not include the newer document in the documents that I analyzed. This search process ultimately yielded

- a report from the PIAAC’s Numeracy Expert Group (2009),
- an overarching framework document describing the constructs of interest in PIAAC (OECD 2012),
- a comprehensive “Technical Report” describing the minutiae of the development process (OECD 2016a),
- a *Reader’s Companion* to the PIAAC’s development (OECD 2016b), and
- a detailed *First Results* document from the 2011-2012 administration of the exam (OECD 2013a).

The number of pages in each of these documents, by order of bullet points, was 67, 62, 1,233, 130, and 466 pages. Specifically unavailable to the public, though, are the fifty-six items used in the Numeracy Assessment. The OECD data request team did not grant me private access to the items (despite stating that I would not share them with others). Five of the fifty-six items (reportedly representative of the larger

set) are available to the public through an informal document⁷ on the PIAAC site and a simulation⁸ of the actual assessment.

Analytical Framework. I answer the first three research questions using data from the sources described above. The means by which I analyzed data to answer those questions are discussed in their respective sections below. The fourth research question—that of the extent to which theory and evidence support interpretations with respect to the proposed assessment uses—invites an evaluative argument based on sources both internal and external to the OECD’s iLibrary. For this last analysis, I drew from relevant sources of validity evidence, as described in five broad categories of the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014). I repeat those evidence sources below, this time parenthetically including commentary specific to the PIAAC numeracy assessment:

1. Assessment content (the numeracy construct description should align with its operationalization via test items; though there are only five publicly available items, these are reported as being representative of the larger set)
2. Response processes (if test developers expect test takers to engage in numeracy in specific ways, evidence should support that questions elicit that behavior)

⁷ See <http://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf> for the sample items.

⁸ The simulation is available at <http://www.oecd.org/skills/ESonline-assessment/>.

3. Internal structure (for example, if assessments items are to be of increasing difficulty, evidence should support that assumption)
4. Relations to other variables (if other variables, such as literacy assessment score, are known to relate to numeracy, then evidence should support that the numeracy assessment differentiates those constructs), and
5. Consequences of the assessment (if there are to be material consequences of an individual or country's score on the numeracy assessment, then evidence should support that those consequences follow from differential scores on the assessment).

As noted earlier, not all five categories may be relevant—the evidence needed will depend on answers to the first three questions.

In following the *Standards*, for assessment content I consider construct validity, i.e., the alignment between the construct and example assessment items (noting the limitation of the analysis); doing this entails examining available assessment items to compare what is assessed to what is intended to be assessed in the construct. For response processes, I discuss whether evidence—such as field testing or pilot studies—is presented by test developers to suggest that test-takers indeed engage in processes expected of numerate behavior. For relations to other variables, I looked within the five key PIAAC documents to see if theoretically related variables such as literacy and mathematical skills (variables which I chose, as explained below) are considered by PIAAC developers in relation to the

numeracy construct. As noted by the authors of the *Standards*, it is important that evidence be provided that demonstrates that the assessment of a construct X theoretically related to another construct Y is indeed measuring X and not Y . Finally, for consequences of score interpretations, I discuss whether evidence is provided by test developers in the five PIAAC documents to justify that score differentials correspond to corresponding actions based on interpretations of those scores. Note that nearly all of these sources of evidence require that I look for their presence in documentation literature concerning the PIAAC. In the relevant parts of the Findings section, I describe how I looked for this specific evidence within PIAAC documentation. Taken together, consideration of these five categories provides evidence of the extent to which we might be persuaded that the score interpretations from the PIAAC assessment are justified in light of the test's proposed uses.

Findings

I organize the findings in relation to the four research questions in two parts: those related to questions one through three, and those related to question four.

Interpreting a Measurement for a Specified Use

What the PIAAC Numeracy Assessment Measures. To answer the first question, that which the PIAAC numeracy assessment attempts to measure, I began by examining an OECD white paper from its PIAAC Numeracy Expert Group (2009) for descriptions of what the numeracy portion of the PIAAC attempts to

measure. I used this document as the primary source of evidence for answering this question, given that it is the sole OECD document delineating the numeracy construct, and is referred to by testmakers in other documents when describing the numeracy portion of the PIAAC. Given the document's organizational structure (described in further detail below), answering this question entailed summarizing the authors' argument, rather than looking through the document for specific codes (for example) related to what the assessment might measure. I referred to other documents, including the Technical Report (OECD 2016a), which describes in detail the test development process, and *Reader's Companion* (OECD 2016b), which outlines the test for those interested in its results, for conflicting information concerning what the numeracy assessment measures. For example, it could have been the case that the test developers decided to include only certain parts of the numeracy construct as outlined by the Numeracy Expert Group. In that sense, conflicting information could manifest as explicit statements suggesting that the construct assessed was distinct from that which the Expert Group described. There were no major deviations in test design or enactment since the Expert Group's (2009) publication.

In the 67-page document, the group situated their conceptualization of the construct of numeracy within those from other groups, assessments, and constructs (e.g., mathematical literacy). Ultimately, the group arrived at a two-pronged definition, with the first prong being that "Numeracy is the ability to access, use,

interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life” (21). The authors noted their intentionality in using the word *engage* in the definition, stating that numeracy necessarily involves dispositional elements beyond just skills. To the authors, these dispositional elements include “positive beliefs and attitudes about mathematics and about oneself as a person capable to cope with mathematical tasks” (24).

Going further, the authors stated that because numeracy is a complex construct, it was essential to add to the definition of the notion of *numerate behavior*. Numerate behavior “involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways” (21). According to the authors, this expansion of the definition allowed for actual operationalization in an assessment, “thereby contributing to the assessment’s validity and interpretability” (21). That is, the expanded definition was an important contributor to the assessment’s validity. Despite this claim, the authors did not discuss validity anywhere else in the document. With that said, the authors did discuss how the introduction of the phrase *numerate behavior* contributed to the assessment’s operationalization. The definition of numerate behavior was then operationalized through questions that drew from

- four categories of *real contexts* (e.g., everyday life, work)
- five types of *responses* (e.g., interpret, communicate)

- four domains of *mathematical content/information/ideas* (e.g., dimension and shape), and
- six venues for *multiple representations* (e.g., maps, tables) which would guide the development of their assessment items.

Importantly, the numerate behavior outlined above hinges on the “activation of” “enabling processes,” which include

- mathematical knowledge and conceptual understanding
- adaptive reasoning and mathematical problem-solving skills
- literacy skills
- beliefs & attitudes
- numeracy-related practices and experience, and
- context/world knowledge. (22)

Where relevant, I expand on the ideas in the two bulleted lists above. The enabling processes will be particularly important for discussing interpretations of scores. For now, I have discussed how the Expert Group used its two-pronged definition to attempt to operationalize numeracy through the notion of numerate behavior.

With that definition in hand, the document then describes how such a framework might manifest through the actual assessment. To that end, it includes a discussion of the limitations of the PIAAC testing environment and how that environment influenced the creation of their assessment item pool. In particular, the eighty-minute test (including all questions, as well as background surveys) was to

be given at home, with a proctor present, using a computer and automated scoring. Those constraints led the Expert Group to create an item pool where principles guiding item creation were that the items cover as many mathematical domains as possible, have “maximal authenticity and cultural appropriateness” (which is a validity claim), be scored automatically, cover different levels of difficulty, require different response actions (e.g., interpret versus compute), be time efficient (i.e., answerable quickly), and adaptable without significant modifications across participating countries (36-37). In my view, the Expert Group faced a tall task, and I discuss the extent to which they worked within and around such constraints in the context of validation later in this paper. An example of an assessment item is provided in Figure 1 below. Other available items are provided in the Appendix.

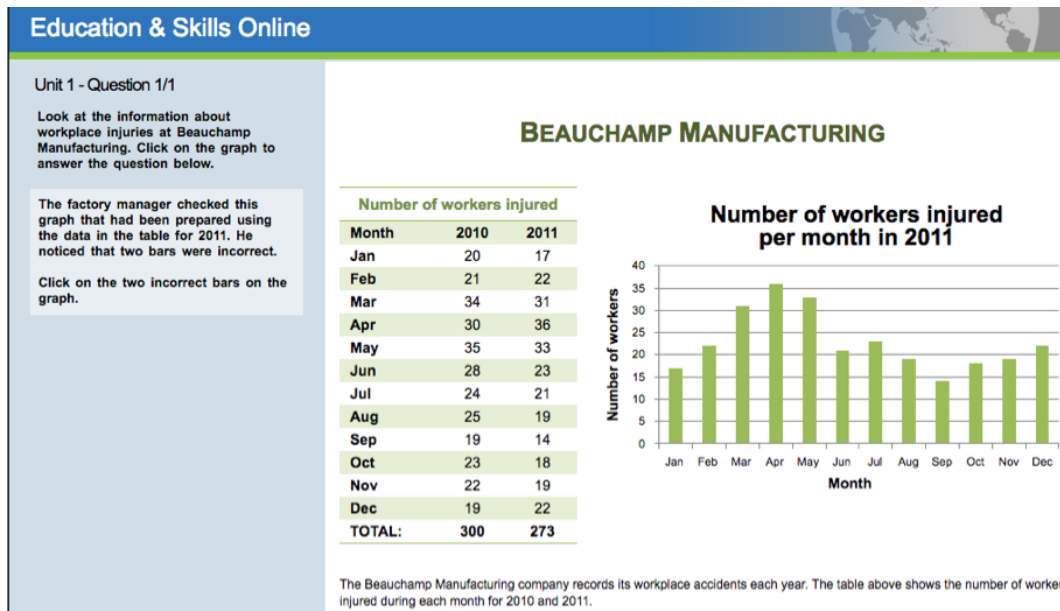


Figure 1: Publicly Available Numeracy Item from the PIAAC.

Source: <http://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf>.

The “Beauchamp Manufacturing” problem requires the test taker to identify two bars on a bar graph that are apparently incorrect in light of the table the data is based on (as opposed, for example, to identifying places where data in the table itself might be incorrect). In relation to the Expert Group’s framework for numerate behavior, note that the context here is work; the response type is interpret and evaluate, as the respondent must interpret the bar graph and then evaluate aspects of its accuracy; the item falls under the grouped mathematical domain of data and chance; and the representation includes both a table and bar graph. The sample item demonstrates the goals the Expert Group discussed in creating problems, as it is quickly answerable, automatically graded, grounded in a potentially authentic context, and adaptable across countries (e.g., bar graphs do not vary significantly in other countries).

The constraints that the Expert Group acknowledge, and that we see manifest in the item in Figure 1, invite critique concerning the apparent disconnect between numeracy as a complex construct—a behavior contingent on enabling processes like beliefs and attitudes—and one that could somehow be operationalized in the manner described above. The authors recognize this issue and include disclaimers throughout their writing. For example, after discussing the constraints above, the Expert Group notes: “As a result of the restrictions discussed above, certain types of numeracy tasks, especially those involving interpretation or evaluation/analysis with communication responses, receive only partial or slight coverage in the first

cycle of PIAAC” (34). As I discuss in the next section, the extent to which this complexity and hedging manifests in other aspects of the test development, such as interpretations of or uses of scores will vary. In summary, to the question of validity, that is what the PIAAC numeracy assessment aims to measure, the answer—subject to hedging—is numerate behavior, which the Expert Group categorizes as falling along dimensions of context, response type, mathematical content, and representation medium.

Uses of the PIAAC Numeracy Assessment. To search for purpose, or the intended uses of the assessment, I examined the five key documents that the previous search process had yielded. In examining those documents, I looked for signaling words such as “purpose” or “objective” and an explicit declaration of that purpose or objective in the context of all of the PIAAC (e.g., not just the literacy portion). Because not all declarations of purpose contained such signal words, though, it was important to read each document more than once for this specific search. For example, in the beginning chapter of *Literacy, Numeracy and Problem Solving in Technology-rich Environments: Framework for the OECD Survey of Adult Skills* (OECD 2012), “Why Assess the Skills of Adults?” the authors opened with the statement:

Understanding the level and distribution of these skills among the adult population in participating countries, as well as the ways such skills are developed and maintained, and the social and economic benefits for individuals, is important for policy makers in a range of areas of social and economic policy. (1)

The statement preceding “is important for” suggests what the OECD attempts to do through its assessment. Specifically, in this OECD document, judgments of validity are tied to “social and economic benefits for individuals.” The primary document that proved fruitful from those five documents was *The Survey of Adult Skills: Reader’s Companion* (OECD 2016b), which had the explicit motive of describing the “‘what’ and ‘how’” of the PIAAC (13). In a manner similar to my approach in answering the first question, I later corroborated my findings by looking for confirmatory and dis-confirmatory evidence in the five sources. I did this by re-reading the five documents to look for statements that suggested a purpose or use either similar to or contradictory to those that I had initially found. Ultimately, I found the purposes bulleted below; these were stated as the major analytical objectives of all of PIAAC:

- Determine the level and the distribution of proficiency in key information-processing skills for certain subgroups of the adult population.
- Better understand factors associated with the acquisition, development, maintenance and loss of proficiency over a lifetime.
- Better understand the relationship of proficiency in information-processing skills to economic and other social outcomes. (36)

These objectives are found somewhat less explicitly in other OECD documents (cf. OECD 2012, 1), but note that none of the documents I examined contained evidence suggesting that these were not the uses of the PIAAC.

The list above concerns objectives of all of the PIAAC (i.e., the assessments of literacy, numeracy, and problem solving in technology-rich environments), and the blanket references to *information-processing skills* suggests that one might read the list with the construct of numeracy explicitly in mind. Note that the PIAAC developers intended to meet the first objective through the three domain assessments, and the second and third objectives through the domain assessments coupled with the background questionnaire, which included closed-response questions about the frequency and use of various skills in one's life, as well as closed-response questions about one's health, occupation status, and other elements related to economic and social outcomes. Beyond these direct uses of the assessment scores, the ultimate goal of PIAAC is to "identify levers" in order to "reduce deficiencies," the rationale being that "Skills transform lives, generate prosperity and promote social inclusion" (OECD 2013b, 4-6). While the notion of identifying levers relates to the bulleted objectives, the task of reducing deficiencies and the rationale for doing so are beyond the scope of what assessment scores can do alone.

Interpreting PIAAC Numeracy Scores. Through the third research question I ask one of the fundamental questions of validity for the PIAAC instrument: In light of the purposes outlined above, how is one to interpret scores on the numeracy assessment? Taken together, the technical report (OECD 2016a) and *Reader's Companion* (OECD 2016b) shed light on score interpretations. The administration

of the PIAAC was a multilateral effort, with dozens of individuals from the ETS, OECD, and partner countries working together to develop and administer the exam. From a methodological standpoint, an important point to note is that—as stated in the first analytical objective above—test developers sought the distribution of skills proficiency among *subgroups* of the adult population—not to report (or even provide) results at the individual level.⁹ Using Item Response Theory scaling and latent regression modeling, test developers created proficiency scales for each of the three domains of interest: literacy, numeracy, and problem solving in technology-rich environments. Each of the scales ranged from 0 to 500 points, and every task in the numeracy domain fell at a point along that scale to indicate its difficulty based on field pilots of the assessment items (OECD 2016a). Test developers then combined item difficulty information with performance information on groups and subgroups within each country, the goal being to develop an “ability distribution” for relevant groups in specified domains (OECD 2016a, 579). To facilitate interpretation of the distributions, each 0-500 scale was broken into six levels: Below Level 1, Level 1, Level 2, and so on until Level 5. Because these proficiency levels are central to how scores are reported, I include those for the numeracy assessment in Table 1 below.

⁹ Individuals did not receive score reports, nor counseling or other resources for improving the skills tested.

Table 1: PIAAC Numeracy Proficiency Levels

Proficiency Level	Description
Below Level 1 (0 to 175)	Tasks at this level are set in concrete, familiar contexts where the mathematical content is explicit with little or no text or distractors and that require only simple processes such as counting, sorting, performing basic arithmetic operations with whole numbers or money, or recognizing common spatial representations.
Level 1 (176 to 225)	Tasks in this level require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit with little text and minimal distractors. Tasks usually require simple one-step or two-step processes involving, for example, performing basic arithmetic operations; understanding simple percents such as 50%; or locating, identifying and using elements of simple or common graphical or spatial representations.
Level 2 (226 to 275)	Tasks in this level require the respondent to identify and act upon mathematical information and ideas embedded in a range of common contexts where the mathematical content is fairly explicit or visual with relatively few distractors. Tasks tend to require the application of two or more steps or processes involving, for example, calculation with whole numbers and common decimals, percents and fractions; simple measurement and spatial representation; estimation; and interpretation of relatively simple data and statistics in texts, tables and graphs.
Level 3 (276 to 325)	Tasks in this level require the respondent to understand mathematical information which may be less explicit, embedded in contexts that are not always familiar, and represented in more complex ways. Tasks require several steps and may involve the choice of problem-solving strategies and relevant processes. Tasks tend to require the application of, for example, number sense and spatial sense; recognizing and working with mathematical relationships, patterns, and proportions expressed in verbal or numerical form; and interpretation and basic analysis of data and statistics in texts, tables and graphs.
Level 4 (326 to 375)	Tasks in this level require the respondent to understand a broad range of mathematical information that may be complex, abstract or embedded in unfamiliar contexts. These tasks involve undertaking multiple steps and choosing relevant problem-solving strategies and processes. Tasks tend to require analysis and more complex reasoning about, for example, quantities and data; statistics and chance; spatial relationships; change; proportions; and formulas. Tasks in this level may also require comprehending arguments or communicating well-reasoned explanations for answers or choices.
Level 5 (376 to 500)	Tasks in this level require the respondent to understand complex representations and abstract and formal mathematical and statistical ideas, possibly embedded in complex texts. Respondents may have to integrate multiple types of mathematical information where considerable translation or interpretation is required; draw inferences; develop or work with mathematical arguments or models; and justify, evaluate and critically reflect upon solutions or choices.

Source: Proficiency descriptions in this table are taken directly from OECD (2016a, 588-591).

Test developers arrived at these proficiency scales for each of the three domains using standard test-norming procedures: upon aggregating performance data and meeting with the domain expert groups to discuss characteristics of the assessment items. Though individuals did not receive their own scores, the developers state that

the score of an individual falling at a particular proficiency level (e.g., Level 4, and in particular, the score 330) indicates that the person would be expected to correctly answer task items with a difficulty level of 330 about 67% of the time.¹⁰ The “Beauchamp Manufacturing” problem from Figure 1 falls into Level 2 from those levels given in Table 1, as it has few distractors (i.e., one column of data is irrelevant), requires only estimation, and does not involve several steps. To provide an example of an interpretation of these scores, I draw from a “Summary of findings and policy recommendations” from *Time for the U.S. to Reskill? What the Survey of Adult Skills Says* (OECD 2013c). The first key finding leading off the document is the following: “Low “basic” skills (literacy and numeracy) are more common in the United States than on average across countries” (11). The statement itself relates to the first purpose of the PIAAC outlined in the three objectives earlier—that of determining “the level and the distribution of proficiency in key information-processing skills for certain subgroups of the adult population” (2016a, 36). The interpretation of this statement is that the percentage of the U.S. adult-aged population scoring at or below Level 1 on the numeracy scale is greater than that of the average across other countries tested. Similar statements can be said about literacy levels.

¹⁰ This quantity, 67%, is referred to as a response probability (RP) value.

In answering questions one through three, I have discussed the construct of numeracy that the PIAAC's developers sought to measure, the stated uses of the numeracy assessment, and the interpretations one is to make based on scores on the numeracy assessment. In an argument-based approach to validation, the core of the validation process is to then consider the extent to which interpretations for those uses are justified in the context of what developers seek to measure. Thus, in the next section, I take this information to answer my research question: to what extent do theory and evidence support interpretations for those uses?

Supporting Interpretations with Theory and Evidence

The first three questions invited summary more than analysis or evaluation. In considering how evidence and theory support interpretations for specified uses, the task transitions to one of making or evaluating claims about support for those interpretations. As one might imagine, the universe of possible interpretations of scores with respect to the three overarching objectives of the PIAAC is vast. Given the reams of work produced by the OECD in describing the PIAAC and its development, any consideration of validity would necessarily be vast as well. I restrict my scope here to interpretations of the PIAAC numeracy assessment scores as they relate to objective one of the PIAAC (determine the level and the distribution of proficiency in key information-processing skills for certain subgroups of the adult population). The rationale for that specific restriction is that objective one centers around the numeracy assessment itself, whereas objectives

two and three focus on its relation to the background questionnaire—a component of PIAAC that, while potentially interesting to study, is not the numeracy assessment itself. In my closing discussion, I will revisit possibilities for future work in relation to opening up the validity discussion to those involving objectives two and three.

I structure this section into parts corresponding to sources of validity evidence discussed in the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014). As I already noted, not all assessments invite the same types of validity evidence, and so some sections will be shorter than others. For example, the category of *internal structure* in this context is not fruitful to explore, because the PIAAC numeracy portion does not include composite or subtest scores to measure different aspects of the numeracy construct. The Numeracy Expert Group (2009) made no claims that the numeracy assessment measures multiple constructs; the only claim made relative to internal structure is that some items were more difficult than others, based on a collection of factors related to item complexity. Such claims were substantiated through pilot evidence and discussions among members of the Expert Group (OECD 2016a), so I do not devote space here to that source of evidence. Rather, I focus here on the categories of assessment content, response processes, relations to other variables, and consequences of score interpretations.

Assessment Content and Response Processes. The PIAAC Numeracy Expert Group (2009) described in detail their conceptualization of numeracy as it should manifest in the assessment item pool. With respect to the operationalization of the construct—that is, the assessment items themselves—the group used the notion of *numerate behavior* to facilitate item development. As noted earlier, numerate behavior “involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways” (21). Built into the expanded version of this definition are response processes (e.g., interpret, communicate), and so I group that category of validity evidence into this discussion as well. The item provided in Figure 1, the “Beauchamp Manufacturing” problem, is an exemplar of the construct of numerate behavior operationalized in an assessment task. Accompanying each of the five publicly available items is a similar mapping from the definition of numerate behavior to an actual task. In combining the developers’ discussion of numerate behavior with the tasks publicly available and the statistical techniques used to determine scores, there are no salient concerns, writ large.

That being said, in light of the test maker’s first objective of determining the level and distribution of numeracy within and across populations, the primary concern that arises in considering the content of the assessment is in *how* the test items purportedly align with the instrument’s stated definition of numerate behavior. In particular, I argue below that the Expert Group’s items do not account

for what it could mean to engage in numerate behavior as delineated by the Expert Group. This failure to account for the possibilities of numerate behavior goes beyond what one might expect of any assessment by virtue of its nature as a proxy. To justify this claim, note that there are three key phrases within the definition of numerate behavior that invite critique here: “managing a situation or solving a problem,” “real context,” and “by responding to.” Below, I expand on how, upon further inspection, these aspects of the construct are not adequately captured in the assessment items.

With respect to “managing a situation or solving a problem,” it is essential to note that judgments about management are inherently bound to a context. Through a social practices lens of numeracy, one would say that for the test taker, the context of these problems is the context of being on a computer and answering questions while being observed by an interviewer (as is the case with any similarly-structured assessment). It is not the case that the test-taker is actually at work and looking for errors in their bar graph. That is, the numeracy event occurs in answering the question, not in actually being in the world described in the question. Consider the question in Figure 1, and that appears in the first row of Table 2—looking at a bar graph for errors in one’s work (or in this case, someone else’s).

Table 2: Context Considerations for Sample PIAAC Numeracy Items

Test Item	Description of Problem	Real-life Factors or Questions to Consider
Beauchamp manufacturing	The test-taker is asked to compare a bar graph with a table that generated that bar graph; the task is to determine which bars on the graph are incorrect.	If the bar graph is generated automatically from the table, is it realistic that only two bars would be incorrect? Would a person in this situation have coworkers that might be interacting with the presentation and that might be responsible for noticing the error as well?
Running shoes	The test-taker is provided with prices for two pairs of shoes, and asked to calculate the cost of the purchase if there is a discount for purchasing both pairs.	When making a purchase online, prices are often automatically calculated in the person's shopping cart. Does successfully managing a shoe purchase require knowing how to calculate this cost? How might a person's goals for the total purchase make this question more complex?
Temperature dial	The test-taker is presented with a temperature dial, and asked what the temperature would be if it were actually 30 fewer degrees Celsius.	Because many temperature gauges are now digital, how might this problem be different? In what context would someone be reading a dial that is incorrect by 30 degrees Celsius, and is it the case that the problem in that context would be knowing what the new temperature would be?

These items are available in the Appendix (see <http://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf>).

The way that one responds to such a “problem” is mediated by a variety of factors, notably including what is expected of them (in being positioned as a test taker, the expectation is that they will answer questions “correctly”). There is no room provided for the test taker to respond to the situation, to ask questions, or to situate their own views, knowledge of the context, beliefs, or habits in relation to the task. They are to simply find two incorrect bars on a graph. In Table 2 above, I raise similar points for two other publicly available questions. These questions are given to test-takers despite the fact that the Numeracy Expert Group (2009), as noted earlier, specifically defined numerate behavior as being contingent upon certain enabling processes, which include beliefs, attitudes, as well as numeracy-related practices and experience (22). Given that extant research suggests that the ways one might attend to this situation would inevitably differ if encountered

outside of this setting (Carraher, Carraher, and Schliemann 1985; Kahan et al. 2017; Lave and Wenger 1989; Tunstall, Matz, and Craig 2018), what is it that we actually learn from seeing what one can do in this restricted context? I offer one potential answer to this question below, but do not fully answer this question in this paper.

It is assumed that one would respond (i.e., the definition of numerate behavior states “by responding to”) by examining the bar graph in comparison to the table to find the error. However, in a context in which this problem actually arose outside of a test-taking setting, one might wonder if the expected mathematics (e.g., examining the bar graph) would be used at all. Given that the graphs were clearly generated by the use of a computer, I question how a computer would make such a mistake if it was relying on inputs from a table; of course, errors can occur, but their possibility does not make this sufficiently authentic in my view. Beyond “managing a situation or solving a problem” and “by responding,” the aforementioned remark speaks to the issue of “real context.” Each of the problems on the PIAAC numeracy assessment is meant to emulate some real context. Through a social practices lens of numeracy, these contexts are real only insofar as they are real in the moment to the test taker. Each task serves as a numeracy event. The extent to which that event occurs with some regularity outside of the PIAAC assessment—that is, for it to be a numeracy practice of the test taker—is not clear. The issue of “real” here may seem to be one of mere semantics, but it is essential to keep in mind that everyone’s lived experiences are different. Of course, it is possible that the assessment

measures certain aspects or components of numerate behavior, but devoid of a fuller context and room for possibility in which that behavior might manifest, one is left to wonder (without any actual evidence) what only partial measurements tell us. It would be misleading then to claim that the assessment measures numerate behavior when the notion of real has not been properly qualified. Furthermore, though culture inevitably influences what is real to each of us, the test developers made clear that they sought contexts that supposedly apply to all cultures, stating: “Item content and questions should appear purposeful to respondents across cultures, although it must be acknowledged that in a large-scale assessment such as PIAAC, not all items and contexts can be personally familiar to all adults within any one country, let alone across all countries” (PIAAC Numeracy Expert Group 2009, 35-36). In the context of what the assessment is supposed to measure, numerate behavior, it is essential to qualify how such statements influence what test scores actually mean. Scores do not measure or tell us what the people in the representative population are doing, or what they might do in a situation, but instead, they tell us how well individuals might respond to a given artificial context to answer a question in a way that has been forced upon them. It does not tell us about the rich possibilities for nuance in response to situations that actually matter to adults. Again, these remarks then beg the question: what does the PIAAC numeracy assessment actually tell us about what people might actually do outside of the assessment setting?

Relations to Other Variables. A salient issue that one might anticipate in attempting to measure numeracy is in distinguishing it from other constructs. In the context of the PIAAC numeracy assessment, the definition of numerate behavior is that it involves using some type of mathematical information to manage a situation or solve a problem in a real context. In light of the discussion above, one might ask how the items used in PIAAC assess more than just the use of mathematical information to solve a problem. Put differently, one might ask, how are we sure that we are measuring numerate behavior and not just mathematical skills in isolation from numerate behavior more broadly? Furthermore, how do we know that the numeracy assessment is not a more elaborate assessment of literacy?

With respect to the former question—one that has been discussed in detail by scholars in quantitative literacy (see Steen et al. 2001)—the Numeracy Expert Group argues that contexts elevate these problems beyond that of context-free mathematics; however, they provide no empirical evidence (e.g., analysis to discern differences in responses to these question types) from the PIAAC or argumentative discussion to substantiate that claim. Across the five key documents that I examined in this study, I found no evidence (which would manifest as a statistical argument) that the numeracy assessment behaves differently than a more traditional mathematics assessment. The Numeracy Expert Group (2009) explicitly acknowledge the latter question (from above), drawing from Baker and Street (1994) to suggest that the two constructs are not mutually exclusive. That being

said, the Expert Group argues that numeracy “is a broad construct with a life of its own” and that its “skill levels are not measured well by literacy measures” (8-9). Ultimately, the Expert Group’s argument is that though numeracy tasks are embedded within texts, the tasks involve more than just reading, and that there are a host of enabling processes specific to numeracy, only one of which is literacy. With literacy, statistical evidence is provided related to the relationship between the numeracy and literacy assessments. Notwithstanding this argument from the Expert Group, the overall disattenuated correlation¹¹ in the initial round of the PIAAC from 2012 between countries’ numeracy and literacy proficiency scores was 0.87 (OECD 2013a; OECD 2016a). Being above 0.85, this is a coefficient that some would suggest is sufficiently high to imply that the two measures are hardly discriminating different constructs (Clark and Watson 1995; Kline 2015). Despite this statistic, upon reporting these correlations, analysts noted, “Literacy and numeracy, nevertheless, constitute distinct skills, each defined by their respective frameworks” (OECD 2013a, 2). The statement inaccurately suggests that divergence in construct definitions is sufficient to establish divergence in construct operationalizations. I comment critically on this argument in further detail in the final section of this paper. In summary, of two important constructs that might covary with performance on the PIAAC numeracy assessment—mathematical skills

¹¹ Through disattenuation, one uses statistical information concerning reliability to correct for errors inherent in the measurement process (Osborne 2008).

more broadly, and literacy as operationalized on the PIAAC—we are not provided with sufficient evidence to support the notion that PIAAC numeracy assessment scores are valid for capturing numerate behavior.

Consequences of the Assessment. The last source of validity evidence discussed in the *Standards* includes consideration of consequences—direct and indirect—stemming from interpretations of scores for a given assessment (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014). As discussed earlier, interpretations of PIAAC numeracy scores are meant to inform policymakers of the proficiencies of their constituents with respect to literacy, numeracy, and problem solving in technology-rich environments. Ultimately, a goal of PIAAC is to “identify levers” in order to “reduce deficiencies,” the rationale being that “Skills transform lives, generate prosperity and promote social inclusion” (OECD 2013b, 4-6). Per the authors of the *Standards*, it is incumbent upon test makers to provide evidence that supports such logic (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014).

In the context of the chain of reasoning above, PIAAC developers would need to demonstrate that (a) interpretations of scores indeed provide evidence of deficiencies in the population of interest, and (b) once those deficiencies are addressed, nations and their ‘more proficient’ constituents will be more prosperous

and socially inclusive. The extent to which the developers demonstrated proposition (a) depends on how we hedge what is measured. As I have argued above, the PIAAC numeracy assessment has validity issues in its attempts to capture numerate behavior, but may indeed have more validity for capturing numeracy skills in isolation of the broader enabling processes associated with those skills. With respect to (b), test developers rely on observational correlations between skills and income (among other metrics) that are based on a static dataset (i.e., the data are limited to one testing period).

If the developers are assuming a causal relation between improvements in PIAAC numeracy scores and metrics related to well-being—an assumption not directly stated, and that I cannot discern in the space of this analysis—then it is reasonable to suggest that they have not provided sufficient evidence toward that relationship. The assessment captures data on participants at one point of time, rather than longitudinally. Furthermore, the data are observational, rather than derived from any sort of controlled experiment. Existing research from scholarship on literacy suggests that a causal mechanism between literacy scores (on other assessments, not the PIAAC) and metrics related to well-being is misguided and not grounded in actual data (Graff 1978; Scribner and Cole 1981).

Finally, it is worth mentioning that validating discussions are typically found in reports of the assessment development process, and that evidence in relation to (a) and (b) are only in OECD score interpretation documents (OECD 2013a; OECD

2013c; OECD 2016b), rather than the development documents themselves (cf. OECD 2016a). Even where they do exist, the evidence in favor of (a) and (b) are never explicitly sectioned off (or even referred to) as validating discussions. This placement is not wholly surprising in the context of other developers' validations. In an analysis of assessments and associated validations from assessment developers, Cizek, Rosenberg, and Koons (2008) found that this source of evidence was largely nonexistent in extant validations, despite the fact that key figures in scholarly discussions of assessment validation had called for its inclusion since 1989 (see Messick 1989).

Discussion and Looking Ahead

The end product of a validation process or study is not a “yes” or a “no,” but instead an inference based on a set of qualified statements about an assessment in the broader context of score interpretations for stated uses (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014; Sireci and Sukin 2013). In this section, I synthesize my work above to make claims about the extent to which interpretations of scores on the PIAAC numeracy assessment are valid for the OECD's stated uses of the assessment. I then offer practical suggestions for those in the *Numeracy* community interested in using or further exploring the PIAAC, or in developing their own assessments.

Beyond Valid or Invalid

Per the Numeracy Expert Group tasked with developing and operationalizing the construct of numeracy for the PIAAC, “Numeracy is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life” (2009, 21). Going further, the Numeracy Expert Group argued that such a definition is inadequate for conveying the construct’s complexity and for operationalizing the construct through assessment items; for this reason, we need the notion of numerate behavior, which “involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways,” and is contingent upon “activation of several enabling factors and processes” which include (among other things) beliefs, attitudes, practices, experiences, and real-world context knowledge (21-22). In each of the five publicly available numeracy items, test makers outline how the construct of numerate behavior manifests in the items.

In the discussion prior to this section, I outlined issues in how this operationalization manifests in an example assessment item, notably including that the assessment item itself (as representative of the others) does not allow for the enabling processes that numerate behavior is purportedly contingent upon. Furthermore, I critiqued the definition of numerate behavior itself, arguing that it assumes a binary notion of correctness in what it means for one to manage a

situation or solve a problem (one that relies on mathematical behavior), and that it assumes a reality that only exists in the assessment itself. Though this critique suggests that the PIAAC assessment does not measure what it sets out to measure, and thus that assessment scores do not represent what was intended, it is important to keep in mind that validity is not just about construct-operationalization alignment, but rather about whether theory and evidence support interpretations of scores for proposed uses. In the context of the PIAAC numeracy assessment, a certain muddiness arises when we begin to consider how scores of the assessment are to be interpreted.

As noted earlier, numeracy scores are reported on the scale of proficiency given in Table 1. This scale was developed using pilot data and the Expert Group's comments on item difficulty. Based on this scale, scores about the construct of interest—numeracy, or numerate behavior—are ultimately then about the extent to which a group collectively answered a set of items varying in difficulty. Assuming that the experts involved in analysis completed their work correctly from a statistical standpoint (which I have no reason to doubt), scores, along with the interpretations provided in Table 1, appear to be valid for the use of describing the skills discussed in those tables. The major caveat is that the numeracy suggested by the heading in the Table, and the construct purportedly measured and operationalized by the test developers, are different. Notwithstanding the potential validity of these specific score interpretations for a specified use, it is essential that

one qualifies statements about the assessment itself so that individuals are not misled. If one examines the *Reader's Companion* (OECD 2016b), one sees in progression an overview of numeracy and numerate behavior, followed by the scoring table; there is no signaling that the two are in conflict. Hence, a potential consequence of score interpretations here is that one could be misled. For this reason, it is reasonable to argue that the validity of score interpretations is compromised.

In summary, the major finding pertaining to validity in this paper is that score interpretations from the PIAAC numeracy assessment may be considered valid for the use of describing distributions of proficiency in subgroups of interest, but

- the construct of interest—real-life numerate behavior—is not what is measured by the instrument,
- evidence distinguishing what is measured from other constructs, such as the OECD's conception of literacy, is largely absent, and
- consequences of the uses of the scores are not adequately justified.

These findings suggest some validity issues, namely that interpretations of scores do not align with descriptions of numerate behavior. Furthermore, they arise from my analysis of existing OECD documents and related literature—not from perusal of any straightforward discussion of validity from the test developers. The dearth of any validity argument from PIAAC test developers is a problem in itself, as it is incumbent upon test developers to clearly outline the evidence and theory that

support interpretations of scores for specified uses (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014).

Towards Caution and Responsibility

To *Numeracy* readers, the notion that results of the PIAAC numeracy assessment invite pause for concern may not come as a surprise. Scholars in our community have taken great strides to develop and report on assessments that invite more than just the capacity to correctly answer multiple-choice or fill-in-the-blank questions, the rationale being that alternative assessments might “show whether students have strengthened a tendency to use that capacity or have developed the skills necessary to deploy the capacity effectively in contexts other than those in the test” (Grawe, Lutsky, and Tassava 2010, 1). Though not specifically grounded in the language of a social practices approach to numeracy, such work—in congruence with that approach—highlights the notion that if we seek to understand what students do (i.e., their practices), we should provide them with the freedom and space to tell us what it is that they do. If the assessments we use to elicit what students do sacrifice that space to account for constraints such as time, efficiency, or culture, then it is imperative that we acknowledge that sacrifice and qualify our work appropriately.

As scholars of numeracy, we know all too well that data is subject to interpretation. The ways that we report our work are informed by a series of decisions that we make, whether conscious or unconscious, and ultimately those

decisions influence how our work might be taken up by others. Just as we desire for our students (Polito 2014), or for journalists (Yarnall and Ranney 2017), to be aware of how quantitative information can be communicated, so too should we take it upon ourselves to consider how the information *we* communicate more broadly can be communicated. In the context of the PIAAC numeracy assessment, I have argued that nontrivial lapses in communication suggest that the assessment measures something that it does not. We should be aware of these lapses by interrogating statistics about test scores, by carefully hedging the ways that we talk about large-scale assessments, and by—as responsible consumers and producers of information—seeking out more information before assuming we have the full story.

Beyond what may seem trite or obvious to some, I hope this analysis has provided information for scholars to consider in developing their numeracy assessments in the future. In particular, I have outlined sources of evidence to consider in making judgments about validity for an assessment (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), including those pertaining to a test’s content, its internal structure, the ways test-takers are to respond, relationships among the variables it aims to measure, and its consequences. Though not all of these sources may be necessary for supporting an interpretation with a given use in mind—especially when the scope or consequences of one’s assessment may be smaller than those of PIAAC—it is imperative that one be aware of where

experts in assessment validation currently stand (Cizek, Rosenberg, and Koons 2008). Awareness of existing scholarship is critical to developing a robust collective literature base around numeracy (Scheaffer 2008), even as our individual understandings and work vary in epistemology, method, and purpose.

APPENDIX

Five publicly available items from the OECD's PIAAC numeracy assessment (see

<http://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf>)

Education & Skills Online

Unit 1 - Question 1/1

Look at the information about workplace injuries at Beauchamp Manufacturing. Click on the graph to answer the question below.

The factory manager checked this graph that had been prepared using the data in the table for 2011. He noticed that two bars were incorrect. Click on the two incorrect bars on the graph.

BEAUCHAMP MANUFACTURING

Number of workers injured		
Month	2010	2011
Jan	20	17
Feb	21	22
Mar	34	31
Apr	30	36
May	35	33
Jun	28	23
Jul	24	21
Aug	25	19
Sep	19	14
Oct	23	18
Nov	22	19
Dec	19	22
TOTAL:	300	273

The Beauchamp Manufacturing company records its workplace accidents each year. The table above shows the number of worker injured during each month for 2010 and 2011.

OECD PIAAC Section

Unit 11 - Question 1/1

Read the article about wind power stations. Using the number keys, type your answer to the question below.

How many wind power stations would be needed to replace the power generated by the nuclear reactor?

Wind Power Stations

In 2005, the Swedish government closed the last nuclear reactor at the Barsebäck power plant. The reactor had been generating an average energy output of 3,572 GWh of electrical energy per year.

Work continues in Sweden on installing large offshore wind farms using wind power stations. Each wind power station produces about 6,000 MWh of electrical energy per year.

For your information:
 Electrical energy is measured in Watt hours (Wh)

1 kWh	= 1 kilo Wh	= 1,000 Wh
1 MWh	= 1 Mega Wh	= 1,000,000 Wh
1 GWh	= 1 Giga Wh	= 1,000,000,000 Wh

Unit 1 - Question 1/1

Look at the shoe sale advertisement. Using the number keys, type your answer to the question below.

How much would you pay during the sale if you purchase the two pairs of shoes shown?

\$

Running Shoes

SALE! Buy one pair - get the second (of equal or lesser value) for half price!



\$29.50



\$34.20

Look at the thermometer. Using the number keys, type your answer to the question below.

If the temperature shown decreases by 30 degrees Celsius, what would the temperature be in degrees Celsius (°C)?

°C

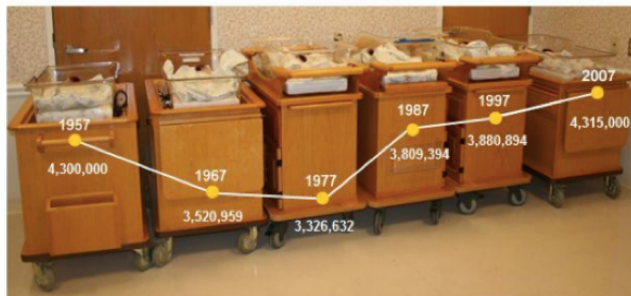


Look at the graph about the number of births. Click to answer the question below.

During which period(s) was there a decline in the number of births? Click all that apply.

- 1957 - 1967
- 1967 - 1977
- 1977 - 1987
- 1987 - 1997
- 1997 - 2007

The following graph shows the number of births in the United States from 1957 to 2007. Data are presented every 10 years.



REFERENCES

REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Baker, Dave, and Brian Street. 1994. "Literacy and Numeracy: Concepts and Definitions." In *Encyclopedia of Education 1994*, edited by Husén, Torsten, and Neville Postlethwaite. Oxford: Pergamon Press.
- Cahoon, Andrew, and Semra Kiliç-Bahi. 2019. "Assessing Quantitative Literacy: Challenges and Opportunities." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 185-96. Washington, DC: Mathematical Association of America.
- Carraher, Terezinha Nunes, David William Carraher, and Analúcia Dias Schliemann. 1985. "Mathematics in the Streets and in Schools." *British Journal of Developmental Psychology* 3 (1): 21-9.
- Cizek, Gregory J., Sharyn L. Rosenberg, and Heather H. Koons. 2008. "Sources of Validity Evidence for Educational and Psychological Tests." *Educational and Psychological Measurement* 68 (3): 397-412.
- Clark, Lee Anna, and David Watson. 1995. "Constructing Validity: Basic Issues in Objective Scale Development." *Psychological Assessment* 7 (3): 309.
- Craig, Jeffrey, and Lynette Guzmán. 2018. "Six Propositions of a Social Theory of Numeracy: Interpreting an Influential Theory of Literacy." *Numeracy* 11 (2): Article 1.
- Emanuel, Gabrielle. 2016. "America's High School Graduates Look Like Other Countries' High School Dropouts." *NPR*, March 10. <https://www.npr.org/sections/ed/2016/03/10/469831485/americas-high-school-graduates-look-like-other-countries-high-school-dropouts>.
- Foote, Mary Q., and Tonya Gau Bartell. 2011. "Pathways to Equity in Mathematics Education: How Life Experiences Impact Researcher Positionality." *Educational Studies in Mathematics* 78 (1): 45-68.
- Foucault, Michel. 1980. *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*. First American ed. New York: Pantheon Books.

- Frankel, Todd C. 2015. "U.S. Millennials Post 'Abysmal' Scores in Tech Skills Test, Lag behind Foreign Peers" *The Washington Post*, March 2. https://www.washingtonpost.com/news/wonk/wp/2015/03/02/u-s-millennials-post-abysmal-scores-in-tech-skills-test-lag-behind-foreign-peers/?utm_term=.b787a5c59ade.
- Gal, Iddo, and Dave Tout. 2014. "Comparison of PIAAC and PISA Frameworks for Numeracy and Mathematical Literacy." *OECD Education Working Papers* 102.
- Gaze, Eric, Aaron Montgomery, Semra Kiliç-Bahi, Deann Leoni, Linda Misener, and Corrine Taylor. 2014. "Towards Developing a Quantitative Literacy/Reasoning Assessment Instrument." *Numeracy* 7 (2): Article 4.
- Gergen, Mary, and Kenneth Gergen. 2000. "Qualitative Inquiry: Tensions and Transformations." In *Handbook of Qualitative Research*, 2nd ed., edited by Denzin, Norman, and Yvonna Lincoln, 1025-46. Thousand Oaks, CA: Sage.
- Graff, Harvey J. 1978. *The Literacy Myth: Literacy and Social Structure in the Nineteenth Century City*. New York: Academic Press.
- Grawe, Nathan D., Neil S. Lutsky, and Christopher J. Tassava. 2010. "A Rubric for Assessing Quantitative Reasoning in Written Arguments." *Numeracy* 3 (1): Article 3.
- Grawe, Nathan D. 2011. "Beyond Math skills: Measuring Quantitative Reasoning in Context." *New Directions for Institutional Research* 149: 41-52.
- Griffin, Patrick, Esther Care, and Mark Wilson, eds. 2018. *Assessment and Teaching of 21st Century Skills: Methods and Approach*. New York: Springer.
- Kahan, Daniel M., Ellen Peters, Erica C. Dawson, and Paul Slovic. 2017. "Motivated Numeracy and Enlightened Self-government." *Behavioural Public Policy* 1 (1): 54-86.
- Kane, Michael. 2012. "All Validity Is Construct Validity. Or Is it?" *Measurement: Interdisciplinary Research & Perspective* 10 (1-2): 66-70.
- Karaali, Gizem, Edwin Villafane-Hernandez, and Jeremy Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9 (1): Article 2.

- Klein, Stephen, Roger Benjamin, Richard Shavelson, and Roger Bolus. 2007. "The Collegiate Learning Assessment: Facts and Fantasies." *Evaluation Review* 31 (5): 415-39.
- Kline, Rex B. 2015. *Principles and Practice of Structural Equation Modeling*. New York: Guilford Publications.
- Kosko, Karl W., and Jesse L. Wilkins. 2011. "Communicating Quantitative Literacy: An Examination of Open-ended Assessment Items in TIMSS, NALS, IALS, and PISA." *Numeracy* 4 (2): Article 3.
- Lather, Patti. 1993. "Fertile Obsession: Validity After Poststructuralism." *The Sociological Quarterly* 34 (4): 673-93.
- Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Messick, Samuel. 1989. "Validity." In *Educational Measurement*, 3rd ed., edited by Robert L. Linn, 13-103. New York: Macmillan.
- Mislevy, Robert J. 2018. *Sociocognitive Foundations of Educational Measurement*. New York: Routledge.
- Mislevy, Robert J., and Geneva D. Haertel. 2006. "Implications of Evidence-Centered Design for Educational Testing." *Educational Measurement: Issues and Practice* 25 (4): 6-20.
- Newton, Paul E. 2012. "Clarifying the Consensus Definition of Validity." *Measurement: Interdisciplinary Research & Perspective* 10 (1-2): 1-29.
- Newton, Paul E., and Jo-Anne Baird. 2016. "The Great Validity Debate." *Assessment in Education: Principles, Policy & Practice* 23 (2): 173-7.
- Organisation for Economic Co-operation and Development (OECD). 2012. *Literacy, Numeracy and Problem Solving in Technology-rich Environments: Framework for the OECD Survey of Adult skills*. Paris: OECD Publishing.
- OECD. 2013a. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013b. *Skilled for Life?: Key Findings from the Survey of Adult Skills*. Paris: OECD Publishing.

- OECD. 2013c. *Time for the U.S. to Reskill? What the Survey of Adult Skills Says*. Paris: OECD Publishing.
- OECD. 2016a. *Technical Report of the Survey of Adult Skills (PIAAC)*. Paris: OECD Publishing.
- OECD. 2016b. *The Survey of Adult Skills: Reader's Companion*, 2nd ed. Paris: OECD Publishing.
- Osborne, Jason W. 2008. "Is Disattenuation of Effects a Best Practice?" In *Best Practices in Quantitative Methods*, edited by Jason W. Osborne, 239-245. Thousand Oaks, CA: SAGE Publications, Inc.
- Oughton, Helen M. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2.
- PIAAC Numeracy Expert Group. 2009. "PIAAC Numeracy: A Conceptual Framework." *OECD Education Working Papers* 35.
- Polito, Jessica. 2014. "The Language of Comparisons: Communicating about Percentages." *Numeracy* 7 (1): Article 6.
- Rear, David. 2019. "One Size Fits All? The Limitations of Standardised Assessment in Critical Thinking." *Assessment & Evaluation in Higher Education* 44 (5): 664-675.
- Rhodes, Terrel L, ed. 2010. *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities.
- Rice, Mark. 2009. "On Education, The U.S. Doesn't Measure Up." *Forbes*, October 22. <https://www.forbes.com/2009/10/22/public-education-funding-oecd-opinions-contributors-mark-rice.html#6548d2e21f26>.
- Roohr, Katrina, HyeSun Lee, Jun Xu, Ou Liu, and Zhen Wang. 2017. "Preliminary Evaluation of the Psychometric Quality of HEIghten™ Quantitative Literacy." *Numeracy* 10 (2): Article 3.
- Scheaffer, Richard L. 2008. "Scientifically Based Research in Quantitative Literacy: Guidelines for Building a Knowledge Base." *Numeracy* 1 (1): Article 3.

- Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Shavelson, Richard J., Julián P. Mariño von Hildebrand, Olga Zlatkin-Troitschanskaia, and Susanne Schmidt. 2019. "Reflections on the Assessment of Quantitative Reasoning." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 163-76. Washington, DC: Mathematical Association of America.
- Shulman, Lee S. 1981. "Disciplines of Inquiry in Education: An Overview." *Educational Researcher* 10 (6): 5-23.
- Sireci, Stephen G., and Tia Sukin. 2013. "Test Validity." In *APA Handbook of Testing and Assessment in Psychology*, Volume 1, 61-84. Washington, DC: American Psychological Association.
- Steen, Lynn A., ed., and National Council on Education and the Disciplines (NCED). 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: NCED.
- Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *U. Chi. L. Rev.* 82 (2): 831-900.
- Terman, Lewis M., Grace Lyman, George Ordahl, Louise Ordahl, Neva Galbreath, and Wilford Talbert. 1915. "The Stanford Revision of the Binet-Simon Scale and some Results from its Application to 1000 Non-selected Children." *Journal of Educational Psychology* 6 (9): 551-62.
- The National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: The National Commission on Excellence in Education.
- Thorndike, Edward L. 1916. *An Introduction to the Theory of Mental and Social Measurements*, 2nd ed. New York, NY: Teachers College, Columbia University.
- Tunstall, Samuel L., Rebecca L. Matz, and Jeffrey C. Craig. 2018. "Quantitative Literacy Courses as a Space for Fusing Literacies." *The Journal of General Education* 65 (3-4): 178-94.

- Tunstall, Samuel L., Vincent Melfi, Jeffrey C. Craig, Richard Edwards, Andrew Krause, Bronlyn Wassink, and Victor Piercey. 2016. "Quantitative Literacy at Michigan State University, 3: Designing General Education Mathematics Courses." *Numeracy* 9 (2): Article 6.
- Vacher, H. L. 2015. "Educational Assessment Is an Enduring Theme of *Numeracy*." *Numeracy* 8 (1): Article 1.
- Wolcott, Harry F. 1990. "On Seeking—and Rejecting—Validity in Qualitative Research." In *Qualitative Inquiry in Education: Continuing the Debate*, edited by Eisner, Elliot, and Alan Peshkin, 121-52. New York, NY: Teachers College Press.
- Woloshin, Steven, and Lisa M. Schwartz. 2002. "Press Releases: Translating Research into News." *JAMA* 287 (21): 2856-8.
- Yarnall, Louise, and Michael Andrew Ranney. 2017. "Fostering Scientific and Numerate Practices in Journalism to Support Rapid Public Learning." *Numeracy* 10 (1): Article 3.
- Zerr, Ryan. 2019. "Assessing Quantitative Literacy as a Cumulatively-Acquired Intellectual Skill." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 177-84. Washington, DC: Mathematical Association of America.
- Zinshteyn, Mikhail. 2015. "The Skills Gap: America's Young Workers Are Lagging Behind." *The Atlantic*, February 17.
<https://www.theatlantic.com/education/archive/2015/02/the-skills-gap-americas-young-workers-are-lagging-behind/385560/>.

Study Two: Critical Discourse Analysis of Relational Links in *Skilled for Life?*

Introduction

In all countries, adults with lower skills are far more likely than those with better literacy skills to report poor health, to perceive themselves as objects rather than actors in political processes, and to have less trust in others.

—Angel Gurría, *Skills Matter: Further Results from the Survey of Adults Skills*

The quote above from then (and present) Secretary-General of the Organisation for Economic Cooperation and Development (OECD), Angel Gurría, is a common refrain in the Forewords of many OECD reports stemming from the Programme for the International Assessment of Adult Competencies (PIAAC) (see OECD 2012, 2013a, 2013b, 2013c, 2013d, 2016a). The PIAAC is an offshoot of another, perhaps more well-known OECD exam, the Programme for International Student Assessment (PISA), but differs most notably from PISA in that it is aimed at individuals aged sixteen to sixty-five, rather than fifteen year-olds—the sole group taking part in PISA. In the sentence excerpted from a PIAAC report above, Gurría establishes a clear link between skills—which in the case of the PIAAC, includes those of literacy, numeracy, and problem-solving in technology-rich environments—and various aspects of well-being measured through PIAAC (e.g., annual income, self-reported health, and self-efficacy in relation to political

processes).¹² Although there is no allusion to *cause* in Gurría’s remark, nor warrant for causal claims based on the PIAAC’s observational studies of various skills alone, there is nonetheless an implied association (in the use of “far more likely”) among the constructs of skills and well-being in his statement. And the implied direction of the association is that skills bring about well-being. What is intriguing about this relationship is that despite eschewing causality throughout the excerpted OECD document (e.g., OECD 2016a, 144), the document ends with calls for policy makers to increase citizens’ skills so that societies and their citizens might reap the benefits of more skills (cf. 148). That is, regardless of whether a causal relationship exists, increases in skills effect benefits for policy makers and their constituents.

Spurred by concerns about the validity of such claims, as well as the foreclosure of other possibilities for what *numeracy* in particular might resemble (e.g., numeracy as a social practice, rather than a skill), in this study I use critical discourse analysis to investigate the construction of a relationship between numeracy skills and well-being in a prototypical OECD policy document, *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b). The goal of this investigation is to both denaturalize taken-for-granted beliefs about the nature of numeracy (hence the critical focus), as well as to better understand—at the micro-linguistic and macro-social levels—how policy documents like *Skilled for*

¹² Although Gurría refers to literacy specifically in the quote, the larger context of the quote suggests that by *skills* he intends all three constructs assessed in the PIAAC.

Life render such a relation coherent. Following a variant of Fairclough's (2003) approach to critical discourse analysis, which includes the study of text at the micro-linguistic and macro-social levels, the question I ask is, "In what ways are relations of association and causality between numeracy and well-being constructed in *Skilled for Life*?" I use the findings from this analysis to raise questions about the type(s) of numeracy found valuable in large-scale assessments such as PIAAC, as well as how mathematics educators can play in disrupting discourses concerning numeracy.

This investigation is situated in what has become a growing realm of socio-political work in mathematics education that draws from social theorists such as Bernstein, Bourdieu, Fairclough, Foucault, and Halliday (among others) to address questions about the broader social milieu in which the practices of mathematics teaching and learning take place (Gutiérrez 2013; Lerman 2000; Morgan 2006; Morgan 2014; Valero and Zevenbergen 2014). While diverse in scope and method, socio-political investigations in mathematics education tend to raise questions about asymmetrical power relations between students and other entities, whether through studies of the functions of mathematics education (Kollosche 2018; Larnell 2016; Pais 2013), identities and opportunities afforded to mathematics learners (Larnell 2016; Straehler-Pohl et al. 2014; Wagner and Herbel-Eisenmann 2008), or the nature of mathematics curricula (Jorgensen, Gates, and Roper 2013; Martin 2019; Wagner 2012), among other things. More specifically from this line of work,

this particular article builds on recent studies in this journal concerning the “promises” of numeracy (Craig 2018, 57) and the regime of international assessments comprised of assessments such as the PIAAC (Evans and Tsatsaroni 2014; Kaner, Morgan, and Tsatsaroni 2014). Methodologically, the starting point of this investigation is social semiotics, which posits (among other things) that the meaning of a text is situated within its context of production, consumption, and broader cultural practices (Morgan 2006). I draw from a particular approach rooted in social semiotics, critical discourse analysis, to examine *Skilled for Life*. The rationale for engaging in critical discourse analysis is that doing so helps us to call attention to and denaturalize specific ways that discourse can frame our ways of being in and thinking about the world. Intervening in discourses centered on numeracy and its “effects” is one way in which we might recapture, or reimagine, its local complexity while recognizing its ability to “travel” across contexts (Brandt and Clinton 2002).

The paper is structured as follows: I begin by describing existing work concerning the relationship between literacy or numeracy skills and benefits such as economic development. The reason for including literacy in addition to numeracy in the beginning is because of the existing literature base in literacy studies on the topic. Moreover, as I discuss in the analysis, literacy and numeracy are grouped as similar constructs in *Skilled for Life?* Following this review of literature, I explain critical discourse analysis, as well as discourse as a means of

conveying relationships of association or causation, both of which inform the methods that I use in the study. I then present an analysis of *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b). Through my analysis, I discuss techniques authors of the document employed at the micro-linguistic level to construct the skills of the PIAAC test taker or participating country as influential on, if not deterministic of, various aspects of the individual or country's economic and social well-being. I further argue that writers of the document elide test takers' agency to effect change in their lives by centering skills as the primary means by which they might participate in society. Notwithstanding the authors' explicit statements eschewing causality in the document—statements that conflict with the message implied elsewhere in the document (i.e., that skills influence well-being)—the relation is nonetheless rendered coherent at the broader socio-political level, in that *Skilled for Life* is intertwined with broader narratives concerning skills and well-being. Furthermore, the authors' deployment of various lexico-grammatical moves homogenizes literacy, numeracy, and problem solving in technology-rich environments, a move that I argue we should interrogate and push back against. This point is especially salient if numeracy is to have any distinction from literacy as a construct distinguishable on its own (Craig and Guzmán 2018).

Linking Literacy and Numeracy with Development

The stated intention behind the OECD's PIAAC is to improve member nations' education systems, the goal being for those nations to be better positioned to compete in what has been referred to as the global economy (OECD 2013c). While only eighty minutes in length, about one-third of which are devoted to setup and a background survey, the PIAAC includes separate assessments of numeracy, literacy, and problem solving in technology-rich environments; in the PIAAC, literacy is “the ability to understand, evaluate, use and engage with written texts,” numeracy is “the ability to access, use, interpret and communicate mathematical information and ideas,” and problem solving in technology-rich environment is “the ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks” (OECD 2013b, 4). Skills in these three domains, to the OECD, “transform lives, generate prosperity and promote social inclusion” (OECD 2013b, 6).

The logic above (that literacy promotes development) is not new, and as noted by Bartlett (2008), “Definitions of literacy are not innocent: they incorporate beliefs and assumptions that have political implications” (739). In the context of literacy in particular, there is a long history of scholarship viewing literacy as an “autonomous”¹³ skill (Street 1993, 5), a unit that one can isolate independent of

¹³ I use *ideological model* and *functional view* synonymously in this article. See Perry et al. (2018), Scribner (1984), or Shaw et al. (2017) for further discussion of these terms and their uptake.

other factors to study and bring to other populations or individuals so that they might enjoy its cognitive or economic benefits. Note that an implicit characteristic of literacy in the autonomous model is that one either has “it” or not; in that sense, literacy is a binary characteristic of the individual (Guadalupe and Cardoso 2011). Early empirical studies from Graff (1978), Scribner and Cole (1981), and Heath (1983), among others, have challenged the autonomous model of literacy in demonstrating that literacy is a diverse and multifaceted construct—a practice that varies between and within communities and individuals (Street 1993). A binary conceptualization of literacy, scholars have argued, does not adequately account for the nuanced ways in which individuals and communities act on, create, or adopt written and spoken texts in their lives. From this, it does not make sense to suggest that functional literacy had a deterministic effect on societal organization or individual well-being, among other things (Graff 1978; Scribner and Cole 1981). Scholars studying literacy in communities have found that literacy is a technology with affordances and potentialities, and that its meaning and uptake are localized and dependent upon the complex web of social and cultural factors present in any given context (Bartlett 2008; Brandt and Clinton 2002). Bartlett (2008), for example, found from her ethnographic work with four literacy programs in Brazil that programs’ “impact” was inextricably linked to the type of programming offered and the way participants understood its potential uptake in their unique social, political, and economic context. Now, though the view of literacy as a social

practice (the ideological model, in Street's terms) may be doxa among sociocultural scholars today, the functional view (or autonomous model) remains prevalent in common policy discourse (Bartlett 2008; Shaw et al. 2017). It is worth pointing out that the two perspectives need not be viewed as mutually exclusive (Green and Howard 2007), but instead as simultaneously important in light of the notion that skills can serve as a form of capital for individuals to draw upon in their interactions with the world.

A functional view is also prevalent among those studying numeracy, where the construct has typically been defined as facility in mathematical skills for use in everyday life—a functional analogue to literacy (Cockcroft 1982, as cited in Karaali, Hernandez, and Taylor 2016). In parallel (to a large extent) to the arguments for literacy in the autonomous model, the common rationale for numeracy (in the functional view) is that, in a world “awash in numbers” (Steen et al. 2001, 1), being numerate makes one a better citizen (Hamman 2017), facilitates better decision-making related to health and risk (Fagerlin et al. 2007; Jasper et al. 2013; Peters et al. 2006), and corresponds to wage increases and higher likelihoods of employment (Eide and Grogger 1995; Murnane, Willett, and Levy 1995; Rivera-Batiz 1992). These and related rationale are employed in documents describing the PIAAC's development (e.g., OECD 2013c; PIAAC Numeracy Expert Group 2009), as well as formal reports describing results from its administration (OECD 2013a, 2013b, 2013d, 2016a).

Although it is tempting to subscribe to a logic that suggests that increases in numeracy skills correspond to associated increases in constructs related to well-being, there are issues with arguments anchored in correlations, even those where regression is used to control for other factors (e.g., years of schooling). In this section, I raise a few of these issues. In particular, I discuss the distinction between numeracy practices and numeracy skills, concerns about regression for connecting causal outcomes, as well as shifting views of what numeracy is. My goal is not to deny the potential for some benefit of numeracy (however operationalized), but to emphasize the complexity of making claims about its “impact.”

A consideration I raise first is the practices/skills distinction: that is, our practices in relation to numbers are contingent upon a number of related factors (e.g., affect, context) that inevitably vary by time, place, and person (Carragher, Carragher, and Schliemann 1985; Kahan et al. 2017; Lave and Wenger 1989; Tunstall, Matz, and Craig 2018). Most recently, Craig and Guzmán (2018), as well as Oughton (2018), have argued that—like literacy—numeracy is best characterized through a social practices perspective. Furthermore, Craig and Guzmán (2018) have made the case that our practices in relation to numbers or quantification are sufficiently distinct from those of literacy to merit characterization on their own. Taken together, all of these studies suggest that assessments of skills alone fail to capture nuance with respect to what we actually do; in relation, they dispute the notion that ability largely determines action, given

that actions are influenced by more than just ability. Put differently, assessments of numeracy may not measure what they purport to measure (Tunstall, this dissertation). With this point in mind, it is imperative to be aware of the proxies used in reports connecting numeracy with measures of well-being. St. Clair (2015, 40-1) exemplifies this critical stance in interrogating a report of the U.S. Department of Education (2014), *Making Skills Everyone's Business: A Call to Transform Adult Learning in the United States*, homing in on the report's claim that literacy skills drive economic growth:

Before we accept the claim, we should reflect on the fact that this work is not peer-reviewed, that it uses data from school-age children to make claims about an entire population's cognitive skills, and that the case for investment in skills development is based upon highly speculative modeling that assumes infinite demand for skills.

The last point that St. Clair raises brings out a related critique to the issue of measurement, which is that of the ability of some techniques in the social sciences to capture the "real." As described by Hirschman and Reed (2014), a classic approach (among others) in the social sciences is to establish a relation between a cause X and an event Y by eliminating or controlling for confounding variables. While the mathematics of regression and related methods remain relatively unchallenged, its application to human behavior has come under fire (Abbott 1988, 2001; Fendler and Muzaffar 2008). Abbott (1988, 2001) refers to the use of regression in the social sciences as taking on a general linear reality (referred to at the beginning of this article), wherein constructs and characteristics remain fixed

in time (e.g., that one’s numeracy score remains fixed) and that cause flows from larger trends to influence specific outcomes, but never the other way around. Abbott (1988) suggests that many social scientists are aware of limitations in models they use; the issue (among other things) is that over time these methods reify a specific theory of social causality, even as other methods emerge that may better tackle the question of interest (181).

In the case of numeracy skills and well-being, for example, one could employ Abbott’s (1988) critique to work like that from Rivera-Batiz (1992), who found that after controlling for measures like years of schooling and literacy scores, a measure of quantitative literacy¹⁴—in this case, a set of multiple-choice arithmetic items grounded in real-world contexts—predicted that those with higher quantitative literacy scores were more likely to be employed than those with lower scores. Rivera-Batiz further argued that “low quantitative literacy appears to be critical in explaining the lower probability of employment of young Black Americans relative to Whites” (313). In the vein of Abbott (1988), a critique of these findings is that they suggest that a construct (i.e., a quantitative literacy score) affects all individuals in a given population across time and space, and that such work might be taken up to reify a particular theory for how quantitative literacy affects the

¹⁴ Scholars in the U.S. context often use quantitative literacy and numeracy synonymously (Karaali, Hernandez, and Taylor 2016). In this case in particular, Rivera-Batiz’s (1992) quantitative literacy aligns with numeracy as used in this article.

individual. This, in some ways, is an issue at play in the use of summary statistics more generally. Notwithstanding the argument that changes in technology have engendered changes in the nature of quantitative literacy (see Craig, Mehta, and Howard 2019), arguments like those from Rivera-Batiz (1992) remain cited in discussions about the need for quantitative literacy today (e.g., Ramirez et al. 2016). The issue with the persistent use of such studies is that if the construct of interest has fundamentally changed in how it manifests among individuals in a population, the findings (regardless of their validity) no longer bear as much meaning.

Beyond concerns about arguments anchored in correlations without any evidence of causality, one can also take issue with the ways that arguments centered on numeracy skills and well-being are taken up in deliberations about what should be included in the curriculum. Subscription to the notion that numeracy skills bring about well-being tends to lead to a human capital approach to education, an approach where teachers and students alike are treated as manipulable capital in service of a nation's economy (Yasukawa and Black 2016). The Quantway curriculum for Quantitative Literacy,¹⁵ developed by the Carnegie Foundation for the Advancement of Teaching, is an exemplar manifestation of such correlations being taken up to justify the content of a curriculum. Though ostensibly centered on quantitative literacy, Quantway is a skills-focused curriculum streamlined for

¹⁵ See <https://www.carnegiemathpathways.org/the-pathways-system/#curricula> for more information.

mass adoption; it is currently used at approximately ninety postsecondary institutions. Even if one is to teach coursework centered on numeracy skills, a focus on skills alone (or any one thing in isolation) can lead to deficit discourses about students, with instructors or policy makers focusing on what students cannot do, rather than what they are already doing, and how to build from there (Oughton 2018; Pardoe 2000; Perry et al. 2018). Of course, all of this writing is not to suggest that basic mathematical skills are not used in daily life, but just as our numeracy practices are complex and ill-defined, so too is the relationship between numeracy skills and well-being. The purpose of synthesizing critiques in this section is not to make an argument that “anything goes” in relation to curriculum centered on numeracy—an argument rebutted by Oughton (2018) (see also Pardoe 2000 in the context of literacy instruction)—but to make the case that discourse which makes an unqualified connection (e.g., the Quantway site above) between numeracy skills and well-being is misguided.

With the aforementioned point in mind, the present study stemmed from my reading of various OECD documents that describe results from the first administration of the PIAAC (OECD 2013a, 2013b, 2013c, 2013d). As noted at the outset of the study, I sensed in my reading of those documents that links were being made between numeracy scores and measures of well-being, notwithstanding comments from the OECD precluding any causal connections. The issues I have just outlined

concerning those connections provide purpose in interrogating an OECD document in further detail.

Method

Before turning to the method used in analysis of a particular OECD document, I discuss the body of scholarship pertaining to discourse analysis that I draw from for this work. I also consider how discourse can convey relationships of association or causation.

Critical Discourse Analysis

In this study, by *discourse* I mean language in use (Gee 2014).¹⁶ Discourse analysis, then, is the study of language in use (Gee 2014; Gee and Handford 2012). While there are many frameworks for thinking about and carrying out discourse analysis, each of them takes as a starting point that we make meaning of our world(s) through language, and that, “through speaking and writing in the world, we make the world meaningful in certain ways and not in others” (Gee and Handford 2012, 5). Moreover, a starting point of social semiotics—a branch within the study of discourse and linguistics—is that language use is always situated in a given context, both locally (e.g., a classroom) and more broadly (e.g., U.S. structural racism)

¹⁶ Gee distinguishes between Discourse and discourse, with "big D discourse" referring to the combination of language and other tools to enact particular types of identities. Gee (2014, 40) notes that his use of the term Discourse relates to Foucault's (1966) use of discourse, Lave and Wenger's (1991) communities of practice, Bourdieu's (1990) practices, Latour's (2005) actor-actant networks, and Hacking's (1986) kinds of people.

(Halliday 1978, as cited in Morgan 2006). The broader approach I use for discourse analysis here is critical discourse analysis (CDA), which centers on how “discourse exercises social power in institutionalizing and controlling ways of thinking and acting” (Berkovich and Benoliel 2018, 5). Critical discourse analysis helps us attend to the notion that discourse in a given event often serves some individuals or groups and dis-serves or even erases others (de Freitas and Zolkower 2009). Furthermore, at the same time that CDA involves critique, its work is both normative and explanatory in that one uses CDA not only to evaluate discourse in the context of what they believe is just, but also goes beyond description of existing realities by providing potential explanations for why we see them.

Insofar as CDA is an approach, and not a prescriptive method, there are a variety of manners and contexts in which one might engage in CDA (e.g., Fairclough 2003; Van Dijk 1993; Wodak and Meyer 2015). In mathematics education, for example, scholars have used CDA by name to study discourse ranging from episodes of dialogue in the classroom (e.g., Brantlinger 2013; de Freitas and Zolkower 2009; Evans, Morgan, and Tsatsaroni 2006; Wagner and Herbel-Eisenmann 2008) to a single word problem in a university textbook (Le Roux 2008; Le Roux and Adler 2016). Others yet have engaged in discourse analysis using tools of Halliday’s (1978) Systemic Functional Linguistics to study issues of authority and power (among other things) without specific reference to CDA (e.g., Herbel-Eisenmann, Wagner, and Cortes 2010; Mesa and Chang 2010;

Morgan 2016; Straehler-Pohl et al. 2014). A thread across these methodologically diverse studies from mathematics education is that the study of discourse can help us to uncover, or better understand, how power circulates in and through discourse.

In this particular study, I adopt Fairclough's (1995; 2003) approach to CDA, as central to his approach is the notion that in a given context, language use is part of a broader social practice, and accordingly, we may view texts and other forms of discourse as social events embedded in these larger networks of practice. For example, one could study a high-school mathematics textbook as an artifact of discourse in the practice of school mathematics. The event/practice view of language use aligns with my broader subscription to numeracy as a social practice, and is why I chose Fairclough's approach to CDA, rather than those of others such as Van Dijk (1993) or Wodak and Meyer (2015).

While the specific analytical strategies used in a given context will vary depending on the object of study and the researcher's positionality in relation to that object, Fairclough (2003) suggests that a researcher attend to three interrelated dimensions of discourse: (1) the discourse object itself, (2) the practices of production and reception of the text, and (3) the socio-political factors related to these processes. Put differently, "CDA involves the investigation of texts, their processing, and their social context through their description, interpretation, and explanation" (Berkovich and Benoliel 2018, 5). Whereas the first of these three dimensions is focused on lexico-grammatical features internal to the text (what I

refer to here as the micro-level), the latter two involve connecting the broader social practices and structures (the macro-level) to understand patterns observed at the micro-level. The process of connecting macro-level patterns with the observations at the micro-level is called interdiscursive analysis.

The tools that one uses for studying the discourse object itself at the micro-level will vary depending on the questions that one asks. In the context of this study, where I am foremost interested in the construction of association or causality at the micro-level, I need analytical tools to understand how such relations might be constructed in discourse. I turn to these in the next section.

Tools for Studying Discourse and Causation

Though as humans we develop understandings of association and causation from an early age (Keil 2003), a number of scholars argue that discourse is a key vehicle for the construction of causal reasoning (Achugar and Schleppegrell 2005; Halliday 1993; Halliday 2014; Kemmer and Verhagen 1994; Sanders and Sweester 2009). Regardless of whether we (as authors or speakers) intend to convey a relationship of causation or not, the English language is particularly known for its implicit conveyance of causal relations, as our sentences frequently have a subject-verb-object structure, and the “noun-subject carries more or less implicitly the meaning of agentive cause” (Kress 2003, 55). As a simple example, in the clause, “Maggie led the group through the forest,” one might infer that Maggie’s actions caused the group to reach the end of the forest. More broadly beyond grammar, “The way

cause is constructed is often implicit and naturalized, without clear marking of causal reasoning” (Achugar and Schleppegrell 2005, 299). This notion has particular import in this study, where notwithstanding the intentions of OECD authors in reporting on results from the PIAAC, discourse has the potential (or the power) to convey specific ideas about how constructs are related.

Much of the work related to causation and language comes from linguistics (e.g., Comrie and Polinsky 1993; Shibatani 2002) or cognitive science (e.g., Duffy, Shinjo, and Meyers 1990; Tapiero, van der Broek, and Quintana 2002), where scholars have analyzed how verbs, predicates, and clauses¹⁷ come together within sentences to convey causal relationships. Outside of linguistics and cognitive science, a separate but related vein of literature on discourse and causation is that of history education (e.g., Coffin 2004; Martin 2002), where causality is a key construct for students and historians alike to grapple with. A particular study in this domain, from Achugar and Schleppegrell (2005), had a significant influence on the methods employed in this analysis.

Through an analysis of history textbooks, Achugar and Schleppegrell identified several ways in which passages from history texts built and conveyed casual relations for historical events and themes. In their analysis, the two scholars acknowledged that causality was often *implicit*, meaning that typical verbs and

¹⁷ Note that (in general) a clause includes a subject and a predicate. It is the smallest unit for expressing a proposition. A clause may not constitute a sentence, which is one reason why I refer to clauses here rather than sentences. See the Glossary in Gee (2014) for definitions of these terms.

conjunctions used to convey causality such as “lead to” and “because” (e.g., Kemmer and Verhagen 1994) were not sufficient for capturing the often implicit ways in which causality was conveyed. Instead, causality was constructed through rhetorical structure, in addition to inter- and intra-clause moves (Martin 1996, 2002). Achugar and Schleppegrell's method for analysis included selecting two passages of interest from history textbooks, and then analyzing them in depth by looking for rhetorical structure, thematic progression, the organization of clauses, and patterns in the flow of clauses. I use a similar technique for my analysis in this study, though a departure will be that the document analyzed was not written as a narrative. Furthermore, this analysis is different in that I do not know the intentions of the authors whose work I analyzed. With that said, in the context of this study, knowing the authors' intentions is not as important as understanding the meaning construed through the document—that is, viewing the text as an objectively-given structure, and providing one interpretation (among other possible ones) of its potential meaning. As noted by Fairclough (2003), texts can have effects, even if they are not regular (i.e., every reader will not necessarily make the same meaning from a document). Furthermore, what is presented in text can undermine, or run contradictory to, intentions that one has in mind (Herbel-Eisenmann 2007).

Present Study

The research question that guided this critical discourse analysis was the following:
In what ways are relations of association and causality between numeracy skills

and well-being constructed in *Skilled for Life*? Note this question centers on numeracy skills and well-being, and not the other constructs assessed (i.e., literacy and problem solving in technology-rich environments) and well-being; this choice, as I will discuss in further detail below, limits the data set that I draw from. Furthermore, this question foregrounds the first dimension of discourse (the discourse object itself) that Fairclough (2003) suggests researchers attend to in any given CDA. I consider the other two dimensions of discourse in Fairclough's framework—the practices of production and reception of the text, and the socio-political factors that influence these processes—in the discussion following the analysis of the text at the micro-level.

With this research question in mind, I explored a single OECD-released report on the 2012 PIAAC administration. The document is considered the unit of analysis, given that it is intended to convey a message or set of messages as a whole. Note that the OECD has published several documents on the PIAAC, including summary reports, method/framework discussions, and working papers. All of these are centrally located through the OECD's iLibrary.¹⁸ In choosing a document for this analysis, I delimited my initial choices to their summary reports, given that those are aimed at a wide audience and tend to include exposition over technical detail. With that decision in mind, I found three choices for documents: *Skilled for*

¹⁸ See <https://www.oecd-ilibrary.org/>.

Life? Key Findings from the Survey of Adult Skills (OECD 2013b), *Time for the U.S. to Reskill? What the Survey of Adult Skills Says* (OECD 2013c), and *Skills Matter: Further Results from the Survey of Adult Skills* (OECD 2013d). Note that none of these three choices (nor any on the site) focus exclusively on results from the numeracy portion of the assessment. The first document, *Skilled for Life?*, was the shortest of the three reports at only thirty-two pages (as opposed to 162 and 108 pages). Framed as *Key Findings*, it is aimed at the largest reading audience of the three documents.

To reduce the data for the fine-grained analysis, I read the entire document, taking notes in a separate document on themes the authors aimed to convey. I then wrote an outline of the document based on the headings and subheadings. Finally, I highlighted every sentence in the document that connects PIAAC numeracy results with some outcome of well-being (e.g., wages, trust, health), then copied and pasted that sentence—as well as the preceding and following sentences—into a separate document. Note that by well-being, I specifically attended to characteristics of well-being described by the OECD authors, including “labour market participation, income, health, and social and political engagement” (OECD 2013b, 4). At times, I had to infer that an outcome in the document related to one of these verbatim outcomes; for example, having less trust in the political process (6) is an outcome related to political engagement, even if the word *engagement* is not used specifically. Taken together, these steps allowed me to capture the

rhetorical structure of the document, as well as explore relations in and across clauses (three sentences at a time), looking for relations of dependency in a manner similar to that from Achugar and Schleppegrell (2005). Note the document containing the sequences of clauses comprised approximately 15% of the total number of words in the report itself.

After having completed that part of the analysis, I returned to the report to read line by line for any salient constructions of causality that may have been missed in my initial focus on themes, the document outline, and the subset of clauses. As part of that additional reading, I found that there were graphical elements in the report that could contribute to a potential causal interpretation of the connection between skills and well-being. Viewing images as an additional element of the discourse object (Fairclough 2003, 3), I decided to include those in my analysis. In the section that follows, I present the results of my analysis. I then connect the findings related to my research question to the practices of production and reception of the text, as well as the socio-political factors that influence these processes.

Findings: Constructions of Association and Causation in *Skilled for Life*?

As discussed earlier, *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b) is a short document. At 32 pages, with four pages having no substantive content (i.e., graphs or text) and several pages having large images and figures, authors of the document provide its message(s) relatively quickly. I refer

to *authors* in this paper, but as I will discuss later, the reality is that we do not know who authored the document; Angel Gurría, Secretary-General of the OECD signs its Foreword, but otherwise there are no authors directly mentioned. The document's structure is the following: Foreword, About the PIAAC, Key Findings, Key Points for Policy, and Policy Challenges. Note that the formatting of the document, reading almost like a magazine, is such that Key Findings and Key Points for Policy are at times interspersed throughout the exposition (though the findings do precede the policy recommendations in the beginning). Furthermore, results pertaining to numeracy are not sectioned off relative to those of literacy and problem solving in technology-rich environments. The linkage between numeracy skills (in addition to those of the other constructs tested) and well-being infuses the document in both explicit and subtle ways. I describe those that were explicit first.

Explicit ways include statements directly indicating some sort of link. Across the document, there were twenty-five instances where there was a sentence or string of sentences that directly linked numeracy skills with some aspect of well-being. Note this count does not include statements that are only in reference to literacy or to problem solving in technology-rich environments (of which there were four beyond the twenty-five); however, it does include statements where the authors more generally refer to *skills* or *competencies*, which in this context include the numeracy portion of the PIAAC. Of these twenty-five statements, only one was in reference to numeracy exclusively—a finding that I circle back to in the discussion.

To demonstrate the breadth of these statements (i.e., the different ways links are presented), five examples are quoted below. The Appendix includes the full list.

If there is one central message emerging from this new Survey of Adult Skills, it is that what people know and what they can do with what they know has a major impact on their life chances. (6)

...per capita incomes are higher in countries with larger proportions of adults who reach the highest levels of literacy or numeracy proficiency and with smaller proportions of adults at the lowest levels of proficiency. (6)

While the causal nature of these relationships is difficult to discern, these links clearly matter, because trust is the glue of modern societies and the foundation of economic behaviour. (7)

Taken together, these results underscore the crucial importance of information-processing skills in adults' participation in the labour market, education and training, and in social and civic life. (7)

Higher levels of literacy and numeracy facilitate learning; therefore people with greater proficiency are more likely to have higher levels of education and be in jobs that demand ongoing training. They may also have the motivation and engagement with work that encourage individuals to learn and/or their employers to support them. All this can create a virtuous cycle for adults with high proficiency—and a vicious cycle for those with low proficiency. (17)

From life chances and participation in the economy, to trust and participation in social and civic life, the authors make clear in the quotes above that data from the PIAAC indicate a link between numeracy skills and various proxies for well-being. Both in the excerpts above and in the list in the Appendix, the majority of these direct links are given in the first half of the document. Those at the beginning of the document were also most straightforward (whereby straightforward I mean the number of clauses needed to convey the link), as evidenced in the distinction in the number of clauses between (for example) the first and fifth quotes above. I will

return to some of these specific quotes when discussing how organizational and grammatical features within the text also contribute to the relationship the authors establish in this work.

Indeed, what most sparked my curiosity to perform this analysis was the subtle ways in which the authors constructed an association. To that end, I begin discussing these subtle ways by analyzing how macro-level features of the document, including headings and overall structure, suggest a link between numeracy skills and well-being. I then discuss how patterns within and across clauses come together, in my view, to convey a link as well.

Construction at the Thematic and Organizational Levels

Various resources are available for authors to construct causal relationships in their writing. Notwithstanding the intention of the author(s), macro-level features of a text, including headers, the ordering of the argument, and introductory or summary statements (among other things) can come together to convey relationships of association or causation (Martin 2002a, as cited in Achugar and Schleppegrell 2005). In the context of *Skilled for Life?* the document is structured so as to introduce the PIAAC, highlight key findings, and then provide policy recommendations in relation to those findings. Recall my observation above that many of the direct statements linking numeracy skills with well-being were provided in the first half of the document. These direct statements were primarily within the introduction of the PIAAC and the initial part describing key findings. It

is notable that the description of the PIAAC itself includes these linking statements without citation, as they establish a relation upfront for the reader—one that is then *confirmed* in the findings. For example, in the Foreword from Angel Gurría, the Secretary-General situates the Survey of Adult Skills in a world of “hyper-connected societies and increasingly knowledge-based economies,” noting that “Governments need a clear picture not only of how labour markets and economies are changing, but of the extent to which their citizens are equipping themselves with the skills demanded in the 21st century, since people with low skills proficiency face a much greater risk of economic disadvantage, a higher likelihood of unemployment, and poor health” (OECD 2013b, 3). Going on to the next page, in the section, About the Survey of Adult Skills (PIAAC), authors write in the first paragraph: “...the Survey of Adult Skills focuses on how adults develop their skills, how they use those skills, and what benefits they gain from using them” (OECD 2013b, 4). Similarly, in the second paragraph, the authors declare: “With this information, the Survey of Adult Skills can help policy makers to...examine the impact of reading, numeracy and problem-solving skills on a range of economic and social outcomes...” (4). By making these connections between skills and well-being outcomes in the first section of the document (following the Foreword), one is to expect that the skills tested will demonstrate some sort of “impact on” outcomes related to well-being. The direction of a potential association is assumed, rather than found.

The authors then develop this theme in the beginning section concerning key findings. In Figure 1 below, we see that the headers commencing the key findings section concern this direct link. Note that in using the term *skills*, the header inherently envelops all of the constructs assessed through the PIAAC (i.e., literacy, numeracy, and problem solving in technology-rich environments). The association between these skills and broad units like *lives* and *economies* is not justified or expanded upon; that is, though the focus implied in the header is sensible given the OECD and its mission of building “better policies for better lives,”¹⁹ what is actually meant by “transform lives” and “drive economies” is unclear.



Skills have a major impact on each individual's life chances.

Skills transform lives, generate prosperity and promote social inclusion. Without the right skills, people are kept at the margins of society, technological progress does not translate into economic growth, and enterprises and countries can't compete in today's globally connected and increasingly complex world. Getting the best returns on investment in skills requires good information about the skills that are needed and available in the labour market. It also requires policies that ensure that skills are used effectively to generate better jobs that lead to better lives. To support these goals, the OECD has begun to measure the skills of adult populations.

in information-processing skills limits adults' access to many basic services, to better-paying and more-rewarding jobs, and to the possibility of participating in further education and training, which is crucial for developing and maintaining skills over the working life and beyond.

...and countries with lower skill levels risk losing in competitiveness as the world economy becomes more dependent on skills.

Those relationships hold not just for individuals; they also apply to countries: per capita incomes are higher in countries with larger proportions of adults who reach the highest levels of literacy or numeracy proficiency and with smaller proportions of

Figure 2: Example of Callout from *Skilled for Life?*

Note: The assertion at the top of the page and the two headers in the text itself. See OECD (2013b, 6).

¹⁹ See <http://www.oecd.org/about/>.

Another aspect of the organization of the findings section that we see in Figure 2 is the strategic use of callouts (in green text) to convey key points. For example, in Figure 2, a callout states, “Skills have a major impact on each individual’s life chances” (6). These callouts allow one to discern the main messages quickly without having to read all of the text. Such callouts continue throughout the section, as demonstrated in Figure 3, which begins the second page of the findings section.

On average, as adults’ proficiency increases, their chances of being in the labour force and being employed increase, as do their wages. Skills proficiency is also positively associated with other aspects of wellbeing.

are more likely than those with better literacy skills to report poor health, to believe that they have little impact on political processes, and not to participate in associative or volunteer

inclusive, and building real engagement with citizens all involve citizens’ skills.

Figure 3: Example of Formatting to Call out Important Information in *Skilled for Life?*

Note: Again, we see the use of separated text to call attention to an important message from the authors. See OECD (2013b, 7).

Of course, not all of the callouts concern links between skills and well-being. Following the discussion of the callout quote seen in Figure 3, the nature of the findings transitions from those about individuals’ skills and well-being to those about variation in skills within and across the countries tested. The remainder of the key findings section, as well as the section concerning policy recommendations, concern this variance and what actions policy makers should take to ensure that workers’ skills are fully utilized. Having established at the beginning of the

document the ostensible economic and social benefits of numeracy skills, the authors' recommendations that follow are justified.

The document ends with remarks about policy challenges, coming full circle to highlight the importance of fostering skills in light of their benefits. In particular, the opening sentence of the final content page—a “concluding synthesis of information” (Achugar and Schleppegrell 2005, 302)—begins: “Since it is costly to develop a population’s skills, countries need to prioritise investment of scarce resources and design skills policies such that investments reap the greatest economic and social benefits” (OECD 2013b, 30). In that quote, the authors suggest that investments in skills are what will bring about the “greatest” benefits, without evidence of what benefits alternatives (e.g., other uses of funds) could bring about. Insofar as “Effective skills policies are everybody’s business” (30), the document ends with an imperative for the reader to take action. The authors have not only given direct indications of a link between numeracy skills and well-being, but at the macro level, through the structure of the document, they have constructed and maintained a cohesive case for promoting numeracy skills to support well-being.

This cohesion is also maintained through lexico-grammatical features of the text at the micro level, which I turn to next.

Thickening Links through Clausal Work

As noted earlier, I highlighted every sentence (or excerpt) in the document that connected PIAAC numeracy results with some outcome of well-being (e.g., wages,

trust, health), copying and pasting that excerpt—as well as the preceding and following sentences—into a separate document. Examining the full excerpts allowed me to home in on how association or causation was constructed at the clausal level. In this section, I report on characteristics of the clauses containing these links, with the primary finding being that test-takers and countries as a whole were most often positioned as objects acted on by skills. That is, the narrative of who takes on what role is characterized by skills as the agents influencing the skill beholders as passive recipients.

The most basic way that we see this is in clauses or clusters of clauses where skills, in deployment with active verbs, bring about some aspect of well-being: “Skills have a major impact on each individual’s life chances. Skills transform lives, generate prosperity and promote social inclusion.” (OECD 2013b, 6). In that excerpt, the verbs *have*, *transform*, *generate*, and *promote* are all driven by skills. Their *impact* is on both individual and the society in which they reside, in that prosperity and social inclusion are both constructs that rely on an individual in their relation to society. Skills have the role of subject in several additional instances in the document:

Skills proficiency is also positively associated with other aspects of wellbeing. (7)

Skills are only of value when they are used—whether in the labour market or in other non-market settings, such as voluntary work, home production or even in leisure activities. (20)

...proficiency levels are independently related to wages. (22)

The first and third of the three quotes above are interesting in the sense that the document's authors had the choice to put either of *proficiency levels*, *skills proficiency*, *other aspects of wellbeing*, or *wages* first. I argue that their choices reflect a particular understanding of how those measures are related. For example, the authors could have written, "Aspects of wellbeing are also positively associated with skills proficiency" and "Wages are independently related to proficiency levels," which I argue suggests that wages and wellbeing are what influence proficiency levels—a different theory of how those measures are related.

Beyond the basic case above, it is notable that even when, from a grammatical perspective, the test taker or country is the subject, skills—deployed as characteristics of the individual or country—still control the outcome in the clause. That is, the authors attribute skills as important in influencing the action. The authors, for example, included the following::

...the Survey of Adult Skills focuses on how adults develop their skills, how they use those skills, and what benefits they gain from using them. (5)

Low-skilled individuals are increasingly likely to be left behind... (6)

...countries with lower skill levels risk losing in competitiveness as the world economy becomes more dependent on skills. (6)

Those with lower skills proficiency also tend to report poorer health, lower civic engagement and less trust. (6)

Many adults with low skills proficiency are outside the workforce. (21)

Whereas in the first of the five quotes above, adults with skills *gain* from using them (notably a direct causal verb), in the remaining four, individuals are *left behind* or located in the periphery of society (e.g., lose competitiveness, have poorer

health); though the authors do not directly attribute this consequence to low skills, insofar as they are the only descriptors used to describe the individual (i.e., “adults with low skills proficiency”), the reader is likely to infer a causal relationship.

A notable feature of *Skilled for Life?* that I called attention to at the beginning of this paper was the sense that, in spite of the report’s authors’ attempts to eschew causal relations between numeracy skills and well-being, there was still an effort being made to construct a unidirectional association between those two measures. In this section, I have shown that micro-level features of the discourse in *Skilled for Life?* function to construe test takers as objects acted on by the “force” of skills. A function of these choices is that they serve to construct a specific theory for how skills and well-being operate in the world. One does not need to explicitly use causal verbs to convey a relationship that implies causation—these elements at the clausal level come together to achieve this feat all the same (Achugar and Schleppegrell 2005).

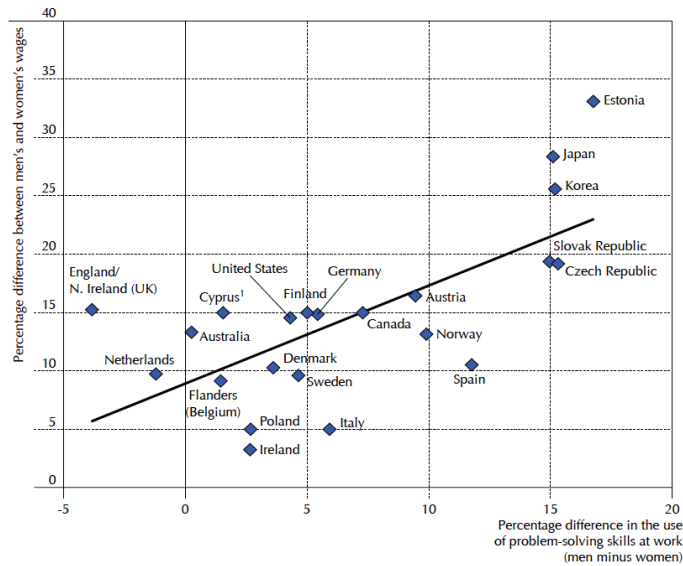
As noted before, after having completed the analysis discussed above, I returned to the report to read line by line for any salient constructions of causality that may have been missed in my initial focus on themes, the document outline, and the subset of clauses. Stemming from that additional reading, a final additional element of the text that I bring up in this analysis is two of the six graphs in the report that may contribute to a potential causal interpretation of the connection

between skills and well-being. The two graphs are provided in Figures 4 and 5 below.



Figure 4: A Graph that Has the Potential for Construing Causality.
Note: The horizontal axis variable is derived from a measure of skills. The vertical axis variable is an outcome related to well-being. See OECD (2013b, 21).

■ Correlation between gender gap in wages and in the use of problem-solving skills at work ■



1. See notes regarding Cyprus on page 2.

Notes: The gender gap in wages is computed as the percentage difference between men's and women's average hourly wages, including bonuses. The wage distribution was trimmed to eliminate the 1st and 99th percentiles. The bold line is the best linear prediction. The sample includes only full-time employees.

Figure 5: A Graph that Suggests an Association Between Variables Derived from Skills and from an Outcome of Well-being.

Note: See OECD (2013b, 29).

Note that in Figures 4 and 5, both graphs depict a bivariate relationship using a scatterplot. In both cases, the horizontal axis variable is derived from a measure of skills, and the vertical axis variable is an outcome related to well-being. Furthermore, readers are provided with a line of best fit to suggest that there is a relationship between the two variables. I do not discuss these graphs in further detail given that they are not specifically related to numeracy, but it is worth pointing out that both (1) a common misconception related to interpreting scatterplots is to mistake correlation with causation, and in relation, (2) that one typically places “independent” variables on the horizontal axis and “dependent”

variables on the vertical axis. More broadly, Lemke (1990) and Porter (1996) argue that graphs and statistics are often used to justify a particular ideology under the guise that that ideology is grounded in the objective analysis containing that graph or statistic.

The Text in Its Broader Milieu

As described in the Methods section, the general approach to CDA for Fairclough involves attention to (1) the discourse object itself, (2) the practices of production and reception of the text, and (3) the socio-political factors related to these processes. The research question that guided my CDA in this study was, “In what ways are relations of association and causality between numeracy and well-being constructed in *Skilled for Life?*” On its own, this research question invites attention to the first dimension of discourse in Fairclough’s framework for CDA. With that said, a theoretical underpinning of social semiotics more generally is that language use is always situated in a given context, both locally and more broadly (Halliday 1978, as cited in Morgan 2006). Furthermore, and more specific to Fairclough, language use is part of a broader social practice, and accordingly, we may view texts and other forms of discourse as social events embedded in these larger networks of practice. With that in mind, it is important to now consider the broader practices and socio-political factors influenced by and related to the OECD’s report.

Whenever possible, I will connect the findings concerning the discourse object itself to these broader practices and socio-political factors.

In considering the practices of production and reception related to PIAAC, it is helpful to step back to consider the larger umbrella of the OECD. The stated mission of the OECD is to “build better policies for better lives” by “establishing international norms and finding evidence-based solutions to a range of social, economic and environmental challenges.”²⁰ At the core of its work is an ongoing battery of international assessments, which currently include PISA, the PIAAC, as well as the Teaching and Learning International Survey, in addition to one-time studies such as the International Early Learning and Child Well-being Study and the Study on Social and Emotional Skills. Beyond these current projects, previous rolling studies in education include the International Adult Literacy Survey, and the Adult Literacy and Life Skills Survey. All of the aforementioned studies intersect with member countries’ processes for education, with targeted age-bands ranging from early childhood to adolescence and adulthood. The purpose of these measurements is to gauge where member countries are with respect to OECD-driven objectives, and from those measurements, to recommend policies to participating countries that steer them closer to where the OECD would like for them to be. Data and reports are presented in multiple venues, including press

²⁰ See <http://www.oecd.org/about/>.

conferences, the OECD’s expansive iLibrary—a digital corpus of all OECD studies, reports, and white papers—as well as formal research conferences such as the PIAAC “Research to Practice” Conference.²¹

The production process of documents such as *Skilled for Life* is contingent upon the work of multiple parties over the course of about a decade. As described in the *The Survey of Adult skills: Reader's Companion* (OECD 2013c), after working groups of the OECD commissioned the assessment, several “expert groups” (e.g., the PIAAC Numeracy Expert Group) assembled both separately and together to develop the assessment. From there, assessment specialists developed the assessment itself, sending teams to then pilot the questionnaire and selected items in participating countries before administering the more formal assessment to various participating countries in the first round of the Survey’s administration (OECD 2013c). Once the assessment itself concluded, statisticians worked with the assessment specialists to begin data analysis and the production of summary reports, from which documents such as *Skilled for Life* were finally produced; for rolling assessments such as the PIAAC, committees then convened to plan for the next administration in light of results from the previous administration. Workers for OECD Publishing, located in Paris, helped with copy-editing and ensuring uniformity across various documents. While some background reports related to

²¹ More information is available at <http://piaacgateway.com/us-piaac-conference-2018>. Also, one can view upcoming events of the OECD at <https://www.oecd.org/newsroom/upcomingevents/>.

the PIAAC have named authors (e.g., PIAAC Numeracy Expert Group 2009), summary reports such as *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b) and *Time for the U.S. to Reskill? What the Survey of Adult Skills Says* (OECD 2013d) do not. The opening pages of *Skilled for Life*, for example, provide no names, instead stating, “This work is published on the responsibility of the Secretary-General of the OECD” (OECD 2013b, 2).

A striking feature of the process of production described above is that accountability for the assessment is distributed across hundreds of individuals, most of whom go unnamed. Indeed, even in the sentence above, Secretary-General of the OECD, Angel Gurría, is only referred to in title. Scholars in Science and Technology Studies argue that such a move to separate the researcher from their product is a key characteristic of practices of knowledge production in science (e.g., Fairclough 2003; Latour 1983; Waidzunus 2012), adding to the ostensible objectivity of a research study or product. Moreover, in dislocating the actors from the action, research and subsequent actions appear inevitable (Fairclough 2003). Taken together, the production process of *Skilled for Life* entailed hundreds of actors whose labor funneled into an ultimately nameless—and thus in some respects, unquestionable—product. As I discussed earlier, many of the claims concerning the relationship between numeracy skills and well-being are provided without citation, and I would argue that because the document is presented as an

official report of the OECD, the report has the potential to reify and render coherent a particular relation of association that is not subject to question.

The practices of reception associated with documents such as *Skilled for Life* have been discussed at length by education researchers. Though *Skilled for Life* has been cited only thirteen instances (according to a June 2019 search in Google Scholar), more broadly researchers have found that results of OECD assessments tend to have “shock” effects, or immediate calls for change in learning of the performance of one’s educational system relative to others (Grek 2012; Mangez and Hilgers 2012; Waldow 2009). Nations in Western Europe and Eastern Asia tend to be among those at the “top” of the list of countries when ranked by average performance. One can certainly find elements of shock in U.S. news media following the PIAAC, with reports abound suggesting that American adults are unable to complete basic mathematics tasks (Emanuel 2016; Frankel 2015; Zinshteyn 2015). Because the ability to perform such tasks is constructed as associated with outcomes of well-being, the call for action following the alarm centers on improving students’ skills. And while the actual effects of such reports are unclear (and in some respects, unknowable), the OECD does play an authoritative role in national and global conversations about education policy (Rautalin, Alasuutari, and Vento 2019). This role connects to larger movements related to politics and education.

There are several socio-political factors at play as part of the broader milieu in which *Skilled for Life* exists. A factor that relates directly to the OECD's production of the document is the Global Education Reform Movement, a term coined to describe neoliberal "policies such as standardization of education, focus on literacy and numeracy, embracing managerialism and other corporate ideas (e.g. marketization and privatization), test-based accountability, and increased control of schools and schooling (e.g. national curricula)" (Berkovich and Benoliel 2018, 2). These policies in education are rooted in broader movements to measure, sort, and surveil citizens—ensuring no one is left "at the margins" (OECD 2013b, 6)—with the assumption that measurement facilitates "soft governance" (Kanes, Morgan, and Tsatsaroni 2014; Ozga 2012), or "voluntary self-control" (Pongratz 2006), as entities engage come to compare their characteristics (e.g., scores) with those of other entities. By their nature, the processes of measurement and comparison lead one to focus on certain things and not on others. Berkovich and Benoliel (2018), for example, argue that the OECD's Teaching and Learning International Survey framework aims to exercise normative control over what counts as quality teaching in providing a very narrow vision of what quality teaching can look like.

Just as Berkovich and Benoliel found OECD authors advocating for a narrow conception of the notion of teacher quality, we see a flattening in *Skilled for Life* of the construct of numeracy. Indeed, as alluded to earlier, numeracy was regularly referred to in tandem with other skills (literacy and problem solving in technology-

rich environments) as connected with outcomes pertaining to well-being. Of the twenty-five instances in which numeracy was connected with well-being, only once was it discussed with an outcome on its own. In the case of the PIAAC numeracy assessment, one could argue that in defining and assessing numeracy in a competence-based model focused on skills, other characterizations of numeracy are lost (Craig 2018; Tsatsaroni and Evans 2014). For example, Craig (2018) suggested that numeracy education might center around problems, rather than on specific disciplinary skills. More broadly in the context of mathematics education, *Skilled for Life* serves to advance a particular official curriculum, one that views the purpose of mathematics education to be that of advancing human capital in what has been referred to as a global economy (OECD 2013b, 30). This purpose stands in contrast to other purposes, such as the learning of mathematics for human flourishing (Tunstall and Ferkany 2017) or for critical literacy (e.g., Frankenstein 1990), among others. What is intriguing about the PIAAC's justification for the development of numeracy skills is that they are not only to serve people in their daily lives, but they also help one in participating in the economy, among other benefits to well-being. Pais (2013) argues that this logic is faulty insofar as it obfuscates the larger role of mathematics education in tracking students into certain careers. In viewing numeracy as a set of social practices related to, but distinct from, literacy practices more generally (Craig and Guzmán 2018), this lumping of numeracy with literacy (and problem solving in technology-rich environments)

obfuscates the distinct possibilities afforded by practices of quantification (e.g., Porter 1996).

Talking Back

As part of their review of literature concerning the OECD's PISA and its uptake among mathematics education researchers, Kaner, Morgan and Tsatsaroni (2014) found that there had been "remarkably little" critical work of the OECD's testing regime, even among scholars whom they referred to as being actively involved in research on social issues (162). Indeed, a search within Google Scholar of the term PIAAC (both spelled out and abbreviated), beginning from 2013, yields that the most cited paper is one from the *European Economic Review* titled "Returns to Skills Around the World: Evidence from PIAAC" (Hanushek et al. 2013); the paper consists of a quantitative analysis of the connection between PIAAC skills and wages across the countries tested. Authors in that paper use the data without critically examining (or at least explicitly saying so) its origins. None of the other papers that appear in the first page of results centers on critique of the assessment. The present paper responds to their call for "systematic and principled" (163) research within mathematics education that aims to disrupt the OECD's regime on mathematics as it is or might be employed in our daily lives.

Tsatsaroni and Evans (2014) make clear that the OECD's regime extends beyond PISA (or TALIS) through the OECD's PIAAC as well, concluding: "The

aim of educational researchers must be to support the development of potentially powerful knowledge, like numeracy, and to prevent its being reduced to a narrow competency” (181). Indeed, given that literacy, numeracy, and problem solving in technology-rich environments were almost exclusively positioned together as *skills* in the document used for this analysis, we see a similar push with PIAAC to that found by Berkovich and Benoliel (2018) for the normative control of constructs like literacy and numeracy.

In this paper I have demonstrated that through lexico-grammatical (micro) features of an OECD text describing findings from the PIAAC, its authors have constructed a specific theory for how numeracy skills and measures of well-being interact among individuals and society more broadly. The relation between numeracy skills and measures of well-being suggests that numeracy is indeed “powerful knowledge,” but that power is derived solely from the skills measured in PIAAC. Insofar as the test taker is provided no agency outside of that world, such knowledge holds little power in the sense described by Tsatsaroni and Evans (2014). I began this research guided by a curiosity, asking: How is it possible that in writing, one can explicitly discount causation while simultaneously reasoning in ways that assume a unidirectional association? An answer I found is that *Skilled for Life* produces this unidirectional association not only through direct statements and through discourse-semantic features of the text such as headings and summary statements, but also through its discursive construction of the test taker as an object

controlled by the power of skills. Though I cannot make claims about the effects a given text will have on individual readers, it would appear important that we continue to interrogate claims put forth by the OECD in favor of its narrow operationalizations of various constructs. That is, that we “talk back” to those speaking on everyone’s behalf.

APPENDIX

Across *Skilled for Life?* (OECD 2013b), there were twenty-five paragraphs that included a sentence or string of sentences that directly linked numeracy skills with some aspect of well-being. Note this count does not include statements that are only in reference to literacy or to problem solving in technology-rich environments; however, it does include statements where the authors more generally refer to *skills* or *competencies*, which in this context include the numeracy portion of the PIAAC. Below is a full list of these instances in order of their appearance.

...the Survey of Adult Skills focuses on how adults develop their skills, how they use those skills, and what benefits they gain from using them. (5)

To this end, the Survey of Adult Skills collects information on...how these skills are related to labour market participation, income, health, and social and political engagement. (5)

With this information, the Survey of Adult Skills can help policy makers to...examine the impact of...numeracy...on a range of economic and social outcomes... (5)

Skills have a major impact on each individual's life chances. Skills transform lives, generate prosperity and promote social inclusion. Without the right skills, people are kept at the margins of society... (6)

If there is one central message emerging from this new Survey of Adult Skills, it is that what people know and what they can do with what they know has a major impact on their life chances. (6)

As the demand for skills continues to shift towards more sophisticated tasks, as jobs increasingly involve analysing and communicating information, and as technology pervades all aspects of life, those individuals with poor literacy and numeracy skills are more likely to find themselves at risk. Poor proficiency in information-processing skills limits adults' access to many basic services, to better-paying and more-rewarding jobs, and to the possibility of participating in further education and training, which is crucial for developing and maintaining skills over the working life and beyond. (6)

...per capita incomes are higher in countries with larger proportions of adults who reach the highest levels of literacy or numeracy proficiency and with smaller proportions of adults at the lowest levels of proficiency. (6)

How literacy skills are distributed across a population also has significant implications on how economic and social outcomes are distributed within the society. The Survey of Adult Skills shows that higher levels of inequality in literacy and numeracy skills are associated with greater inequality in the distribution of income, whatever the causal nature of this relationship. If large proportions of adults have low reading and numeracy skills, introducing and disseminating productivity-improving technologies and work-organisation practices can be hampered; that, in turn, will stall improvements in living standards. (6)

Those with lower skills proficiency also tend to report poorer health, lower civic engagement and less trust. (6)

On average, as adults' proficiency increases, their chances of being in the labour force and being employed increase, as do their wages. Skills proficiency is also positively associated with other aspects of wellbeing. (7)

While the causal nature of these relationships is difficult to discern, these links clearly matter, because trust is the glue of modern societies and the foundation of economic behaviour. (7)

Taken together, these results underscore the crucial importance of information-processing skills in adults' participation in the labour market, education and training, and in social and civic life. (7)

The survey results offer vital insights for policy makers working to tackle the challenges involved in developing skills, activating the supply of skills, and putting skills to more effective use so as to achieve better outcomes for individuals and societies. While the survey only shows correlations, these results, when combined with the wealth of OECD policy analysis, can inform improvements to skills systems. (7)

The fact that the countries with the greatest social inequities in the OECD Programme for International Student Assessment (PISA) are also those with low rates of social mobility as observed in the Survey of Adult Skills

suggests that the relationship between social disadvantage and lower skills proficiency may be established early in individuals' lives. (10)

Higher levels of literacy and numeracy facilitate learning; therefore people with greater proficiency are more likely to have higher levels of education and be in jobs that demand ongoing training. They may also have the motivation and engagement with work that encourage individuals to learn and/or their employers to support them. All this can create a virtuous cycle for adults with high proficiency – and a vicious cycle for those with low proficiency. (17)

More adults will be tempted to invest in education and training if the benefits of improving their skills are made apparent to them. For example, governments can provide better information about the economic benefits, including wages net of taxes, employment and productivity, and non-economic benefits, including self-esteem and increased social interaction, of adult learning. (18)

Skills are only of value when they are used – whether in the labour market or in other non-market settings, such as voluntary work, home production or even in leisure activities. (20)

To the extent that workers' productivity is related to the knowledge and skills they possess, and that wages reflect such productivity, individuals with more skills should expect higher returns from labour market participation and would thus be more likely to participate...Employed adults also tend to have higher mean proficiency scores in literacy and numeracy than unemployed adults, who score higher, in turn, than those outside the labour force. (20)

Many adults with low skills proficiency are outside the workforce. (21)

The large shares of low-skilled adults outside the labour force present additional challenges to policy makers because these adults' lack of skills is likely to be closely linked to their prospects for employment... Yet a lack of skills presents a formidable obstacle to employment for these adults; tackling these skills deficits will be important to enhance their longer-term employment prospects... (21)

Earnings increase with proficiency, but to very different degrees across countries. (22)

...both education, whether measured in years or in attainment level, and proficiency levels are independently related to wages. (22)

Skills will only translate into better economic and social outcomes if they are used effectively...developing skills and making them available to the labour market will not translate into better social and economic outcomes if those skills are not used effectively on the job. (24)

Since it is costly to develop a population's skills, countries need to prioritise investment of scarce resources and design skills policies such that investments reap the greatest economic and social benefits. (30)

Seeing skills as a tool to be honed over an individual's lifetime will also help countries to better balance the allocation of resources to maximise economic and social outcomes. (30)

REFERENCES

REFERENCES

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6 (2): 169-86.
- Abbott, Andrew. 2001. *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.
- Achugar, Mariana, and Mary J. Schleppegrell. 2005. "Beyond Connectors: The Construction of Cause in History Textbooks." *Linguistics and Education* 16 (3): 298-318.
- Bartlett, Lesley. 2008. "Literacy's Verb: Exploring what Literacy Is and what Literacy Does." *International Journal of Educational Development* 28 (6): 737-53.
- Berkovich, Izhak, and Pascale Benoliel. 2018. "Marketing Teacher Quality: Critical Discourse Analysis of OECD Documents on Effective Teaching and TALIS." *Critical Studies in Education*: 1-16.
- Brandt, Deborah, and Katie Clinton. 2002. "Limits of the Local: Expanding Perspectives on Literacy as a Social Practice." *Journal of Literacy Research* 34 (3): 337-56.
- Brantlinger, Andrew. 2014. "Critical Mathematics Discourse in a High-School Classroom: Examining Patterns of Student Engagement and Resistance." *Educational Studies in Mathematics* 85 (2): 201-220.
- Bourdieu, Pierre. 1990. *The Logic of Practice*. Stanford, CA: Stanford University Press.
- Cockcroft, Sir Wilfred H. 1982. *Mathematics Counts. Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr. Wilfred H. Cockcroft*. London: Her Majesty's Stationery Office.
<http://www.educationengland.org.uk/documents/cockcroft/cockcroft1982.html>.
- Coffin, Caroline. 2004. "Learning to Write History: The Role of Causality." *Written Communication* 21 (3): 261-89.
- Comrie, Bernard, and Maria Polinsky. 1993. *Causatives and Transitivity*. Vol. 23. Amsterdam, The Netherlands: John Benjamins Publishing.

- Craig, Jeffrey. 2018. "The Promises of Numeracy." *Educational Studies in Mathematics* 99 (1): 57-71.
- Craig, Jeffrey, and Lynette Guzmán. 2018. "Six Propositions of a Social Theory of Numeracy: Interpreting an Influential Theory of Literacy." *Numeracy* 11 (2): Article 1.
- Craig, Jeffrey, Rohit Mehta, and James P. Howard III. 2019. "Quantitative Literacy to New Quantitative Literacies." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 15-26. Washington, DC: Mathematical Association of America.
- de Freitas, Elizabeth, and Betina Zolkower. 2009. "Using Social Semiotics to Prepare Mathematics Teachers to Teach for Social Justice." *Journal of Mathematics Teacher Education* 12 (3): 187-203.
- Duffy, Susan A., Makiko Shinjo, and Jerome L. Myers. 1990. "The Effect of Encoding Task on Memory for Sentence Pairs Varying in Causal Relatedness." *Journal of Memory and Language* 29 (1): 27-42.
- Emanuel, Gabrielle. 2016. "America's High School Graduates Look Like Other Countries' High School Dropouts." *NPR*, March 10.
<https://www.npr.org/sections/ed/2016/03/10/469831485/americas-high-school-graduates-look-like-other-countries-high-school-dropouts>.
- Evans, Jeff, Candia Morgan, and Anna Tsatsaroni. 2006. "Discursive Positioning and Emotion in School Mathematics Practices." *Educational Studies in Mathematics* 63 (2): 209-226.
- Fagerlin, Angela, Peter A. Ubel, Dylan M. Smith, and Brian J. Zikmund-Fisher. 2007. "Making Numbers Matter: Present and Future Research in Risk Communication." *American Journal of Health Behavior* 31 Suppl 1 (1): S47-56.
- Fairclough, Norman. 1995. *Critical Discourse Analysis: The Critical Study of Language*. London: Longman.
- Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Fendler, Lynn, and Irfan Muzaffar. 2008. "The History of the Bell Curve: Sorting and the Idea of Normal." *Educational Theory* 58 (1): 63-82.

- Foucault, Michel. 1966. *The Order of Things: An Archaeology of Human Sciences*. New York: Random House.
- Frankel, Todd C. 2015. "U.S. Millennials Post 'Abysmal' Scores in Tech Skills Test, Lag behind Foreign Peers" *The Washington Post*, March 2. https://www.washingtonpost.com/news/wonk/wp/2015/03/02/u-s-millennials-post-abysmal-scores-in-tech-skills-test-lag-behind-foreign-peers/?utm_term=.b787a5c59ade.
- Frankenstein, Marilyn. 1990. "Incorporating Race, Gender, and Class Issues Into a Critical Mathematical Literacy Curriculum." *The Journal of Negro Education* 59 (3): 336–347.
- Gee, James Paul. 2004. "Discourse Analysis: What Makes it Critical?" In *An Introduction to Critical Discourse Analysis in Education*, edited by Rebecca Rogers, 19-50. Mahwah, NJ: Lawrence Erlbaum.
- Gee, James Paul. 2014. *An Introduction to Discourse Analysis: Theory and Method*. 4th ed. New York: Routledge.
- Gee, James Paul, and Michael Handford, eds. 2012. *The Routledge Handbook of Discourse Analysis*. London: Routledge.
- Graff, Harvey J. 1978. *The Literacy Myth: Literacy and Social Structure in the Nineteenth Century City*. New York: Academic Press.
- Green, Alix, and Ursula Howard. 2007. "Skills and Social practices: Making Common Cause." *National Research and Development Centre for Adult Literacy and Numeracy*.
- Grek, Sotiria. 2012. "What PISA Knows and Can Do: Studying the Role of National Actors in the Making of PISA." *European Educational Research Journal* 11 (2): 243-254.
- Grogger, Jeff, and Eric Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *The Journal of Human Resources* 30 (2): 280-310.
- Guadalupe, Cesar, and Manuel Cardoso. 2011. "Measuring the Continuum of Literacy Skills among Adults: Educational Testing and the LAMP Experience." *International Review of Education* 57 (1): 199-217.
- Gutiérrez, Rochelle. 2013. "The Sociopolitical Turn in Mathematics Education." *Journal for Research in Mathematics Education* 44 (1): 37-68.

- Hacking, Ian. 1986. "Making Up People." In *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*, edited by Heller, Thomas C., Morton Sosna, and David E. Wellbery, 222-36. Stanford, CA: Stanford University Press.
- Halliday, M. A. K. 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. Baltimore, MD: University Park Press.
- Halliday, M. A. K. 1993. "Towards a Language-based Theory of Learning." *Linguistics and Education* 5 (2): 93-116.
- Halliday, M. A. K. 2014. *An Introduction to Functional Grammar*. Revised by Christian M. I. M. Matthiessen. 4th ed. London: Arnold.
- Hamman, Kira. 2017. "Rethinking the Numerate Citizen: Quantitative Literacy and Public Issues—Discussion." *Numeracy* 10 (2): Article 12.
- Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2015. "Returns to Skills Around the World: Evidence from PIAAC." *European Economic Review* 73: 103-30.
- Heath, Shirley Brice. 1983. *Ways with Words: Language, Life, and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Herbel-Eisenmann, Beth A. 2007. "From Intended Curriculum to Written Curriculum: Examining the 'Voice' of a Mathematics Textbook." *Journal for Research in Mathematics Education* 38 (4): 344-369.
- Herbel-Eisenmann, Beth, David Wagner, and Viviana Cortes. 2010. "Lexical Bundle Analysis in Mathematics Classroom Discourse: The Significance of Stance." *Educational Studies in Mathematics* 75 (1): 23-42.
- Hirschman, Daniel and Isaac Ariail Reed. 2014. "Formation Stories and Causality in Sociology." *Sociological Theory* 32 (4): 259-282.
- Jasper, John D., Chandrima Bhattacharya, Irwin P. Levin, Lance Jones, and Elaine Bossard. 2013. "Numeracy as a Predictor of Adaptive Risky Decision Making: Numeracy and Risky Decision Making." *Journal of Behavioral Decision Making* 26 (2): 164-73.

- Jorgensen, Robyn, Peter Gates, and Vanessa Roper. 2014. "Structural Exclusion Through School Mathematics: Using Bourdieu to Understand Mathematics as a Social Practice." *Educational Studies in Mathematics* 87 (2): 221-239.
- Kanes, Clive, Candia Morgan, and Anna Tsatsaroni. 2014. "The PISA Mathematics Regime: Knowledge Structures and Practices of the Self." *Educational Studies in Mathematics* 87 (2): 145-65.
- Karaali, Gizem, Edwin Villafane-Hernandez, and Jeremy Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9 (1): Article 2.
- Keil, Frank. 2003. "Categorisation, Causation, and the Limits of Understanding." *Language and Cognitive Processes* 18 (5-6): 663-92.
- Kemmer, Suzanne., and Arie Verhagen. 1994. "The Grammar of Causatives and the Conceptual Structure of Events." *Cognitive Linguistics* 5 (2): 115-56.
- Kollosche, David. 2018. "Social Functions of Mathematics Education: A Framework for Socio-political Studies." *Educational Studies in Mathematics* 98 (3): 287-303.
- Kress, Gunther R. 2003. *Literacy in the New Media Age*. London: Routledge.
- Larnell, Gregory V. 2016. "More than Just Skill: Examining Mathematics Identities, Racialized Narratives, and Remediation Among Black Undergraduates." *Journal for Research in Mathematics Education* 47 (3): 233-269.
- Latour, Bruno. 1983. "Give Me a Laboratory and I will Raise the World." In *Science Observed: Perspectives on the Social Study of Science*, edited by Mulkay, Mike, and Karin Knorr Cetina, 141-70. *Beverly Hills: Sage Publications*.
- Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Lemke, Jay L. 1990. *Talking Science: Language, Learning, and Values*. Norwood, N.J: Ablex Pub. Corp.

- Le Roux, Kate. "A Critical Discourse Analysis of a Real-world Problem in Mathematics: Looking for Signs of Change." *Language and Education* 22 (5): 307-326.
- Le Roux, Kate, and Jill Adler. 2016. "A Critical Discourse Analysis of Practical Problems in a Foundation Mathematics Course at a South African University." *Educational Studies in Mathematics* 91 (2): 227-246.
- Lerman, Stephen. 2000. "The Social Turn in Mathematics Education Research." In *Multiple perspectives on Mathematics Teaching and Learning*, edited by Boaler, Jo: 19-44. New York: Ablex.
- Danny Bernard Martin. 2019. "Equity, Inclusion, and Antiracism in Mathematics Education." *Race Ethnicity and Education* 22 (4): 459-478.
- Mangez, Eric and Mathieu Hilgers. 2012. "The Field of Knowledge and the Policy Field in Education: PISA and the Production of Knowledge for Policy." *European Educational Research Journal* 11 (2): 189-205.
- Martin, James R. 1996. "Waves of Abstraction: Organizing Exposition." *The Journal of TESOL France* 3 (1): 87-104.
- Martin James R. 2002a. "Meaning Beyond the Clause: SFL Perspectives." *Annual Review of Applied Linguistics* 22: 52-74.
- Martin, James R. 2002b. "Writing History: Construing Time and Value in Discourses of the Past." In *Developing Advanced Literacy in First and Second Languages: Meaning With Power*, edited by Schleppegrell, Mary J., and M. Cecilia Colombi, 87-118. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mesa, Vilma, and Peichin Chang. 2010. "The Language of Engagement in Two Highly Interactive Undergraduate Mathematics Classrooms." *Linguistics and Education* 21 (2): 83-100.
- Morgan, Candia. 2014. "Understanding Practices in Mathematics Education: Structure and Text." *Educational Studies in Mathematics* 87 (2): 129-143.
- Morgan, Candia. 2016. "Studying the Role of Human Agency in School Mathematics." *Research in Mathematics Education* 18 (2): 120-141.
- Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *The Review of Economics and Statistics* 77 (2): 251-66.

- Organisation for Economic Co-operation and Development (OECD). 2012. *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013a. *First Results from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013b. *Skilled for Life? Key Findings from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013c. *The Survey of Adult skills: Reader's Companion*. Paris: OECD Publishing.
- OECD. 2013d. *Time for the U.S. to Reskill? What the Survey of Adult Skills Says*. Paris: OECD Publishing.
- OECD. 2016a. *Skills Matter: Further Results from the Survey of Adults Skills*. Paris: OECD Publishing.
- OECD. 2016b. *Technical Report of the Survey of Adult Skills (PIAAC)*. 2nd ed. Paris: OECD Publishing.
- Oughton, Helen M. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2.
- Ozga, Jenny. 2012. "Governing Knowledge: Data, Inspection and Education Policy in Europe." *Globalisation, Societies and Education* 10 (4): 439-55.
- Pais, Alexandre. 2013. "An Ideology Critique of the Use-value of Mathematics." *Educational Studies in Mathematics* 84 (1): 15-34.
- Pardoe, Simon. 2000. "Respect and the Pursuit of 'Symmetry.'" In *Situated Literacies. Reading and Writing in Context*, edited by Barton, David, Mary Hamilton, and Roz Ivanič, 149-166. London: Routledge.
- Perry, Kristen H., Donita M. Shaw, Lyudmyla Ivanyuk, and Yuen San Sarah Tham. 2018. "The "Ofcourseness" of Functional Literacy: Ideologies in Adult Literacy." *Journal of Literacy Research* 50 (1): 74-96.

- Peters, Ellen, Daniel Västfjäll, Paul Slovic, C. K. Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. "Numeracy and Decision making." *Psychological Science* 17 (5): 407-13.
- PIAAC Numeracy Expert Group. 2009. "PIAAC Numeracy: A Conceptual Framework." *OECD Education Working Papers* 35.
- Pongratz, Ludwig A. 2006. "Voluntary Self-control: Education Reform as Governmental Strategy." *Educational Philosophy and Theory* 38 (4): 471-82.
- Porter, Theodore M. 1996. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Ramirez, Gerardo, Hyesang Chang, Erin A. Maloney, Susan C. Levine, and Sian L. Beilock. 2016. "On the Relationship Between Math Anxiety and Math Achievement in Early Elementary School: The Role of Problem Solving Strategies." *Journal of Experimental Child Psychology* 141: 83-100.
- Rautalin, Marjaana, Pertti Alasuutari, and Eetu Vento. 2019. "Globalisation of Education Policies: Does PISA Have an Effect?" *Journal of Education Policy* 34 (4): 500-22.
- Rivera-Batiz, Francisco L. 1992. "Quantitative Literacy and the Likelihood of Employment among Young Adults in the United States." *The Journal of Human Resources* 27 (2): 313-28.
- Rogers, Rebecca, Elizabeth Malancharuvil-Berkes, and Melissa Mosley. 2005. "Critical Discourse Analysis in Education: A Review of the Literature." *Review of Educational Research* 75 (3): 365-416.
- Sanders, Ted, and Eve Sweetser. 2009. *Causal Categories in Discourse and Cognition*. Vol. 44. Berlin: De Mouton Gruyter.
- Schleppegrell, Mary J. 2012. "Systemic Functional Linguistics." In *The Routledge Handbook of Discourse Analysis*, edited by Gee, James Paul, and Michael Handford, 47-60. London: Routledge.
- Scribner, Sylvia. 1984. "Literacy in Three Metaphors." *American Journal of Education* 93 (1): 6-21.
- Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.

- Shaw, Donita, Kristen H. Perry, Lyudmyla Ivanyuk, and Sarah Tham. 2017. "Who Researches Functional Literacy?" *Community Literacy Journal* 11 (2): 43-64.
- Shibatani, Masayoshi. 2002. *The Grammar of Causation and Interpersonal Manipulation*. Amsterdam, The Netherlands: John Benjamins Publishing.
- St. Clair, Ralf. 2015. "House of Cards: Analyzing 'Making Skills Everyone's Business.'" *Journal of Research and Practice for Adult Literacy, Secondary, and Basic Education* 4 (2): 37-42.
- Straehler-Pohl, Hauke, Sainza Fernández, Uwe Gellert, and Lourdes Figueiras. 2014. "School Mathematics Registers in a Context of Low Academic Expectations." *Educational Studies in Mathematics* 85 (2): 175-199.
- Street, Brian V. 1993. *Cross-cultural Approaches to Literacy*. Vol. 23. Cambridge: Cambridge University Press.
- Tapiero, Isabelle, Paul van den Broek, and Marie-Pilar Quintana. 2002. "The Mental Representation of Narrative Texts as Networks: The Role of Necessity and Sufficiency in the Detection of Different Types of Causal Relations." *Discourse Processes* 34 (3): 237-58.
- Tsatsaroni, Anna, and Jeff Evans. 2014. "Adult Numeracy and the Totally Pedagogised Society: PIAAC and Other International Surveys in the Context of Global Educational Policy on Lifelong Learning." *Educational Studies in Mathematics* 87 (2): 167-86.
- Tunstall, Samuel Luke. 2016. "Words Matter: Discourse and Numeracy." *Numeracy* 9 (2): Article 5.
- Tunstall, Samuel Luke, and Matthew Ferkany. 2017. "The Role of Mathematics Education in Promoting Flourishing." *For the Learning of Mathematics* 37 (1): 25-28.
- U.S. Department of Education, Office of Career, Technical, and Adult Education. 2014. *Making Skills Everyone's Business: A Call to Transform Adult Learning in the United States*. Washington, DC: U.S. Department of Education, Office of Career, Technical, and Adult Education.
- Valero, Paola and Robyn Zevenbergen. 2004. *Researching the Socio-Political Dimensions of Mathematics Education: Issues of Power in Theory and Methodology*. Vol. 35. Boston: Kluwer Academic Publishers.

- Van Dijk, Teun A. 1993. "Principles of Critical Discourse Analysis." *Discourse & Society* 4 (2): 249-283.
- Wagner, David. 2012. "Opening Mathematics Texts: Resisting the Seduction." *Educational Studies in Mathematics* 80 (1-2): 153-169.
- Wagner, David, and Beth Herbel-Eisenmann. 2008. "'Just Don't': The Suppression and Invitation of Dialogue in the Mathematics Classroom." *Educational Studies in Mathematics* 67 (2): 143-57.
- Waidzunus, Tom. 2012. "Young, Gay, and Suicidal: Dynamic Nominalism and the Process of Defining a Social Problem with Statistics." *Science, Technology, & Human Values* 37 (2): 199-225.
- Waldow, Florian. 2009. "What PISA Did and Did Not Do: Germany After the 'PISA-shock.'" *European Educational Research Journal* 8 (3): 476-483.
- Wodak, Ruth, and Michael Meyer, eds. 2015. *Methods of Critical Discourse Studies*. London: Sage.
- Yasukawa, Keiko, and Stephen Black. 2016. *Beyond Economic Interests: Critical Perspectives on Adult Literacy and Numeracy in a Globalised World*. Vol. 18. Amsterdam, The Netherlands: Sense Publishers.
- Zinshteyn, Mikhail. 2015. "The Skills Gap: America's Young Workers Are Lagging Behind." *The Atlantic*, February 17.
<https://www.theatlantic.com/education/archive/2015/02/the-skills-gap-americas-young-workers-are-lagging-behind/385560/>.

Study Three: College Students' Numeracy Events and Discussion of Public Issues in Focus Groups

Introduction

In 2015, there were just over 3.4 million enrollments in mathematics courses at or below the calculus level at two- and four-year post-secondary institutions in the U.S. (Blair, Kirkman, and Maxwell 2018). Given that roughly 85% of U.S. college students graduated contemporaneously with degrees outside the fields of science, technology, engineering, and mathematics (STEM) (Snyder, de Brey, and Dillow 2018), it is reasonable to infer that the majority of those 3.4 million students enrolled in such courses to fulfill a general education graduation requirement. Though the labels for such general education graduation requirements for mathematics differ (e.g., quantitative reasoning, quantitative literacy, or simply mathematics), the courses that tend to fulfill the requirement traditionally include College Algebra, Precalculus, Mathematics for the Liberal Arts, Calculus, or Statistics, among others.

A common rationale for requiring students to complete at least one such course as part of their general education is so that they might learn mathematics or quantitative skills that they can draw on for use in their professional, day-to-day, or civic lives (Hamman 2017; Steen et al. 2001). With that said, given that College Algebra and Precalculus may be viewed as preparation for Calculus (Gaze 2018), there has been pushback from students and faculty in the past three decades

concerning the extent to which College Algebra and Precalculus are actually relevant to students' careers or daily lives (Hastings 2006; Steen 1997; Steen 2001; Tunstall 2018). It is in this milieu of policy and decision-making that credit-bearing courses specifically centered on numeracy (though perhaps labeled differently) have begun to emerge at small (e.g., Gaze 2014) and large institutions alike (e.g., Tunstall et al. 2016).

Because students often enroll in a general education mathematics course based on their intended major, rather than just their prior mathematical preparation, the student population that enrolls in a numeracy-focused course tends to be quite diverse in terms of mathematical background, career path, and intended degree, among a host of other forms of diversity. With that point in mind, a question that bears relevance to educators involved in the teaching and development of those courses is how a numeracy-focused course can serve students whose diverse experiences and incoming skill sets have led them to such a course. Indeed, a common setup for a numeracy-focused course is for students to learn skills related to mathematics or statistics (e.g., proportional reasoning, basic data analysis and visualization) and to connect those skills with issues of relevance to students. For this reason, as students come to a course centered on issues related to their everyday lives, it is important that the course meet students where they are—not just in terms of mathematical content, but also in the way that students already think about or relate to those everyday issues. While there exists some research concerning how

students in numeracy courses reason with personal or socially relevant issues, there is a need for research that specifically considers the types of numeracy events students engage in when reasoning with such issues outside of a formal classroom setting (Baker 1996; Craig and Guzmán 2018). Note that a numeracy event is an event (whether spoken, written, or thought) that is mediated in some way by quantification; insofar as these events form a basis for the broader practices that many educators seek for students to engage in, it is important that we understand the nature and types of numeracy events that students already tend to engage in. With that idea in mind, the purpose of the present study was to explore how a small group of undergraduate students engaged with public issues—a specific context for numeracy practices—in a focus group setting. By public issues, I mean issues of likely interest to a given community and that potentially affect that community's functioning or well-being, such as a county vote on school funding, or a neighborhood debate on allowing electric scooters on sidewalks.

Building from previous work examining how students in Quantitative Literacy courses reasoned with public issues during semi-structured interviews (Tunstall, Matz, and Craig 2018; Tunstall et al. 2016), this study explores the following research questions: How do students who have the option to enroll in a Quantitative Literacy course discuss public issues in a focus group setting? Do numeracy events occur as they articulate their reactions? If so, what are the characteristics of these numeracy events? Note that I use Baker (1996) and Craig and Guzmán's (2018)

notion of numeracy events here to help identify (though not guarantee) broader numeracy practices. Exploring questions such as these may support work in the field of numeracy education by (a) specifically attending to ways students think about issues in informal settings (i.e., outside of a classroom context where they might be expected to use quantitative reasoning), (b) considering the nature of numeracy events in such settings, which could provide educators with insights into how to build from such events, and (c) utilizing a practices perspective for research (which is mostly novel to the U.S. numeracy community).

The next section of this paper begins with a brief review of literature concerning public issues and numeracy practices. In that review of literature, I elaborate on existing work at the intersection of numeracy practices and public issues, specifically homing in on what we know about how individuals use quantitative reasoning (or not) in the consideration of everyday topics, and more specifically public issues. I then present the study's data collection and analysis methods. Then, I elaborate on these two key results from students in these specific groups: (1) students leveraged their background experiences and knowledge in articulating standpoints on public issues, and in relation, students actively expressed critiques and questions as they engaged with the three artifacts; and (2) numeracy events, when they did occur, were primarily centered around students' acknowledgements of the importance of numbers, more so than their active articulation in building on that importance through group conversation. As I will

discuss later, an implication of (1) is that when discussing public issues in mathematics courses, it appears to be important to use an interdisciplinary lens and to elicit students' preexisting beliefs and understandings about the issue at hand. The result of (2) indicates that there is both promise in, and a need for, structured means of exploring issues through a quantitative lens with students in classrooms centered on quantitative literacy.

Numeracy and Public Issues

The message presented in many calls for numeracy or quantitative literacy is that because the world of the twenty-first century is “awash in numbers” (Steen et al. 2001, 1), it is imperative that our citizenry be numerate for their own sake, that of their communities, as well as for democracy itself. While there are many diverse contexts in which one might practice numeracy (e.g., a grocery store, browsing social media), a common context of interest to those involved in numeracy education is civic life (Briggs 2018; Erickson 2016; Hamman 2017; Mellow 2018; Steen et al. 2001). Indeed, with quantitative rhetoric saturating advertisements, political discourse, and many of our everyday conversations (Porter 1995), this interest is not surprising. Stepping aside from the question of the necessity of numeracy for participation in civic life—a topic that has generated debate among numeracy scholars (see the discussion between Hamman 2017 and Erickson 2017)—I argue that it is important that we understand how students tend to discuss

public issues if we seek to engage with students in the discussion of public issues in coursework centered on numeracy. Regardless of whether one takes a functional or practices-focused approach to numeracy,²² the use of numeracy for citizenship, and in particular, the discussion of public issues, is worthy of our attention.

Outside of the normative work alluded to above regarding the import of numeracy for deliberating public issues, there have only been a few studies to date concerning *how* individuals use numerical information or quantitative reasoning in interacting with public issues (whether alone or in a group setting) outside of a classroom setting. Moreover, while that research does help us in understanding the *how* of my first research question, it does not specifically address individuals' interactions with issues through a numeracy events lens—the subject of my second and third research questions. I make this claim from having searched for literature with the following terms in the abstract (using pairs of terms, with one from the first three and one from the last three, for a total of six searches): numeracy, quantitative literacy, quantitative reasoning, public issues, decision making, and deliberation. Note that my rationale for caring about whether such work is done in a formal classroom setting is that prior research suggests that the ways in which students interact with quantitative or mathematical information outside of

²² Note that numeracy has traditionally been defined as facility in mathematical skills for use in everyday life—a functional analogue to literacy (Cockcroft 1982, as cited in Karaali, Hernandez, and Taylor 2016). The term practices comes from literacy studies, and conveys the patterned (or regular) ways in which individuals take up literacy in their lives (Barton and Hamilton 2000).

classrooms is often different from how such interaction might be supported or taught in a classroom setting. For example, research from Carraher, Carraher, and Schliemann (1985) found that street vendors in Brazil could perform complex mental calculations using procedures distinct from those taught in school; those same vendors performed significantly worse when completing those problems if required to use equivalent formulations typical to school mathematics. In a different context, Murtaugh (1985) found that problems encountered by grocery shoppers—those which could be formulated as basic arithmetic calculations—were not viewed by the shoppers as such, but rather were approached in manners consonant with the situation and the specific problem the shoppers desired to solve.

The research that has been done concerning numeracy and interaction with public issues of relevance to this study has been in the last five years and completed by those studying decision making and psychology. For example, in a study of the interaction between numeracy scores (as measured by a multiple-choice test and Likert-scale questions) and hypothetical decision-making when presented with data on various topics, Kahan and colleagues (2017) found that individuals with higher numeracy scores were more likely to make correct inferences when the topic was of seemingly little public importance (in this case, skin rashes); however, when presented with data on the controversial topic of gun control, nearly all participants made inferences aligned with their existing political beliefs, with the more numerate individuals using quantitative reasoning selectively to reach their

conclusion. Nurse and Grant (2019) found similar results in the context of study participants' decisions surrounding climate change. These studies contribute to a burgeoning study of the concept of motivated reasoning in psychology (Redlawsk 2002), which is the notion that when making new decisions, individuals tend to arrive at decisions that adhere to their existing belief structures, notwithstanding their existing numeracy skills (among other things). While valuable, all of this research from psychology and decision-making studies has (by necessity) been done using large sample sizes and purely quantitative measures. At the same time that such research provides us with potential patterns we might expect among the general public in thinking about public issues, it does not capture the nature of numeracy events as (or if) they occur in individuals' reasoning. Moreover, it does not tell us what types of evidence individuals gravitate towards when interacting and discussing artifacts that contain a variety of information formats (e.g., headers, pictures, graphs, text, author information). The research that I do here allows for individuals to respond in a variety of ways (e.g., on paper, and through group discussion) to artifacts, and so opens up the possibilities of what we might learn from participants in the research process.

Outside of this work in psychology and public policy, the only other study found in my search concerning numeracy and public issues is the one from which the present study builds. Stemming from a desire to better understand how students in a quantitative literacy course at Michigan State University were *already* thinking

about public issues before taking the courses, in 2016 I engaged in semi-structured interviews with five students enrolled in the course. In the interviews, we discussed their reaction to several media artifacts (e.g., a tweet, newspaper article, or a video). The specific question that we explored as part of the study was the following: How do the students in our interviews reason quantitatively—if at all—when asked to articulate their reactions to artifacts concerning public issues? In our analysis, we found that students discussed the artifacts in a variety of ways, sometimes using quantitative reasoning and other times discussing the issue through a completely different lens. That is, the extent to which students referenced or used mathematics or quantitative reasoning varied from student to student (some not using it at all). For example, two students discussed the legality of stop-and-frisk, whereas one student talked about its disproportionate impact on certain groups. Moreover, we found that students tended to leverage their background experiences—whether related to ethnicity, age, or religion, among other things—to argue certain standpoints on public issues (Tunstall, Matz, and Craig 2018).

We did not use the framework of motivated reasoning to analyze the interviews, but the results of our study did suggest that individuals' reasoning with public issues were informed by what they thought of the issue before coming to participate in the study. There were also several areas for improvement in that study. One aspect of the design was that students were discussing the public issues with just me, rather than with peers. Discussing issues with me, rather than with a peer,

could lead students to respond in ways that they believe I would approve of as a graduate student and researcher. Of course, that problem does not go away with a focus group, but I believe it is mitigated by emphasizing to students that the group is meant to be a conversation among peers. Another drawback with that study is that students did not have space to write or collect their thoughts before discussing the issues; they also did not have the opportunity to influence our analysis after the interviews had taken place. The present study builds upon that study by explicitly using a numeracy practices and events framework, and by improving the methods used for collecting data.

Taken together, this existing work in anthropology (Murtaugh 1985), mathematics education (Tunstall, Matz, and Craig 2018), and psychology (Carraher, Carraher, and Schliemann 1985; Kahan et al. 2017; Nurse and Grant 2019), suggests that students' prior experiences and beliefs would likely contribute to how individuals discuss public issues in focus groups discuss. What it does not tell us, though, is how we might expect numeracy events to occur as students discuss issues in a group setting. As part of the previous study that I engaged in, only some students (and for some artifacts) used quantitative reasoning (and thus broached numeracy events) when thinking out loud about the artifacts. Now, in a group setting, it remains unclear what will happen if one student begins a numeracy event for others to then build on (or not). There are added layers of complexity that

this study provides by having students engage in focus groups and by generating multiple forms of data (to be discussed below).

Method

Research questions that guided this study were the following: How do students who have the option to enroll in a Quantitative Literacy course discuss public issues in a focus group setting? Do numeracy events occur as they articulate their reactions? If so, what are the characteristics of these numeracy events? To examine these questions, I formed focus groups of students (Liamputtong 2011) and followed a semistructured, think-aloud protocol (Thelk and Hoole 2006). As noted by Liamputtong (2011), a focus group methodology is useful in “exploring and examining what people think, how they think, and why they think the way they do about the issues of importance to them without pressuring them into making decisions or reaching a consensus” (7). Though I could not guarantee that the artifacts I chose would be of interest to the students, I did (as discussed below) intentionally choose topics that I thought would be of importance to college students in Michigan. The purpose of having students in focus groups, rather than engaging in conversations one-on-one or having a formal group interview, was to simulate the reality that we rarely think about public issues in a vacuum. While the focus group method does lead to a potential loss in the depth of information obtained from any single respondent, the benefit is that it captures what is more

likely to occur in a deliberative setting. In relation, another rationale for having focus groups was that students might be more comfortable in expressing their feelings amongst individuals whom they view as peers, as opposed to me, whom they may view as a figure with power (Madriz 2003).

Participants

Insofar as there was no intention in this study to generalize from the focus groups, and I wanted to make sure that all students were compensated for their time, I chose to hold two focus groups with four students in each group. Having obtained approval for the study from the Institutional Review Board at Michigan State University (MSU), in Summer 2018 I worked with MSU's Registrar's Office to obtain a list of students who were enrolled in their first mathematics or statistics course in Fall Semester 2018 and had not completed the University's mathematics requirement (this means students placing out of mathematics altogether were not included in the selection pool). The rationale for this delimitation of the selection pool was that I did not want to include students who may have already been in a college course focused on mathematics and public issues. Using the Registrar's list, I reached out to students via email, offering them a \$25 Amazon gift-card for their time (a maximum of two hours), and telling them that we would be discussing public issues for a dissertation project. Given the large student body enrolled at MSU, it did not take long to have enough students interested, and so I scheduled the two groups for the days preceding the start of the fall 2018 semester. In Table

3 below, I provide a list of the students (using pseudonyms). The column, “Self-described identity markers,” is based on information I asked of students through a Google Form at the end of the focus group; I expand on the rationale for that request when discussing the Protocol in a later section.

Table 3: Focus Group Participants

Name	Major/minor	Year	Self-described identity markers
Group 1			
Pashiel	Human Biology	Sophomore	Logical, liberal, and open-minded
Alexa	International Relations	Junior	Victim, phoenix, and advocate
Layla	Political Science/African Studies	Junior	Black, woman, and socialist
Savannah	Experience Architecture	Sophomore	Christian, female, and student
Group 2			
Teta	Nursing	Sophomore	Female, activist, and Rwandese
C	Theater and English Education	Junior	Liberal, feminist, and environmentalist
Jayla	English	Sophomore	Female, Black, and Detroit native
Ash	Political Science	First year	Passionate, gamer, and Michigan native

As one can see in Table 3, participants varied in their program of study and the year they were in at MSU. Next, I discuss the characteristics of the focus groups, including the protocol used and the data collected.

Focus Groups

Protocol. For both focus groups, the meeting began with a few minutes of informal (and unrecorded) time for folks to get to know one another and to have light refreshments. During this time, I explained to students in general terms the purpose of the study, noting that I would provide more information—notably my research

questions and positionality in relation to the work—after we had finished talking about the public issues. Once recording began, we went around more formally to introduce ourselves. We then went through three cycles of examining and discussing a media artifact. For each cycle, students (1) read or watched the artifact alone, (2) independently filled out a section on a Google Form, and then (3) discussed the artifact with the whole group.

In relation to (1), students were instructed to annotate their copy of the article by circling things they found interesting or important, and by writing down any questions that came up for them; for the video, students had a blank piece of paper to write down their thoughts. Students knew that their work would be collected and later scanned. For (2), students used a laptop or tablet to answer the question, “What’s your general reaction to what you’ve just read or watched? Please include as much detail as you can—your response does not have to be polished or ‘academic.’” Following (1) and (2), our group share-out consisted of students first telling the group what they wrote down on the Google Form (each one in turn, going in a random order, so a different person spoke first for each artifact), and then discussing a question I had prepared specific to the artifact itself. After going through three cycles of artifact discussion, students independently filled out a final Google Form and received their compensation. This final Google Form allowed for me to confirm students’ fall semester mathematics enrollment and to ask for a few identity markers that were important to them. I explained to students in person that

that might be helpful for me in thinking about why they responded in the ways that they did. I then told students that, if possible and they had time, I would be following up with them once I had completed my analysis to see if they had any questions, concerns, or suggestions about the ways I had interpreted their participation.

Media Artifacts. As noted above, there were three artifacts that each focus group examined. The two articles are reproduced in the Appendix A, and the news video is provided in a footnote. Students were presented the artifacts in full, with no information (e.g., source, date, author, etc.) excluded from it.

The first artifact was an article from *Mother Jones* (Sonde 2018) concerning a recent lawsuit between Monsanto and a man who had been diagnosed with terminal cancer; the article's title was, "The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal." I chose the article because it was short (less than one front-and-back page printed), it took a clear stance on the topic (as evidenced through opinionated statements throughout), it had the potential for students to use quantitative reasoning in discussing it (there were several statistics used as part of the argument), and it would likely be of interest to students. Based on personal experience, I have found that college students talk about the safety of their foods; given that this article covered a potentially controversial topic, I thought that it might generate hearty discussion. As part of our group share-out, the specific prompt that I posed to the group was the following: Should we—the general

public—be concerned about “glyphosate” in products like cereals? What about in products like RoundUp? Explain why or why not.

The second artifact was a three-minute video²³ from *Vice News* entitled, “Charter vs. Public Schools And The Kids In The Middle.” In the video, correspondent Gianna Toboni traveled to Michigan to discuss charter and public schools with select parents and students, a school administrator, and the husband of U.S. Secretary of Education Betsy DeVos, Dick DeVos. Similar to the first artifact, I chose this video because it was short, it had a clear message, and its setting would likely engage the students in my focus group, given that many of them were from Michigan and had likely heard about Betsy DeVos in both local and national contexts. There were also a few statistics provided in the video that students might call attention to in their discussion. The prompts I asked in our group share-out were the following: What are factors that come to mind when you think about whether charter schools are a viable option to send one’s child to? Would you send your child to a charter school? Why or why not?

The last artifact was a brief Fox News article titled, “White Police Officers Don’t Unfairly Target Black Suspects, Study Says” (Derespina 2016). Similar to the first two artifacts, I chose the third artifact because it was short, contained a clear stance, and contained multiple statistical statements that students might utilize

²³ See <https://www.youtube.com/watch?v=cVYUTkYW4u4>.

in engaging with it. Given that it covered the topic of race and crime, I predicted that it might be the most provocative of the three (though notably, I did not know the race of students whom I would be working with before meeting them in person). The prompt I asked in our group share-out was the following: Do you believe that police officers should be required to wear body cameras? Situate your response in some way within the context of this article.

As one can discern from the protocol and the artifacts that I chose, a key aspect of the focus groups was that students were not prompted to use quantitative reasoning; the rationale for this was that I wanted to see how students would think about issues if they encountered them in a more realistic context outside the classroom or focus group space. If prompted to use numbers or quantitative reasoning in the focus group, students might respond in ways potentially different from how they might in daily life, as has been noted by others (e.g., Boersma and Klyve 2013). Of course, such an extrapolation (from a focus group to “daily life”) is impossible here, and I do not intend to suggest that the results can be divorced from the contexts of the focus groups. Nonetheless, I still aimed to simulate the ecology in which we think about public issues.

Data Collection. Taken as a whole, the data that I collected as part of the study included: (1) audio of the entire focus group, (2) students’ individual annotations of each artifact, (3) students’ Google Form responses to each artifact, (4) students’ responses to final questions on the Google Form (concerning the identity markers

and any final comments), and (5) responses to emails that I sent in following up on the study. Note that the group conversation in response to each artifact is captured through (1). Students' individual sense-making or private views (that they were willing to share with me) were captured in (2) and (3). The purpose of (4) was to gather some information about what students thought was important about themselves, and the purpose of (5) was to see if students had any comments or questions about the way I had analyzed the data, discussed in further detail below.

Analysis

The three research questions guiding this study invite a description of what occurred over the course of the two focus groups. Upon scanning students' annotated articles, compiling students Google Form responses, and transcribing the audio files of the focus groups, I read each of them multiple times. I then wrote a summary memo for each focus group, and a summary memo of each student's participation in the focus group. The summary memos for the whole group outlined my understanding of how the group went (e.g., students' engagement) and general reactions they had to each of the three artifacts; they were approximately one single-spaced page. As an example of what I mean by engagement, I noted that for the first focus group, students were initially quiet, seemingly because they had not met me nor each other before; I further went on to note that students interacted with each other more as they noticed connections with one another (e.g., I asked students what classes they were in before the formal commencement of the group). As an

example of what I mean by “general reaction” to the artifacts, I noted that the first group let out an audible sigh when they saw Betsy DeVos featured in the Vice News video.

Summary memos for the individuals also included a short note about the student’s engagement (e.g., Was the student outwardly enthusiastic or passionate about a given topic?), but focused more on my understanding of how students responded to each of the three artifacts, as evidenced in their in-group discussion and their Google Form responses. For example, for Savannah and the RoundUp article, I wrote:

In the group discussion and in her annotation, Savannah expressed concern about her consumption of certain foods in the past year that appeared to be related to Cheerios, such as oatmeal and breakfast bars. She stated on the Google Form that she sometimes felt she could taste the chemicals on her food, and she did not necessarily trust companies to keep her safety in mind. Savannah noted in the Group Discussion that her grandmother had always told her to rinse off certain foods before eating them due to the presence of pesticides on them. Taken together, she expressed uncertainty about what this article meant for her future food consumption.

I included in the memo numeracy events that I had identified, given that (as discussed below) I necessarily interpreted students’ thought processes in order to identify them. They were approximately half of a single-spaced page each. I emailed a copy of the individual and group summary memos to each of the eight students in December 2018, asking that they briefly check it to see if they had any questions or suggestions about what I had written. Five of the eight students

responded that the memos looked fine based on what they remembered, and three students did not respond.

The next stage of data analysis consisted of examining the memos, generating themes, and then revisiting the original data to look for any disconfirming evidence in relation to those themes. I discuss each of these steps below.

To initially arrive at themes, I examined the memos and wrote out—for each artifact and each student—what came across as characteristic of, or essential to, how students responded to the specific artifact in group discussion and in writing. The reason for using the memos as the primary source for generating themes is that they captured, in one place, all of the evidence sources related to students' responses. That is, they combined a student's annotation, Google Form response, and group discussion participation into one central source from which to look at the student's response to the artifact. Continuing with the example of Savannah and the first artifact, I wrote that characteristic of her response to the artifact was the following: Suspicious of companies; Unsure of some topics in the article; Discussed family member's experiences. I acknowledge that what I viewed as characteristic of a student's response may not be exactly how the student would have self described their response; with that said, by sending the summary memos to students, I did aim to reduce the possibility of inaccuracies that might emerge in subsequent interpretations. Savannah was one of the students who responded and

noted that she did not have any concerns about the summary memo from her participation.

This reduction in data allowed for me to create a table of characteristics of students' responses, which are provided in Appendix B. After having identified these characteristics, I looked for similarities among them. For example, for many students their prior experiences with family informed their responses; another characteristic was prior experiences more broadly with respect to race. Both of those characteristics (in addition to others) could be grouped under the larger category of "Prior experiences," which is why I used that label as a broader theme. I am aware of the possibility that different researchers may have developed different categorizations of the data. I acknowledge this in my analysis by discussing the nature of students' contributions as specifically as possible so that their individuality is not lost (e.g., highlighting race, as opposed to just referring to it as part of a student's background). Furthermore, though I aimed to examine only those memos to come up with these themes (i.e., to let the themes emerge from the data of this study), I was nonetheless inevitably influenced by the prior study (Tunstall, Matz, and Craig 2018) that I had completed with colleagues. To reduce the impact of that study on this analysis, I avoided reading that work (or its associated data) for three months leading up to this analysis.

With respect to the second and third research questions, data analysis was informed by Baker (1996) and Craig and Guzmán's (2018) definitions of a

numeracy event, which is an event mediated in some way by quantification. In light of Baker's (1996, 81) notion that, in a numeracy event, quantification is "integral to the nature of the participants' interactions and their interpretive processes," there was necessarily some interpreting that I had to do to identify numeracy events. This is why it was important that I include numeracy events in my memos to students. In going through all of the data generated from the focus groups, I identified numeracy events in relation to their source (e.g., an annotation of the news article, or a student's statement in the group discussion), and then considered moments before and after that source that seemed to have an impact on student's writing or the group's conversation. For example, I noticed in Pashiel's annotation of the *Mother Jones* article that she had circled two quantities, and that she followed up by briefly alluding to the discrepancy between them in her written response; this was a numeracy event, as it was an instance of quantification influencing a participant's interpretation of an artifact. Identifying this in the data triggered me to make sure that I checked to see if she followed up on this in the group discussion of the artifact (which as I will discuss below, she did not). Note that I did this for all participants and all numeracy events. These identified events served as the unit of analysis which I describe in the Findings section.

Note that, based on the definition of a numeracy event, numeracy events are not always observable, as an event can be mediated by quantification without there being a remnant of that mediation in the form of audio, video, or text. For example,

a student's reaction to the RoundUp article could be heavily influenced by a statistic mentioned at the beginning of the article, and yet the student might not state that in their subsequent individual reflection or in the group discussion. Hence, a limitation of the approach outlined above in identifying numeracy events is that I could not ascertain events mediated by quantification that did not manifest in students' writing or audio through an explicit reference to quantitative information. For example, a student may have read a passage containing data which then influenced their response in a specific direction; I do not capture that as a numeracy event if the response itself did not include quantitative information. Though this limitation is nontrivial, I did aim to circumvent it by providing students numerous opportunities to voice their interpretations of (and responses to) the artifacts. An additional limitation to bring up here is that numeracy events are not the same as numeracy practices, which are patterned (or repeated) things individuals tend to do in numeracy events. We discern practices through events, but from these data alone I cannot make claims about students' actual numeracy practices; I can only make inferences about what those practices might be, if we assume that students do such things regularly or in similar contexts in their day-to-day lives.

Findings

I have outlined the results from the study by research question. Within each research question's section, I organize the discussion by themes that emerged from

analysis of the data. Note that I do not claim that any emergent themes are to be generalized for these students to other contexts, or from these students to other students. I am referring to the specific students in this study discussing the issues from these focus groups. I encourage the reader to make sense of the data dialectically in relation to their own experiences.

Question One: How Students Discussed Public Issues

Leveraging Prior Experiences. Students' prior experiences had a salient impact on how they talked about the three public issues in writing and with peers in the focus group. This theme was robust across the eight students and the three artifacts. Where prior experiences had left marked impressions, students' views appeared to be stronger and more adamant.

With respect to the first artifact (concerning glyphosate), students' existing experiences and understandings of chemicals, foods, and cancer influenced how they responded to the issue discussed in the article. For example, Alexa made clear in her annotation (Fig. 6), her writing, and her dialogue in the focus group that she was not in favor of the use of pesticides in foods: "If you ask me, pesticide consumption is just as bad as eating paint or super glue or hand sanitizer."

The EPA, which considers Roundup "not likely to be carcinogenic to humans," deems an exposure rate of 2 milligrams per kilogram of body weight per day acceptable. EWG, however, calculates that the highest permissible level of exposure, based on the Food Quality Protection Act's margin of safety for childhood risk, is 0.01 milligram per kilogram of body weight per day.

any is unacceptable

Figure 6: A Snapshot of Alexa's Annotation of the *Mother Jones* Article on Glyphosate in Foods.

As one can see in Figure 6, her comments in the focus group aligned with those from her annotation of the article. Though Alexa did not describe why she had the views that she did, note that they were likely influenced by prior experiences (i.e., not just the present article itself), given that the article (on its own) would not have led her to make the comparison to paint or super glue.

Pashiel appeared to leverage her positionality as a working-class person in discussing the article with the group, stating: “I wonder why they don't care about our health. Do they put pesticides in cereal because they know that working-class people eat it and they want to do everything in their power to kill us slowly?” Similar to Pashiel, C also expressed the view that the corporations in the article were engaged in negligent behavior. C, having had prior encounters with this topic before, also noted (in her written reflection) that this was not the first time that she had read about Monsanto:

I dislike Monsanto as a corporation due to their shady business deals involving GMO seed and their fellow farmers. I automatically identify Monsanto as the evil-doer in this story, I do not believe their statements that the glyphosate is safe, I think that there should be tighter regulations on the types of pesticides used on crops. While Mother Jones is not (in my opinion) a credible news outlet, I do believe that the topic of pesticides is being filtered out of our news along with other climate reality issues that are being forced out of public scrutiny in order for the post-Trump administration to continue to profit while the environment is destroyed. There is an obvious money link between these corporations when scientific research is discredited in order to continue to make profit.

C not only held a strong conviction about Monsanto, but also utilized her understanding of the current political climate, as evidenced both in the quote and in her written in annotation of the article (Fig. 7).

Agency may not — Post-Trump EPA

The EPA, which considers Roundup "not likely to be carcinogenic to humans," deems an exposure rate of 2 milligrams per kilogram of body weight per day acceptable. EWG, however, calculates that the highest permissible level of exposure, based on the Food Quality Protection Act's margin of safety for childhood risk, is 0.01 milligram per kilogram of body weight per day.

Figure 7: An Excerpt of C's Annotation of the *Mother Jones* Article on Glyphosate in Foods.

Furthermore, in her quote above, C also broaches the notion of the source of the artifact being an important consideration. When C brought up a similar sentiment in the focus group itself, none of the other folks in the group—Teta, Jayla, or Ash—had heard of *Mother Jones* before, and so they did not discuss the artifact source any further. Instead, all three expressed similar views to that of Alexa from above about the presence of pesticides in foods.

In the context of the first artifact, another source of prior experience that two students brought up was family members' experiences with related issues. Both Savannah and Ash referred to family members at least twice during the course of the focus group. Savannah, in thinking about the first artifact in particular, noted to the group:

I instantly thought about my grandmother again. She, uh, she rinses off her fruit with soap and water, or apple cider vinegar...I would ask like, you know, why are you doing all this? And she's like, you know, they put these pesticides on them and they linked back to cancer. And then, you know, when I read this and they're like, you know, it's possibly in our granola bars and oatmeal and all these other things. I just thought about

last year where, um, I was a freshman in every morning, you know, I looked for a quick fix and it was like oatmeal every other morning or cereal...How are we supposed to rinse that stuff off of cereal? I feel like that's just another way where they're trying to cover things up.

Layla, in building off of Savannah's remark and alluding to an earlier theme from

Pashiel concerning the working class, then added:

I don't know what rich people do, but I know that people in the working class, they always buy cereal, you know, and I would assume that, you know, if you, if you are a rich person, you have money to make an organic breakfast every morning so you don't have to eat cereal. So I kinda connected cereal to, like what tax bracket you're in and what you can afford in the grocery market.

When I followed up with Layla asking about this statement and her identity markers of Black, Woman, and Socialist, Layla noted that her identity as a socialist connected to that remark.

Students in the groups also drew upon their experiences strongly in discussing the third artifact (concerning race and crime). Prior personal experiences both with racism and with Fox News evoked strong reactions among the students in the focus groups. I discuss this artifact next because of its similarity to the first in eliciting student reactions. Given that all of the students disagreed with the argument of the article (both in their private written response and their public group response), students sought ways to speak back to the article, and prior experiences were among the first things students turned to. Five quotes representative of this theme are below:

Savannah: I am an African American, and even if I wasn't, I would be able to tell this article is based on vague evidence...I have witnessed 4-5

police officers tackling and killing a Black, and reportedly unarmed, male. Why, when you have tasers, guns, and professional training, did this man not see another day?

Jayla: This article is from Fox News, which has consistently proven to have a racist and right-wing bias. It overshadows the problem of police brutality as a whole by saying it is not the fault of White cops. Even if White cops don't have a 'personal, irrational bias,' racial targeting is something that's been institutionalized in American society.

Pashiel: Police have been killing Black people ever since the police force was created. There are plenty of books, movies, and witness testimonies to provide truth to that statement. Police have also been recorded killing Blacks and a majority of the time, they are not punished for it so it would make sense that in 2018, with racism still existing in America, a police officer would not be afraid to kill a Black person on camera. There has also been an increase in the amount of 'false alarm' calls that Whites have been making on Blacks to the police. There have been calls made to the police because the neighbors were too loud, a Black child was selling hot dogs outside, a family was cooking, etc. so I would like to know exactly what they consider in their data when they say that Blacks in areas of high crime are more likely to get killed.

Ash: In many of the recent police killings, the victims have been unarmed, innocent, queer, not of legal age. I think this article shows stereotypical upper-class, White news outlet trying to sweep the issue under the rug and does not illuminate the actual horrors being carried out by poorly trained officers of the law.

Teta: This is not so true I think. It makes me sad and really makes me uncomfortable that the news claims that White police do not unfairly target Black suspect.

Across these five quotes from students' written responses, we see evidence that students are employing prior experiences—whether related to eyewitness accounts (Savannah), knowledge from other media encounters (Pashiel and Ash), knowledge

from other academic spaces (Jayla), or combined understandings from all of these (Teta²⁴)—to express their take on the artifact.

As with the first artifact, it was also the case that family background influenced one student's written response. This time, though, it was Alexa, who was more reserved in discussing this artifact. In particular, Alexa noted that though she was Black, her uncle (who is also Black) was a police officer, and that the issue (according to him) is more complex than the article made it seem. To that end, in her written response she noted:

Even if it is subconscious, racism is an ongoing issue and sometimes we unintentionally make assumptions or act a certain way because of it. I don't really know how to solve this debate between whether or not police officers are unfairly targeting black suspects, and I don't know if body cameras are beneficial, but I respect police officers for the work that they do every day, assuming that they are all well-intentioned.

Later, I will return to Alexa's remark, as I was curious as to why she stated that she did not know if body cameras were beneficial. For now, the primary point to note from this quote is that she was drawing upon her experience with her uncle to understand the issue discussed in the article.

I discuss the second artifact (the *Vice News* clip on charter schools) last because it was somewhat less impactful on students in eliciting responses. In writing about and engaging in dialogue about this issue, students again drew on their prior

²⁴ Teta was somewhat less outspoken given that English was not her first language and she had only been living in the U.S. for two years. In a follow-up email with her, I asked her why she wrote what she did, and her response encompassed elements of the other three types of prior experiences I wrote about in that sentence.

experiences, but this time the overall lack of familiarity with the issue appeared to dampen students' contributions on the topic. As I broached the artifact to students after watching the video and had each student go around to give a brief overview of what they wrote, I asked students to also state for the group if they had ever attended a charter school; I also asked that they tell us if they felt confident about what the differences are between public, private, and charter schools. In going around one by one to share, two of the eight students (Savannah and C) stated that they felt confident about the differences between school types.

One (Savannah) of the eight students had actually attended a charter school. Subsequently, in responding to the group's prompt (concerning factors that are important for determining if a charter school is a viable option for one's child), Savannah drew from her experiences in a charter school to frame her written response:

I went to a Charter school from kindergarten all the way through 12th grade. I can be honest and say there was a good and bad...The video made it as though charter schools think they're "better," but why wouldn't you want better? Not all charter schools cost to attend (mine did not), and not all charter schools ignore parent input (mine did not).

And while Savannah did mention these factors in her participation in the group conversation, she—following suit with other students who were less familiar with charter schools—was also quick to point out that she did not agree with what she perceived to be Betsy DeVos's platform on charter schools, and that she was saddened to see so many public schools in Detroit shutting down. Jayla, though not

wholly familiar with the differences between school types, positioned herself as a Detroit native in her identity markers, and this came through in her spoken response for the focus group:

Betsy DeVos probably has no idea how an education facility should perform for its students. A lot of the children in the video seemed to be White and obviously able to afford the charter schooling. A large percentage of Detroit families are people of color and usually fall under the poverty line...yet they're coerced into false "choices" to either send their children to an expensive charter school that isn't actually proven to be better, or create another option for themselves once all of the nearby public schools close down.

Of the remaining six students not quoted with this artifact, five made similar (but seemingly less impassioned) negative comments about Betsy DeVos like Jayla. Teta, being unfamiliar with the U.S. school context, expressed reservations about the utility of charter schools, and stated that she did not know enough information to provide a substantive response. In some ways, this response from Teta serves as a fitting transition to the other major theme concerning how the students in these focus groups discussed the three public issues: they expressed curiosity, asked questions, and interrogated the messages of the artifact's creators.

Asking Questions. Whereas with the first theme, the broad message was that students leveraged prior experiences, understandings, and other elements of their background to discuss the public issues, the broad message of the second theme is that students pushed conversations forward by asking questions and interrogating the artifact at hand. Here, the annotated articles and Google Form inputs were

helpful for capturing students' responses, as they were more conducive to eliciting reflective reactions than in-the-moment verbal responses with the group. Indeed, recall that for each of the artifacts, I asked the students in the Google Form to answer the following prompt: "What's your general reaction to what you've just read or watched? Please include as much detail as you can—your response does not have to be polished or 'academic.'" It is telling, then, that across all eight students and 24 responses to that prompt, each student generated—without direction provocation—at least three questions as part of their reaction. With prompting (as part of the general process of annotation), students also asked several questions on their written annotation of the two articles.

With respect to the first artifact (concerning glyphosate), students' questions centered around three major points: the meaning of *carcinogenic*, the actual amount of glyphosate in their foods, and the discrepancy in the article between the Environmental Protection Agency (EPA) and the Environmental Working Group (EWG). Five of the eight students actually circled the phrase "probably carcinogenic" the first time that it appeared; two examples of this are provided in the annotations of Figure 8 below.

interesting choice of words

etienne voss/Getty

In a landmark ruling last Friday, a groundskeeper after a jury agreed terminal cancer. The jury also found effects of the chemical, glyphosate

Farmers and landscapers the world over, the International Agency for Research on Cancer in 2015, the International Agency for Research on Cancer found the chemical "probably carcinogenic to humans." Environmental Health Hazard Assessment

Probably carcinogenic to humans

etienne voss/Getty

In a landmark ruling last Friday, a California groundskeeper after a jury agreed that a terminal cancer. The jury also found the effects of the chemical, glyphosate. The

Farmers and landscapers the world over, the International Agency for Research on Cancer in 2015, the International Agency for Research on Cancer found the chemical "probably carcinogenic to humans," and Environmental Health Hazard Assessment

Figure 8: A Portion of the Annotations from Layla and Jayla of the Beginning of the *Mother Jones* Article.

Note: Here, both Layla (left) and Jayla (right) call attention to the phrase "probably carcinogenic."

Note that, as the facilitator of the focus group, I avoided telling students the meaning of carcinogenic when it came up as part of the group share-out, and instead asked if anyone else in the group knew the meaning of the word. In both focus groups, students appeared to converge in agreement that a substance was either cancer-causing or not, and that the phrase "probably carcinogenic" was a ploy from the World Health Organization to avoid forcing companies to stop using glyphosate in their foods.

Beyond a concern about the phrasing of statements in the article, students were also concerned—once they assumed any amount of glyphosate was cancer-causing—that the EPA and EWG had differing recommendations for an allowable amount of glyphosate in food. For example, before the group share-out, Ash noted:

I think I'll have to start rethinking what my foods are where they come from for now on. If one group thinks that more than 0.01 milligrams per kilogram of body weight is too much, why should another group be so adamant on saying the human body can allow more than that?

Ash shared a similar view during the group discussion. During the group's discussion, Teta's concern was similar to that of Ash, as she stated:

I am a little bit surprised about all of it. If glyphosate was a cause of cancer to a grown-up man, what's gonna happen to these kids who are eating cereals everyday of their lives? I think public health and everybody should start looking into this seriously.

Just as with the first theme from above, students' responses appear to be more impassioned (as evidenced in the number of questions asked) in relation to the third artifact. Because students did not have substantial prior knowledge or experiences about the context associated with the second artifact (charter schools), their questions were more general (e.g., Why are schools closing down?). I do not include those here.

It was in the third artifact that students' questions were most interrogative (as evidenced in their framing as leading) and numerous. As with the *Mother Jones* article, some questions related to the sources or groups mentioned in the article. For example, C questioned what the Crime Prevention Research Bureau (CPRB) was,

which had served as the leader of the study. C was the only one of the eight students who called attention to the CPRB in the written response or the focus group. Beyond her question about the source, most of the students' questions reflected a clear disagreement with the message of the article, and came across as interrogative of the author and his stance. For example, Pashiel was curious about how the author could make a claim about the measurement of the construct of racism (Fig. 9).

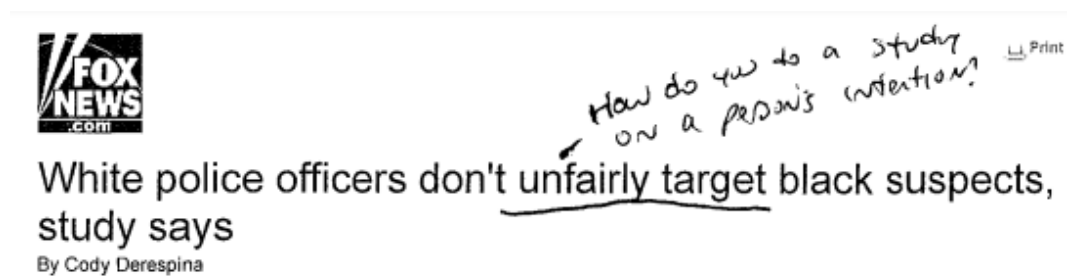


Figure 9: An Excerpt of Pashiel’s Annotation of the *Fox News* Article.

Note: Here, Pashiel questions how a study could tackle a construct that is dubious to measure.

During the focus group itself, Pashiel expanded on this concern, noting: “Like how do you know what they're telling you? What's actually true? Anybody can lie anyway, and of course you're going to lie if your job is on the line, of course you're going to say, oh no, it wasn't racially motivated.” In a similar vein, Layla built upon Pashiel’s remarks to question what was actually being studied. In particular, Layla noted that the study was only measuring police killings, not other metrics:

I'm going to say that this. There's a lot of inconsistency with the article. I don't understand why the title says target and then we're talking about shooting...There are different forms of targeting people. There's

harassment, clearly, but there are other, there are many other ways to do that...We're not mentioning whether they were armed or not, whether you know, what the threat of them actually was...And um, about the fact they said, there being higher crime rates in black neighborhoods. They didn't say what the crimes in the neighborhoods were like. They can say what type of crimes are crimes, general crimes aren't threats to your life. Like, you know, like robbery and all that. We could be measuring this in so many ways. Is this study just to support Fox's agenda?

In her quote, Layla raises a number of points as she speaks back to the author of the article (and the study itself). While most of her comments are not questions per se, they do raise questions and promote conversation, which is the nature of this theme in how students reasoned with public issues.

Taken together, across the two focus groups and the three artifacts, students leveraged their prior experiences and backgrounds to write about and discuss the three public issues; they also asked questions, often interrogating the artifact itself. Though not all that frequent, students also engaged in numeracy events, which I turn attention to in the next section.

Questions Two and Three: Numeracy Events and Their Characteristics

As described earlier, numeracy events are events mediated in some way by quantification (Craig and Guzmán 2018). In this study, such events are observable insofar as they are either written (in annotation or through the Google Form response) or spoken. In this section, I report findings related to the questions: Do numeracy events occur as students articulate their reactions? If so, what are the characteristics of these numeracy events?

The answer to the question, “Do numeracy events occur as they articulate their reactions?” is yes. In both focus groups and in relation to all three artifacts, numeracy events occurred. However, they were more commonly centered around students’ annotations of the two written artifacts than they were in the group discussion; furthermore, the importance of quantification or of numbers was more pronounced in the numeracy events when the quantification or numbers were central to the artifacts themselves (i.e., in this case, the *Mother Jones* and *Fox News* articles).

When and Where Numeracy Events Occurred. Both media articles contained quantities that students had the opportunity to make use of as part of their annotation, their Google Form response, or their reactions in the group share out. Such “latching on to” is characteristic of a numeracy event. An intriguing finding in this study was that these quantities manifest in students annotations more so than their Google Form response or verbal sharing. To expand on this finding, note that with the *Mother Jones* article, key quantities present in the article were the consumption recommendations for glyphosate in foods. These quantities included 2 milligrams of glyphosate per kilogram of body weight per day (the EPA’s recommendation), and 0.01 milligrams of glyphosate per kilogram of body weight per day (the EWG’s recommendation). All eight students in their written annotation of the article circled or commented on these quantities in some way. For example, whereas Savannah queried how much *she was eating* in her foods (Fig. 10), Pashiel

noted that we would not accept the same amount of “poop in our food” (Fig. 11); Layla, on the other hand, circled the two quantities and questioned why they were different (Fig. 12), as noted in the previous section.

products are safe, using the EPA's acceptable levels.
 “Our products are safe and without question they meet regulatory safety levels. The EPA has researched this issue and has set rules that we follow as do farmers who grow crops including wheat and oats,” wrote General Mills spokeswoman Kelsey Roemhildt in an email. “We continue to work closely with farmers, our suppliers and conservation organizations to minimize the use of pesticides on the crops and ingredients we use in our foods.”

Figure 10: Snapshot of Savannah’s Annotation of the *Mother Jones* Article.
Note: In this snapshot, we see that she asks how much glyphosate is present in the foods she is eating.

How much am I eating.
 If there is 0.01 mg of poop in our food, we're not gonna want to eat it, but it's okay for pesticides to be in our food

Figure 11: Snapshot of Pashiel’s Annotation of the *Mother Jones* Article.
Note: Pashiel wrote out (below the text, which is not included here) that while we allow 0.01 milligrams of glyphosate in our foods, we would not accept a similar amount of a different substance (in this case, “poop”).

A report released this morning by the Environmental Working Group (EWG) focused specifically on residue in food products containing oats. It found that many common breakfast cereals and snacks contain levels of glyphosate that it deems harmful, though the Environmental Protection Agency may not. so why's there conflicting views?
 The EPA, which considers Roundup “not likely to be carcinogenic to humans,” deems an exposure rate of 2 milligrams per kilogram of body weight per day acceptable. EWG, however, calculates that the highest permissible level of exposure, based on the Food Quality Protection Act’s margin of safety for childhood risk, is 0.01 milligram per kilogram of body weight per day. Very different again

Figure 12: Snapshot of Layla’s Annotation of the *Mother Jones* Article.
Note: In this portion of Layla’s annotation of the *Mother Jones* article, we see that Layla circled the quantities as important, noting they were “very different.”

From Figures 10, 11, and 12, we see students participating in an individual numeracy event insofar as they are interfacing with quantities to engage on their own with the article. The other five students engaged in similar ways with the quantities. As students transitioned from the annotation phase to their typed

responses concerning the article, only one of the eight students (Ash) brought up the actual quantities themselves; in that case, as quoted earlier, Ash had written: “If one group thinks that more than 0.01 milligrams per kilogram of body weight is too much, why should another group be so adamant on saying the human body can allow more than that?” Beyond this, though, across the two focus groups, the quantities themselves appeared only once in the group discussion (coming from Savannah).

When students shared their reactions to the article, Savannah noted: “...there's like a magic threshold. Oh well you can eat at least two milligrams per kilogram of body weight per day and they'll be fine. It's like, well no, I don't want to eat any.” The rest of the discussion in this focus group centered on the presence or not of glyphosate in the foods, and all of the discussion in the other focus group (which included Ash) centered around the presence or not of glyphosate in foods—not on the actual amounts in the recommendation (e.g., how much food one would need to consume to surpass the recommendation, based on their own weight).

With the *Fox News* article, this pattern—that of latching on to quantities or quantification in annotation, but not in the typed response or group discussion—appeared again, though there were a few notable exceptions. In Figures 13 and 14, Alexa and Jayla demonstrate (as examples of the pattern across the eight students) that quantities or quantification were important for engaging with the article itself.

Among the findings of the investigation released Thursday by the Crime Prevention Research Center: White police officers are not significantly more likely to shoot black suspects; body cameras have had little effect on decreasing police killings; the more cops at the scene, the less likely it is a suspect will be shot.

POLICE SEE HIGHER THREAT OF AMBUSH AFTER YEAR OF UNREST

The study examined data from 2013 to 2015, a time period which almost perfectly bookends the police killing of Michael Brown in Ferguson, Mo., in August 2014, an episode that helped lead to the rise of the Black Lives Matter movement and placed increased scrutiny on police shootings. Yet, the research team of John Lott and Carlisle Moody found the percentage of black suspects killed in the 19 months before Brown's death (24.8 percent) was almost exactly the same as the percentage killed in the 16 months after Brown's death (25 percent).

Figure 13: Alexa's Annotation of the *Fox News* Article.

Note: Here in annotating the article, Alexa first circles the statement (a quantifying statement) that more cops at a scene were associated with a suspect being less likely to be shot. We then see below that she circled two critical percentages presented in the article.

The study examined data from 2013 to 2015, a time period which almost perfectly bookends the police killing of Michael Brown in Ferguson, Mo., in August 2014, an episode that helped lead to the rise of the Black Lives Matter movement and placed increased scrutiny on police shootings. Yet, the research team of John Lott and Carlisle Moody found the percentage of black suspects killed in the 19 months before Brown's death (24.8 percent) was almost exactly the same as the percentage killed in the 16 months after Brown's death (25 percent). *Just because the percentage hasn't changed doesn't mean it's not happening at all*

Figure 14: Jayla's Annotation of the *Fox News* Article.

Note: Jayla remarks that just because the percentages (24.8 and 25) are relatively similar, that similarity does not justify the conclusion the author makes in the article.

Note that the annotations from Alexa and Jayla are exemplars of those of five of the other six students. While these numeracy events occurred as individuals annotated, students primarily focused in their typed responses and in their group sharing on disputing the claim of the article in other ways that did not allude to quantities or quantification from the article itself (e.g., bringing up personal experiences with racism, discussing recent events from the news). The notable exception to this pattern was from Layla, whose group share out in relation to this article I quoted earlier. Layla homed in specifically on the nature of what was

measured in her annotated article, her typed response, and in the group sharing.

Below is the typed response she provided after annotating the article:

...Committing a crime does not make you a high threat or that you're supposed to be shot. We need to look at other things such as injury of black people by police, harassment, shooting of unarmed black people and so forth. Not simply the number of those being shot. The mention of black police in the article, too, was also irrelevant. There was no follow through; it was only briefly mentioned as they didn't supply any data of what black police officers were doing. There are also inconsistencies in this article that were not in the other ones. For example, they lead with the mention of saying race plays no factor in police shooting, only to say that black people commit more crime and that's why they're shot more.

In the quote above, Layla not only interrogates the claims that the author of the article made, but also provides other directions for one to consider if they were to aim to follow up on the conversation with further study. Layla gave similar remarks in the group share out (quoted earlier). Though Layla does not reference quantities specifically in her remark, she notes what quantities should be considered (i.e., what should be measured), and disputes quantitative statements (e.g., “black people commit more crime”) that the author of the article made. In those respects, her typed response is an example of a numeracy event, as quantification mediated the way she approached the news artifact. Again, to summarize, across the discussions of the *Fox News* article in the two focus groups, Layla’s comments constituted the sole numeracy event. This remark is not to suggest that students’ discussions were not vibrant or engaged, but to highlight that they were mediated around other ways of talking about the issue. I discuss this finding in further detail in the next section.

I end this section by commenting on numeracy events surrounding the second news artifact (the video on charter schools from *Vice News*). Whereas with the two written media articles, quantities and quantification were arguably central to the authors' respective arguments, in the *Vice News* video, there were only two quantities mentioned: the number of charter schools opened in Detroit since 1995 (approximately 100), and the number of public schools closed in Detroit since 1995 (approximately 200). Students were instructed that since there was no article for them to annotate, they could write down comments about the video on a blank piece of paper as they watched. That being said, given the short length of the video and the fast-paced nature of the reporting, no students wrote down remarks on the piece of paper. Instead, they typed their responses in the Google Form, and shared them verbally during the group discussion.

With those two spaces for sharing in mind, note that there was only one numeracy event that occurred during either of the two focus groups. The numeracy event itself was from Layla's written response to the video, where she noted:

I liked that we were able to hear from both sides. The parent who felt forced to send her child to a charter school versus the man who worked for the school who argued that it was a choice and that they're being chosen by the people. It's interesting to see how the people who have the power choose to frame their words- how they speak about themselves. They say they're of the people. We heard two students speak but it makes me wonder what other children are saying, as well as how we can get more data on how beneficial charter schools are or are not. They said that test scores are the same, but isn't something else we can measure such as child surveys? It's also important to consider that the children may feel a sense

of superiority because of the freedom they have from their parents as well as what they're being taught by the higher ups about the schools.

In this response to the open-ended prompt, Layla describes how one could take action—in this case, through collecting data—to learn more about the extent to which charter schools are beneficial. Similar to that related to the *Fox News* article, the numeracy event here is again about the ways one measures a construct to analyze a situation. Note that when sharing out her remarks during the group discussion, Layla did not talk about these measurement questions; instead, in a manner similar to that of other students in the group, her comments were related to being confused about the nature of charter schools.

To summarize the results of this section, note that numeracy events did occur as students engaged with the public issues represented in the media artifacts. In both focus groups, students engaged with quantities pertinent to the artifact in their annotation more so than in their general typed response or their verbal sharing with the group. When numeracy events occurred in students' annotations, they were primarily to highlight that a specific quantity or statement was important in some way. Though the number of artifacts is too small to make any substantial claim, one should note that there were more numeracy events when quantities or quantification were central to the artifacts themselves (i.e., in the *Mother Jones* and *Fox News* articles). If a numeracy event encompassed more than highlighting as part of annotation, it was from Layla, who discussed measurement and how one might go about analyzing the issue in further detail. In the Discussion section below, I will

comment on why this may have been the case, and what these results suggest about broaching public issues in postsecondary courses centered on quantitative literacy.

Discussion

The research questions that guided this study centered around (1) *how* students in the focus groups discussed public issues, and (2) in relation *if*—and in what ways—numeracy events occurred as they engaged with those issues. The goal of (1) and (2) was to identify potential leverage points for discussing public issues with students in numeracy-focused courses . To that end, in this section, I discuss how the numeracy events from these focus groups might inform courses centered on quantitative literacy.

Numeracy Events

The numeracy events that occurred across these two focus groups demonstrate that students were attentive to the importance of quantities or quantification when exploring the two media artifacts where quantities or quantification were made apparent by the authors. However, as shown in students' Google Form responses and group discussions, the initial analyses that followed students' annotations was not centered around examining the issues from a quantitative lens. The one exception to this pattern was in the contributions of Layla, who consistently engaged with the issues from the perspective of measurement. Instead, most students' subsequent reactions were guided by students' interrogations of the

artifacts through other lenses, such as personal experiences, family connections, and other aspects of students' identities. In responding to the *Fox News* article to the entire group, Alexa, for example, originally stated that she was not sure if body cameras were beneficial, using her personal experiences to justify her reasoning, rather than engaging directly with the arguments put forth by the author about why they were beneficial (per his argument). While there is nothing inherently wrong with the ways students chose to approach the three media artifacts, from a normative standpoint an educator would be remiss if they were to suggest that all of students' reactions are ideal. Instead, all reactions—whether ideal or “incorrect”—can be didactical for us as educators (Pardoe 2000). For example, it is apparent that students might benefit from learning more about the issues themselves (e.g., carcinogens, charter schools), or more abstract concepts related to the issues (e.g., measurement in education) to develop informed analyses of the artifacts. At the same time, we as educators can learn from students' existing ways of approaching such issues, benefitting in particular from the rich information that students' family experiences and prior backgrounds can inform ways that we approach such issues in the classroom.

Insofar as a central goal of coursework centered around numeracy is to understand and analyze real-world issues from a quantitative lens — simultaneously recognizing the importance and limitations of reasoning quantitatively (Steen et al. 2001)—these findings suggest that students will benefit

from, and likely thrive in, coursework that allows them to build from their existing practices while engaging in learning about disciplinary norms for understanding real world issues through a quantitative lens. Though the construct of quantitative literacy is commonly said to defy any single discipline (Madison 2019), I use the phrase *disciplinary literacy* here because its meaning need not suggest that there is some platonic discipline one is becoming literate in, but rather, as noted by Moje (2015), that there are certain commonly accepted ways of approaching issues through the lens or tools of a specific discipline. In this sense, even if quantitative literacy is not tied to any specific discipline, one can still teach for quantitative literacy by exploring how both disciplines and broader modes of reasoning (e.g., scientific reasoning, historical reasoning) might approach issues from a quantitative lens. Engaging in this work is no trivial task for an educator.

With that in mind, Moje's (2015) 4Es framework for disciplinary literacy learning—engage, elicit/engineer, examine, and evaluate—may be of particular interest for those interested in drawing from students' existing knowledge and experiences (in the sense of what students leveraged in the focus groups in this study) to engage in the analysis of public issues through a quantitative lens. According to Moje, the 4Es serve as a guiding heuristic for how one might go about doing this. To that end, note that the first E, *engage*, is to suggest that classroom practices should resemble those of experts in the discipline under study; in the context of a quantitative literacy classroom, that engagement might resemble the

analysis process of a sociologist or a statistician, among other possibilities. The second E, *elicit/engineer*, is really two-pronged, and meant to convey the importance of eliciting students' existing knowledges and beliefs about a given topic (as discussed in Tunstall, Matz, and Craig 2018), and then engineering with that new knowledge aligned with how an expert might the public issue. For example, based on how students approached the *Fox News* article in this case, an educator could create a jigsaw activity in which students explored specific statistical concepts discussed in the article aligned with what students had found problematic when they first read it.

The third E, *examine*, is to remind us to carefully consider with students the discourse practices of experts in a given field. Continuing with the *Fox News* example, a class could consider the challenges journalists face in presenting quantitative arguments to the general public. Finally, the fourth E, *evaluate*, similar to the third E, entails thinking through with students specifically how certain discourse practices and ways of approaching issues are potentially more valued within a given discipline. Within the *Fox News* example, this would likely mean discussing with students how anecdotal evidence has limitations for making arguments that are scrutinized by the general public. Furthermore, in the context of a quantitative literacy classroom, that might mean explicitly discussing the import of certain ways of approaching issues from a quantitative lens, and the inherent limitations of doing so. Ultimately, one hope might be for students to recognize the

limitations of merely relying on anecdotal evidence (as some students did in these focus groups) to make publicly scrutinizable arguments pertaining to public issues. A future direction for work in the study of quantitative literacy could be to document this process suggested by Moje in the context of a course centered on quantitative literacy.

Conclusion

The study contributes to existing literature on numeracy while raising new ideas and questions in the nascent context of practices related to numeracy at the postsecondary level. There were two broad findings from this study: (1) students leveraged their background experiences and knowledge in articulating standpoints on public issues, and in relation, students actively expressed critiques and questions as they engaged with the three artifacts; and (2) numeracy events, when they did occur, were primarily centered around students' acknowledgement of the importance of numbers, more so than actively referring to or using that importance through group conversation, or using such quantitative information to perform calculations or further analysis.

The first finding, which aligns with results in prior work (Tunstall, Matz, and Craig 2018) poses promise for those teaching coursework in numeracy insofar as it demonstrates that students are likely to be engaged and excited to discuss public issues in postsecondary classrooms; this engagement should come as little surprise

for readers. In the context of coursework centered around numeracy or quantitative literacy, this finding makes clear that it is critical to acknowledge students' existing beliefs and practices as they relate to discussing public issues; without doing so, one loses out on leveraging students' rich, existing ways of thinking about such issues, which, as evidenced in the way students discussed issues in the focus groups, they are likely to do in more realistic (i.e., out-of-class contexts) ways. One also might attend to the idea of motivated reasoning discussed earlier in this paper, given that it is possible that students may use numeracy skills to arrive at conclusions aligned with pre-existing beliefs. The finding also suggests that it is important to explicitly discuss background information about various contexts; though at first this recommendation may seem difficult for mathematics instructors who may not have the background knowledge themselves for understanding various public issues, it is nonetheless critical given that public issues are not solely about mathematics—there are a host of factors to consider. The work of those involved in (for example) teaching mathematics for social justice (e.g., Bartell 2013; Brantlinger 2013; Nasir and Royston 2013), ethnomathematics (e.g., Meaney, Trinick, and Fairhall 2013), or culturally-relevant pedagogy more generally (e.g., Morrison, Robbins, and Rose 2008), can attest to this challenge and provide insights for those involved in teaching for quantitative literacy on how to approach common challenges that instructors face.

The second finding of this study indicates that there is both promise in, and a need for, structured means of exploring issues through a quantitative lens with students in classrooms centered on quantitative literacy. The numeracy events that occurred in the two focus groups of this study were primarily in students' annotations of articles, showing that students were aware of the potential importance of quantities or quantification as they went about completing the annotations. Nonetheless, because most of the students did not engage quantitatively with the artifacts in their typed responses or their group discussions, this suggests that they may benefit from structured experiences with engaging with public issues from a quantitative lens. Of course, given that I did not broach the public issues in these focus groups with any explicit quantitative focus, it is unwarranted to say that students could not do so if prompted.

Future work building from this study could explore student numeracy events in further detail, whether by explicitly prompting students to engage in using quantitative reasoning, or taking a different approach and reporting on how students engage in the process of learning about the analysis of exploring issues through a quantitative lens. The framework given by Moje (2015) reported in the Discussion is one structured means of doing so with students. As noted at the outset of this study, in 2015 there were just over 3.4 million enrollments in mathematics courses at or below the calculus level at two- and four-year post-secondary institutions in the U.S. (Blair, Kirkman, and Maxwell 2018). Because the majority of these

enrollments are for students to complete a mathematics requirement—one which often has the goal for students to engage with public issues through a quantitative lens—this study suggests that there is research and curriculum development to be done so that students have a robust and engaged experience in completing such a course.

APPENDICES

Appendix A: Focus Group Artifacts

8/20/2018

The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal – Mother Jones

Mother Jones

The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal

Glyphosate—deemed “probably carcinogenic”—sticks around on snacks and breakfast foods.

KARI SONDE AUG. 15, 2018 6:00 AM



etienne voss/Getty

In a landmark ruling last Friday, a California Superior Court ordered Monsanto to pay \$289 million in damages to a groundskeeper after a jury agreed that a key chemical in Monsanto's Roundup herbicide was a significant cause of his terminal cancer. The jury also found that Monsanto “acted with malice” in failing to warn him about the potential health effects of the chemical, glyphosate. The ruling paves the way for hundreds of similar cases pending across the country.

Farmers and landscapers the world over use glyphosate to kill weeds, but the chemical has long been controversial. In 2015, the International Agency for Research on Cancer, a part of the World Health Organization, classified glyphosate as “probably carcinogenic to humans,” and again defended that report this year. In 2017, California's Office of Environmental Health Hazard Assessment added it to their list of cancer-causing chemicals.

A series of independent studies in the past few years have revealed that small amounts of glyphosate residue end up on

And a series of independent studies in the past few years have revealed that small amounts of glyphosate residue end up on the packaged food sold in every aisle of the grocery store. “I have brought wheat crackers, granola cereal and corn meal from home and there's a fair amount in all of them,” wrote a

<https://www.motherjones.com/environment/2018/08/roundup-monsanto-glyphosate-cheerios-quaker-oats-cancer-1/>

1/3

8/20/2018

The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal – Mother Jones

the packaged food sold on every aisle of the grocery store.

Food and Drug Administration researcher in an email last year, according to the *Guardian*.

A report released this morning by the Environmental

Working Group (EWG) focused specifically on residue in food products containing oats. It found that many common breakfast cereals and snacks contain levels of glyphosate that it deems harmful, though the Environmental Protection Agency may not.

The EPA, which considers Roundup “not likely to be carcinogenic to humans,” deems an exposure rate of 2 milligrams per kilogram of body weight per day acceptable. EWG, however, calculates that the highest permissible level of exposure, based on the Food Quality Protection Act’s margin of safety for childhood risk, is 0.01 milligram per kilogram of body weight per day.

The EWG tested 29 products, including Cheerios and Quaker Oats granola bars, and found that many of them surpassed the level of glyphosate residue the organization deems safe. The study has limitations: The researchers relied on a small sample size and acknowledged there is a possibility of cross-contamination from nearby crops where the oats were grown. But given that these small amounts of residue accumulate, especially when delivered through foods as ubiquitous as breakfast cereals, the study suggests that parents may have a reason to be cautious.

“When you think about how people eat these oats,” says a senior science adviser at the EWG, “that’s when those numbers start to come into concern.”

“That does not mean, ‘Stop eating this bowl of Cheerios right now!’ But then when you think about how people eat these oats,” says Olga Naidenko, senior science adviser at the EWG, “that’s when those numbers start to come into concern.”

An FDA scientist presented data in 2016 that showed glyphosate on several packaged oat products, including some

for infants. An independent report by Food Democracy Now also points to Cheerios as containing higher amounts of glyphosate, though not at levels that the EPA would identify as unsafe.

When reached for comment, both Quaker Foods and General Mills, which markets Cheerios, responded that their products are safe, citing the EPA’s acceptable levels.

“Our products are safe and without question they meet regulatory safety levels. The EPA has researched this issue and has set rules that we follow as do farmers who grow crops including wheat and oats,” wrote General Mills spokeswoman Kelsey Roemhildt in an email. “We continue to work closely with farmers, our suppliers and conservation organizations to minimize the use of pesticides on the crops and ingredients we use in our foods.”

An email from a Quaker Foods spokesperson notes similar reasoning: “Any levels of glyphosate that may remain are significantly below any limits and well within compliance of the safety standards set by the Environmental Protection Agency (EPA) and the European Commission as safe for human consumption.”

Monsanto (which is now owned by Bayer) maintains that glyphosate does not cause cancer, and intends to appeal the California court’s decision.

This post has been revised.



KARI SONDE

Kari Sonde is a Ben Bagdikian editorial fellow based in the San Francisco office. Reach her at ksonde@motherjones.com or follow her on Twitter [@kari_sonde](https://twitter.com/kari_sonde).

<https://www.motherjones.com/environment/2018/08/roundup-monsanto-glyphosate-cheerios-quaker-oats-cancer-1/>

2/3



White police officers don't unfairly target black suspects, study says

By Cody Derespina

Published November 16, 2016

Fox News

The day after a Minnesota cop was charged in the July shooting death of a black man, an in-depth study purported to show race generally does not play a role in police shootings.

Among the findings of the investigation released Thursday by the Crime Prevention Research Center: White police officers are not significantly more likely to shoot black suspects; body cameras have had little effect on decreasing police killings; the more cops at the scene, the less likely it is a suspect will be shot.

POLICE SEE HIGHER THREAT OF AMBUSH AFTER YEAR OF UNREST

The study examined data from 2013 to 2015, a time period which almost perfectly bookends the police killing of Michael Brown in Ferguson, Mo., in August 2014, an episode that helped lead to the rise of the Black Lives Matter movement and placed increased scrutiny on police shootings. Yet, the research team of John Lott and Carlisle Moody found the percentage of black suspects killed in the 19 months before Brown's death (24.8 percent) was almost exactly the same as the percentage killed in the 16 months after Brown's death (25 percent).

"You might have imagined that if people were shooting black suspects for some type of racial animus, you might imagine there would be a change before and after Ferguson," Lott told FoxNews.com. "And yet [cops] seem to shoot, for black suspects, it's perfectly flat between the two periods of time."

Lott and Moody examined 2,699 police killings from more than 1,500 cities during the three-year time span -- including 1,333 killings not recorded in the FBI's official data. To gather the most precise information, the researchers scoured LexisNexis, Google, official police data and online databases.

The resulting spreadsheet also listed the race for suspects and officers. Putting together the racial component proved to be one of the more difficult tasks.

"News stories tend to not mention the race of the officer when the officer is black, because most of the black officers we found we found by looking at department photos, not news stories," Lott said. "That was not true of white officers."

Once the data was in place, the team controlled for various factors, including total population, racial makeup and, importantly, violent crime levels. A Washington Post article that gained popularity in the spring asserted that since black suspects made up 25 percent of those killed by police, but only 12 percent of the population, the disparity proved black people were more likely to be shot by police.

"However, crime rates differ across neighborhoods," the Lott and Moody study countered. "Black neighborhoods tend to experience higher crime rates. Therefore, race-neutral police randomly assigned to neighborhoods will encounter more criminal activities in black neighborhoods. As such, they can be expected to employ lethal force against a higher proportion of black suspects."

In regards to body cameras, which proponents have claimed would boost transparency and perhaps lead to fewer shootings, the new study concludes the devices had no effect.

"The additional evidence provided by the camera does not seem to alter the behavior of officers," the study stated. "This is also consistent with police not having a personal, irrational bias against black suspects."

Factors found to have increased a black suspect's odds of being killed included the violent crime rate of the area, the suspect's involvement in a property crime and police unionization. The duo concluded black suspects were 65 to 140 percent more likely to be shot by unionized police officers.

URL

<http://www.foxnews.com/us/2016/11/16/white-police-officers-dont-unfairly-target-black-suspects-study-says.html>

<http://www.foxnews.com/us/2016/11/16/white-police-officers-dont-unfairly-target-black-suspects-study-says.print.html>

Appendix B: Characteristics of Participants' Responses to the Three Artifacts

Table 4: Characteristics of Participants' Responses to the Three Artifacts

Name	Artifact One	Artifact Two	Artifact Three
Group 1			
Pashiel	Suspicious of government Concerned about working class Asks questions out of concern for health	Concerned about charter schools	Adamantly against message of the article Questions author's argument Cognizant of systemic racism
Alexa	Against chemical products Asks questions out of concern for health	Proponent of public schools Questions use of charter schools by parents	Recognizes difficulty of police work due to family connection Cognizant of systemic racism
Layla	Unsure of why there is disagreement between groups mentioned in the article	Wonders if there is a better way to measure issues discussed in video	Offended by some of the authors' statements Questions how the author could make such an argument
Savannah	Suspicious of companies Unsure of some topics in the article Discussed family member's experiences	Attended charter schools previously Doesn't necessarily agree with the framing of the video given her past experiences	Acknowledges flaws in logic of author Acknowledges past experiences related to racism
Group 2			
Teta	Concerned about children and their development, both locally and in home country	Unsure of charter schools and how to view them given that they are not prevalent in her home country	Believes author is not making a logical argument Notes this does not align with previous experiences with racism
C	Suspicious of companies, Monsanto specifically, and of political agendas more broadly	Dislikes charter schools due to previous coursework and knowledge about them Notes concern about Betsy DeVos	Concerned about friends and specific groups that are consistently targeted in hate crimes Asks questions about nature of data
Jayla	Unsure of why there is disagreement between groups mentioned in the article	Concerned about Betsy DeVos Suspicious of outsiders influencing public education in Detroit	Adamantly against message of the article Questions author's argument Cognizant of systemic racism
Ash	Brings up family history of purchasing products related to Monsanto Asks questions about issues discussed in article	Notes family members went to charter schools Unsure of which is best given that he went to public schools	Against the author's logic, asking questions about how specific claims were made Notes racism is an obvious issue in the U.S.

REFERENCES

REFERENCES

- Baker, Dave. 1996. "Children's Formal and Informal School Numeracy Practice." In *Challenging Ways of Knowing in English, Maths and Science*, edited Baker, Dave, John Clay, and Carol Fox, 80-8. London: Falmer Press.
- Bartell, Tonya Gau. 2013. "Learning to Teach Mathematics for Social Justice: Negotiating Social Justice and Mathematical Goals." *Journal for Research in Mathematics Education* 44 (1): 129-63.
- Barton, David, and Mary Hamilton. 2000. "Literacy Practices." In *Situated Literacies. Reading and Writing in Context*, edited by Barton, David, Mary Hamilton, and Roz Ivanič, 7-15. London: Routledge.
- Blair, Richelle, Ellen E. Kirkman, and James W. Maxwell. 2018. *Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States: Fall 2015 CBMS Survey*. Providence, RI: American Mathematical Society.
- Boersma, Stuart, and Dominic Klyve. 2013. "Measuring Habits of Mind: Toward a Prompt-less Instrument for Assessing Quantitative Literacy." *Numeracy* 6 (1): Article 6.
- Brantlinger, Andrew. 2013. "Between Politics and Equations: Teaching Critical Mathematics in a Remedial Secondary Classroom." *American Educational Research Journal* 50 (5): 1050-80.
- Briggs, William. "Quantitative Literacy and Civic Virtue." *Numeracy* 11, Iss. 2 (2018): Article 7.
- Carraher, Terezinha Nunes, David William Carraher, and Analúcia Dias Schliemann. 1985. "Mathematics in the Streets and in Schools." *British Journal of Developmental Psychology* 3 (1): 21-9.
- Cockcroft, Sir Wilfred H. 1982. *Mathematics Counts. Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr. Wilfred H. Cockcroft*. London: Her Majesty's Stationery Office.
<http://www.educationengland.org.uk/documents/cockcroft/cockcroft1982.html>.

- Craig, Jeffrey, and Lynette Guzmán. 2018. "Six Propositions of a Social Theory of Numeracy: Interpreting an Influential Theory of Literacy." *Numeracy* 11 (2): Article 1.
- Derespina, Cody. 2016. "White Police Officers Don't Unfairly Target Black Suspects, Study Says." *Fox News*, November 16.
<https://www.foxnews.com/us/white-police-officers-dont-unfairly-target-black-suspects-study-says>.
- Erickson, Ander W.. "Rethinking the Numerate Citizen: Quantitative Literacy and Public Issues." *Numeracy* 9, Iss. 2 (2016): Article 4. DOI:
<http://dx.doi.org/10.5038/1936-4660.9.2.4>
- Erickson, Ander W.. "Rethinking the Numerate Citizen: Quantitative Literacy and Public Issues – Reply." *Numeracy* 10, Iss. 2 (2017): Article 13. DOI:
<http://doi.org/10.5038/1936-4660.10.2.13>
- Gaze, Eric. 2014. "Teaching Quantitative Reasoning: A Better Context for Algebra." *Numeracy* 7 (1): Article 1.
- Gaze, Eric. "Quantitative Reasoning: A Guided Pathway from Two- to Four-Year Colleges." *Numeracy* 11, Iss. 1 (2018): Article 1.
- Hamman, Kira. 2017. "Rethinking the Numerate Citizen: Quantitative Literacy and Public Issues—Discussion." *Numeracy* 10 (2): Article 12.
- Hastings, Nancy B., ed. 2006. *A Fresh Start for Collegiate Mathematics: Rethinking the Courses below Calculus*. Vol. 69. Washington, DC: Mathematical Association of America.
- Heath, Shirley Brice. 1983. *Ways with Words: Language, Life, and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Kahan, Dan M., Peters, Ellen, Dawson, Erica C., and Paul Slovic. 2017. "Motivated Numeracy and Enlightened Self-government." *Behavioural Public Policy* 1 (1): 54-86.
- Karaali, Gizem, Edwin Villafane-Hernandez, and Jeremy Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9 (1): Article 2.
- Liamputtong, Pranee. 2011. *Focus Group Methodology: Principle and Practice*. Great Britain: Sage Publications.

- Madison, Bernard L. 2019. "Quantitative Literacy: An Orphan No Longer." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 37-46. Washington, DC: Mathematical Association of America.
- Madriz, Esther. 2003. "Focus Groups in Feminist Research." In *Collecting and Interpreting Qualitative Materials*, edited by Norman K. Denzin & Yvonna S. Lincoln, 363–388. Thousand Oaks, CA: Sage.
- Meaney, Tamsin, Tony Trinick, and Uenuku Fairhall. 2013. "One Size Does NOT Fit All: Achieving Equity in Māori Mathematics Classrooms." *Journal for Research in Mathematics Education* 44 (1): 235-63.
- Mellow, Gail O.. "Quantitative Literacy: Now More Than Ever." *Numeracy* 11, Iss. 2 (2018): Article 1. DOI: <https://doi.org/10.5038/1936-4660.11.2.1>
- Moje, Elizabeth Birr. 2015. "Doing and Teaching Disciplinary Literacy with Adolescent Learners: A Social and Cultural Enterprise." *Harvard Educational Review* 85 (2): 254-78.
- Morrison, Kristan A., Holly H. Robbins, and Dana Gregory Rose. 2008. "Operationalizing Culturally Relevant Pedagogy: A Synthesis of Classroom-Based Research." *Equity & Excellence in Education* 41 (4): 433-52.
- Murtaugh, Michael. 1985. "The Practice of Arithmetic by American Grocery Shoppers." *Anthropology & Education Quarterly* 16 (3): 186-192.
- Nasir, Na'ilah Suad, and Maxine McKinney de Royston. 2013. "Power, Identity, and Mathematical Practices Outside and Inside School." *Journal for Research in Mathematics Education* 44 (1): 264-87.
- Nurse, Matthew S., and Will J. Grant. "I'll See It When I Believe It: Motivated Numeracy in Perceptions of Climate Change Risk." *Environmental Communication* (2019): 1-18.
- Oughton, Helen M. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2.
- Pardoe, Simon. 2000. "Respect and the Pursuit of 'Symmetry.'" In *Situated Literacies. Reading and Writing in Context*, edited by Barton, David, Mary Hamilton, and Roz Ivanič, 149-166. London: Routledge.

- Redlawsk, David P. "Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making." *The Journal of Politics* 64, no. 4 (2002): 1021-1044.
- Royer, Dan W., and Russell D. Baker. 2018. "Student Success in Developmental Math Education: Connecting the Content at Ivy Tech Community College." *New Directions for Community Colleges* 182: 31-8.
- Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Snyder, Thomas D., Cristobal de Brey, and Sally A. Dillow. 2018. "Digest of Education Statistics, 2016." Retrieved from https://nces.ed.gov/programs/digest/d16/tables_3.asp#Ch3Sub18.
- Sonde, Kari. 2018. "The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal." *Mother Jones*, August 15. <https://www.motherjones.com/environment/2018/08/roundup-monsanto-glyphosate-cheerios-quaker-oats-cancer-1/>.
- Steen, Lynn Arthur. 1997. *Why Numbers Count: Quantitative Literacy for Tomorrow's America*. New York: College Entrance Examination Board.
- Steen, Lynn A., ed., and National Council on Education and the Disciplines (NCED). 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: NCED.
- Street, Brian V. 1995. *Social Literacies: Critical Approaches to Literacy in Development, Ethnography, and Education*. London: Longman.
- Thelk, Amy D., and Emily R. Hoole. 2006. "What Are You Thinking? Postsecondary Student Think-alouds of Scientific and Quantitative Reasoning Items." *The Journal of General Education* 55 (1): 17-39.
- Tunstall, Samuel L., Rebecca L. Matz, and Jeffrey C. Craig. 2018. "Quantitative Literacy Courses as a Space for Fusing Literacies." *The Journal of General Education* 65 (3-4): 178-94.
- Tunstall, Samuel L., Vincent Melfi, Jeffrey C. Craig, Richard Edwards, Andrew Krause, Bronlyn Wassink, and Victor Piercey. 2016. "Quantitative Literacy at Michigan State University, 3: Designing General Education Mathematics Courses." *Numeracy* 9 (2): Article 6.

Tunstall, Samuel L. 2018. "College Algebra: Past, Present, and Future." *PRIMUS*
28 (7): 627-40.

Conclusion

Circling Back: A Milieu for Exploring Numeracy

The nature of this dissertation precludes grandiose claims. Findings from the studies have implications, but insofar as the findings themselves are qualified and subject to interpretation, the implications too should be viewed as contingent. With those remarks in mind, rather than present a *conclusion* per se, I instead use this space to meditate on the curiosities that drove me to this work, ways my findings inform those interests, and questions that have emerged over the course of completing the three studies. At the same time that I hope this coda will connect the work of these three studies, I also aim to elucidate what I perceive to be a certain “cloudiness” that persists in my understanding of numeracy proxies and practices.

I began this dissertation by describing the social milieu in which I position this work. Albeit I did not fasten the work to any specific context, I did note that my population of interest was adults affected by general education mathematics at the postsecondary level. Such a population includes the majority of adults in postsecondary education. The social milieu I described (rephrased succinctly) was a world “awash in numbers” (Steen et al. 2001, 1) and rife with practices of quantification (Porter 1995); the educational backdrop was one of a “totally pedagogized society” (Bernstein 2000, as cited in Tsatsaroni and Evans 2014) where organizations at various levels seek to influence mathematics curricula in

their own interests. Though there is an ostensible lack of agency ascribed to individuals based on the description above, note that I also described a reality in which individuals engage regularly in numeracy practices (Craig and Guzmán 2018; Oughton 2018), or patterned ways in which they appropriate, fashion, and use numbers or quantification in their diverse personal, professional, and public lives. This was an exciting atmosphere for the study of numeracy—one that, in light of its novelty and shifting characteristics, engendered curiosity for me.

Phrased as questions, the curiosities that I outlined in the opening of this dissertation included: What are numeracy practices, and to what extent are they captured through processes of measurement or quantification? How might we attend to numeracy practices in the context of general education mathematics at the postsecondary level? My rationale for focusing on *proxies* for numeracy in particular were that they are commonly used, yet in tension with a social theory of numeracy. A social theory of numeracy foregrounds the social contexts in which numeracy exists, as well as the practices individuals engage in as they interface with numbers or quantification in those contexts. Over the course of the three studies, I arrived at responses to those questions, yet as I will describe later, those answers are hedged, and I remain cloudy about a few things that relate numeracy proxies and practices.

In Response to Curiosity

I chose to pursue those questions by structuring this dissertation into three parts, with each part constituting a standalone examination of some aspect of those questions.

Study One: Validity Analysis of the PIAAC's Numeracy Component

Structured as a validity analysis, the first study broached what I view as a core tension between numeracy proxies and practices: the alignment between what is measured through a proxy and what is “real,” whereby real I mean “actually” representative of what one does outside of the setting of a proxy assessment. The reason that I use quotation marks around *real* and *actually* is that there is no representation of reality that would adequately represent it. As noted by Fairclough (2003): “Reality (the potential, the actual) cannot be reduced to our knowledge of reality, which is contingent, shifting, and partial” (14). Some representations (or proxies), I believe, are better than others, but I do not wish to broach that area of inquiry here.

The study that I completed was a validity examination of the numeracy portion of a well-known international assessment: the OECD's Programme for the International Assessment of Adult Competencies (PIAAC). It is worth noting that when I had first proposed the study as a part of the dissertation, I characterized it as a validity critique, rather than validity examination. Given my positionality in

relation to the work, I was approaching the first study with a clear—if misguided—idea of what I would find. With the guidance of my Dissertation Committee, I shifted focus to (among other things) be more open-minded to what I might find. In following the path set out by the *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014), the formal questions that I pursued included: (1) What does the PIAAC numeracy assessment claim to measure? (2) What are the intended uses of the assessment? (3) How are we to interpret scores with those uses in mind? And (4) to what degree do evidence and theory support interpretations for those uses? Though dependent on the first three, the fourth question was the validity question.

A foremost finding preceded the study itself, in that it was a part of the literature review for it. This was that the result of the validity examination would not be one of valid or invalid, but rather, a set of qualified statements about the PIAAC in the broader context of score interpretations for stated uses (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 2014; Sireci and Sukin 2013). With that point in hand, the main finding was that score interpretations from the PIAAC numeracy assessment may be valid for the use of describing distributions of proficiency in subgroups of interest, but

- the construct of interest—numerate behavior—is not what is measured,

- evidence distinguishing what is measured from other constructs, such as the OECD's conception of literacy, is largely absent, and
- consequences of the uses of the scores are not justified.

I discussed within the study that that finding in and of itself is not altogether surprising, especially in light of the existing literature concerning the difficulty of capturing numeracy as some sort of behavior (Grawe 2011; Shavelson et al. 2019).²⁵ What was of greater interest from the study was that discussions of the exam appeared to overstate what was measured. Furthermore, literature citing results of the PIAAC as justification for numeracy education often appeared to fail to acknowledge the complexity of what the results of the exam might actually tell us. As I will discuss later, an implication of this study and the two others is the need for awareness of what proxies centered on skills actually tell us about what individuals do.

Study Two: Critical Discourse Analysis of Relational Links in Skilled for Life? (OECD 2013b)

Though not necessarily linked to the first study, the second study proceeded well from the first in that involved a careful analysis of the ways in which makers of the PIAAC describe results and recommendations stemming from its administration to several OECD member countries. The curiosity that led to the formulation of the specific study arose from a sense of dissonance. In my reading of several OECD

²⁵ I do not use the word *practice* here, given that such literature does not.

reports (OECD 2013a, 2013b, 2013c), I had a sense that numeracy—as measured through the PIAAC's proxy—was constructed as causally linked with measures of well-being such as wages and health. What was intriguing to me was that the writers explicitly discounted any sort of causal relationship between numeracy skills and well-being, and yet I still felt that the writing exuded messages of causation, or at the very least, a unidirectional association. This led me to the question: In what ways are relations of association and causality between numeracy and well-being constructed in *Skilled for Life* (OECD 2013b)?

I engaged in a critical discourse analysis (Gee 2004) drawing from Achugar and Schleppegrell's (2005) work concerning causal construction in history texts. Following a variant of Fairclough's (2003) approach to critical discourse analysis I examined the text itself and the broader sociopolitical context in which it is situated. The text in this case was *Skilled for Life? Key Findings from the Survey of Adult Skills* (OECD 2013b). This approach allowed me to understand how the text itself and broader sociopolitical context come together to convey a link between skills and well-being.

Though the context of interest is altogether different, the findings of the study are somewhat similar to those of Achugar and Schleppegrell. Authors of *Skilled for Life?* used discourse-semantic moves like headers, summary statements, and begging the question (e.g., assuming a relationship, and then finding one), as well as lexico-grammatical tacks for generating agency-eliding clause structures (e.g.,

conveying statements about test takers that center around their skills proficiency), to construct a unidirectional association between numeracy skills and well-being. The latter strategy provides one answer to the question I began with concerning how the document remains coherent in spite of remarks from the authors concerning the lack of any causal association. Another finding, gleaned in looking for numeracy as part of the analysis of the document, was that the authors *flattened* the constructs of literacy, numeracy, and problem solving in technology-rich environments beneath an umbrella construct of skills, thereby making the case for the benefits of all of the constructs assessed in PIAAC while simultaneously reducing the vibrant uniqueness of each of them.

At a basic level, one contribution of the second study is that it responds to calls from Kanes, Morgan, and Tsatsaroni (2014) and Tsatsaroni and Evans (2014) to contest what the former author group has referred to as a “regime” of testing pushed forth by the OECD. Among other things, this regime aims to commandeer through veneers of objectivity notions of what counts as mathematics or as numeracy; calling attention to the explicit and subtle ways the OECD discursively constructs the “(in)numerate subject” (Jablonka 2015) as deficient—“left behind” (OECD 2013b, 6), even—is a key part of disrupting the power held by the Organisation. Beyond responding to calls for interrogations of international assessments, this study also raises new questions about the nature of claims of association or causation in research around mathematics education. In particular, a new question

that has arisen for me, and that I discuss further below, concerns what is lost or overlooked when we commence work with an assumption that perturbations in some construct *A* (e.g., numeracy skills) are associated with, or impact/influence/cause shifts in some construct *B* (e.g., some measure of well-being).

Study Three: College Students' Numeracy Events and Discussion of Public Issues in Focus Groups

Circling back to the driving questions that steered this dissertation as a whole, it is reasonable to say that the first two studies centered around an exemplar numeracy proxy, issues surrounding its use, and related discourse about its administration. The third study stepped away from numeracy proxies to consider numeracy practices.

Leading my formulation of the third study was a desire to understand postsecondary students' numeracy practices in relation to public issues, and how those might bear relevance on the development and facilitation of numeracy-focused coursework. The questions I asked were: How do students who have the option to enroll in a Quantitative Literacy course discuss public issues in a focus group setting? Do numeracy events occur as they articulate their reactions? If so, what are the characteristics of these numeracy events? The finding pertaining to the first question was that students leveraged their background experiences and knowledge in articulating standpoints on public issues (corroborating our finding

in Tunstall, Matz, and Craig 2018). Students also actively expressed critiques and questions as they engaged with the three artifacts. In relation to the second research question, I found that numeracy events, when they did occur, were primarily centered around students' acknowledgements of the importance of numbers, more so than their active articulation in building on that importance through group conversation. Because most of the students did not engage quantitatively with the artifacts in their typed responses or their group discussions, this suggests that they may benefit from structured experiences with engaging with public issues from a quantitative lens. Future work building from this study could explore student numeracy events in further detail, whether by explicitly prompting students to engage in using quantitative reasoning, or taking a different approach and reporting on how students engage in the process of learning about the analysis of exploring issues through a quantitative lens. The framework given by Moje (2015) reported in the Discussion is one structured means of doing so with students.

Lingering and Emergent Clouds

The curiosities that guided me to this work centered on the ideas of *numeracy*, *practices*, and *measurement*. In narrowing from these three ideas, the broader questions that I sought to explore in completing the three studies of this dissertation included: What are numeracy practices, and to what extent are they captured through processes of measurement or quantification? How might we attend to

numeracy practices in the context of general education mathematics at the postsecondary level? Insofar as these were not formal research questions, the task of exploring these was a personal goal. My view is that I have answered the question of “What are numeracy practices?” over the course of my review of literature and my engagement with them in the third study. With respect to the second and third questions above, my findings suggest xxx... research can analyze what is captured, as well as what might be lost, in analyzing the PIAAC numeracy assessment (or other assessments). At the same time, there are a few aspects concerning the relation between numeracy proxies and practices that either remain or newly emerge as cloudy to me.

Across the first two studies here, I see a need for attention to the affordances and limitations of proxies for numeracy. With that said, given that proxies for numeracy are unlikely to simply go away in the coming years, one might ask the question: How should one measure numeracy, especially in a context where resources and time are limited? Unfortunately, my initial response of, “We should do so carefully, aware of the complexity and implications of such measurement,” does not address the heart of the question. While I am inclined to push back against attempts to measure and sort students (e.g., see Fendler and Muzaffar 2008), I do believe that there are concrete ways that we can acknowledge and embrace students’ numeracy practices while simultaneously aiming to “capture” in some sense what students do “in the wild” (a metaphor borrowed from Hutchins 1995).

One practice to avoid, as I have argued in this dissertation, is the unfettered use of closed-response mathematics questions without attention to their limitations as representations of what individuals do. Instead, if one seeks to understand what students do, it seems imperative to ask or observe them directly; this might manifest through the use of student-driven projects, or even compilations of student artifacts in the form of portfolios (Schneider 2009). Results from the third study of this dissertation suggest (as what should be of little surprise) that students are likely to be engaged in discussions about topics that matter to them; finding these topics, or letting students tell us about them directly, and then allowing students to demonstrate their brilliance through open-ended projects that address disciplinary literacy skills (Moje 2015), is an approach to education centered around numeracy or quantitative literacy that has much promise. Insofar as I do not condone quick or simplistic measurements for something that is evidently complex, the question of how one might create a numeracy measurement with limited resources and time remains cloudy for me. That is, I believe that the question, “How should one measure numeracy, especially in a context where resources and time are limited?” remains open.

Other questions have emerged over the course of this dissertation that I did not have before I began. Some are theoretical and relate to the study of the “benefits” of numeracy; these include: Who benefits from coursework centered around numeracy, and in what ways? How do such benefits change as we focus on

numeracy practices, rather than skills? What do we potentially miss as we consider the benefits “of” numeracy, rather than (for example) how numeracy benefits or changes as its users change? That is, can we switch the direction of association or causation to generate new questions or new understandings about the nature of numeracy? Would doing so be reasonable? Some of these questions may be too distanced from existing discourses about numeracy to even be sensible. At the very least, I hope that they generate new ideas and questions for others as they grapple with the notion of numeracy. Given that it would be straightforward to substitute *numeracy* with *mathematics* or *mathematical literacy* in the phrasing of those questions, I believe they have import for a wider community to consider than just that centered on numeracy.

Other emergent questions from this dissertation relate to how students reason with public issues. Before presenting those questions, I first want to share a brief portion of dialogue that I had when I engaged in a semi-structured interview with an instructor of Quantitative Literacy at Michigan State University. In addition to asking about their experiences in teaching the courses, I also requested that they engage with the same artifacts and prompts I had used in the focus groups with students. This interview occurred at the beginning of the fall 2018 semester (after the focus groups); the instructor consented to be involved in the research, and was compensated in the same way that the students were. A portion of the dialogue is

transcribed below; this segment followed their reading of the *Mother Jones* article (Sonde 2018).

Luke: Do you think an article like this might be relevant for the Quantitative Literacy class?

Instructor: I feel like the rates—the two milligrams—like the difference between the two milligrams and the hundredth of a milligram. That’s just per kilogram of body weight. So how much would that mean for me at my weight? And also converting from pounds to kilograms, since most of us are not accustomed to thinking in terms of kilograms. I don’t know that if they read, “two milligrams per kilogram of body weight per day,” they would know what that means for a person who weighs 150 pounds.

Luke: Right. That’s a hard question. Would you expect for students or anyone reading the article in the courses to do that in their hand, or like on paper?

Instructor: That’s what I was personally thinking as I read it. I don’t know that I would personally sit down and do the calculations, but it was one of the first thoughts that kind of flickered through my head: OK, so how much does this mean for the average person? This is just something that the article doesn’t really address, right? So it’s in milligrams per kilogram of body weight per day. And then, so you’ll have the Cheerios or whatever, and that’s just a recommendation of what not to surpass...So here is what I would want. It’s an ideal goal, but for students to be not just consumers, but also to put their editor hats on. So they’re like, “Well that wasn’t really presented well.” I don’t know how far we are away from that, but I want to give my students confidence. They would say, “You know what? This could be clearer. There needs to be more detail.” That, versus, “Oh, I’m depressed. I don’t understand what they are saying.”

The numeracy event above was illuminating for me, as it broached a tension I have felt for some time in thinking about what it is that individuals involved in numeracy education (myself included) want for our students. In particular, I have wondered why I ask students to do calculations as part of a course if I felt that it was unlikely that they (or I) would ever actually do them outside of the classroom.

The statistic in the article, two milligrams of Cheerios per kilogram of body weight per day, was an excellent example to consider. The conversation I had with the instructor revealed that they, in a similar fashion, were unlikely to engage in a calculation in reading that specific article; what mattered to them was that students engage in the practice of critically interrogating the article as an artifact that is presented to consumers. To the instructor, a Quantitative Literacy course might facilitate a student's skills as an editor, rather than as a calculator. Of course, rewriting that component of the article would require calculations—but it would also necessitate an awareness of how information be presented in a suitable manner for readers. I believe that this focus in a course aligns well with the principles discussed by Moje (2015) in describing a framework for fostering adolescents' disciplinary literacy practices, as it encourages students to take on and participate in—rather than just consume from—the discourse communities of various disciplines.

Based on this conversation, as well as those that I had in the focus groups, new questions have emerged. These include: Where might a postsecondary course that positions students in the manner described above live at a university? What stakeholders might have the privilege to design and teach it, and how can we best support them to do so? Logistically, to what extent are students (mis)served if such a course is given in a lecture format, rather than in a more intimate setting with fewer students? On a different note, given that students need not be at the

postsecondary level to engage with the world as editors, what might a curriculum look like at the secondary level that fosters numeracy practices such as interrogating the quantitative reasoning present in media? Does it already exist? How might it situate within the existing mathematics curriculum? Some of these questions are not necessarily novel within extant discussions of numeracy (cf. Madison 2015), but the impetus driving them (i.e., a social practices perspective of numeracy) is indeed nascent. Nonetheless, insofar as we still live in a world “awash in numbers” (Steen 2001, 1), our work is united by a fundamental commonality. My hope is that we may continue to learn from one another’s perspectives as we strive for a just world.

REFERENCES

REFERENCES

- Achugar, Mariana, and Mary J. Schleppegrell. 2005. "Beyond Connectors: The Construction of Cause in History Textbooks." *Linguistics and Education* 16 (3): 298-318.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Bernstein, Basil B. 2000. *Pedagogy, Symbolic Control, and Identity: Theory, Research, Critique*. No. 4. Lanham, MD: Rowman & Littlefield.
- Craig, Jeffrey, and Lynette Guzmán. 2018. "Six Propositions of a Social Theory of Numeracy: Interpreting an Influential Theory of Literacy." *Numeracy* 11 (2): Article 1.
- Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Fendler, Lynn, and Irfan Muzaffar. 2008. "The History of the Bell Curve: Sorting and the Idea of Normal." *Educational Theory* 58 (1): 63-82.
- Gee, James Paul. 2004. "Discourse Analysis: What Makes it Critical?" In *An Introduction to Critical Discourse Analysis in Education*, edited by Rebecca Rogers, 19-50. Mahwah, NJ: Lawrence Erlbaum.
- Grawe, Nathan D. 2011. "Beyond Math skills: Measuring Quantitative Reasoning in Context." *New Directions for Institutional Research* 149: 41-52.
- Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT press.
- Jablonka, Eva. 2015. "The Evolvement of Numeracy and Mathematical Literacy Curricula and the Construction of Hierarchies of Numerate or Mathematically Literate Subjects." *ZDM* 47 (4): 599-609.
- Kanes, Clive, Candia Morgan, and Anna Tsatsaroni. 2014. "The PISA Mathematics Regime: Knowledge Structures and Practices of the Self." *Educational Studies in Mathematics* 87 (2): 145-65.
- Madison, Bernard. 2015. "Quantitative Literacy and the Common Core State Standards in Mathematics." *Numeracy* 8 (1): Article 11.

- Moje, Elizabeth Birr. 2015. "Doing and Teaching Disciplinary Literacy with Adolescent Learners: A Social and Cultural Enterprise." *Harvard Educational Review* 85 (2): 254-78.
- Organisation for Economic Co-operation and Development (OECD). 2013a. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013b. *Skilled for Life? Key Findings from the Survey of Adult Skills*. Paris: OECD Publishing.
- OECD. 2013c. *Time for the U.S. to Reskill? What the Survey of Adult Skills Says*. Paris: OECD Publishing.
- Oughton, Helen M. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2.
- Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Schneider, Carol Geary. 2009. "The Proof Is in the Portfolio." *Liberal Education* 95 (1): 1-2.
- Shavelson, Richard J., Julián P. Mariño von Hildebrand, Olga Zlatkin-Troitschanskaia, and Susanne Schmidt. 2019. "Reflections on the Assessment of Quantitative Reasoning." In *Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education*, edited by Tunstall, Samuel Luke, Gizem Karaali, and Victor Piercey, 163-76. Washington, DC: Mathematical Association of America.
- Sireci, Stephen G., and Tia Sukin. 2013. "Test Validity." In *APA Handbook of Testing and Assessment in Psychology*, Volume 1, 61-84. Washington, DC: American Psychological Association.
- Sonde, Kari. 2018. "The Roundup Chemical Found Responsible for Cancer Might Also Be in Your Cereal." *Mother Jones*, August 15. <https://www.motherjones.com/environment/2018/08/roundup-monsanto-glyphosate-cheerios-quaker-oats-cancer-1/>.
- Steen, Lynn A., ed., and National Council on Education and the Disciplines (NCED). 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: NCED.

Tsatsaroni, Anna, and Jeff Evans. 2014. "Adult Numeracy and the Totally Pedagogised Society: PIAAC and Other International Surveys in the Context of Global Educational Policy on Lifelong Learning." *Educational Studies in Mathematics* 87 (2): 167-86.

Tunstall, Samuel L., Rebecca L. Matz, and Jeffrey C. Craig. 2018. "Quantitative Literacy Courses as a Space for Fusing Literacies." *The Journal of General Education* 65 (3-4): 178-94.