

8-2016

# Teachers and the Development of Student Noncognitive Skills

Albert Cheng

*University of Arkansas, Fayetteville*

Follow this and additional works at: <http://scholarworks.uark.edu/etd>

 Part of the [Developmental Psychology Commons](#), [Educational Leadership Commons](#), and the [Education Policy Commons](#)

---

## Recommended Citation

Cheng, Albert, "Teachers and the Development of Student Noncognitive Skills" (2016). *Theses and Dissertations*. 1657.  
<http://scholarworks.uark.edu/etd/1657>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu), [ccmiddle@uark.edu](mailto:ccmiddle@uark.edu).

Teachers and the Development of Student Noncognitive Skills

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy in Education Policy

by

Albert Cheng  
University of California, Berkeley  
Bachelor of Arts in Mathematics, 2006  
Biola University  
Master of Arts in Education, 2012

August 2016  
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

---

Dr. Patrick J. Wolf  
Dissertation Director

---

Dr. Gema Zamarro  
Committee Member

---

Dr. Jay P. Greene  
Committee Member

---

Dr. Robert Maranto  
Committee Member

## **Abstract**

Scholars of education policy are increasingly aware of the independent role that noncognitive skills (e.g., self-regulation, social skills, and other personality or character traits) play in long- and short-run student well-being. However, little is known about how these skills are effectively developed. One theory is that noncognitive skills are developed through role modeling by teachers. A student, by virtue of observing and sharing a social connection with his or her schoolteachers, begins to emulate noncognitive skills that they exhibit. In this dissertation, I test this theory. I focus specifically on noncognitive skills related to conscientiousness and measure them using new behavioral proxies based upon survey effort. In chapter 2, I use panel data and techniques to show that students experience declines in conscientiousness in years when they are taught by teachers who exhibit less conscientiousness. Assuming that students are not systematically sorted to teachers with varying degrees of conscientiousness based upon time-varying student characteristics, the relationship between student and teacher conscientiousness can be interpreted as causal. In chapter 3, I corroborate these findings in an analysis that possesses greater internal validity. Using data in which teachers were randomly assigned to classrooms, I find that students become more conscientious when they are taught by more conscientious teachers. If teacher noncognitive skills are transmitted to students, particularly by role modeling, it may be desirable for school leaders to ensure that their school communities exhibit coherence over a values system that they desire students to embrace. In chapter 4, I discuss a policy proposal for this end. Specifically, I test whether school leaders who possess more flexibility in hiring and dismissing teachers are more adept at building a staff that embodies a core set of values. I provide descriptive evidence suggesting that schools exhibit more values coherence when principals have more autonomy over personnel decisions. In conclusion,

education policy has only recently given increasing attention to noncognitive skills. More effort must be devoted to studying this topic as carries normative implications not only for the means and aims of existing educational institutions but also for the labor of reforming and improving them.

## **Acknowledgements**

To the EDRE Faculty – Pat, Jay, Gema, Bob, Gary, and Bob – I don't know how to appreciate all of you enough for the ways in which you've served me the past several years. I think it's safe to say that my outcomes aren't only due to selection, though I recognize the irony of such a statement. But then again, the fact that I recognize the irony might make my comments doubly ironic, which only further illustrates why I am indebted to you all. In the end, a simple and genuine "Thank You" is all I can offer here in writing. I hope it at least somewhat does justice to the gratitude I feel for your labor.

I must also acknowledge numerous journal editors and anonymous referees who have reviewed article versions of Chapters 2 and 4. Their comments and suggestions have been invaluable for shaping these chapters – both for those who voted to publish and those who voted to reject the manuscripts that I submitted. I would also like to thank Matthew Kraft, Kata Mihaly, and participants at the Association of Education Finance and Policy's 41st Annual Conference for their comments on earlier versions of Chapter 3. Gema Zamarro also deserves special recognition for her intellectual contributions to Chapter 3.

And let me not discount the credit due to current EDRE students and the students who have gone before me. Your comments during the Brown Bag series or bantering in informal hallway discussions have undoubtedly shaped these papers and my own thinking. Thanks for making it an office atmosphere that will be difficult to leave, not just because of our serious conversations but because of our light-hearted ones too. I would especially like to single out Collin Hitt who has been my indispensable partner in crime throughout our mutual forays into this noncognitive skills world. It's only fitting that we started and are now earning our degrees...what's the word, contemporaneously. Congratulations.

To my family – grandparents, parents, aunts, uncles, and cousins. To this day, I still have trouble explaining everything that I do to all of you, particularly those with whom I usually don't speak English. But as the adage goes: Show, don't tell. So here's a glimpse into what your support has wrought. You don't have to get it all, but I think you get enough of it to see why I owe you many thanks.

Finally, to my fellow forerunning friends. You know who you are. Grin.

## **Dedication**

To my John-Locke-quoting Blue A-kong, who searched for Truth. To my music-loving A-kong, who cherished Beauty. And to the both of you, who sought the Good of others through your respective vocations. May I continue in the heritage you have left for me.

## Table of Contents

Chapter 1: Introduction.....	1
References.....	6
Chapter 2: Teacher effects on Student Conscientiousness: Evidence from Panel Data .....	9
Literature Review.....	12
Teacher Impacts on Student Cognitive Ability.....	12
Teacher Impacts on Student Noncognitive Ability.....	13
How Do Teachers Affect Student Noncognitive Ability? .....	15
Issues with measuring noncognitive skills.....	17
Methods.....	18
Data.....	18
Response Rate as a Proxy for Noncognitive Skills.....	19
Empirical Strategy .....	20
Results.....	26
Validation of Student Item Response Rate .....	26
Main Analysis .....	27
Discussion and Conclusion.....	37
References.....	42
Chapter 3: Teacher Effects on Student Conscientiousness: Experimental Evidence .....	46
Literature Review.....	49
Measuring Teacher Quality.....	49
The Role of Teacher Noncognitive Skills.....	53
Measuring of Noncognitive Skills .....	55
Research Questions.....	58
Methods.....	60
Data.....	60
Survey-Effort Measures of Teacher Conscientiousness .....	61
Traditional Measures of Teacher Quality .....	63
Empirical Strategy .....	66
Teacher impacts on student outcomes. ....	67
Results.....	73
Relationships between Measures of Teacher Quality.....	73
Teacher Impacts on Student Noncognitive Skills.....	75



Teacher Noncognitive Skills and Student Noncognitive Skills.....	79
Traditional Measures of Teacher Quality and Student Noncognitive Skills. ....	82
Discussion and Conclusion.....	86
Summary.....	86
Implications for Understanding Teacher Quality .....	90
Implications for Future Research and Practice .....	92
References.....	96
Appendix: Items on Student Effort Scale .....	102
Chapter 4: Fostering School Mission Coherence for Noncognitive-Skill Development.....	103
Literature Review and Theoretical Framework .....	104
The Importance of Organizational Mission Coherence .....	104
Hiring and Dismissal as a Mission Building Mechanism.....	107
Personnel Policies and Organizational Effectiveness: Mission Coherence as Mediator? ..	109
Research Methodology .....	113
Data.....	113
Measure of Mission Coherence .....	114
Measure of Administrator Influence over Hiring Decisions.....	115
Measure of Barriers to Dismissal.....	115
Empirical Strategy .....	120
Results.....	122
Discussion.....	124
Conclusion .....	128
References.....	129
Chapter 5: Conclusion.....	135
References.....	138

## List of Tables

Student Item Response Rates and Teacher Return Rates .....	21
Student Response Rate, Test Scores, Years of Education, and Employment.....	27
Variation in Teacher Return Rates.....	28
Variation in whether Teacher Returns Survey.....	29
Transition Probabilities for whether Teacher Returns Survey.....	30
Associations between Teacher Return, Student Item Response, and Student Test Scores.....	33
Anderson-Hsiao Estimates.....	34
Arellano-Bond Estimates .....	35
Variation in Teacher Return Rates (Anderson-Hsiao Models).....	36
Variation in whether Teacher Returns Survey (Anderson-Hsiao Models).....	36
Transition Probabilities for whether Teacher Returns Survey (Anderson-Hsiao Models).....	37
Student Outcomes Correlation Matrix .....	70
Relationships between Different Measures of Teacher Quality .....	74
Post-Attrition Baseline Covariate Balance across Randomly Assigned Teachers within Randomization Block.....	76
Post-Attrition Baseline Covariate Balance between Student Characteristics and Randomly Assigned Teacher Characteristics .....	78
ITT Estimates of Teacher Effects (Based on Survey-effort Measures of Conscientiousness) on Student Outcomes .....	80
IV Estimates of Teacher Effects (Based on Survey-effort Measures of Conscientiousness) on Student Outcomes .....	81
ITT Estimates of Teacher Effects (Based on Traditional Measures of Teacher Quality) .....	84
IV Estimates of Teacher Effects (Based on Traditional Measures of Teacher Quality) .....	85
Summary Statistics and Two-Factor Loadings of Barriers-to-Dismissal Items .....	116
Summary Statistics for Independent Variables.....	118
Coefficient Estimates of Empirical Model .....	122
Personnel Policy Differences between Charter and other Public Schools.....	124

## Chapter 1: Introduction

Noncognitive skills refer to personality traits and character attributes, like conscientiousness, openness to new experience, or neuroticism, that are not easily captured by scores on standardized tests of cognitive ability. Some scholars have also referred to noncognitive skills as social and emotional skills or soft skills (West et al., 2016). There has been a recent resurgence in the study of noncognitive skills in the field of educational policy primarily due to research from economists and psychologists demonstrating the ways in which these skills affect short- and long-run student outcomes. For instance, more conscientious students are more likely to complete their homework, more academically successful, and less likely to commit disciplinary infractions (Borghans, Meijers, & ter Weel, 2008; MacCann, Duckworth, & Roberts, 2009; Trautwein et al, 2006). In the long run, more conscientious individuals are more likely to reach higher levels of educational attainment, to be employed, have higher incomes, and to be healthier. They are also less likely to engage in criminal or risky behaviors such as drug use (Almlund et al., 2006; Heckman, Stixrud, & Urzua, 2006; Jones et al., 2006; Kaestner & Callison, 2011; ter Weel, 2008). Noncognitive skills matter for well-being in life.

Prior to this interest in noncognitive skills, education policy research has primarily emphasized, as a key outcome of interest, student achievement on standardized tests of cognitive ability. This focus has been prevalent particularly since the 1980s when a report titled *A Nation at Risk* was released by the US federal government. This report documented many ways in which achievement in reading, math, and other academic content areas were declining among American students. It asserted, for example, that “the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people.” Alarmist in its tone, this report galvanized numerous school reform efforts and, among

many proposals, called for implementation of curricular standards and standardized testing. In 2001, these efforts culminated in the federal No Child Left Behind Act, which mandated all states to regularly administer standardized tests in math and reading to ensure basic levels of academic proficiency (Vinovskis, 2009).

This requirement together with the availability of standardized tests of cognitive ability that have long been under development by psychometricians and the means to collect student performance data at scale led to a proliferation of education research that evaluated educational programs and policy based upon how they affected student achievement. Whether explicitly mentioned or tacitly implied, educational effectiveness meant effectiveness at improving student achievement.

The emphasis on measuring and improving student cognitive ability is not entirely unfounded. For many decades, human capital research has pointed out the important role that academic achievement and cognitive ability plays in long-run life outcomes. Greater cognitive ability is associated with higher levels of educational attainment and better labor-market outcomes (Becker, 1964; Hanushek, 2011). More recent work shows that students realize larger gains in future income and educational attainment when they are taught by teachers who are more effective at improving student achievement on standardized tests of cognitive ability (Chetty, Friedman, & Rockoff, 2014).

However, scholars and policymakers are beginning to recognize the limits of relying exclusively on student achievement outcomes. Not only do noncognitive skills affect long-run life well-being net of the effect of cognitive skills but the benefits of many educational interventions and policies are understated when noncognitive skill outcomes are overlooked. Consider the research on charter schools, private school vouchers, and early-childhood education

programs. Using experimental research methods, scholars have often found that these interventions do not dramatically alter student achievement, yet they have large impacts on other long-run life outcomes such as educational attainment, criminal behavior, health, employment, and earnings. Presumably, these interventions affected student noncognitive skills but not cognitive skills (Booker et al., 2014; Cowen et al., 2013; Heckman, Pinto, & Savelyev, 2013; Wolf et al., 2013). More complete appraisals and a more accurate understanding of educational interventions and policy requires inquiry into their effects on cognitive and noncognitive outcomes.

For this dissertation, I present three papers investigating the formation of student noncognitive skills. In the first two papers, I estimate teacher impacts on student noncognitive skills, paying particular attention to how teacher conscientiousness affects their student's conscientiousness. This analysis is done in two data sets that allow for different degrees of causal inference. In chapter 2, I use panel data and student fixed effects to examine year-to-year changes in student conscientiousness as they sort through teachers of varying levels of conscientiousness during their secondary school years. In chapter 3, I use data from the Measures of Effective Teaching Project where teachers were randomly assigned to classrooms to estimate causal impacts of teachers with differing levels of conscientiousness on student conscientiousness. I find that students experience gains in conscientiousness when they have teachers who are more conscientious. Curiously, teachers who are effective at improving student test scores do not have commensurate effects on this set of student noncognitive skills.

In these two studies, I also utilize a novel performance measure of conscientiousness based upon the effort that teachers and students exert to complete surveys. Specifically, I use the frequency with which respondents skip questions, appear to provide careless answers, or

altogether fail to even begin the survey as behavioral proxies for conscientiousness. This approach to measuring noncognitive skills is not prone to social desirability bias, reference-group bias, and other limitations endemic to self-reported measures of noncognitive skills (Duckworth & Yeager, 2015). Recent research has validated most of these survey-effort measures of noncognitive skills related to conscientiousness (Hitt, Trivitt, & Cheng, forthcoming; Hitt, 2015; Zamorro et al., 2016). The analyses in Chapter 2 and Chapter 3 showcase the application of these innovative measures. I describe these measures in greater detail later but, for now, highlight them as a valuable contribution to the measurement and future study of noncognitive skills.

For the third paper, I investigate the potential of a policy proposal to improve the development of student noncognitive skills. Based upon the prior two studies, it appears that teachers transmit noncognitive skills to their students by modelling them. That is, teachers embody a particular value system which students seem to emulate. Indeed, psychologists, sociologists, and philosophers have theorized that norms and values are socially learned in this fashion (Bandura, 1977; Bandura & Walters, 1963; MacIntyre, 2007). Schools are communities that possess an ethos as well as conceptions of the means and ends of education. These foundational features of school communities guide the personal and collective behavior of their members. Thus, schools with greater coherence in values are likely to be more effective at inculcating the noncognitive skills associated with their respective values systems (Coleman & Hoffer, 1987; see also Bellah et al., 1985).<sup>1</sup>

---

<sup>1</sup> For example, scholars point out that Catholic schools are largely successful at improving student achievement and educational attainment (Altonji, Elder, & Taber, 2005; Evans & Schwab, 1995; Grogger & Neal, 2000; Hoffer, Greeley, & Coleman, 1985; Neal, 1997). They attribute this success to the well-defined mission and values system in Catholic school communities. Catholic school teachers and leaders are able to instill in their students a particular

How, then, can school leaders promote values coherence among their school communities to enhance the development of particular noncognitive skills? I propose and test the hypothesis that granting school leaders greater autonomy over personnel decisions will enable them to more effectively foster values coherence within their schools. Research from public administration and organizational management suggests that this flexibility would enable school leaders to ensure alignment between organizational values and the values of individual personnel (Downs, 1967; Maranto & Wolf, 2013; Wilson, 1989). Using a nationally representative sample of public schools, I descriptively show that there is greater agreement and coherence over school mission and values when principals have more autonomy over hiring and dismissing teachers. Identifying ways for schools and other educational programs to more clearly and consistently embody a set of organizational values may further enhance the development of student noncognitive skills, though directly testing this hypothesis is outside the scope of this analysis.

These papers illustrate the variety of ways in which schools and educational interventions affect students. In particular, they deepen the understanding of the mechanisms behind the formation of noncognitive skills and suggest a way forward for policymakers and practitioners who seek to improve noncognitive skill outcomes – all of which bear upon the wellbeing of student who are being served.

---

set of noncognitive skills that are consistent with the Catholic school mission and value system, which in turn influence student outcomes (Coleman & Hoffer, 1987; Bryk, Lee, & Holland, 1993). That being said, the theory that greater values coherence enhances the development of particular student noncognitive skills can be empirically tested. However, I am unaware of any studies that have directly tested it, nor am I able to test it with my data.

## References

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. D. (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 1–181). Amsterdam: Elsevier.
- Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1), 151–184.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A., & Walters, R.H. (1963). *Social learning and personality development*. New York: Holt, Rinehart and Winston.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York, NY: National Bureau of Economic Research.
- Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1996). *Habits of the Heart: Individualism and Commitment in American Life*. Berkeley, CA: University of California Press.
- Booker, K., Sass, T., Gill, B., & Zimmer, R. (2014). *Charter high schools' effect on long-term attainment and earnings*. Retrieved from [http://www.mathematica-mpr.com/publications/pdfs/education/charter\\_long-term\\_wp.pdf](http://www.mathematica-mpr.com/publications/pdfs/education/charter_long-term_wp.pdf)
- Borghans, L., Meijers, H., & Bas, ter W. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46(1), 2–12.
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679.
- Coleman, J. S., & Hoffer, T. (1987). *Public and Private High Schools*. New York, NY: Basic Books.
- Cowen, J. M., Fleming, D. J., Witte, J. F., Wolf, P. J., & Kisida, B. (2013). School Vouchers and Student Attainment: Evidence from a State-Mandated Study of Milwaukee's Parental Choice Program. *Policy Studies Journal*, 41(1), 147–167.
- Downs, A. (1967). *Inside Bureaucracy*. Boston, MA: Little, Brown and Company.
- Evans, W. N., & Schwab, R. M. (1995). Finishing High School and Starting College: Do Catholic Schools Make a Difference? *The Quarterly Journal of Economics*, 110(4), 941–974.



- Grogger, J., & Neal, D. (2000). Further Evidence on the Effects of Catholic Secondary Schooling [with Comments] (pp. 151–201). Washington, DC: Brookings Institution Press.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. *Educational Researcher*, 44(4), 237–251.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466–479.
- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6), 2052–2086.
- Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Hitt, C. (2015). Just filling in the bubbles: Using careless answer patterns on surveys as a proxy measure of noncognitive skills (EDRE Working Paper 2015-06). Fayetteville, AR: Department of Education Reform, University of Arkansas.
- Hitt, C., Trivitt, J.R., Cheng, A. (forthcoming). When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*.
- Hoffer, T., Greeley, A. M., & Coleman, J. S. (1985). Achievement Growth in Public and Catholic Schools. *Sociology of Education*, 58(2), 74–97.
- Jones, D. E., Greenberg, M., & Crowley, M. (n.d.). Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. *American Journal of Public Health*.
- Kaestner, R., & Callison, K. (2011). Adolescent Cognitive and Non-cognitive Correlates of Adult Health. *Journal of Human Capital*, 5(1), 29–69.
- MacCann, C., Duckworth, A.L., & Roberts, R.D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, 19(4), 451–458.
- Macintyre, A. (2007). *After Virtue: A Study in Moral Theory* (3rd ed.). Notre Dame, IN: University of Notre Dame Press.
- Maranto, R., & Wolf, P. J. (2013). Cops, teachers, and the art of the impossible: Explaining the lack of diffusion of innovations that make impossible jobs possible. *Public Administration Review*, 73(2), 230–240.
- Neal, D. (1997). The Effect of Catholic Secondary Schooling on Educational Attainment. *Journal of Labor Economics*, 15(1), 98–123.

- ter Weel, B. (2008). The Noncognitive Determinants of Labor Market and Behavioral Outcomes: Introduction to the Symposium. *Journal of Human Resources*, 43(4), 729–737.
- Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, 98(2), 438–456.
- West, M., Kraft, M. A., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148–170.
- Wilson, J. Q. (1991). *Bureaucracy: What Government Agencies Do and why They Do it*. New York, NY: Basic Books.
- Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC. *Journal of Policy Analysis and Management*, 32(2), 246–270.
- Vinovskis, M. (2009). *From A National Risk to No Child Left Behind: National Education Goals and the Creation of Federal Education Policy*. New York, NY: Teacher's College Press.
- Zamarro, G., Cheng, A., Shakeel, M., & Hitt, C. (2016). Comparing and validating measures of character skills: Findings from a nationally representative sample. Paper presented at the Association for Education Finance and Policy 41st Annual Conference. Denver, CO.

## **Chapter 2: Teacher effects on Student Conscientiousness: Evidence from Panel Data**

School reform and improvement efforts are often judged by how well they improve student cognitive ability as measured by student achievement on standardized test scores. This emphasis is not entirely misplaced. Student achievement growth on test scores is a meaningful predictor of later-life outcomes such as educational attainment and income (Chetty, Friedman, & Rockoff, 2014; Hanushek, 2011; Murnane et al., 2000). Schools, therefore, play an important role in developing cognitive skills among their students because such skills pertain to their students' future wellbeing.

However, schools do not only convey content knowledge and develop cognitive skills. They also convey value systems and social norms that may shape their students' personality, behavioral tendencies, and character. These traits are referred to as noncognitive skills, or skills that are not easily captured by test scores in math or reading (Borghans et al., 2008; Duckworth & Yeager, 2015; West et al., 2016). Economists, psychologists, and other social scientists are paying greater attention to noncognitive skills as emerging research demonstrates that noncognitive skills are positively linked to student outcomes such as health, crime, educational attainment, income, and employment (Almlund et al., 2011, Heckman, Stixrud, & Urzua, 2006). Importantly, the relationship between noncognitive skills and student outcomes holds above and beyond the impact of cognitive skills on these outcomes. Such a result has spurred additional inquiry into how schools and other educational institutions can inculcate these noncognitive skills for their students' wellbeing.

Existing research shows that teachers play a large role in affecting cognitive skills (Rivkin, Hanushek, & Kain 2005; Rockoff, 2004). Given that teachers have impacts on student cognitive skills, it is reasonable to expect that teachers may also have impacts on student

noncognitive skills. However, there is less research showing the extent to which teachers affect student noncognitive skills (but see Blazar & Kraft, 2015; Gershenson, 2016; Jackson, 2012; Jennings & DiPrete, 2010; Koedel, 2008; Kraft & Grace, 2016). There are largely two reasons for this gap in the research. First, the little research of teacher impacts on student noncognitive skills is partly due to the predominant focus by researchers and policymakers on cognitive skill development. A second reason for the lack of research on noncognitive skills is the difficulty in measuring noncognitive skills. Unlike standardized tests, self-reported surveys of noncognitive skills are rarely administered on a regular basis. And even if those surveys are regularly administered, they are prone to nontrivial measurement issues such as social desirability bias or reference group bias (Duckworth & Yeager, 2015).

This study is motivated by the lack of research into the role that teachers play in their students' noncognitive skill development. Providing additional empirical evidence of the ability for teachers to influence student noncognitive skills would provide a broader picture of the effects that teachers have on their students. If so, further work to identify observable teacher characteristics associated with effects on student noncognitive skills would advance the understanding of the noncognitive skill development process. For instance, one could ask why a particular type of teacher influences student noncognitive skills more than others and use that insight to develop theories about the development of noncognitive skills. Such insight would also be useful for developing effective educational interventions aimed at improving student noncognitive skills. Evidence that teachers affect student noncognitive skills would also carry implications for conceptualizations of teacher quality, which at the moment is typically limited to a teacher's ability to improve student cognitive ability.

In this study, I use a five-year, longitudinal dataset of students with student fixed effects to track student noncognitive skill development. Moreover, I avoid typical issues when it comes to measuring noncognitive skills by using a behavioral measure, as opposed to a self-reported measure, of noncognitive skills. Specifically, I use item response rates on surveys as a proxy for student noncognitive skills related to conscientiousness. There are theoretical reasons to believe that survey response patterns are not random but are related to certain noncognitive skills. Indeed, some research has demonstrated this proposition (Robinson-Cimpian, 2014). Other work has explicitly validated item response rates as a measure of survey effort as a proxy for noncognitive skills related to conscientiousness (Hitt, Trivitt, & Cheng, forthcoming).

As I show below, under straightforward ordinary least squares estimation techniques, item response rates for students in my data are predictive of their educational attainment and employment status measured nearly 20 years later. More importantly, using student-fixed effects to leverage the longitudinal nature of the data, I find that students experience increases in item response rates when they have teachers that are diligent enough to complete and return their own surveys for data collection. This pattern may suggest that teachers with a particular set of noncognitive skills instill similar noncognitive skills into their students. This is the first study to demonstrate a relationship between teacher and student performance on similar behavioral tasks. Interestingly, teachers that influence student noncognitive skills do not appear to influence student cognitive skills as measured by test scores.

The remainder of the article is divided into four sections. In the subsequent section, I review the research on noncognitive skills, paying particular attention to the theory of how they can be inculcated in students. I then detail the methods of this study in the second section and present the results in the third section. I discuss these findings in greater detail and conclude in

the fourth section. Overall, I interpret these results as evidence that teachers play a role in inculcating certain noncognitive skills that are important for students' long-run life outcomes. Furthermore, different teachers have varying effects on the development of their student noncognitive as well as cognitive skills.

## **Literature Review**

### **Teacher Impacts on Student Cognitive Ability**

Teacher quality is the most important specific school factor for improving student cognitive ability (Rivkin et al., 2005; Rockoff, 2004). Some research indicates that high-quality teachers, as measured by their ability to raise student math and reading test scores, improve longer-run outcomes such as their students' educational attainment and employment income (Chetty et al., 2014).

Nonetheless, scholars are generally unable to identify high quality teachers based upon observable characteristics absent measures of student achievement. For instance, years of teaching experience are generally uncorrelated with teacher quality after the first three to five years of teaching (Hanushek & Rivkin, 2006; Clotfelter, Ladd, & Vigdor, 2006; Goldhaber, 2007). Teacher licensure is likewise not strongly correlated with a teacher's ability to raise student scores on achievement tests (Buddin & Zamarro, 2009; Hanushek & Rivkin, 2006; Kane, Rockoff, & Staiger 2008; Podgursky, 2005). Although there is some evidence that having more content knowledge, as measured by the number of courses taken in that content area, is associated with higher teacher quality, this relationship largely holds mostly for secondary school teachers in math or science (Clotfelter et al., 2006). There is also a lack of evidence that pedagogical knowledge for a specific content area is linked with student achievement (but see Hill, Rowan, & Ball, 2005). Although some research has demonstrated that achievement is higher for students with teachers that have higher cognitive ability, as measured by their

performance on the Praxis or other standardized licensure tests (Goldhaber, 2007; Clotfelter et al., 2006), other work finds no relationship between teacher cognitive ability and student achievement (Buddin & Zamarro 2009). Finally, Duckworth, Quinn, and Seligman (2009) provide suggestive evidence that some teacher noncognitive abilities (e.g., grit and life satisfaction) are positively correlated with student gains in cognitive ability. However, their analysis is based upon a convenience sample of an atypical group of first- and second-year Teach for America teachers.

In summary, research suggests that teacher quality matters for student wellbeing, but it is difficult to predict teacher quality solely based on teacher inputs and observable characteristics. This has led to some proposals to relax the selection of teachers based upon inputs (e.g., credentials) and to evaluate teachers based upon their outputs or actual performance (e.g., student achievement) (Podgursky, 2005; Kane et al., 2008; Hanushek, 2011). Notably, these proposals all define teacher quality as the ability to increase student achievement, or cognitive ability, as measured by growth on standardized test scores in math and reading. This approach raises the important issue of whether teachers are able to raise student noncognitive ability and whether these increases in noncognitive ability yield benefits for students in the long-run net of increases in cognitive ability. If so, there may be reason to consider teacher impacts on student noncognitive ability when conceptualizing teacher quality, especially if teachers who have impacts on student noncognitive ability are not the same teachers who have impacts on student test scores.

### **Teacher Impacts on Student Noncognitive Ability**

Whether teachers have impacts on student noncognitive ability has received little empirical attention. One reason for the lack of this research is the infrequent systematic collection of noncognitive skill measures. Consider student achievement data, or measures of

cognitive skills. Given mandates for annual testing and data systems that link student data to data about their teachers, researchers can estimate value-added models and identify causal effects of teachers on student cognitive skills, though there is debate about the validity of these models (Gaurino, Reckase, & Wooldridge, 2015; Koedel & Betts, 2011; Rothstein, 2009). In contrast, measures of noncognitive skills are not sufficiently developed for accountability purposes (Duckworth & Yeager, 2015). Nor are measures of noncognitive skills typically administered to students, much less linked to the students' teachers.

There are, however, exceptions to the dearth of research into teacher impacts on student noncognitive skills. Using data from North Carolina, Jackson (2012) estimates a factor model based on a set of non-test score outcomes (i.e., GPA, classroom attendance, suspensions, on-time grade progression) to proxy for a student's noncognitive skills. He finds that teachers have demonstrable effects on this measure of student noncognitive skills net of their impacts on student test scores (i.e., cognitive skills). In other work, Koedel (2008) provides evidence that variation in teacher quality explains differences in high school dropout rates. This finding can be interpreted as teachers having differential effects on student noncognitive skills that lead to different attainment outcomes, assuming that educational attainment is driven by noncognitive skills, as Heckman and Rubenstein (2001) suggest. However, it is not clear how much of the association between teacher quality and dropout rates is driven by teacher impacts on student cognitive skills as cognitive skills are also important determinants of educational attainment. It is also possible that improvements to noncognitive skills lead to improvements in cognitive skills, which in turn, lead to positive student outcomes (Heckman, 2000). Moreover, teachers that have impacts on noncognitive skills may not necessarily be the same ones that have impacts on



cognitive skills. Jackson (2012) finds this to be the case in his work, as do Jennings and DiPrete (2010).

Other research indirectly suggests that schools and teachers have impacts on student noncognitive skills. Several educational interventions have not demonstrably improved student test scores yet have improved other student outcomes attributable to gains in noncognitive skills. For example, evaluations of several school choice programs, such as charter schools and private-school vouchers, have little to no impact on test scores but do have impacts on educational attainment (Booker et al., 2013; Chingos & Peterson, 2015; Wolf et al., 2013). Likewise, gains in cognitive ability from being randomly assigned to early-childhood interventions, such as the Perry Preschool Project, are known to dissipate when children enter elementary school. Yet students who participated in these early childhood interventions realize improvements in labor-market and health outcomes in adulthood as well as reductions in the incidence of criminal behavior (Heckman, Pinto, & Savelyev, 2013). The effects that these educational interventions have on educational attainment, health, crime, and labor market outcomes together with *the lack* of corresponding gains in cognitive ability suggests that schools and teachers have important impacts on student noncognitive skills.

### **How Do Teachers Affect Student Noncognitive Ability?**

Although schools and their teachers appear to have impacts on student noncognitive ability, the channels through which they have such impacts are unclear. Character education and other similar formal curricula, for example, are aimed at improving noncognitive skills but there is little understanding of how they alter noncognitive skills. In fact, there is little evaluation of whether character education programs even alter noncognitive skills in the first place (Berkowitz & Bier, 2004).

Other work has investigated whether particular instructional approaches are more effective at improving noncognitive skills, but no strong relationship between the two has been found (Jennings & DiPrete, 2010). Presumably, different pedagogical practices could generate different experiences for students and lead to the development of particular noncognitive skills (Ames, 1992; Dweck, 2006; Yeager et al., 2014). Still, research on noncognitive skills is relatively nascent and has merely established the importance of noncognitive skills for student outcomes. Scholars have not empirically demonstrated systematic ways to improve noncognitive skills for students in primary and secondary schools at scale. Nor do they clearly understand the mechanisms behind the development of noncognitive skills.

It is also possible that noncognitive skill development occurs in less technical ways. Psychologists and sociologists have long proposed that learning is social (Bandura, 1977). Some have more specifically argued that individuals learn group norms by observing the behaviors of other group members, called social referents, in specific situations. A social referent helps individuals discern what types of behaviors are acceptable or unacceptable by allowing them to observe what behaviors are rewarded or sanctioned within the group (Sherif & Sherif, 1964). Teachers are particularly well-situated to act as role models, instilling a set of traits derived from a certain value system into their students (Berkowitz & Bier, 2004). Indeed, some scholars believe this mechanism partially explains why Catholic schools have been successful at improving student outcomes. Catholic schools are rich communities with a well-defined value system that is embodied by their teachers and other school workers. The values, in turn, are inculcated into their students and play a role in the formation of particular character traits and personality dispositions (Bryk et al., 1993; Coleman & Hoffer, 1987).

## **Issues with measuring noncognitive skills**

Previously described evidence from school choice and early childhood education research supports the prevalent intuition that schools affect student noncognitive skills. Nonetheless, there have been very few direct empirical tests of whether individual teachers influence student noncognitive skills. This study fills this gap and uses a novel method to measure noncognitive skills to do so.

As mentioned earlier, one reason for the lack of this evidence is the infrequent collection of noncognitive skill measures. But even if psychometric scales were regularly administered to students, researchers have another problem: Scales designed to measure noncognitive skills are prone to social desirability bias, satisficing, and similar problems endemic to survey data. The potential for this type of systematic error is always present in self-reported data (Krosnick, Narayan, & Smith, 1996).

Even if students honestly answer items on a survey, another problem remains. Consider an item that asks a student to specify how hard-working he is. Although that student may be honest in his response, his assessment of what it means to be a hard worker is relative to some external standard. This problem is called reference group bias and may be the reason behind paradoxical research results where students who experience improved outcomes (e.g., test scores, educational attainment, criminal behavior) rate themselves lower on the very noncognitive skills that are supposedly positively correlated with those outcomes (Duckworth & Yeager, 2015; Dobbie & Fryer, 2015; West et al., 2016).

To circumvent the limitations of using self-reported measures of noncognitive skills, I use a performance task to solicit a behavioral measure of noncognitive skills. Specifically, I argue that completing a survey can be viewed as a performance task. Surveys of sufficient length are tedious tasks and much like homework assignments. Completing them and refraining from

skipping items requires a great deal of diligence and persistence. Students need to heed instructions, respect those who are assigning the task, and exert basic effort to respond to the items. In this respect, survey effort is a measure of a particular set of noncognitive skills (Robinson-Cimpian, 2014). In the analysis below, I use item response rates — or the extent to which students do not shirk and skip questions — as measures of noncognitive skills to examine whether individual teachers are able to alter their students' noncognitive skills. Using six large-scale longitudinal data sets Hitt et al. (forthcoming) have validated using item response rates on surveys as a measure of persistence and effort. They find, for instance, that individuals who have higher item response rates in adolescence complete more years of schooling, even after controlling for measures of cognitive ability. So there is not only theoretical reason to believe that item response rates capture traits associated with conscientiousness but also empirical evidence to do so as the measure predicts later-life outcomes in the same way as expected from those noncognitive skills. In the next section, I present the use of item response rates and other methods of my analysis in greater detail.

## **Methods**

### **Data**

Data for this analysis come from the Longitudinal Study of American Youth (LSAY). In 1987, a nationally representative sample of public school seventh and tenth graders was selected to participate in the panel. LSAY was intended to provide descriptive data about students from adolescence into adulthood. In particular, LSAY focused on gathering information about students' attitudes towards science and math, their career prospects in those fields, and opinions regarding their math and science classes.

This study focuses on LSAY's seventh-grade cohort, which consists of approximately 3,000 students. These students were biannually surveyed and annually completed standardized

tests in math and science through their twelfth-grade year in 1994. Surveys and standardized tests occurred in separate 50-minute sessions during the school day and were administered by research coordinators. LSAY additionally surveyed each student's respective math and science teachers in each year of the study. Teachers received surveys and were provided the means to return completed surveys via the postal service. Because teachers were not surveyed when this student cohort was in twelfth grade, the analysis is based upon data from the seventh through eleventh grades. Students were subsequently surveyed as adults from 2007-2011, also on an annual basis (Miller, 2014).

### **Response Rate as a Proxy for Noncognitive Skills**

Item response rates on student surveys are used as behavioral measures for student noncognitive skills related to conscientiousness. In each wave of the LSAY, students generally faced between 150 and 360 items on the questionnaires. Such a lengthy survey lends additional credence for interpreting item response rates as behavioral indications of noncognitive skills. Item response rates are simply the proportion of items that students answer out of the total number of questions students were asked to answer. Item response rates are computed for each wave of LSAY. Table 1 displays summary statistics of item response rates for students from seventh through eleventh grade. Most students complete a majority of the surveys they are asked to fill out, though there is still variation in the number of items that they skip. For the analysis, student item response rates are standardized by year as differences over time and across annual survey instruments may induce year-to-year differences in item response rate (e.g., see the drop in item response rates for 10th grade due to different item formats on the questionnaire). Students who were absent during data collection periods are not included in the analysis, as opposed to imputing a response rate of 0 percent.

Initially, I computed teacher item response rates just as I calculated student item response rates. However, unlike students who were required to complete surveys during an assigned time in school, teachers were mailed surveys via the postal service and asked to return completed forms. This data collection method resulted in a large number of teachers who never completed a survey. Moreover, those who did complete a survey completed between 90 and 100 percent of the items, yielding little variation in teacher item response rate and little study power. Consequently, I use a binary variable indicating whether a teacher did or did not return the survey during each respective wave of LSAY. The assumption is that this binary variable also captures each teacher's level of conscientiousness, which was required to complete and return the survey via postal mail.<sup>2</sup> To summarize, whether a teacher returns the survey is the measure of teacher noncognitive skills, while item response rate is the measure of student noncognitive skills. Table 1 shows summary statistics for teacher return rates.

### **Empirical Strategy**

**Validation of Item Response Rates.** Before conducting the analysis to determine if teachers affect student noncognitive ability, it would be helpful to empirically validate student item response rate as measures of noncognitive ability instead of only relying on aforementioned theoretical reasons and prior research of other datasets to argue that item response rates capture noncognitive skills. To provide empirical validation of response rate as a proxy for noncognitive

---

<sup>2</sup> One may worry that a binary indicator for whether a teacher returns a survey is a noisy measure of teacher conscientiousness. It is possible that such an indicator has low year-to-year correlation for individual teachers. Because hardly any teachers were asked to complete a survey in consecutive years — they were only asked to complete a survey if they had a student in the LSAY sample — such correlations cannot be calculated. Nevertheless, a low year-to-year correlation in the dummy variable indicating whether a teacher returns the survey leads to conservative hypothesis tests. Thus, one can be more confident that statistically significant relationships between this variable and other variables are material and not spurious, should such relationships be found.

Table 1

*Student Item Response Rates and Teacher Return Rates*

	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
<i>Student Item Response Rates (%)</i>					
Mean	97.6	98.3	98.0	86.5	96.8
Standard Deviation	5.5	4.2	4.2	28.9	6.2
Minimum	5.5	3.5	18.8	1.1	9.6
Average Number of Questions Faced	178	265	282	149	288
<i>Math Teacher Return Rates (%)</i>					
Students with Teachers who do not Return the Survey	15.6	18.2	10.6	25.4	26.2
<i>Science Teacher Return Rates (%)</i>					
Students with Teachers who do not Return the Survey	17.9	23.4	16.8	31.0	30.8

Note: Maximum student item response rates for each semester are 100 percent. Teachers were only surveyed in spring semesters.

skills, I follow Hitt et al. (forthcoming) and run a series of regression models using item response rates to predict long-run life outcomes. It is sufficient to show that item response rate in adolescence mimics other measures of noncognitive skills by being an important and independent determinant of long-run life outcomes (Almlund et al., 2011). In particular, I average each student's survey response rates from seventh through ninth grade<sup>3</sup> and use them as an independent variable in regressions to predict educational attainment and future employment, as shown in the following model:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \mu_i, \quad (1)$$

In Equation 1,  $Y_i$  is a long-run outcome for student  $i$  (e.g., educational attainment, employment status) and  $S_i$  is student  $i$ 's average response rate from seventh through ninth grade.  $X_i$  is a vector of student background characteristics, such as student gender, race, mother's education and the urbanicity and US region of the student's school. I also average each student's math and science test scores from the seventh to ninth grade to control for cognitive ability. Outcome variables are measured about 24 years later when the respondents are in their late thirties.

Unlike student response rates, the data do not allow me to conduct a similar validation exercise for the behavioral measure of teacher conscientiousness. Whether an indicator for whether a teacher returns or does not return a survey captures teacher conscientiousness relies on theoretical assumptions. Discussion of this limitation as it pertains to the results is presented later in the article.<sup>4</sup>

---

<sup>3</sup> Item response rates and test scores from multiple years are incorporated to improve the precision of these estimates. Point estimates are robust to using fewer years of data. I also do not incorporate information from grades 10 and 11, the point where many students drop out of school, to avoid compositional effects.

<sup>4</sup> In chapter 3, I use other data to validate whether teachers fail to turn in their surveys captures meaningful characteristics. As it turns out, teachers who fail to complete surveys also receive



**Main analysis.** To investigate whether particular types of teachers affect student noncognitive ability, I estimate models that examine changes in students' noncognitive skills as they are taught by teachers with varying levels of noncognitive skills over their secondary schooling experience. I take advantage of the longitudinal nature of the LSAY by including student fixed effects to control for time-invariant heterogeneity across students. In doing so, I examine within-student changes in conscientiousness as they face teachers with different levels of conscientiousness. I estimate models of the form:

$$Y_{it} = \beta_0 + \beta_1 N_{it} + \beta_2 X_{it} + \gamma_{it} + \nu_i + \epsilon_{it}. \quad (2)$$

In equation (2),  $Y_{it}$  is student  $i$ 's response rate in the spring semester of time period  $t$ . Values for  $Y_{it}$  are expressed as standard deviations and standardized by year.  $N_{it}$  is an indicator equal to one if student  $i$ 's teacher did not return a survey in time period  $t$  and equal to zero otherwise.  $X_{it}$  is a vector of time-varying student characteristics such as student cognitive ability as measured by standardized math and science tests.<sup>5</sup> Meanwhile,  $\gamma_{it}$  is a vector of year dummies that capture secular time trends,  $\nu_i$  is a student-fixed effect, and  $\epsilon_{it}$  is the time-varying error term. This model is run twice: once where  $N_{it}$  represents the student's math teacher and once where  $N_{it}$  represents the student's science teacher. The coefficient of interest,  $\beta_1$ , captures the influence that teachers who either return or fail to return their survey have on student noncognitive skills based on year-to-year variation in student item response rates or test scores.

---

lower ratings on other indicators of teacher quality, such as those based upon formal classroom observations and student ratings.

<sup>5</sup> Standardized test scores in reading and language arts are not available in the data. One may question, therefore, the validity of the measure of cognitive ability. However, it should be pointed out that the concern is likely overstated, given that scores on standardized math and science tests are highly correlated with scores on reading tests.

I also run an additional model to assess marginal changes in student item response rate in years where the student has a math teacher and a science teacher who both do not return the survey. In particular, I estimate:

$$Y_{it} = \beta_0 + \beta_1 N^a_{it} + \beta_2 N^b_{it} + \beta_3 X_{it} + \gamma_{it} + \nu_i + \epsilon_{it}, \quad (3)$$

where  $N^a_{it}$  is an indicator equal to one if either student  $i$ 's math or science teacher but not both failed to return the survey in time period  $t$ .  $N^b_{it}$  is an indicator equal to one if both student  $i$ 's math and science teachers did not return the survey in time period  $t$ . The other variables are as they are in equation (2).

One difficulty in conducting research of teacher impacts on student outcomes is accounting for the nonrandom assignment of students to teachers. Principals may systematically assign students to teachers. Without accounting for such patterns, one cannot be confident that causal impacts have been identified (Rothstein, 2009). The inclusion of student fixed effects in this study helps to address this concern by controlling for time-invariant student characteristics that may be correlated with the assignment of students to teachers. But insofar as such assignment is correlated with time-*invariant* factors, these results do not have a causal interpretation. For instance, more conscientious parents may improve their child's conscientiousness while also making sure that their child is assigned to a more conscientious teacher. Unless such parental efforts are constant over time, student fixed effects models will confound teacher effects with parent effects. Similar caveats must be considered when interpreting the results.

The use of longitudinal data together with the inclusion of student fixed-effects also effectively controls for any observable and unobservable time-invariant student characteristics (e.g., gender, race, socioeconomic status, family background or upbringing) that could affect

outcomes of interest. Put differently, item response rates for each student are compared to item response rates for the same student in different years of the data and after accounting for annual trends in response rate over time. Thus, the coefficient estimates of the indicator for whether a teacher completed the survey or not captures annual increases or decreases in student item response rate (i.e., conscientiousness) as the student sorts through teachers who either return or fail to return the teacher-level survey (i.e., teachers with varying levels of conscientiousness) over time.

I also run analogous models where the dependent variable is student test scores instead of student item response rate. These models examine whether teachers with the noncognitive skills captured by whether or not they respond to surveys have a differential impact on student cognitive ability relative to teachers who exhibit those noncognitive skills to a lesser extent.

**Additional models.** Note that the aforementioned models do not control for student item response rates from prior years due to a methodological issue that arises when estimating student fixed-effects models that control for lagged measures of response rate. A key identifying assumption for obtaining valid estimates with student fixed-effects is strict exogeneity, which requires that all model covariates are uncorrelated with the time-varying error term in all periods. But including lagged measures of the dependent variable — here, student response rate — mechanically introduces such correlation. To address this issue and to control for prior-year item response rates, one can first difference equation (2) and instrument for the lagged change in student response rate with twice-lagged student response rate. This instrumental variables technique was originally proposed by Anderson and Hsiao (1981). Alternatively, one could use a general method of moments estimation technique proposed by Arellano and Bond (1991), which essentially uses additional lags of student response rate to instrument for lagged student response

rate. I use both of these techniques to estimate models that include prior-year measures of student response rates, which capture how students' year-to-year growth in conscientiousness changes as they encounter teachers with varying levels of a similar noncognitive skill throughout secondary school.

## Results

### Validation of Student Item Response Rate

As mentioned earlier, it is worthwhile to provide evidence that item response rate is a legitimate measure of noncognitive skills in LSAY. Table 2 shows the results of regressions where student item response rates and test scores are used to predict two long-run life outcomes — educational attainment and employment. Item response rates are generally associated with these long-run life outcomes much in the same way as noncognitive skills above and beyond the contribution of cognitive ability. As shown in columns 1 and 2, increasing item response rate or test scores by one standard deviation is associated with completing about one additional year of education. Moreover, column 3 indicates that increasing item response rates by one standard deviation is associated with completing almost an additional half of a year of education net of the impact of test scores.<sup>6</sup>

Item response rates in adolescence is also positively correlated with labor-market outcomes, specifically, future employment. As shown in column 4 of Table 2, an increase of one standard deviation in item response rate is associated with approximately a six-percent increase

---

<sup>6</sup> One could also use multinomial logistic regression to estimate the relationship between item response rates and educational attainment specified as a categorical variable indicating the highest level of education completed (e.g., high school dropout, high school diploma, bachelor's degree, etc.). Doing so does not change the results. These estimates are available from the author upon request.

in the likelihood of being employed.<sup>7</sup> Meanwhile, an increase of one standard deviation in cognitive ability as measured by standardized test scores is associated with a four-percent increase in the likelihood of being employed (see column 5). Finally, item response rate remains correlated with employment above and beyond the impact of cognitive ability (see column 6).

Table 2

*Student Response Rate, Test Scores, Years of Education, and Employment*

	Dependent Variable					
	Years of Education			Employment		
	(1)	(2)	(3)	(4)	(5)	(6)
Item Response Rate	0.99** (0.16)		0.49** (0.14)	0.059** (0.016)		0.045* (0.019)
Test Scores		0.98** (0.09)	0.90** (0.10)		0.038** (0.015)	0.029† (0.015)
R <sup>2</sup>	0.24	0.32	0.33	n/a	n/a	n/a

Notes: N = 1,556. Linear regression coefficients are reported for educational attainment outcomes. Marginal effects after logit estimation are reported for employment outcomes. Explanatory variables are expressed in standard deviations and are averaged using the respondent's seventh through ninth grade data. Educational attainment and employment outcomes measured 24 years after initial wave of LSAY data collection. Control variables include student's race, gender, and mother's educational attainment as well as the urbanicity and US region of the student's school. †p<0.1, \* p<0.05, \*\* p<0.01.

**Main Analysis**

**Variation in Independent Variable of Interest.** I now return to answering the primary objective of this study, which is to examine whether teachers influence the noncognitive

<sup>7</sup> If one controls for educational attainment in these models, the coefficients for item response rates and test scores attenuate given that educational attainment, according to other research, mediates their relationships with labor market outcomes (Cawley et al, 2011; Hitt et al., in press). The coefficient for item response rate becomes approximately 0.038 with a standard error of 0.020 and is marginally significant at the 90 percent confidence level. The coefficient for test scores becomes approximately 0.010 with a standard error of 0.015 and is statistically indistinguishable from zero.

development of their students. Because equations (2) and (3) are estimated using student fixed effects, it is useful to examine the sources of variation in the independent variable of interest, namely, the dummy variable indicating whether a teacher returns the survey or not. Specifically, how much of this variation occurs across students, and how much occurs within students over time? I present this information in Tables 3 through 5.

The variance decomposition in the dummy variable indicating whether the teacher returns his or her survey is shown in Table 3. More variation in this variable occurs within students than between students over time, which is desirable for the student fixed-effects analysis. About 61 percent of overall variation in whether a math teacher returns the survey occurs within students, with the remaining 39 percent due to variation between students. Likewise, about 66 and 34 percent of overall variation in whether a science teacher returns the survey occurs within and between students, respectively.

Table 3

*Variation in Teacher Return Rates*

	Standard Deviation	Percent of Overall Variance (%)
<b>Math Teachers</b>		
Overall	0.49	n/a
Within Students	0.39	61.1
Between Students	0.31	38.9
<b>Science Teachers</b>		
Overall	0.50	n/a
Within Students	0.41	66.2
Between Students	0.29	33.8

Other ways to portray variation in teacher response rate are shown in Table 4. The first column of Table 4 shows the proportion of student-year observations that fell into each category. For example, about 59 percent of student-year observations had math teachers that returned the

teacher-level survey, while the math teachers for the other 41 percent of student-year observations did not. Analogous figures for science teachers are similar. The second column shows the proportion of students who at least once throughout the panel had teachers that returned the survey or did not return the survey. About 95 percent of students in the data had a math or science teacher who returned the survey. Meanwhile, about 80 and 87 percent of students had math or science teachers who did not return the survey, respectively. Finally, column 3 shows the proportion of students who only had a teacher in a single category. About 62 percent of students who ever had math teachers who returned the teacher-level survey always had math teachers who returned the survey. In contrast, about half of the students who ever had math teachers who did not return the survey always had such teachers. Similarly, approximately 57 percent of students always had a science teacher that returned the teacher-level survey, while about 54 percent of students always had a science teacher that did not return the teacher survey.

Table 4

*Variation in whether Teacher Returns Survey*

Category	Proportion of student-year observations that fell into the given category	Proportion of students who fell into the given category at least once throughout the panel	Proportion of students who never switch out of the given category
<b>Math Teachers</b>			
Teacher Returns Survey	58.7	94.9	61.9
Teacher Does not Return Survey	41.3	79.5	51.9
<b>Science Teachers</b>			
Teacher Returns Survey	53.4	94.6	56.5
Teacher Does not Return Survey	46.6	86.8	53.7

Note: All numbers are percentages.

Table 5

*Transition Probabilities for whether Teacher Returns Survey*

		Panel A: Math Teachers	
		<u>Status in Year t</u>	
		Math Teacher Returns Survey	Math Teacher Does not Return Survey
<u>Status in Year t-1</u>	Math Teacher Returns Survey	68.7	31.3
	Math Teacher Does not Return Survey	24.9	75.1

---

		Panel B: Science Teachers	
		<u>Status in Year t</u>	
		Science Teacher Returns Survey	Science Teacher Does not Return Survey
<u>Status in Year t-1</u>	Science Teacher Returns Survey	61.1	38.9
	Science Teacher Does not Return Survey	35.0	75.0

Table 5 displays year-to-year transition probabilities across categories of teacher item response rates. Teacher item response rate categories for prior year are listed down the rows, while teacher item response rate categories for subsequent year are listed across the columns. The diagonal shows the proportion of students who remained the same category in teacher item response rate from one year to the next. As shown in the first entry of the upper panel in Table 5, about 69 percent of students who, for a particular year, had a math teacher that returned the teacher-level survey also had, for the subsequent year, a math teacher that returned the survey. About 75 percent of students who, for a particular year, had a math teacher that failed to return the survey also, for the subsequent year, had a math teacher that failed to return the survey. Figures in the off-diagonal indicate the proportion of students who had different types of teachers in consecutive years. For example, about 30 percent of students who, in a particular year, had math teachers that returned the survey then had, in the subsequent year, math teachers that did not return the survey. About one quarter of students switched teacher types in the opposite direction. Corresponding figures for science teachers are shown in Panel B.



Figures in Table 3 through 5 depict the variation in the dummy variable indicating whether or not a teacher returns the teacher-level survey. Although Table 3 shows more variation over time (within-student variation) than across students (between-student variation), the last column of Table 4 demonstrates that well over half of the students who encounter one type of teacher (i.e., those that return the survey or those that do not) always have the same type of teacher for all time periods in the data. Further, a relatively low percentage of students switch between the two types of teachers in consecutive years, as depicted in Table 5. Transition probabilities involving math teachers are in the upper panel, while transition probabilities involving science teachers are in the lower panel. These patterns suggest that student-fixed effect methods, which rely on within-student variation, may not be a highly efficient estimation technique. Moreover, limitations of the external validity of this method must be acknowledged as a slight majority of students never switch across the two types of teachers.

That being said, results based on student-fixed effects estimation represent conservative tests of a relationship between student item response rate and whether or not the student's teacher returns the teacher-level survey. The low within-student variation in the dummy variable indicating whether or not a teacher returns the teacher-level survey means inflates standard errors, decreasing study power.<sup>8</sup> The models invoking Anderson-Hsiao and Arellano-Bond estimation techniques would be even more conservative because the use of instrumental

---

<sup>8</sup> On the other hand, low within-student variation in the independent variable of interest might inflate coefficient estimates. It is possible that the increase in the magnitude of the coefficient estimates is sufficient to overcome their imprecision, leading to statistically significant results. To check this possibility, I run the analysis using pooled OLS while clustering standard errors at the student level. If the coefficients from the student fixed effects analysis are not noticeably larger in magnitude than the coefficient estimates from the pooled OLS analysis, then the low variation is not inflating coefficient estimates in the fixed-effect analysis. Indeed, I find this result. In fact, pooled OLS estimates are over twice the size of fixed effects estimates.

variables introduces statistical noise and controlling for lagged measures of the dependent variable results in the loss of additional periods in the panel, further lowering variation in the dummy variable indicating whether or not a teacher returns his or her survey. Thus, should a relationship between the two variables exist, one can be even more confident that a relationship between the two variables is material and not the result of statistical chance.

**Regression Results.** Table 6 displays regression results of the main analysis. Indeed, there exists a relationship between a student's item response rate on the student-level survey and whether his or her teacher returns the teacher-level survey. Column 1 demonstrates that students experience a drop in their item response rates by about 0.08 standard deviations in years where they have a math teacher who fails to complete the teacher survey. Likewise, column 2 shows that students experience a drop in their item response rates by about 0.09 standard deviations in years where they have a science teacher who fails to complete the teacher survey. Students with math and science teachers who both do not respond to surveys experience a drop of 0.13 standard deviations in their item response rates. In contrast, having teachers who fail to complete the teacher survey does not appear to be related to student test scores (Table 6, Columns 4 through 6).

Table 6

*Associations between Teacher Return, Student Item Response, and Student Test Scores*

	Dependent Variable					
	Student Item Response Rate			Student Test Scores		
	(1)	(2)	(3)	(4)	(5)	(6)
Math Teacher Does Not Return to Survey	-0.08** (0.03)			0.01 (0.01)		
Science Teacher Does not Return to Survey		-0.09** (0.03)			-0.00 (0.01)	
Both Teachers do not Return Survey			-0.13** (0.04)			-0.00 (0.01)
Student Observations	3,040	3,040	3,040	2,918	2,918	2,918

Notes: †p&lt;0.1, \*p&lt;0.05, \*\*p&lt;0.01.

Results based on the Anderson-Hsiao and Arellano-Bond estimators are shown in Table 7 and Table 8, respectively. The correlations between a student's item response rate and whether his or her math teacher returns the teacher-level survey that were found using the student-fixed effects models are not robust to the Anderson-Hsiao and Arellano-Bond specifications. In contrast, the correlation between student's item response rate and whether his or her science teacher returns the teacher-level survey remains statistically significant in the Anderson-Hsiao and Arellano-Bond specifications. In years when students have a science teacher that fails to return the teacher-level survey, their item response rate decreases by 0.08-0.09 standard deviations.

Table 7

*Anderson-Hsiao Estimates*

	Dependent Variable					
	Student Item Response Rate			Student Test Scores		
	(1)	(2)	(3)	(4)	(5)	(6)
Math Teacher Does Not Return Survey	-0.01 (0.03)			-0.01 (0.01)		
Science Teacher Does not Return Survey		-0.08** (0.03)			-0.00 (0.01)	
Both Teachers do not Return Survey			-0.07 (0.05)			-0.01 (0.01)
Student Observations	2,215	2,215	2,215	2,826	2,826	2,826

Notes: †p<0.1, \*p<0.05, \*\*p<0.01.

Table 8

*Arellano-Bond Estimates*

	Dependent Variable					
	Student Item Response Rate			Student Test Scores		
	(1)	(2)	(3)	(4)	(5)	(6)
Math Teacher Does Not Return Survey	-0.01 (0.03)			-0.01 (0.01)		
Science Teacher Does not Return Survey		-0.09** (0.03)			-0.00 (0.01)	
Both Teachers do not Return Survey			-0.09† (0.05)			-0.01 (0.01)
Student Observations	2,215	2,215	2,215	2,826	2,826	2,826

Notes: †p<0.1, \*p<0.05, \*\*p<0.01.

The prevalence of statistically insignificant results are not surprising given the additional loss of variation in the independent variable of interest due to the use of the Anderson-Hsiao and Arellano-Bond techniques. Empirical evidence of this claim is displayed in Tables 9 through 11, which are analogous to Tables 3 through 5 but are restricted to displaying summary statistics for periods in the panel that are used in computing Anderson-Hsiao estimators. As shown in the tables, within-student variation in the dummy variable indicating whether a teacher returns a survey is much lower. For example, whereas there was more variation over time (within-student variation) than across students (between-student variation) in the dummy variable in the student-fixed effects models, there is actually less variation over time within students than across students in the Anderson-Hsiao models (see Table 9). Overall variation is much lower for math teachers than for science teachers, explaining why results based on science but not math teachers are robust to the Anderson-Hsiao and Arellano-Bond specifications.

Table 9

*Variation in Teacher Return Rates (Anderson-Hsiao Models)*

	Standard Deviation	Percent of Overall Variance (%)
Math Teachers		
Overall	50.0	n/a
Within Students	30.8	38.1
Between Students	39.3	62.0
Science Teachers		
Overall	50.0	n/a
Within Students	32.7	44.3
Between Students	36.7	55.7

Table 10

*Variation in whether Teacher Returns Survey (Anderson-Hsiao Models)*

Category	Proportion of student-year observations that fell into the given category	Proportion of students who fell into the given category at least once throughout the panel	Proportion of students who never switch out of the given category
Math Teachers			
Teacher Returns Survey	48.0	68.3	70.4
Teacher Does not Return Survey	52.0	74.5	69.7
Science Teachers			
Teacher Returns Survey	40.7	64.4	63.2
Teacher Does not Return Survey	59.3	83.7	70.9

Note: All numbers are percentages.

Table 11

*Transition Probabilities for whether Teacher Returns Survey (Anderson-Hsiao Models)*

		Panel A: Math Teachers	
		<u>Status in Year t</u>	
		Math Teacher Returns Survey	Math Teacher Does not Return Survey
<u>Status in Year t-1</u>	Math Teacher Returns Survey	65.7	34.8
	Math Teacher Does not Return Survey	14.7	85.2
		Panel B: Science Teachers	
		<u>Status in Year t</u>	
		Science Teacher Returns Survey	Science Teacher Does not Return Survey
<u>Status in Year t-1</u>	Science Teacher Returns Survey	55.9	31.3
	Science Teacher Does not Return Survey	17.7	75.1

Returning to the remaining columns in Table 7 and Table 8, in years when a student's math and science teachers both fail to return the teacher-level survey, his or her item response rate decreases by about 0.09 standard deviations based on the Arellano-Bond estimator, though the result is only significant at  $\alpha = 0.1$ . There is no such relationship based upon the Anderson-Hsiao estimator. Finally, note that the relationship between test scores and whether a teacher returns a survey remains null. In general, results based upon these two estimation techniques lend additional confidence in the results of the student-fixed effects models.

### Discussion and Conclusion

The purpose of this analysis was twofold: (a) to determine if teachers have impacts on student noncognitive skills and (b) if so, to attempt to shed light onto what kinds of teachers have such impacts. I adopt the approach suggested by Hitt et al. (forthcoming) and use measures of survey effort as a behavioral measure of noncognitive skills related to conscientiousness. Consistent with other research, my results indicate that students who possess higher levels of

these noncognitive skills, as measured by item response rates, ultimately have higher levels of educational attainment and are more likely to be employed, even after controlling for cognitive ability (Almlund et al., 2011; Heckman et al., 2006).

More central to the original purposes of the analysis, the results suggest that students realize gains in these noncognitive skills when they are taught by teachers who also possess a greater degree of similar noncognitive skills. Presumably, teachers require some minimum level of noncognitive skills associated with conscientiousness in order to complete surveys and return them to the data collection agency via postal mail. Students exhibit lower item response rates in years when they have teachers who fail to return surveys and higher item higher response rates in years when they have teachers who do return them. At the same time, the teachers who return surveys and are having impacts on student noncognitive skills, as captured by item response rates, are not the same teachers that have impacts on student test scores. Such a result is consistent with the emerging research on how teachers contribute to cognitive and noncognitive student outcomes (Jackson, 2012; Jennings & DiPrete, 2012). Certain teachers are benefiting their students in ways that are (a) nontrivial for longer-run life outcomes and (b) not fully captured by test scores.

However, some study limitations are worth considering. First, results hinge on the assumption that teacher return rate captures traits related to conscientiousness. Unlike student item response rates, validating this proposition is not possible in the data. There are a variety of other reasons why some teachers complete surveys while others do not. Second, there were several factors that led to Type II errors. Low within-student variation in teacher conscientiousness, as measured by the binary proxy, decreases the precision of estimates and attenuates coefficient estimates in the empirical models. Likewise, about half of the students did



not experience variation in teacher conscientiousness, decreasing the generalizability of the results.

In the next chapter, I address several of these limitations using better data. With this other data set, I am able to present empirical evidence that teachers who do not return surveys appear worse than those who do on a variety of measures of teacher quality. Such findings lend more support to the claim that the indicator of whether a teacher completes a survey or not captures noncognitive skills related to conscientiousness. Moreover, the analysis in the next chapter does not rely on year-to-year variation within students to identify teacher effects, enabling me to use an estimation technique that is more efficient and generalizable to a broader population of students.

Furthermore, the analysis presented in this current chapter is unable to ascertain how, exactly, noncognitive skills are transmitted from students to teacher. Results are consistent with the theory that such skills are transmitted through role modeling. Teachers with certain proclivities that reflect a lack of conscientiousness may actively or passively transmit those proclivities to their students (Bandura, 1977; Berkowitz & Bier, 2004; Sherif & Sherif, 1964). Alternatively, teachers with particular noncognitive skills may utilize pedagogical approaches conducive to fostering those noncognitive skills. For instance, teachers with growth mindset may teach students in a way that also fosters growth mindset (Dweck, 2006). Such a theory is also consistent with the results. Ultimately, much more inquiry is needed to better understand what phenomena and mechanisms underlie these results.

It would also be useful to determine what other observable teacher characteristics, if any, predict a teacher's ability to raise student noncognitive skills. Unfortunately, LSAY data are too limited to investigate whether, for example, teacher credentials or experience is correlated with

teacher noncognitive skills and their impacts on student noncognitive skills. As a result, restraint should be exercised before drawing policy implications. For instance, teachers with a particular noncognitive skill set appear to have impacts on student noncognitive skills, which in turn, affect student educational attainment net of what is predicted by cognitive skills. One might suggest that schools ought to recruit such teachers. Yet without knowledge of what observable characteristics are correlated with having those noncognitive skills, it is not clear how such teachers can be readily identified. It is possible that observable characteristics, such as teacher credentials or years of experience, are not strongly correlated with a teacher's ability to improve student noncognitive skills, just as they are not strongly correlated with a teacher's ability to improve student cognitive skills (Goldhaber, 2008).

Likewise, one might suggest that appraisals of teacher quality should incorporate teacher impacts on student noncognitive skills. This proposal is not unreasonable given results from this study and other evidence that teachers have impacts on student outcomes not captured by test scores. Still, it is unclear how to effectively measure this dimension of teacher quality in practice, even though it may be desirable to do so. As discussed earlier, self-reported scales have serious sources of bias and survey item response rates, like other behavioral measures, can be misinterpreted and may lack external validity. Both types of measures can easily be corrupted in high-stakes settings as well (Duckworth & Yeager, 2015).

Nonetheless, broadening the understanding of teacher quality to include teacher noncognitive skills and teacher impacts on student noncognitive skills is worthy of more discussion. Likewise, schools play a pivotal role in the development of many children. They do more than deliver content knowledge and improve cognitive skills. Schools and other educational interventions communicate values and influence noncognitive skills. Overlooking

this facet of education may result in an incomplete picture of how educational institutions and interventions affect students. Noncognitive-skills is a field ripe for research and rife with unanswered questions. For instance, to what extent are noncognitive skills malleable? How can they be developed or how are they transmitted (Borghans et al., 2008; Heckman, 2000)? Complete appraisals of educational policies and institutions hinge on answers to many of these questions. This study draws some attention into this otherwise vastly understudied yet important topic.

## References

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. D. (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 1–181). Amsterdam: Elsevier.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Anderson T.W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, 76(375), 598-609.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277–297.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Berkowitz, M. W., & Bier, M. C. (2004). Research-Based Character Education. *The Annals of the American Academy of Political and Social Science*, 591(1), 72–85.
- Booker, K., Sass, T., Gill, B., & Zimmer, R. (2014). *Charter high schools' effect on long-term attainment and earnings*. Retrieved from the Mathematical Policy Research website: [http://www.mathematica-mpr.com/publications/pdfs/education/charter\\_long-term\\_wp.pdf](http://www.mathematica-mpr.com/publications/pdfs/education/charter_long-term_wp.pdf)
- Borghans, L., Meijers, H., & Bas, ter W. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46(1), 2–12.
- Blazar, D., & Kraft, M. A. (2015). Teacher and teaching effects on students' academic behaviors and mindsets (Working Paper No. 41). Cambridge, MA: Mathematica Policy Research.
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115.
- Cawley, J., Heckman, J., & Vytlačil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4), 419–422.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.
- Chingos, M. M., & Peterson, P. E. (2015). Experimentally Estimated Impacts of School Vouchers on College Enrollment and Degree Attainment. *Journal of Public Economics*, 122, 1–12.

- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778–820.
- Coleman, J. S., & Hoffer, T. (1987). *Public and Private High Schools*. New York, NY: Basic Books.
- Dobbie, W., & Fryer, R. G. (2015). The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy*, 123(5), 985–1037.
- Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, 4(6), 540–547.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251.
- Dweck, C. (2006). *The psychology of success*. New York, NY: Random House Inc.
- Goldhaber, D. D. (2007). Everyone's doing it but what does teacher testing tell us about teacher effectiveness. *Journal of Human Resources*, 42(4), 765–794.
- Goldhaber, D. D. (2008). Teachers matter, but effective teacher quality policies are elusive. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 146–165). New York, NY: Routledge.
- Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125–149.
- Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can Value-Added Measures of Teacher Performance Be Trusted? *Education Finance and Policy*, 10(1), 117–156.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466–479.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1052–1078). Amsterdam: Elsevier.
- Heckman, J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3–56.
- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052–2086.
- Heckman, J., & Rubenstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), 145–149.

- Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 372–406.
- Hitt, C.E., Trivitt, J.R., Cheng, A. (forthcoming). When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*.
- Jackson, K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.
- Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, 83(2), 135–159.
- Kane, T., Rockoff, J. E., & Staiger, D. (2008). What does certification tell us about teacher effectiveness. *Economics of Education Review*, 27(6), 615–631.
- Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics*, 64(3), 560-572.
- Koedel, C., and Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, 6(1), 18–42.
- Kraft, M.A., & Grace, S. (2016). Teaching for tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies (Working Paper). Brown University: Providence, RI.
- Krosnick, J.A., Narayan, S., & Smith, W.R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, 1996(70), 29–44.
- Miller, J.D. (2014). *Longitudinal Study of American Youth: User's manual*. Inter-university Consortium for Political and Social Research, University of Michigan: Ann Arbor, MI.
- Murnane, R. J., Willet, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4), 547–568.
- Podgursky, M. J. (2005). Teacher licensing in U.S. public schools: The case for simplicity and flexibility. *Peabody Journal of Education*, 80(3), 15–43.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.

- Robinson-Cimpian, J.P. (2014). Inaccurate estimation of disparities due to mischievous responders: Several suggestions to assess conclusions. *Educational Researcher*, 43(4), 171-185.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537–571.
- Sherif, M., & Sherif, C. (1964). *Reference groups: Exploration into conformity and deviation in adolescents*. Chicago, IL: Henry Regnery Company
- West, M., Matthew A. Kraft, Amy S. Finn, Rebecca Martin, Angela L. Duckworth, Christopher F.O. Gabrieli, & John D.E. Gabrieli. (2016). Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148-170.
- Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management*, 32(2), 246–270.
- Yeager, D.S., Henderson, M.D., Paunesku, D., Walton, G.M., D'mello, S., Spitzer, B.J., & Duckworth, A.L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology*, 107(4), 559-580.

### **Chapter 3: Teacher Effects on Student Conscientiousness: Experimental Evidence<sup>9</sup>**

There is no doubt in the literature that teachers play an important role in improving student performance. Multiple researchers have shown wide variation in teacher effectiveness towards improving student achievement on standardized tests of math and reading (Rivkin, Hanushek, & Kain 2005; Rockoff, 2004). Teachers who are effective at improving student achievement have also been found to have impacts on their students' long-run life outcomes such as educational attainment and employment income (Chetty et al, 2014).

However, teachers also affect other student outcomes besides test scores. A growing body of research documents the meaningful impacts that teachers have on their students' noncognitive skills, which refer to personality and character attributes such as diligence, self-control, and a propensity to engage in prosocial behaviors. Indeed, this is what I demonstrate in this analysis. Moreover, this research suggests that teacher impacts on student achievement are weakly correlated with teacher impacts on noncognitive skills. That is, teachers who are most effective at improving student test scores are not necessarily the most effective at improving student noncognitive skills and vice-versa (Backes & Hansen, 2015; Blazar & Kraft, 2015; Cheng, 2015; Gershenson, 2016; Jackson, 2012; Jennings & Diprete, 2010; Koedel, 2008; Kraft & Grace, 2016; Ruzek et al., 2014). It appears that teacher quality is multidimensional, comprising more than just the ability to improve student achievement.

It is, therefore, possible that some dimensions of teacher quality are left unmeasured if policymakers and practitioners solely rely on traditional value-added scores or other similar measures of teacher quality that are derived from student achievement on standardized tests (Grissom, Loeb, Doss, 2016). Some schools and researchers have responded by engaging in

---

<sup>9</sup> This chapter is co-authored with Gema Zamarro.



efforts to develop alternative measures of teacher quality in an attempt to capture other aspects of teacher effectiveness and effective teaching practice. Some evaluators have asked students to rate teachers on surveys (Ferguson, 2012) and many others have relied on evaluations of teachers through formal classroom observations (Danielson, 2007; Pianta & Hamre, 2009). Notably, measures of teacher quality based upon these alternative measures are only modestly correlated with value-added scores, suggesting that different measures of teacher quality capture distinct dimensions of teacher quality (Kane, McCaffrey, & Staiger, 2012).

Although it appears that teachers benefit their students in a variety of ways, it is not clear which kinds of teachers are most effective at improving particular student outcomes. Observable teacher characteristics are largely uncorrelated with teacher value-added scores (Buddin & Zamarro, 2009; Goldhaber, 2008; Hanushek & Rivkin, 2006; Jacob, 2007). Similarly, the research identifying teacher impacts on student noncognitive skills has generally not yet pinpointed the observable teacher characteristics that are associated with such impacts.

In this study, I address this research gap. I investigate whether teacher noncognitive skills may be correlated with teacher effectiveness, particularly in improving student noncognitive skills. Research has not thoroughly investigated this possibility because measures of noncognitive skills are seldom available. One exception is Rockoff et al., (2011), who found among a sample of novice teachers that noncognitive skills such as self-efficacy, extraversion and conscientiousness are marginally correlated with impacts on student achievement and teacher retention (see also Duckworth et al., 2009). I aim to improve upon this observational work in two ways. First, I use data from the Measuring of Effective Teaching (MET) Project where teachers were randomly assigned to classrooms of students. This feature allows me to account for the bias that arises from the systematic sorting of students to teachers (Rothstein,

2009) and then to estimate causal impacts of teachers on their students. Second, I rely upon novel performance-task measures of conscientiousness. Following Hitt, Trivitt, and Cheng (forthcoming) and Zamarro et al. (2016), I utilize a type of performance-task measure of conscientiousness that is based upon levels of engagement and effort that respondents exert in completing surveys. I henceforth refer to my measures as *survey-effort measures of conscientiousness*. These survey-effort measures of conscientiousness are readily available in my data and address common limitations of the self-reported measures of noncognitive skills that Rockoff et al. (2011) have used (e.g., social desirability bias, reference group bias, see Duckworth & Yeager [2015]). Furthermore, this approach enables me to create measures of student noncognitive skills, expanding the range of student outcomes for study.

I first show that my measures of teacher conscientiousness are correlated with some though not all traditional measures of teacher quality, including ratings based upon student surveys, formal classroom observation protocols, and subjective principal ratings. I then show that teachers with higher levels of conscientiousness are more effective at improving student conscientiousness but not student test scores. These are important findings as prior research demonstrates the independent role that conscientiousness plays in influencing educational attainment, job performance, employment, income, health, criminal behavior, and other indicators of well-being in life (Almlund et al., 2011; Dalal, 2005; Duckworth et al., 2007; Roberts et al., 2007).

Moreover, I find that not all traditional measures of teacher quality are correlated with impacts on student conscientiousness. Specifically, variation in value-added scores and ratings based upon some formal classroom observations does not explain variation in teacher impacts on student conscientiousness, but variation in other formal classroom observation protocols,

principal subjective ratings, and student ratings does.<sup>10</sup> I interpret these findings to suggest that teacher quality is multidimensional: Different teachers affect different student outcomes to varying degrees, and an assortment of teacher quality measures are needed to identify distinct effects on a set of relevant student outcomes. If policymakers and school leaders wish to improve the quality of the teacher workforce and student outcomes, they must think more clearly, holistically, and precisely about the multifaceted nature of teacher quality.

The remainder of the article is divided into four sections. I first review the relevant literature associated with teacher quality, emphasizing the ways in which it is currently measured and proposing that teacher noncognitive skills could be relevant to the matter. I also provide background information related to my novel survey-effort measures of teacher conscientiousness. Next I describe my data, how I construct the variety of measures of teacher quality as well as conscientiousness, and my empirical strategy. I then present the results in the third section and, in the final section, discuss their implications.

## **Literature Review**

### **Measuring Teacher Quality**

Teachers play a crucial role in improving student short-run outcomes such as achievement as well as longer-run outcomes such as educational attainment and employment income (Chetty et al., 2014; Koedel, 2008; Rivkin, Hanushek, & Kain 2005; Rockoff, 2004). Despite the importance of identifying, hiring, and retaining effective teachers, policymakers and school leaders are largely unable to consistently do so.

One reason for this difficulty is that many observable characteristics such as educational background and certification are generally uncorrelated with a teacher's ability to raise student

---

<sup>10</sup>Using the same dataset as I do here, Kraft and Grace (2016) find a similar result based upon self-reports of grit as a student outcome measure.

achievement on standardized tests (Buddin & Zamarro, 2009; Goldhaber, 2008; Hanushek & Rivkin, 2006; Jacob, 2007). School leaders cannot use such readily-available information to identify and to hire the most effective teachers. Although there is some evidence that teachers improve with more years of experience, school leaders cannot use this fact for hiring novice teachers. Moreover, the returns to experience appear to attenuate after three to five years (Clotfelter, Ladd, & Vigdor, 2006; Hanushek & Rivkin, 2006; but see Wiswall, 2013; Papay & Kraft, 2015).

Due to these limitations, other scholars have alternatively proposed to assess teacher quality based upon observing teacher effectiveness after they have started their career and then making personnel decisions given this new information (Podgursky, 2005; Kane et al., 2008). Typically, this includes estimating how much teachers improve their students' test scores, but it is not always possible to calculate these value-added measures with validity (Rothstein, 2009; Koedel & Betts, 2011). Different approaches for isolating causal effects of teachers upon student achievement are available, though how suitable each approach is depends on a variety of contextual details (Gaurino, Reckase, & Wooldridge, 2014; Zamarro et al., 2015). This is not to mention that value-added scores are sometimes impossible to calculate because students in many grades and subjects are not tested.

But even assuming that value-added scores are valid, approaches that assess teachers solely based upon their ability to improve test scores may fail to capture all important dimensions of teacher quality. Although raising test scores is important for improving long-run life prospects for students (Chetty et al., 2014), other factors play a large and independent role in influencing the future well-being of students. A growing body of research in economics and psychology demonstrates the importance of noncognitive skills in determining outcomes such as

educational attainment, employment, income, health, and criminal behavior, even after controlling for performance on tests of cognitive ability (Almlund et al., 2014; Heckman, Stixrud, & Urzua, 2006). Insofar as value-added scores and similar measures based upon student test performance do not capture teacher effects on student noncognitive skills, they will misstate the benefits that teachers provide to their students.

Alternatives for measuring teacher quality have been proposed, presumably because they capture aspects of effective teaching that value-added scores do not. Formal classroom observations and student surveys of teachers, for example, potentially provide finer-grained contextual details about a teacher's classroom environment and instructional practices that may bear upon student outcomes (Pianta & Hamre, 2009). Curiously, however, these alternative measures have mainly been validated based upon how strongly they are correlated with student performance on standardized tests (Garret & Steinberg, 2015; Kane et al., 2012). The underlying assumption is that teacher quality is unidimensional and only concerns a teacher's ability to improve student achievement. That is, measures derived from classroom observations and student ratings are useful insofar as they more fully capture a teacher's ability to improve test scores than value-added measures alone can. The practical implication, then, is to combine alternative measures with value-added scores to form a more valid and more reliable composite measure of teacher quality since each measure captures independent information about a teacher's ability to raise test scores (Kane et al., 2012; Mihaly et al, 2013).<sup>11</sup>

---

<sup>11</sup> As it turns out, student and formal classroom observation ratings are only modestly correlated with student achievement on standardized tests (Kane et al., 2012). In a separate study using data from the MET Project, Garrett and Steinberg (2015) find that teacher ratings based on classroom observations that used Danielson's (2007) Framework for Teaching are positively correlated with student test scores but much of this is due to the systematic sorting of higher-achieving students to teachers who have higher classroom observation ratings.

However, this approach could be misguided if teachers affect their students in meaningful and measurable ways that are not captured by test scores. Indeed, this is what the literature of teacher impacts on student noncognitive skills suggests. Emerging research shows that teachers who have large effects on test scores do not necessarily have large effects on noncognitive skills that, in turn, contribute to the future well-being of students. Similarly, teachers who have large effects on student noncognitive skills do not necessarily have equally sizable effects on student test scores (Blazar & Kraft, 2015; Cheng, 2015; Gershenson, 2016; Jackson, 2012; Jennings & Diprete, 2010; Koedel, 2008; Ruzek et al., 2014). Failing to consider a variety of student outcomes may lead to misclassifications of teacher effectiveness as teachers may benefit students on outcomes that are unconsidered or unobserved. Of particular relevance to my work is Kraft and Grace's (2016) analysis of the MET data; they find weak relationships between teacher effects on test scores and teacher effects on noncognitive skills. In short, there are reasons to doubt the assumption that different measures of teacher quality collectively capture a unidimensional factor of teacher effectiveness. There is not only variation between teachers in the ability to improve a specific student outcome but also variation within a teacher in his or her ability to improve a variety of student outcomes.

In related work, several scholars have found that subjective ratings of teachers given by principals are only moderately correlated with teacher value-added scores. Although these ratings are most strongly correlated with value-added scores among the least and the most effective teachers, they are unable to differentiate teachers within the middle of the distribution of value-added scores (Harris, & Sass, 2014; Jacob & Lefgren, 2008; Rockoff et al., 2012). Notably, Harris, Ingle, and Rutledge (2014) find that principals base their ratings not only upon a

teacher's ability to improve test scores but also upon teacher noncognitive skills, particularly the amount of effort they exert in their everyday work.

### **The Role of Teacher Noncognitive Skills**

Harris et al.'s (2014) finding that principals form judgments based upon teacher noncognitive skills suggests that teacher noncognitive skills may be a key component of teacher effectiveness. Indeed, research from labor economics and psychology demonstrates an association between worker productivity and noncognitive skills. More conscientious workers, for example, exhibit better job productivity (Dalal, 2005; Heckman et al., 2006; Roberts et al., 2007). Different combinations of noncognitive skills may be required to improve worker productivity across different types of occupations (Borghans, ter Weel, & Weinberg, 2008). Although it is possible that more conscientious teachers are more highly-valued by principals, as Harris et al. (2014) suggest, it remains unclear how teacher noncognitive skills relate to other measures of teacher quality and how they directly affect students.

There are only a few instances when scholars collected measures of teacher noncognitive skills and studied their relationship to educational outcomes. For example, Duckworth et al. (2009), using self-reported measures of teacher noncognitive skills, have found that teacher self-reports of grit and life satisfaction are predictive of student test scores among Teach for America teachers (Duckworth et al., 2009). In a sample of novice elementary and middle school math teachers, Rockoff et al. (2011) demonstrate that self-reports of conscientiousness, extraversion, and self-efficacy are correlated with subjective ratings that are given by their mentor teachers but are only weakly correlated with student achievement and teacher job retention.

While Duckworth et al. (2009) and Rockoff et al. (2011) focus on how teacher noncognitive skills affect student achievement, other work has focused on their impacts on student noncognitive skills. For example, Blazar and Kraft (2015) find that fourth- and fifth-

grade students exhibit more self-efficacy when they have teachers who are more adept at lending emotional support to students in their interactions and through fostering a safe, positive classroom environment. Presumably, teachers who are more effective at improving student self-efficacy engage in certain classroom practices and processes that are conducive to realizing these outcomes (Bargagliotti, Gottfried, & Guarino, 2016; Pianta & Hamre, 2009). On the other hand, some research indicates that variation in pedagogical practices and teacher observable characteristics fails to explain variation in noncognitive skill outcomes (Jennings & Diprete, 2010).

In the previous chapter of this dissertation, I use longitudinal data to show that students receive increases in particular noncognitive skills in years when they have teachers who have higher levels of those same skills. I drew upon social learning theory to posit that students may learn noncognitive skills through observing role models such as teachers (Bandura, 1977; Bandura & Walters, 1963). This theory explains why teachers with a particular set of noncognitive skills may be more effective at improving the same set of noncognitive skills among their students. At the very least, students appear sensitive to and influenced by teacher behaviors. In fact, Blazar and Kraft (2015) posit that teachers who provide more social and emotional support improve student self-efficacy and happiness not because of a particular pedagogical approach but because of the behaviors that they model in providing such support. I extend this line of research by further studying the relationship between teacher noncognitive skills, specifically conscientiousness, and other measures of teacher quality and their role in improving student cognitive and noncognitive skills. Furthermore, instead of relying upon self-reported measures of noncognitive skills, I rely upon newly developed survey-effort measures, which I describe next.



## Measuring of Noncognitive Skills

The aim of developing and exploiting innovative measures of conscientiousness based upon survey effort is to capture levels of conscientiousness among teachers and students with more validity. Noncognitive skills data can be collected via performance-task measures. These types of measures begin by asking individuals to complete a task; researchers then observe variation in the individuals' behaviors as they complete it. My novel survey-effort measures of conscientiousness are a type of performance-task measure and are constructed by observing how much effort teachers exert towards responding to a survey. Completing such clerical tasks requires careful attention to detail and persistence to avoid skipping or providing thoughtless, inaccurate answers (Hitt et al., forthcoming; Jackson et al., 2010). In other words, I view completing the survey itself as a task that requires conscientiousness and use three approaches to parameterize survey effort and to create measures of teacher conscientiousness: (a) item response rate, (b) careless answering patterns, and (c) survey omission.

**Item nonresponse rate.** Respondents sometimes demonstrate low effort by altogether skipping items or thoughtlessly providing answers of "I don't know." Item nonresponse rate is parameterized as the proportion of questions on a survey that an individual neglects to answer out of the number of questions he was supposed to answer. Hitt et al. (forthcoming) validate item nonresponse rate as a performance-task measure of noncognitive skills related to conscientiousness. In six nationally-representative, longitudinal datasets of US secondary school students, item nonresponse rates are found to be predictive of educational attainment, which in turn influence labor market outcomes, in adulthood (see also Chapter 2 of this dissertation). In my data, I use item nonresponse rate among teachers as a measure of conscientiousness, provide validation that it captures meaningful teacher attributes, and explore whether it captures a

meaningful dimension of teacher quality by investigating whether it is a determinant of student outcomes.

**Careless answering patterns.** When asked to complete surveys, some individuals begin the survey and do not skip items but still exert low effort by hastily providing thoughtless and random answers. This behavior results in careless answering patterns, which is a behavior that can be detected and parameterized (Hitt, 2015; Meade & Craig, 2012). This measure has been validated as a proxy for conscientiousness in two nationally-representative samples of US adolescent schoolchildren and a nationally-representative sample of US adults as it is positively correlated with educational attainment, employment income, a greater likelihood of being employed in a high-skilled job, and self-reported measures of conscientiousness, even after controlling for cognitive ability (Hitt, 2015; Zamarro et al., 2016). I describe the construction of this measure in greater detail in the methods section below.

*Survey omission.* Rather than skipping items or responding thoughtlessly, some individuals exhibit low survey effort by entirely ignoring a survey even after they are asked to complete it. Although there may be numerous reasons for why teachers, in particular, ignore a survey despite volunteering to participate in the study that requires its completion, there is no reason not to rule out survey omission as a manifestation of low survey effort. In other words, while some teachers exert low effort by beginning the survey but skipping items or providing inaccurate answers, others do not even begin the survey. In the previous chapter, I use a binary variable of whether a teacher completes or does not complete a survey as a measure of teacher conscientiousness, but I was not able to provide empirical validation for it because no data were available to conduct a validation test.

One study, however, shows that among new teachers who were invited to fill out a survey, those who did were rated as higher-quality teachers by their mentors than those who did not respond. This result suggests that, on balance, teachers who overlook surveys which they have been invited to complete may also tend to be of lower quality (Rockoff et al., 2011). My data enable me to provide additional validation of survey omission as a meaningful measure of teacher quality. I further explore whether refraining from completing a survey that one volunteered to do through prior agreement is systematically related to other measures of teacher quality and predictive of student outcomes.

**Why use performance-task measures?** It is much more commonplace to rely upon self-reported measures of noncognitive skills. Researchers usually execute this approach by administering surveys and using responses to a series of Likert-type items. Yet there are some drawbacks; I highlight two. First, self-reported measures are relatively convenient to collect, but they are prone to social desirability and reference group bias (Duckworth & Yeager, 2015; Dobbie & Fryer, 2015; Paulhus 1991; West et al., 2016). In contrast, my survey-effort measures do not face the same threats to validity, especially because respondents typically do not know they are being observed on how much effort they exert to complete a survey. Their behavior while completing the survey then reveals something about their noncognitive skills without being colored by sources of bias endemic to self-reported measures.<sup>12</sup> The second limitation of self-reported measures is more practical in nature. Self-reported measures of noncognitive skills are

---

<sup>12</sup> This critique is not intended to call all survey research into question. Self-reported measures of noncognitive skills have been validated in a variety of circumstances and scholars have gleaned much knowledge from this research approach. That being said, there are incidences where sources of bias endemic to self-reported measures have possibly distorted results (Dobbie & Fryer, 2015; West et al., 2016). All approaches to measurement have unique strengths and weaknesses.

rarely collected for teachers. This is why teacher quality research that uses large scale data sets rarely examines the topic of noncognitive skills. My data, which come from the MET study, is no exception. However, my survey-effort measures of teacher conscientiousness can be readily constructed from any dataset that has administered surveys to teachers. Latent information about a respondent's noncognitive skills can be recovered using my approach within any data set, opening new avenues to research in this understudied area.

### **Research Questions**

In summary, research has found that educational background, years of experience, and other observable teacher characteristics are weak predictors of teacher quality as measured by impacts on student achievement. There is likewise little understanding of how other observable teacher characteristics predict teacher impacts on student noncognitive skills. After all, researchers have only recently begun to study teacher impacts on student noncognitive skills, finding that the teachers who are effective at improving them do not necessarily comprise the same group of teachers who are effective at improving student achievement.

This literature, however, has not extensively studied the possibility that teacher noncognitive skills explain variation in teacher effectiveness. Teacher noncognitive skills could be related to existing measures of teacher quality and also play an important role in improving both cognitive and noncognitive student outcomes. I address these issues in my study and contribute to the understanding of teacher quality by answering two specific research questions about teacher conscientiousness.

First, how are my survey-effort measures of teacher conscientiousness correlated with existing measures of teacher quality? Answering this question lends some validity to my measures by showing they are meaningful and systematically correlated to established measures of teacher quality. It also provides insight into what, exactly, current measures of teacher quality

are actually capturing. Quite possibly, classroom observation protocols, student ratings, and value-added scores may not only provide an accurate account of the practices that teachers use for teaching and relating to students but also measure personality traits that are also crucial for student outcomes. In other words, I examine the extent, if any, that teacher conscientiousness is captured by value-added scores and ratings based upon classroom observations, subjective principal opinions, or student surveys.

Second, how are my survey-effort measures of teacher conscientiousness correlated with student cognitive and noncognitive outcomes? I pay particular attention to whether more conscientious teachers are more effective at increasing student achievement and student conscientiousness based upon self-reported and survey-effort measures. If so, I may have uncovered an observable teacher characteristic that is linked to teacher effectiveness — a result that has largely eluded researchers in teacher quality. As a point of comparison, I also examine the extent to which higher-performing teachers, as judged by traditional measures of teacher quality (e.g., student ratings, principal subjective ratings, formal classroom observations, and value-added scores) affect student conscientiousness as captured by my survey-effort measures. Kraft and Grace (2016) have already shown that student ratings of their teachers are positively correlated with teacher impacts on grit, while traditional value-added scores and ratings based upon formal classroom observations are uncorrelated with them. Other prior research using MET data has already established the predictive power of teacher value-added scores and, to a much lesser degree, student ratings and formal classroom observations to forecast student achievement outcomes (Kane et al., 2012; Mihaly et al., 2013). Whether traditional measures of teacher quality also forecast survey-effort measures of student conscientiousness is an unanswered question which I address.

Answers to these two questions provide a more refined picture of teacher quality by describing how teacher conscientiousness is linked to a variety of student outcomes. I describe my analytical methods next.

## Methods

### Data

Data for this study come from the MET Project. Six large, urban public school districts participated in the MET project, which lasted two school years from 2009-2011. The districts involved are New York City Department of Education, Charlotte-Mecklenburg Schools, Denver Public Schools, Memphis City Schools, Dallas Independent School District, and Hillsborough County Public Schools. In the second year of the study over 1,500 teachers from nearly 300 schools were randomly assigned within schools and grades to classrooms of students ranging from fourth to ninth grade (White & Rowan, 2012). I leverage this random assignment to estimate the causal effect of teachers on a variety of student outcomes observed at the end of the second year.

My measures of teacher quality are constructed based on data from the first year of the MET study.<sup>13</sup> In other words, I assume that teacher quality is captured with validity in the first

---

<sup>13</sup> Whenever possible, I prefer to build measures of teacher quality based upon data from the first year of the MET study. I refrain from using data from the second year of the MET study as it may confound the causal direction of the relationships between measures of teacher quality and student outcomes – the latter which are measured in the second year. That is, a teacher's randomly assigned classroom may influence the teacher's behavior and, ultimately, measures of teacher quality. In such a case it is unclear to what extent the teacher is affecting student outcomes or vice-versa. The only exception is that I use data from both years of the MET study to create teacher value-added scores since research has documented that value-added measures based on one year of data are unstable (McCaffrey et al., 2009). In contrast, other work suggests that other measures based upon classroom observations or student ratings possess greater within-year stability (Polikoff, 2015; Pianta et al., 2008). It is also worth mentioning that MET teachers did not receive their teacher quality ratings from the first year of the study so that their behaviors in the second year of the study are uninfluenced by any such feedback (Polikoff, 2015).

year of the MET study, and then I estimate how teachers of different quality affect a variety of student outcomes measured in the second year. This strategy of predicting how teachers will affect student performance in a given year based upon performance in other years with a different group of students has been used in prior research (Chetty et al., 2014; Jackson, 2012). Note too that I assume that teacher noncognitive skills are stable traits, particularly for adults – a claim that possesses some empirical evidence (Cobb-Clark & Schurer, 2012).

### **Survey-Effort Measures of Teacher Conscientiousness**

I begin by describing how I construct my survey-effort measures of teachers' conscientiousness: item nonresponse rate, careless answering patterns, and survey omission.

**Item nonresponse rate.** To compute item nonresponse rates for teachers, I use the Teacher Working Conditions Survey that teachers participating in the MET study completed during the first year of the study. Teachers were asked to answer 144 items on this survey, and new teachers were asked to answer an additional 39 items. Items on the questionnaire were designed to capture unique constructs such as time use, school leadership, the management of student conduct, and other aspects of the teacher's school and professional life. All surveys were administered through a confidential online system (Rowan & White 2012). Dividing the total number of questions that teachers did not answer by the total number of questions that they were supposed to answer yields my first survey-effort measure of teacher conscientiousness, item nonresponse rate. On average, teachers skipped or responded "Don't know" to 9 percent of the items.<sup>14</sup>

---

<sup>14</sup>Not counting responses of "Don't know" as an instance of nonresponse does not substantively change the results, a pattern consistent with other research validating the use of item nonresponse (see Hitt et al., forthcoming).

**Careless answering patterns.** My second proxy of teacher conscientiousness identifies inconsistent answering patterns to create a measure of survey effort. Methods for building this measure can be found in Hitt (2015) and Zamarro et al., (2016), which is a generalization of methods found in Huang et al. (2012), Johnson (2005), and Meade & Craig (2012). I provide a sketch of the method below. In the present study, teacher careless answering is derived from the Teacher Working Conditions Survey.

We use response data from eight scales on the survey to run a series of bivariate regressions where the dependent variable is the response to an individual item in a given scale and the independent variable is an average of the responses to the remaining items in the scale. The residual in this regression captures the deviations between a teacher's actual response to an item and his expected response to the item based upon his as well as the sample's responses to the other items on the same scale. I obtain residuals from regressions for all items from each of the eight scales that I use. Then, I average the absolute values of the residuals within each scale and standardize each average to have a mean equal to 0 and standard deviation equal to 1. These values capture levels of careless answering within each of the eight scales. Higher values indicate more careless answering as larger residuals, in terms of absolute value, indicate greater deviations from expected responses. In other words, higher values on this measure indicate lower levels of conscientiousness. Finally, I compute an overall level of careless answering by averaging levels of careless answering from each of the eight scales and, once again, standardizing these values to have a mean equal to 0 and standard deviation equal to 1. Importantly, Cronbach's alpha for the scales I use range from 0.81 to 0.94. These figures suggest a high level of internal consistency within each scale and lend credence to my assumption that



deviations in expected responses to an item are attributable to a lack of survey effort and not random measurement error in the scale.

As it turns out, the measures based upon careless answering patterns and those based upon item nonresponse are uncorrelated ( $\rho = -0.03$ ). This low correlation does not necessarily imply that my measures are capturing distinct latent traits. It could simply indicate that different individuals exhibit low survey effort in distinct ways – either by skipping questions or by hastily providing thoughtless answers (Zamarro et al, 2016). I return to a discussion about this possibility in the final section of this article.

**Survey omission.** I also use the Teacher Working Conditions Survey to build my final survey-effort measure of conscientiousness: an indicator of survey omission. This variable takes on a value equal to 1 if a teacher who volunteered to be in the MET study never submitted nor even started the Teacher Working Conditions Survey as asked. This variable equals 0 for teachers who responded to the survey as asked. About 27 percent of teachers in my sample failed to respond to the Teacher Working Conditions Survey. Because teachers who did not begin the survey do not provide responses from which I construct item nonresponse and careless answering measures, I cannot report correlations between my measure of survey omission and other two measures of teacher conscientiousness.

### **Traditional Measures of Teacher Quality**

Having described the derivations of my survey-effort measures of teacher conscientiousness, I now describe the derivations of traditional measures of teacher quality, which include value-added scores, scores based upon formal classroom observation rubrics, student ratings of their teachers, and subjective ratings made by principals.

**Value-Added Scores.** I use two years of student test scores based upon state assessments to compute value-added scores for each teacher. Test scores are standardized by district, grade,

subject, and year. Value-added scores are computed by estimating models that include teacher fixed effects and then implementing an empirical Bayes adjustment to mitigate measurement error due to variation in the number of student observations that are available for computing each teacher's effect.

In particular, I estimate

$$Y_{ijt} = \alpha Y_{i(t-1)} + \mathbf{X}_{it}\boldsymbol{\beta} + \delta \bar{Y}_{j(t-1)} + \bar{\mathbf{X}}_{jt}\boldsymbol{\gamma} + \boldsymbol{\theta}_i + \epsilon_{it}. \quad (1)$$

In equation (1)  $Y_{it}$  is the test score for student  $i$  in classroom  $j$  during year  $t$ , while  $Y_{i(t-1)}$  is the test score for student  $i$  in the prior year.  $\mathbf{X}_{it}$  is a vector of demographic characteristics for student  $i$  including age and indicators for gender, race, free and reduced-priced lunch status, English language learner status, gifted status, and special education status.  $\bar{Y}_{j(t-1)}$  and  $\bar{\mathbf{X}}_{jt}$  represent measures of prior-year test scores and student demographic characteristics, respectively, averaged across all students in classroom  $j$ .  $\boldsymbol{\theta}_j$  is a vector of teacher fixed effects and  $\epsilon_{ijt}$  is the error term. Value-added scores for each teacher are computed by taking estimates of  $\boldsymbol{\theta}_j$  and, following Tate (2004) applying an empirical Bayes adjustment. Let  $\hat{\mu}$  and  $\hat{\sigma}^2$  be the mean and variance, respectively, of the estimated distribution of teacher value-added scores across the sample and let  $\hat{\sigma}_j^2$  be the estimated variance of the estimated value-added score for teacher  $j$ . Bayes-adjusted teacher value-added scores,  $\theta_j^{EB}$  are thus given by:

$$\theta_j^{EB} = \theta_j \lambda_j + \hat{\mu}(1 - \lambda_j), \text{ where } \lambda = \hat{\sigma}^2 / (\hat{\sigma}^2 + \hat{\sigma}_j^2).^{15}$$

---

<sup>15</sup>After computing my value-added scores, I discovered that the data provided to me by the MET study also contained pre-constructed value-added scores. Correlations between my value-added scores and the pre-constructed value-added scores are 0.90 for math and 0.88. Replicating my analysis using the pre-constructed value-added scores rather than my data does not substantively change the results.

**Formal classroom observations.** Researchers in the MET Project video-recorded multiple lessons for each classroom section that a participating teacher taught during both years of the study. These videos were then shown to evaluators trained in the use of one of two classroom observation rubrics. Some evaluators were trained to rate lessons based on the Classroom Assessment Scoring System (CLASS) developed by Pianta, La Paro, and Hamre (2008). The CLASS instrument is designed to capture the extent to which teachers support student learning and emotional growth through fostering a safe and positive classroom climate, managing classroom time and student behavior, engaging students, and using effective pedagogy. Other evaluators were trained to rate lessons based upon Danielson's (1996) Framework for Teaching rubric, hereafter FFT. This rubric is similar to CLASS as it is also designed to capture the extent to which teachers cultivate a classroom environment that is conducive to learning and whether they use effective instructional techniques that promote student learning. All videos were rated by these evaluators, and composite CLASS and FFT scores were created by averaging scores on the various components of each respective rubric. A teacher's overall classroom observation score, whether it is based upon the CLASS or the FFT instrument, is constructed by averaging his composite scores across all of his raters.

**Student perceptions.** Students of teachers participating in the MET Project were annually administered the Tripod survey developed by Ferguson (2012). Based upon the student's responses, the Tripod survey captures seven dimensions of effective teaching. For example, the dimension named Care captures the extent to which teachers foster a sense of safety, belonging, and support in the classroom for their students. The extent to which teachers push students to work hard, exert greater effort to learn, and to think critically or deeply about a topic is captured by the dimension named Challenge. Other dimensions capture other

instructional practices that support student engagement and learning. A teacher's overall Tripod score is created by averaging responses at the individual-student level and then averaging these scores again at the classroom level.

**Principal Subjective Ratings.** As part of the original MET study, principals in participating schools were asked to rate up to twelve teachers who were also a part of the MET study. Principals rated these teachers on a six-point ordinal scale, which included the following categories: Exceptional, Very Good, Good, Fair, Poor, and Very Poor. In my model specifications, I dichotomize principal subjective ratings. The variable takes on a value equal to 1 for teachers receiving any of the highest three ratings in one category, while the variable takes on a value equal to 0 for teachers receiving any of the lowest three ratings as another. This categorization approximately divided the teachers into equal halves along the distribution of principal ratings. I elected to dichotomize this variable to facilitate the interpretation of my results as well as my estimation techniques, which I now describe.

### **Empirical Strategy**

Using the aforementioned survey-effort and traditional measures of teacher quality, I conduct a series of analyses to answer my two research questions. Recall that I first ask how my survey-effort measures of teacher conscientiousness are correlated with traditional measures of teacher quality. I then, pertaining to my second research question, examine whether my survey-effort measures are correlated with student cognitive and noncognitive outcomes. That is, do more conscientious teachers affect students in different ways than less conscientious teachers?

**Relationships between measures of teacher quality.** To address my first research question, I investigate whether my three survey-effort measures of teacher conscientiousness are correlated with other measures of teacher quality. Although prior work has validated my survey effort measures as proxies for conscientiousness among adolescents and adults, including

teachers (see Hitt, 2015; Hitt et al, forthcoming; Rockoff et al., 2011; Zamarro et al., 2016), this analysis provides additional evidence of whether my behavioral measures are merely random noise or actually meaningful signals of teacher quality. If my survey-effort measures are indeed meaningful signals, this analysis provides a sense of what, exactly, my measures capture with respect to other widely-used measures of teacher quality.

In this analysis, I run a series of bivariate regressions where the dependent variable is a traditional measure of teacher quality (e.g., value-added score, classroom observation score, student ratings, principal ratings) and the independent variable is one of my survey-effort measures of teacher conscientiousness (e.g., item nonresponse rate, survey completion, and careless answering patterns). I express all variables, except my dichotomous indicators of survey completion and principal ratings, in terms of standard deviations for ease of interpretation.<sup>16</sup>

**Teacher impacts on student outcomes.** For my second analysis, I examine whether each survey-effort measure of teacher noncognitive skills is predictive of student outcomes. Specifically, I consider teacher impacts on a variety of cognitive and noncognitive outcomes: (a) test scores in math and reading, (b) a student self-reported measure of grit, (c) a student self-reported measure of effort, (d) student item nonresponse, and (e) student careless answering patterns on a survey. I discuss each of these measures in turn.

Test scores in math and reading are based upon state-mandated assessments administered during the second year of the MET study. All test scores are standardized by district, year, and grade to have a mean and standard deviation equal to 0 and 1, respectively.

---

<sup>16</sup> One could conduct an identical analysis by computing correlation coefficients, but I opt for using bivariate regressions so that standard errors and confidence intervals are also provided.

During each school year, students also completed the Student Perceptions Survey administered by MET researchers. This survey included the Tripod instrument as well as items soliciting basic demographic information. In the second year of the MET study, the Student Perceptions Survey included the Duckworth and Quinn (2009) Grit Scale and several items designed to measure the amount of effort that the student exerts in class.

Duckworth and Quinn (2009) define grit as “perseverance and passion for long-term goals” and have found it to be positively correlated with academic outcomes such as retention and grade point average in postsecondary students (p. 172). For my data, I create scale scores by averaging each student’s responses to the eight Likert-type items on the Grit Scale, reverse coding items when necessary. Scale scores range from 1 to 5, with higher values indicating higher levels of grit, and have a mean of 3.56 with a standard deviation of 0.67.

Items for the self-reported measure of student effort are shown in the appendix to this chapter. The scale consists of five and four Likert-type statements for secondary- and primary-school students, respectively, to answer. I reverse-code items as necessary and then average the responses. Higher values indicate that the student exerts more effort in class. Scale scores range from 1 to 5 with an average of 4.07 and a standard deviation of 0.67.

The remaining measures of student noncognitive skills are two survey-effort measures of conscientiousness – item nonresponse and careless answering.<sup>17</sup> These measures are based upon student effort on the Student Perceptions Survey administered in the second year of the MET study. Elementary school students were asked to complete 75 items, while secondary school students were asked to complete 83 items. The average item nonresponse rate was 2.7 percent

---

<sup>17</sup>There is no measure of survey omission for students as they were compelled to complete the survey once consenting to participate. Teachers, in contrast, consented to the study but could freely decide whether or not to comply with the study by completing their respective survey.

with a standard deviation of 9.0 percent. Values for the careless answering measure are, by construction, standardized to have a mean equal to 0 and standard deviation equal to 1. I build the measure of student careless answering based upon six of the seven subscales in the Tripod Survey administered in the second year of the MET study. I use the Care, Control, Clarify, Challenge, Captivate, and Confer subscales.<sup>18</sup> These scales have Cronbach's alphas that range from 0.68 to 0.85 for secondary school students and 0.63 to 0.84 for elementary school students. The Tripod survey has a seventh subscale but I opted to omit it due to its apparent lack of reliability as indicated by a low Cronbach's alpha value.

I am unable to examine whether my survey-effort measures of student conscientiousness are predictive of later-life outcomes in the MET data as such data are unavailable. Instead, I rely on other research that has validated these measures. Lower item nonresponse rates and lower levels of careless answering as measured in adolescence are associated with greater levels of educational attainment and a greater likelihood of employment when measured in adulthood, even after controlling for cognitive ability as measured by standardized test scores (Cheng, 2015; Hitt et al., forthcoming; Hitt, 2015). Moreover, other studies of schoolchildren demonstrate that conscientiousness is associated with academic and labor-market success (Duckworth et al., 2007;

---

<sup>18</sup> Ferguson's (2012) Tripod instrument, which is a measure of teacher quality, is also included on the Student Perceptions Survey from which I derive several student outcome variables such as self-reported grit, self-reported effort, item nonresponse, and careless answering. However, it is important to reemphasize that measures of teacher quality are all based upon surveys administered in the first year of the study, while student outcome measures come from surveys administered in the second year of the study. Thus, student item nonresponse or careless answering on the Tripod survey during the second year of the MET study does not distort measures of teacher quality, which are based upon the Tripod survey administered in the first year of the MET study. Moreover, students only enter the dataset if they have a teacher participating in MET, and because students in the data typically do not have the same teacher for two consecutive years, my measures are not affected by same-source bias. That is, responses on the Tripod instrument used to construct measures of teacher quality are not provided by the same students who provide responses from which I build my student outcome measures.

MacCann, Duckworth & Roberts 2009; Poropat, 2009; Trautwein et al. 2006; Tsukayama, Duckworth & Kim 2013; Kraft & Grace, 2016; Roberts et al., 2007).

Table 1 displays the correlations between each student outcome measure. One can easily observe that measures of student noncognitive skills are, if anything, modestly correlated with student test scores. The magnitudes of the correlation coefficients never surpass 0.25. Interestingly, the two self-reported measures of student noncognitive skills (i.e., grit and effort) are more strongly correlated with each other than they are to test scores or survey-effort measures of conscientiousness. Meanwhile, item nonresponse appears uncorrelated with all other measures. In fact, as is the case among teachers, there is essentially no association between item nonresponse and careless answering among students. On the other hand, careless answering seems to be equally correlated with test scores and self-reported measures of noncognitive skills, albeit modestly.<sup>19</sup>

Table 1

*Student Outcomes Correlation Matrix*

	Math Test Scores	English Test Scores	Self-Reported Grit	Self-Reported Effort	Item Nonresponse
Math Test Scores	-				
English Test Scores	0.72	-			
Self-Reported Grit	0.14	0.09	-		
Self-Reported Effort	0.22	0.23	0.44	-	
Item Nonresponse	0.08	0.09	0.04	0.04	-
Careless Answering	-0.25	-0.25	-0.21	-0.23	-0.07

<sup>19</sup> It is interesting that correlations between self-reported measures of noncognitive skills are stronger than correlations between self-reported and performance-task measures. This is a pattern which other work has also found (Zamarro et al., 2016). The higher correlations among self-reported measures could be driven by the common mode in which the measures are collected, namely, through self-reports. Whether this is evidence of bias that is common across self-reported measures is unclear.



We leverage the random assignment of teachers to classrooms in the second year of the MET Project to estimate causal impacts of teachers, who vary in my survey-effort measures of conscientiousness, on each of the student outcomes. I compute both intent-to-treat (ITT) estimates and local average treatment effect (LATE) estimates using an instrumental variables (IV) strategy.

For the ITT analysis, I use ordinary least squares to estimate models of the form

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 \mathbf{X}_i + \phi_i + \epsilon_i \quad (2)$$

where  $Y_i$  is one of the outcomes of interest for student  $i$  measured in the second year of the MET study,  $\mathbf{X}_i$  is a vector of student demographic characteristics as in equation (1) but now also includes prior-year test scores in math and English, and  $\epsilon_i$  is the usual error term but clustered at the classroom level.  $T_i$  is the independent variable of interest and represents one of my survey-effort measures of conscientiousness for the teacher to whom student  $i$  was randomly assigned. Again, these measures of teacher characteristics are obtained based upon data in the prior (first) year of the MET study. The associated coefficient,  $\beta_1$ , captures the magnitude of the effect that teachers with varying levels of conscientiousness have on a particular student outcome. To better capture the experimental design and randomization process, I include randomization-block fixed effects,  $\phi_i$ . In the MET study, teachers teaching in the same school, grade, and subject were first placed into blocks and then randomized to classrooms (Rowan & White, 2012). Note, then, that the randomization block fixed effect also controls for unobserved variation within schools, grade-levels, and the subject area taught by the teacher.

I employ a two-stage least squares approach to compute my IV estimates. In this framework, I use survey-effort measures of teacher conscientiousness for each student's randomly assigned teacher as an instrument for the same measure of teacher conscientiousness

for the student's actual teacher. Again, teacher characteristics are measured based on data from the first year of the MET study. In particular, I estimate the following two-stage model:

$$T_i^A = \gamma_0 + \gamma_1 T_i^R + \gamma_2 \mathbf{X}_i + \phi_i + \mu_i \quad (3)$$

$$Y_i = \beta_0 + \beta_1 \hat{T}_i^A + \beta_2 \mathbf{X}_i + \phi_i + \epsilon_i \quad (4)$$

where  $T_i^R$  is a measure of teacher conscientiousness for the teacher to whom student  $i$  was randomly assigned and  $T_i^A$  is the corresponding measure of teacher conscientiousness for the teacher whom student  $i$  actually had during the school year<sup>20</sup>, and  $\hat{T}_i^A$  represents fitted values of  $T_i^A$  based upon estimations of Equation 3. The other variables are as they are in equation (2).<sup>21</sup> These models, together with the random assignment of teachers to students, provide causal estimates of how student outcomes are altered when they have teachers with varying levels of conscientiousness. Again, standard errors are clustered at the classroom level.

As an additional analysis, I estimate the same models substituting my survey-based measures of teacher conscientiousness with traditional measures of teacher quality (e.g., value-added scores, student ratings, scores on formal classroom observations, principal subjective

---

<sup>20</sup> MET data that was made available to me contained multiple teacher identification numbers for each student: (a) a randomly assigned teacher identifier, (b) an actual teacher in October identifier, (c) an actual teacher in May identifier, and (d) a global teacher identifier. Curiously, the global teacher identifier and the identifiers for actual teachers in October and May were not always consistent, so it was sometimes unclear which teacher was the student's actual teacher. I ran my analysis using each of these identifiers and found that my results are not sensitive to the choice of identifier. I present results for the specification that uses the identifier for the student's actual teacher in October.

<sup>21</sup> Strictly speaking, it is not absolutely necessary to include variables to control for prior-year measures of my student outcome variables because I rely on the random assignment of teachers to students. Including them would be useful to improve the precision of my estimated coefficients. However, doing so is not possible in the data since I do not observe all students in both years of the MET study. Students are only part of the MET study during years where they have MET teachers. The only exception to this rule is the inclusion of prior-year test scores for all students because they are provided in the MET data.

ratings<sup>22</sup>). These models provide a point of comparison for my survey-effort measures. In other words, they reveal how well the survey-effort measures of teacher conscientiousness predict and explain variation in student outcomes relative to traditional measures of teacher quality, further honing similarities and distinctions between each measure of teacher quality. At the very least, I would like to examine the ability of traditional measures of teacher quality based upon data from the first year of the MET study to predict survey-effort measures of student conscientiousness in the subsequent year.

## Results

### Relationships between Measures of Teacher Quality

I begin by presenting relationships between my survey-effort measures of teacher conscientiousness with traditional measures of teacher quality. Table 2 lists coefficient estimates and standard errors from bivariate regressions where my behavioral measures of teacher quality are independent variables and traditional measures of teacher quality are dependent variables. In general, my behavioral measures appear to be uncorrelated with value-added scores but are correlated with other indicators of teacher quality such as those based upon formal classroom observations, student ratings, and principal ratings. Teachers who are less conscientious as captured by my survey-effort measures have worse scores on other measures of teacher quality.

For instance, a one standard deviation increase in the item nonresponse rate (i.e., lower conscientiousness) is associated with between a 0.05 to 0.10 standard-deviation decrease in FFT scores, CLASS scores, and Tripod ratings. Similarly, I observe that teachers who more

---

<sup>22</sup>Although the principal subjective ratings variable is binary, I use linear probability models in the first stage of the IV models to predict ratings that principals gave to students' actual teachers. I ran specifications where I estimate probit models in the first stage. Results whether I use linear probability models or probit models do not substantively differ. Moreover, an ITT analysis using a nonlinear specification that include dummies for each of the six principal ratings yielded findings similar to those presently shown in the results section.

extensively engage in careless answering appear to have lower FFT scores. A one standard deviation increase in careless answering is associated with a decrease in FFT scores by 0.05 standard deviations, though the result is only significant at the 90 percent confidence level ( $p = 0.06$ ).

Table 2

*Relationships between Different Measures of Teacher Quality*

	(1) Value- added in Math	(2) Value- added in English	(3) FFT Score	(4) CLASS Score	(5) Student Tripod Ratings	(6) Principal Subjective Ratings
Item	-0.011	0.014	-0.059**	-0.068**	-0.085***	-0.006
Nonresponse	(0.034) N = 813	(0.033) N = 904	(0.029) N = 1,217	(0.028) N = 1,238	(0.022) N = 1,960	(0.013) N = 1,385
Careless Answering	-0.015 (0.033) N = 813	0.043 (0.030) N = 904	-0.054* (0.029) N = 1,217	-0.037 (0.028) N = 1,238	-0.001 (0.022) N = 1,960	-0.022 (0.013) N = 1,385
Survey Omission	0.036 (0.067) N = 1,109	0.056 (0.064) N = 1,240	-0.141** (0.061) N = 1,555	-0.185*** (0.061) N = 1,580	0.017 (0.046) N = 2,597	-0.046* (0.027) N = 1,852

Note: Value-added scores are based upon two years of data. FFT, CLASS, and Student Tripod ratings are based on lessons evaluated in the first year of the MET study. All measures, except the binary Survey Omission and Principal Subjective Ratings variables, are standardized to have a mean equal to 0 and standard deviation equal to 1. Columns 1 through 5 report coefficients from bivariate regressions where survey-effort measures of noncognitive skills are the independent variable. Column 6 reports marginal effects after estimating an analogous probit model. Standard errors are in parenthesis and sample sizes are written below. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ .

I finally consider relationships between survey-effort measures of conscientiousness and my indicator for whether a teacher overlooks the Teacher Working Conditions Survey. As shown in the last row of Table 2, teachers who fail to begin the survey, on average, have FFT and CLASS scores that are lower than teachers who complete the survey. The differences in scores based upon these observation protocols are approximately 0.14 and 0.19 standard deviations,

respectively. Moreover, teachers who fail to complete the survey are about 5 percent more likely to receive one of the three lower subjective ratings from their principals rather than one of the three higher ratings. However, there are no discernable differences in value-added scores between teachers who do or do not begin the survey.

### **Teacher Impacts on Student Noncognitive Skills**

Turning attention to my second research question, I find that my survey-effort measures of teacher conscientiousness are important determinants of certain student outcomes but not others. Again, I leverage the randomized assignment of teachers to students to estimate the causal effects that teachers with varying levels of conscientiousness have on student conscientiousness and test scores.

**Post-Randomization and Post-Attrition Covariate Balance.** Before presenting results, I check if the randomization process yielded covariate balance. If so, this would lend credence to the claim that the randomization of teachers to students was properly executed so as to remove the bias in my estimates that are attributable to systematic sorting of students to teachers. As can be expected, a substantial amount of attrition and noncompliance with random assignment occurred in the execution of the original MET study (Kane et al., 2013; Rowan & White, 2015). However, the ITT and LATE estimates only partially address the noncompliance issue. Put more explicitly, some non-complying students transferred districts, switched to non-participating schools in a participating district, or remained at the same participating school but transferred to a classroom taught by a teacher who was not participating in the MET study. Importantly, outcomes for non-complying students who were not taught by a MET teacher attrite and are not observed in my data. For this reason, my checks for post-randomization covariate balance are based upon data that have been collected post-attrition. Detecting covariate balance in this data would lend more credence that my estimates can legitimately be interpreted as causal.

We check for post-attrition, post-randomization covariate balance in two ways. First, I test to see if classrooms within randomizations blocks differed along student demographic characteristics and student prior-year test scores. To do this, I first demean each student characteristic or prior-year test score by randomization blocks. I then regress demeaned values of a particular student characteristic or a prior-year test score on a vector of dummy variables that indicate a student's randomly assigned teacher. Finally, I conduct an F-test to see if I can reject the null hypothesis that the estimated coefficients on the set of teacher dummy variables are jointly equal to each other. Failing to reject the null hypothesis would provide evidence that classrooms were balanced across student covariates within randomization blocks and suggest that the randomization process occurred appropriately so as to eliminate the systematic sorting of students to teachers.

Table 3

*Post-Attrition Baseline Covariate Balance across Randomly Assigned Teachers within Randomization Block*

<i>Student Characteristic</i>	F-Statistic	P-Value
Age	0.82	1.000
Male	0.57	1.000
English-language Learner	1.28	0.000
Special Education	0.93	0.953
Gifted	1.55	0.000
Free or reduced-priced lunch	0.67	1.000
Black	0.68	1.000
Hispanic	0.72	1.000
Asian	0.69	1.000
White	0.71	1.000
Prior Year Math Test Scores	1.60	0.000
Prior Year English Test Scores	1.50	0.000

Notes: I estimated a model that used teacher fixed effects to predict each student characteristic, while also controlling for randomization blocks. This table displays F-statistics and p-values from tests that coefficients estimates for teacher fixed effects are jointly equal to each other. Sample sizes in each of these estimated models range from 16,069 to 18,872 depending on missing data.

Results for these estimates are shown in Table 3. Although I do not find evidence of imbalance along a majority of student characteristics across classrooms within the same randomization blocks, there is some evidence of disparity. For instance, English language learners, gifted students, and students with higher test scores are more likely to be assigned to particular teachers within the same randomization blocks. To address this issue, I include controls for a wide range of classroom-level characteristics, including but not limited to the average prior-year test scores and proportions of students who are classified as gifted or as English-language learners.

As a second test for covariate balance, I run a series of bivariate regressions where I use teacher quality measures of a student's randomly-assigned teacher to predict each observable student demographic characteristic. If the randomization process was implemented appropriately, I should not observe any explanatory variables attaining statistical significance in these regressions. Results are shown in Table 4. Though I find statistically significant differences in student characteristics in a few of my regressions, I cannot clearly rule out the possibility that these have occurred by chance given the number of statistical tests that were conducted. More precisely, I find 8 statistically significant covariates out of 72 tests, when I should expect about 7 by chance (assuming a significance level of  $\alpha = 0.1$ ). Again, I control for all available observable individual-level and classroom-level demographic characteristics and prior-year test scores in my analyses.

Ultimately, given the inclusions of this wide range of covariates together with evidence of post-randomization and post-attrition balance across a majority of student covariates, I maintain that I have sufficiently eliminated systematic sorting of students to teachers. In other words, I have strong reason to believe that my estimates are both causal and valid.

Table 4

*Post-Attrition Baseline Covariate Balance between Student Characteristics and Randomly Assigned Teacher Characteristics*

	Tripod Score	FFT Score	CLASS Score	Survey Omision	Item Non- response	Careless Answering
<i>Student Characteristic</i>						
Age	0.022 (0.020)	-0.011 (0.019)	-0.002 (0.009)	-0.018 (0.014)	-0.013 (0.065)	0.008 (0.005)
Male	0.006 (0.020)	0.030 (0.019)	0.013 (0.011)	-0.003 (0.012)	-0.088 (0.062)	0.011* (0.005)
ELL	-0.010 (0.011)	-0.025† (0.014)	-0.005 (0.007)	0.012 (0.013)	0.033 (0.050)	0.003 (0.003)
Special Education	0.003 (0.010)	-0.006 (0.010)	-0.001 (0.006)	0.007 (0.008)	0.014 (0.034)	0.002 (0.003)
Gifted	0.004 (0.011)	0.003 (0.015)	0.002 (0.006)	0.021 (0.009)	0.046 (0.055)	-0.008† (0.004)
FRL	0.006 (0.016)	-0.014 (0.020)	-0.016† (0.009)	0.012 (0.012)	0.053 (0.059)	-0.004 (0.005)
Black	-0.002 (0.012)	0.002 (0.014)	-0.001 (0.006)	-0.018† (0.010)	-0.023 (0.042)	0.003 (0.003)
Hispanic	0.010 (0.014)	-0.011 (0.019)	0.001 (0.001)	-0.002 (0.012)	0.039 (0.052)	0.001 (0.004)
Asian	-0.001 (0.009)	0.014 (0.014)	0.004 (0.006)	0.009 (0.007)	0.029 (0.023)	-0.001 (0.002)
White	-0.006 (0.013)	-0.007 (0.014)	-0.006 (0.006)	-0.015* (0.008)	0.034 (0.044)	-0.004 (0.004)
Prior Year Math Scores	-0.037 (0.043)	0.067 (0.047)	0.017 (0.023)	0.001 (0.032)	-0.037 (0.168)	-0.030* (0.013)
Prior Year English Scores	-0.059 (0.044)	0.021 (0.048)	-0.006 (0.023)	-0.023 (0.033)	0.111 (0.156)	-0.028* (0.133)

Notes: ELL = English language learner. FRL = Free or Reduce-priced lunch. This table displays coefficient estimates from separate bivariate regressions where the dependent variable is a student characteristic and the independent variable is a teacher quality measure. Standard errors are in parenthesis. All regressions control for randomization block. \*p<0.05; †p<0.1.



**Teacher Noncognitive Skills and Student Noncognitive Skills.** ITT and IV estimates are shown in Tables 5 and 6 respectively. I first discuss the ITT results. Each panel in Table 5 displays coefficients and standard errors that are estimated from Equation 2 for one of my three survey-effort measures of teacher conscientiousness and each student outcome. For example, the first panel under column 5 suggests student item nonresponse increases by about 2.4 percent of a standard deviation when a student is randomly assigned to a teacher who, all else equal, is one standard deviation higher on the distribution of teacher item nonresponse.

In fact, I find that all three survey-effort measures of teacher conscientiousness are predictive of student item nonresponse (Column 5) but are not predictive of student achievement in math (Column 1) or English (Column 2), student self-reported effort (Column 4), and careless answering (Column 6). As shown in column 5, effect sizes range from 2 to 6 percent of a standard deviation in item nonresponse. Focusing on results for grit in Column 3, I find that students self-report lower levels of grit when they are randomly assigned to teachers who fail to start the Teacher Working Conditions survey. In other words, students randomly assigned to a less conscientious teacher appear to become less conscientious, primarily according to the item-nonresponse proxy and self-reported grit measures. In contrast, student test scores do not seem to be affected when they are randomly assigned to teachers of varying levels of conscientiousness.

Table 5

*ITT Estimates of Teacher Effects (Based on Survey-effort Measures of Conscientiousness) on Student Outcomes*

	Student Outcome					
	(1) Math Test Scores	(2) English Test Scores	(3) Self-Reported Grit	(4) Self-Reported Effort	(5) Item Nonresponse	(6) Careless Answering
Teacher Item Nonresponse	-0.008 (0.013)	-0.008 (0.010)	0.021 (0.013)	-0.015 (0.015)	0.024* (0.014)	-0.010 (0.017)
R <sup>2</sup>	0.70	0.67	0.21	0.14	0.21	0.18
Observations	5,712	7,247	10,518	10,682	10,706	10,648
Teacher Careless Answering	-0.013 (0.010)	-0.006 (0.007)	0.017 (0.012)	0.021 (0.014)	0.035*** (0.012)	0.017 (0.014)
R <sup>2</sup>	0.70	0.67	0.21	0.14	0.21	0.18
Observations	5,712	7,247	10,518	10,682	10,706	10,648
Survey Omission	-0.051 (0.033)	-0.009 (0.029)	-0.058* (0.035)	0.084 (0.037)	0.064** (0.030)	0.059 (0.043)
R <sup>2</sup>	0.68	0.67	0.20	0.15	0.21	0.18
Observations	7,003	8,814	12,809	13,022	13,051	12,991

Note: All regressions control for student gender, race, age, special education status, free or reduce-priced lunch status, gifted status, English learner status, prior-year test scores, and randomization blocks, as well as classroom composition (i.e., average prior year test scores, proportion students in the class of a particular gender, race, special education status, free or reduce-priced lunch status, gifted status, and English learner status). Standard errors clustered at the classroom level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table 6

*IV Estimates of Teacher Effects (Based on Survey-effort Measures of Conscientiousness) on Student Outcomes*

	Student Outcomes					
	(1) Math Test Scores	(2) English Test Scores	(3) Self-Reported Grit	(4) Self-Reported Effort	(5) Item Nonresponse	(6) Careless Answering
Teacher Item Nonresponse	-0.011 (0.020)	-0.012 (0.015)	0.031 (0.023)	0.021 (0.026)	0.033 (0.023)	-0.019 (0.028)
R <sup>2</sup>	0.70	0.67	0.21	0.14	0.21	0.18
Observations	5,712	7,247	10,102	10,253	10,277	10,229
Teacher Careless Answering	-0.023 (0.017)	-0.011 (0.012)	0.030 (0.024)	0.040 (0.026)	0.059*** (0.022)	0.026 (0.027)
R <sup>2</sup>	0.70	0.67	0.21	0.14	0.21	0.18
Observations	5,712	7,247	10,102	10,253	10,277	10,229
Survey Omission	-0.086 (0.055)	-0.015 (0.050)	-0.086 (0.064)	-0.146** (0.068)	0.137*** (0.052)	0.134* (0.074)
R <sup>2</sup>	0.68	0.67	0.20	0.13	0.21	0.18
Observations	7,003	8,814	12,228	12,425	12,454	12,404

Note: All regressions control for student gender, race, age, special education status, free or reduce-priced lunch status, gifted status, English learner status, prior-year test scores, and randomization blocks, as well as classroom composition (i.e., average prior year test scores, proportion students in the class of a particular gender, race, special education status, free or reduce-priced lunch status, gifted status, and English learner status). Standard errors clustered at the classroom level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Results based upon the IV estimates are shown in Table 6 and generally comport with results based upon the ITT analysis. As shown in the second panel of Column 5, students experience increases in item nonresponse rates when they have teachers who exhibit more carelessness while answering surveys. In the third panel under Column 5, one can also see that relative to students who have teachers that complete the Teacher Working Conditions Survey, students with teachers who do not complete the Teacher Working Conditions Survey have item nonresponse rates that are approximately 0.14 standard deviations higher, all else being equal. In contrast, the associations between (a) student item nonresponse with teacher item nonresponse and (b) student self-reported grit and teacher survey omission, which were significant under the ITT models, are no longer statistically significant under the IV models. The coefficient estimates are positive and larger than the ITT estimates but imprecisely estimated. Furthermore, the IV estimates indicate that students self-report less effort in their classes if they are taught by teachers who refrain from responding to the Teacher Working Conditions Survey. Relative to students with teachers who do indeed respond to the survey, these students' self-reported effort ratings are about 0.15 standard deviations lower.

**Traditional Measures of Teacher Quality and Student Noncognitive Skills.** To shed additional light onto my survey-effort measures of teacher conscientiousness, I now present estimates of the relationships between student noncognitive skill outcomes and traditional measures of teacher quality. ITT results are shown in Table 7 and are displayed in a fashion analogous to Table 5 but with traditional measures of teacher quality rather than my survey-effort measures of teacher noncognitive skills. Findings shown in the first panel demonstrate that teachers who have been rated more highly by their students on the Tripod survey during the first year of the MET study are more effective at improving conscientiousness in their subsequent set

of students during the second year of the MET study. Students who are randomly assigned to a teacher whose Tripod rating is one standard deviation higher have self-reported grit scores that are about 0.03 standard deviations higher, self-reported effort scores that are about 0.05 standard deviations higher, and careless answering scores that are 0.09 standard deviations lower. Higher quality teachers as judged by student ratings do not appear to have an effect on student item response rates.

In the second panel, one can also observe that students who are randomly assigned to teachers who receive higher principal ratings exhibit less careless answering. That is, these students become more conscientiousness. Relative to students assigned to teachers who received one of the three lower categories of principal ratings, students assigned to teachers who received one of the three higher categories of principal ratings have careless answering scores that are almost 10 percent of a standard deviation lower.

Turning to the third and fourth panels, variation in teacher quality as measured by classroom observations protocols sometimes explains variation in teacher effectiveness at improving student noncognitive skills. While FFT ratings are uncorrelated with such impacts, CLASS ratings are modestly predictive of student grit and item nonresponse but slightly more strongly predictive of careless answering. A one standard deviation increase in a teacher's rating based upon the CLASS protocol is associated with an increase in grit and item nonresponse by about 0.03 standard deviations. The corresponding effect size for careless answering is 0.04 standard deviations.

Finally, I observe in the last two panels that teacher value-added scores show no association with impacts on student noncognitive skills.

Table 7

*ITT Estimates of Teacher Effects (Based on Traditional Measures of Teacher Quality)*

	Student Outcomes			
	(1) Self-Reported Grit	(2) Self-Reported Effort	(3) Item Nonresponse	(4) Careless Answering
Student Tripod Ratings	0.034*** (0.011)	0.051*** (0.014)	-0.003 (0.017)	-0.091*** (0.017)
R <sup>2</sup>	0.20	0.13	0.19	0.18
Observations	12,364	12,563	12,587	12,527
Received Higher Principal Ratings	-0.030 (0.032)	0.024 (0.034)	-0.006 (0.026)	-0.098*** (0.037)
R <sup>2</sup>	0.20	0.13	0.25	0.18
Observations	9,873	10,042	10,064	10,010
FFT Score	0.002 (0.016)	-0.000 (0.018)	0.018 (0.020)	-0.009 (0.021)
R <sup>2</sup>	0.21	0.13	0.21	0.18
Observations	11,786	11,988	12,016	11,956
CLASS Score	0.027* (0.016)	-0.010 (0.017)	0.027* (0.014)	-0.039** (0.020)
R <sup>2</sup>	0.21	0.13	0.21	0.18
Observations	11,786	11,988	12,016	11,956
Teacher Value Added (English)	0.013 (0.018)	0.016 (0.021)	-0.020 (0.019)	-0.027 (0.024)
R <sup>2</sup>	0.21	0.12	0.24	0.18
Observations	7,190	7,302	7,324	7,310
Teacher Value Added (Math)	-0.008 (0.024)	0.028 (0.025)	0.051 (0.033)	-0.001 (0.034)
R <sup>2</sup>	0.22	0.13	0.12	0.18
Observations	6,762	6,886	6,894	6,844

Note: Principal ratings is a dichotomous variable where the omitted category represents teachers in approximately the lower half of the distribution of principal ratings. All regressions control for student gender, race, age, special education status, free or reduce-priced lunch status, gifted status, English learner status, prior-year test scores, and randomization blocks, as well as classroom composition (i.e., average prior year test scores, proportion students in the class of a particular gender, race, special education status, free or reduce-priced lunch status, gifted status, and English learner status). Standard errors clustered at the classroom level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table 8

*IV Estimates of Teacher Effects (Based on Traditional Measures of Teacher Quality)*

	Student Outcomes			
	(1) Self-Reported Grit	(2) Self-Reported Effort	(3) Item Nonresponse	(4) Careless Answering
Student Tripod Ratings	0.054** (0.021)	0.075*** (0.025)	-0.009 (0.029)	-0.146*** (0.028)
R <sup>2</sup>	0.20	0.13	0.19	0.18
Observations	11,840	12,024	12,048	11,998
Received Higher Principal Ratings	-0.052 (0.056)	0.041 (0.059)	-0.011 (0.044)	-0.171*** (0.064)
R <sup>2</sup>	0.21	0.13	0.24	0.18
Observations	9,873	10,042	10,064	10,010
FFT Score	0.008 (0.027)	0.019 (0.030)	0.013 (0.033)	-0.045 (0.033)
R <sup>2</sup>	0.21	0.13	0.21	0.18
Observations	11,263	11,450	11,478	11,428
CLASS Score	0.045 (0.028)	0.021 (0.029)	0.036 (0.024)	-0.077** (0.034)
R <sup>2</sup>	0.21	0.13	0.21	0.18
Observations	11,263	11,450	11,478	11,428
Teacher Value Added (English)	0.024 (0.042)	0.033 (0.050)	-0.067* (0.039)	-0.081 (0.049)
R <sup>2</sup>	0.21	0.12	0.24	0.18
Observations	6,895	6,995	7,017	7,011
Teacher Value Added (Math)	-0.012 (0.058)	0.059 (0.060)	0.113 (0.078)	-0.003 (0.079)
R <sup>2</sup>	0.22	0.13	0.12	0.19
Observations	6,385	6,504	6,512	6,464

Note: Principal ratings is a dichotomous variable where the omitted category represents teachers in approximately the lower half of the distribution of principal ratings. All regressions control for student gender, race, age, special education status, free or reduce-priced lunch status, gifted status, English learner status, prior-year test scores, and randomization blocks, as well as classroom composition (i.e., average prior year test scores, proportion students in the class of a particular gender, race, special education status, free or reduce-priced lunch status, gifted status, and English learner status). Standard errors clustered at the classroom level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Analogous estimates based upon IV models are largely consistent with ITT results, where teachers who are rated higher by their students are more effective at improving student conscientiousness as measured by student self-reported grit, self-reported effort, and careless

answering. Likewise, students exhibit less careless answering when they have teachers who receive higher ratings from their principals or based on the CLASS protocol. Yet there are differences between the ITT and IV results. Teachers rated more highly on the CLASS protocol no longer appear more effective at improving student self-reported grit and item nonresponse. Notably, these relationships were only significant at the 90 percent confidence level in the ITT models. Lastly, students taught by English teachers with higher value-added scores also appear to experience decreases in item nonresponse rate, though the result is only significant at the 90 percent confidence level.

## **Discussion and Conclusion**

### **Summary**

In this article, I have aimed to measure teacher noncognitive skills and to assess how they affect similar noncognitive skills in students. Little research has investigated the role that teacher noncognitive skills play in educational outcomes (but see Duckworth et al., 2009; Rockoff et al., 2011). The paucity of research is attributable to the fact that teacher noncognitive skills data are rarely collected and available. I overcome this data limitation by utilizing three new survey-effort measures of conscientiousness that can be constructed using data collected from teacher surveys: (a) item nonresponse, (b) careless answering, and (c) survey omission. Based upon conceptual reasons from survey methods research and empirical evidence, I view these behavioral measures as proxies for noncognitive skills related to conscientiousness (Hitt et al., forthcoming; Hitt, 2015; Krosnick, 1991; Smith, 1982; Zamarro et al., 2016). Furthermore, I build survey-effort measures of conscientiousness for students, which typically are unavailable in many datasets, to expand the range of student outcomes to beyond test scores.

I first find that my survey-effort measures of teacher conscientiousness are correlated with some but not all existing measures of teacher quality, namely those based upon formal



classroom observations, student ratings, and principal ratings but not value-added scores. This result provides additional validation for the use of item nonresponse, careless answering, and survey omission as proxies for meaningful teacher characteristics. These survey-effort measures may represent important observable teacher characteristics that can be easily obtained from existing data and used for further research of teacher quality.

Specifically, I maintain that I have captured levels of teacher conscientiousness with my survey-effort measures, as other work suggests (Hitt, 2015; Hitt et al, forthcoming, Zamarro et al., 2016). In fact, findings from prior research suggest that ratings based upon principals, students, or other classroom observations may capture teacher conscientiousness. For instance, Rockoff et al. (2011) find that novice teachers who have higher self-reported levels of conscientiousness also received higher subjective ratings from their mentor teachers. Interestingly, Rockoff et al. additionally found that novice teachers who did not complete their survey were rated lower by the mentor teacher than those who completed the survey. The intention of this comparison was not a test of teacher conscientiousness but a robustness check to address the issue of missing data. Perhaps these researchers, by happenstance, actually uncovered more evidence that novice teachers with lower conscientiousness receive lower subjective ratings. The proposition that ratings based upon observations of teachers capture personality traits has also been raised by Harris et al. (2014) and is worth further investigation. Certainly, it would be valuable for scholars and practitioners to reflect upon what, exactly, ratings by principals, students, and other observers of teachers in their classrooms actually capture. Additional inquiry into whether survey-effort or other related measures capture other noncognitive skills would be worthwhile.

Second, I find that teachers who exhibit more conscientiousness as measured by my survey-effort measures are more effective at improving student conscientiousness. Yet I do not find evidence that teacher conscientiousness is tied to student achievement. Leaning upon the random assignment of teachers to students in my data, I interpret my results as causal. Students exhibit higher item nonresponse rates but not higher test scores when they have teachers who exhibit higher item nonresponse rates<sup>23</sup>, greater levels of careless answering, or fail to complete a survey when asked to do so. There is also some suggestive evidence that students self-report lower levels of grit when they are assigned to teachers who do not complete surveys and self-report lower levels of effort when they are actually taught by those teachers.

Third, I do not find much evidence that teachers with high formal classroom observation or value-added scores are effective at improving student conscientiousness. The exception to this result is that higher-quality teachers as measured by the CLASS observation protocol appear to be more effective at improving student conscientiousness as measured by careless answering. However, the traditional measures of teacher quality that are most strongly predictive of student conscientiousness are student ratings of teachers — similar to what Kraft and Grace (2016) have found — and principal subjective ratings.<sup>24</sup> Taken together, these results indicate that existing measures of teacher quality do not fully capture all the relevant ways in which teachers influence their students. In particular, my survey-effort measures of conscientiousness capture impacts that

---

<sup>23</sup> Again this result was only significant at the 90 percent confidence level in the ITT analysis and imprecisely estimated in the IV analysis.

<sup>24</sup> The FFT, CLASS, and Tripod instruments also have a variety of subscales designed to capture different dimensions of teacher quality. I estimate models using scores on these subscales in addition to models that use scores based on the entire scale. Results when using separate scores from the seven subscales of the Tripod instrument all reflect overall results. Results when using separate scores from subscales of the FFT and CLASS were mixed with no obvious patterns.

teachers have upon student item nonresponse — something traditional measures of teacher quality are less able to do in my data.

Overall, I interpret my estimates as lower bounds for the impacts that teachers have upon student conscientiousness. Arguably, each of my survey-effort measures capture latent traits that are not even necessarily strongly correlated with conscientiousness. For instance, both conscientiousness and problems with survey administration could explain variation in my measure of survey omission. A variety of logistical reasons may influence why some teachers did not begin the survey. Likewise, variation in careless answering could partially reflect genuine variation in teacher response patterns or natural measurement error in the scales (even though I already selected scales with higher levels of Cronbach's alpha) in addition to a lack of conscientiousness. All of these issues generate random measurement error and, at worst, cause my results to attenuate. Furthermore, while student test scores are arguably mostly influenced by one teacher in a particular content area, student noncognitive skills can be influenced by multiple teachers. For secondary students that have more than one teacher, this possibility introduces additional random noise in my analysis. For these reasons, I interpret my results to be conservative estimates. That I find systematic relationships between survey-effort measures of teacher conscientiousness and student outcomes, therefore, is even more striking.

Although I cannot provide evidence for the mechanisms for why these relationships exist, these findings are consistent with social learning theory where students learn to be more conscientious by observing their more conscientious role models (Bandura, 1977; Bandura & Walters, 1963). These patterns are similar to results in the previous chapter where students, over the course of their secondary schooling, become more conscientious (as measured by item nonresponse) during school years where they have more conscientious teachers (as measured by

survey omission). It is also possible that more conscientious teachers tend to be more effective at implementing certain classroom management or instructional techniques that could be conducive to fostering student conscientiousness (Blazar & Kraft, 2015). Indeed, my survey-effort measures of teacher conscientiousness are correlated with teacher quality measures based upon formal classroom observations and the CLASS instrument. Still, I do not find that measures of teacher quality based upon the FFT or value-added scores are predictive of student outcomes in conscientiousness. Again, more work investigating what teacher traits, exactly, are captured by classroom observation protocols or how classroom interactions between teachers and students affect student noncognitive outcomes will be useful.

### **Implications for understanding teacher quality**

My findings suggest that teacher quality is multidimensional. Some teachers are effective at improving student test scores while others are more effective at improving student conscientiousness. Moreover, it appears that teachers who are themselves more conscientious are more effective at improving student conscientiousness but not necessarily student test scores. This finding tracks with Kraft and Grace (2016) who use MET data to find that teacher effects on student noncognitive skills and achievement are weakly correlated. Indeed, these findings align with those from a growing body of teacher quality research (Blazar & Kraft, 2015; Cheng, 2015; Gershenson, 2016; Jackson, 2012; Jennings & Diprete, 2010; Koedel, 2008).

This paper joins this body of work studying teacher impacts on student noncognitive outcomes, all of which suggest evaluations of teachers, schools, and other educational interventions need to consider both student achievement and noncognitive skill outcomes. The former have been the focus and standard by which educational programs are evaluated, but research overlooks impacts on noncognitive skills by solely relying on cognitive measures as outcome variables. Such oversight is not inconsequential because noncognitive skills are

important determinants for later-life outcomes, even after accounting for cognitive ability (Heckman, Stixrud, & Urzua, 2006). Without considering impacts on noncognitive skills, the benefits that teachers impart to their students will likely be misstated (Heckman, Pinto, & Savelyev, 2013). Teachers that realize large gains in student noncognitive skills will likely be categorized as ineffective if evaluation systems only rely on gains in student achievement (Grissom et al., 2016). In fact, it is worthwhile to reiterate that even my survey-effort measures of noncognitive skills are predictive of student educational attainment and labor-market outcomes; this finding cannot be replicated in the MET data but has been documented in several longitudinal analyses (Hitt et al., forthcoming; Hitt, 2015; Cheng, 2015).

Our findings also speak to prior research from the original MET study. Kane et al. (2012) find that classroom observations, student ratings, and value-added scores are correlated with student test scores, albeit weakly, and recommend creating a composite score that aggregates all these measures to gauge teacher impacts on student achievement. The composite score, as they argue and demonstrate, is more reliable than each measure alone and is more predictive of teacher value-added scores (see also Mihaly et al, 2013). However, the assumption behind Kane et al.'s (2012) recommendation is that teacher quality is a unidimensional construct (e.g., the ability to improve student achievement) and that different measures of teacher quality capture mutually exclusive parts of that construct. My results provide reason to dispute that assumption. Not only is teacher quality multidimensional but different measures of teacher quality capture different aspects of teacher quality. For instance, my survey-effort measures of teacher conscientiousness clearly predict student conscientiousness but not student achievement, and student ratings of teachers are predictive of student grit, effort, and careless answering as well as test scores. Aggregating different measures of teacher quality to ascertain their predictive power

to forecast student achievement may mask teacher effects on other student outcomes and obscure the multifaceted ways in which different teachers benefit their students.<sup>25</sup>

### **Implications for Future Research and Practice**

Identifying effective teachers has been elusive. Observable characteristics such as years of experience, educational background, and licensure are at best only weakly correlated with student achievement outcomes (Buddin & Zamarro, 2009; Goldhaber, 2008; Jacob, 2007). In this study, I follow Rockoff et al.'s (2011) approach by first hypothesizing that teacher noncognitive skills play an important role in shaping student outcomes. My innovation, however, is to present evidence supporting this hypothesis by using survey-effort measures of teacher noncognitive skills.

An increasing number of studies find that teachers vary in their effectiveness at improving a variety of student outcomes, but it still remains unclear what types of teachers are effective at improving particular outcomes (Hanushek & Rivkin, 2006). I am only aware of four other studies that have examined this issue. First, Rockoff et al. (2011) provides some evidence that teacher self-efficacy is related to student test scores in math, but the correlation is modest. Second, Blazar and Kraft (2015) show that teachers who are rated highly on the CLASS rubric appear to be more effective at improving student self-efficacy. Third, the previous chapter demonstrates that students experience gains in conscientiousness in years when they have more

---

<sup>25</sup> Certainly, one could undergo the empirical exercise of recalculating ideal weights as in Kane et al. (2012) and Mihaly et al. (2013) to create composite measures of teacher quality by including my survey-effort measures of teacher and student noncognitive skills. One of the intents behind Kane et al. (2012) and Mihaly et al. (2013) was to provide guidance for practitioners who desire to utilize traditional measures of teacher quality in a systematic way. However, I caution that my survey-effort measures of teacher quality are currently suitable for research purposes only, not for use by schools and policymakers in their everyday operations and practice.

conscientious teachers. Fourth, Duckworth et al. (2009) show that teacher grit and life satisfaction are predictive of student achievement.

More work in the same vein as these studies and this present study needs to be undertaken to better understand what kinds of teachers are effective at improving both student cognitive and noncognitive outcomes. Testing for associations between teacher personality traits, teaching practices, and student outcomes, as this work and Rockoff et al. (2011) have done, could be a more promising avenue to uncovering the elusive observable teacher characteristics that are predictive of student outcomes. More generally, identifying the kinds of teachers that effectively develop student cognitive and noncognitive skills is a task that warrants additional scholarly attention.<sup>26</sup>

For now, this work as well as other studies that find heterogeneity in the ways that teachers affect their students, suggests that scholars, policymakers, and practitioners need to think more critically about the ways in which they conceptualize teacher quality. It is not wholly unreasonable to construe teacher quality as a teacher's ability to improve student achievement. After all, student achievement and cognitive ability are crucial components of human capital development and play an important role in determining a student's future educational attainment, employment, earnings, and other long-run life outcomes (Becker, 1964; Chetty et al., 2014). However, noncognitive skills also play their own role in determining the same outcomes, above and beyond the role that student achievement plays (Heckman et al., 2006; Almlund et al., 2011). It is also unclear to what extent scores on tests of cognitive ability are driven by student effort or

---

<sup>26</sup> Bargagliott et al., (2016) show that kindergarten teachers who utilize certain pedagogical practices to teach math appear to be more effective at improving particular noncognitive skills in their students. Research investigating teaching practices associated with improving student noncognitive skills is equally important, though some studies fail to find a relationship between pedagogical approaches and student outcomes (Jennings & Diprete, 2010).

student content knowledge (Borghans & Schils, 2013; Mendez et al., 2015; Hitt, Zamarro, & Mendez, 2016). A conception of teacher quality that only focuses on impacts on student achievement are therefore incomplete, overlooking the nontrivial ways in which teachers benefit students.

My use of survey-effort measures of teacher and student conscientiousness also is a key contribution to research. These measures can be readily constructed in most data sets that rely on self-reports and provide viable research strategy to answer questions about student and teacher noncognitive skills – a topic that is receiving increasing attention in several academic fields. More importantly, I have shown in this study that these survey-effort measures are not simply random noise. They are related to other traditional measures of teacher quality and predictive of student outcomes. Even if one disputes their validity as measures of conscientiousness, such relationships demand explanation.

More study of these survey-effort measures needs to be completed. Little is known, for example, about the stability of such measures and how they will behave in different survey contexts. Indeed, in other work, I have found that the ways respondents shirk on surveys varies depending upon whether the survey is compulsory or voluntary and whether the survey is administered via pencil-and-paper, with an interviewer present, or via a computer (Zamarro et al., 2016; Hitt et al., forthcoming). In fact, the differences in survey mode may explain why I found that my measures of teacher conscientiousness were predictive of student item nonresponse but not of student carelessness. Item nonresponse may have been a more expedient way to shirk than providing careless answers on the particular survey that students were tasked to complete — a pattern found elsewhere (Zamarro et al., 2016; Hitt et al., forthcoming). This need for additional research also explains why I do not support the use of my survey-effort measures



for high-stakes teacher evaluation. For now, I maintain that these measures provide useful information that can be used for research purposes that will enhance the understanding of teacher quality and how to improve student outcomes. Different teachers shape their students in many ways that affect their long-run life prospects and future well-being. It behooves researchers and policymakers to better understand the underlying mechanisms behind these developmental and formative processes.

## References

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. D. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 1–181). Amsterdam: Elsevier.
- Backes, B., & Hansen, M. (2015). Teach for America Impact Estimates on Nontested Student Outcomes. (CALDER Working Paper No. 146). Washington, DC: National Center for Analysis of Longitudinal Data in Education.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A., & Walters, R.H. (1963). *Social learning and personality development*. New York: Holt, Rinehart and Winston.
- Bargagliotti, A., Gottfried, M.A., Guarino, C. (2016). The effects of kindergarten mathematical instructional practices on young children’s noncognitive development. Paper presented at the Association for Education Finance and Policy 41st Annual Conference. Denver, CO.
- Blazar, D., & Kraft, M. A. (2015). Teacher and teaching effects on students’ academic behaviors and mindsets (Working Paper No. 41). Cambridge, MA: Mathematica Policy Research.
- Borghans, L., & Schils, T. (2012). The leaning tower of PISA. *Unpublished Manuscript*. Available at <http://www.sole-jole.org/13260.pdf> (Accessed 30 March 2016).
- Borghans, L., ter Weel, B., & Weinberg, B.A. (2008). Interpersonal styles and labor market outcomes. *Journal of Human Resources*, 43(4), 815-858.
- Buddin, R., & Zamarro, G. (2009). Teacher Qualifications and Student Achievement in Urban Elementary Schools. *Journal of Urban Economics*, 66(2), 103–115.
- Cheng, A. (2015). Like teacher, like student: Teachers and the development of student noncognitive skills (EDRE Working Paper No. 2015-02). Fayetteville, AR: Department of Education Reform, University of Arkansas.
- Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679.
- Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, 41(4), 778–820.
- Cobb-Clark, D.A., & Schurer, S. (2012). The stability of big-five personality traits. *Economic Letters*, 115(1), 11-15.
- Dalal, R.S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, 90(6): 1241-1255.

- Danielson, C. (2007). *Enhancing professional practice: A framework (2nd ed)*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Dobbie, W., & Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*, *123*(5), 985–1037.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (Grit-S). *Journal of Personality Assessment*, *91*(2), 166–174.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. *Educational Researcher*, *44*(4), 237–251.
- Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, *4*(6), 540–547.
- Ferguson, R.F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, *94*(3), 24-28.
- Garrett, R., & Steinberg, M. P. (2015). Examining Teacher Effectiveness Using Classroom Observation Scores: Evidence from the Randomization of Teachers to Students. *Educational Evaluation and Policy Analysis*, *37*(2), 224–242.
- Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, *10*(1), 117–156.
- Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125-149.
- Goldhaber, D. D. (2008). Teachers Matter, But Effective Teacher Quality Policies are Elusive. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 146–165). New York, NY: Routledge.
- Grissom, J. A., Loeb, S., & Doss, C. (2016). The multiple dimensions of teacher quality: Does value-added capture teacher's nonachievement contributions to their schools? In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation Systems: Making the most of multiple measures* (pp. 37–50). New York, NY: Teachers College Press.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher Quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1052–1078). Amsterdam: Elsevier.
- Harris, D. & Sass, T. (2014). Skills, productivity, and the evaluation of teacher performance. *Economics of Education Review*, *40*, 183-204

- Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability. *American Educational Research Journal*, 51(1), 73–112.
- Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6), 2052–2086.
- Hitt, C.E. (2015). Just filling in the bubbles: Using careless answer patterns on surveys as a proxy measure of noncognitive skills (EDRE Working Paper 2015-06). Fayetteville, AR: Department of Education Reform, University of Arkansas.
- Hitt, C.E., Trivitt, J.R., Cheng, A. (forthcoming). When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*.
- Hitt, C.E., Zamarro, G., & Mendez, I. (2016) What if students don't care? Reexamining international differences in achievement and non-cognitive skills. Paper presented at the Association for Education Finance and Policy 41st Annual Conference. Denver, CO.
- Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114.
- Jacob, Brian A. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children* 17(1): 129–53.
- Jacob, B., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101–136.
- Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the behavioral indicators of conscientiousness (BIC). *Journal of Research in Personality*, 44(4), 501–511.
- Jackson, K. (2012). *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina* (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.
- Jennings, J. L., & DiPrete, T. A. (2010). Teacher Effects on Social and Behavioral Skills in Early Elementary School. *Sociology of Education*, 83(2), 135–159.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39, 103–129.

- Kane, T., Rockoff, J. E., & Staiger, D. (2008). What Does Certification Tell Us about Teacher Effectiveness. *Economics of Education Review*, 27(6), 615–631.
- Kane, T., McCaffrey, D.F., & Staiger, D.O. (2012). *Gathering feedback for teaching: Combining High-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T.J., McCaffrey, D.F., Miller, T., and Staiger, D.O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Seattle, WA: Bill and Melinda Gates Foundation.
- Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics*, 64(3), 560–572.
- Koedel, C., and Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, 6(1), 18–42.
- Kraft, M.A., & Grace, S. (2016). Teaching for tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies (Working Paper). Brown University: Providence, RI.
- MacCann, C., Duckworth, A.L., & Roberts, R.D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, 19(4), 451–458.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The Intertemporal Variability of Teacher Effect Estimates. *Education Finance and Policy*, 4(4), 572–606.
- Meade, A., & Craig, S.B. (2012). Identifying careless responses survey data. *Psychological Methods*, 17(3), 437-455.
- Mendez, I., Zamarro, G., Clavel, J.G., & Hitt, C.E. (2015). Non-cognitive abilities and Spanish regional differences in student performance in PISA 2009 (EDRE Working Paper No. 2015-05). University of Arkansas: Fayetteville, AR.
- Mihaly K., McCaffrey D. F., Staiger D., & Lockwood J. R. (2013). A composite estimator of effective teaching. Seattle, WA: Bill & Melinda Gates Foundation.
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Pianta, R.C., & Hamre, B.K., (2009) Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational*

- Researcher*, 38(2), 109–119.
- Pianta, R.C., LaParo, K.M., & Hamre, B.K. (2008). Classroom Assessment Scoring System Manual, Pre-K. Baltimore, MA: Brookes Publishing Co.
- Podgursky, M. J. (2005). Teacher licensing in U.S. public schools: The case for simplicity and flexibility. *Peabody Journal of Education*, 80(3), 15–43.
- Polikoff, M.S. (2015). The stability of observation and student survey measures of teaching effectiveness. *American Journal of Education*, 121, 183-212.
- Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322-338.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.
- Roberts, B.W., Harms, P.D., Caspi, A., & Moffitt, T.E. (2007). Predicting the counterproductive employee in a child-to-adult prospective study. *Journal of Applied Psychology*, 92(5), 1427–1436.
- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2), 247–252.
- Rockoff, J. E., Jacob, B., Kane, T., & Staiger, D. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43–74.
- Rockoff, J. E., Staiger, D., Kane, T., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, 102(7), 3184–2313.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537–571.
- Ruzek, E.A., T. Domina, A.M. Conley, G.J. Duncan, & Karabenick, S.A. (2014). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, 35(5-6), 852–882.
- Tate, R. (2004). A cautionary note on shrinkage estimates of school and teacher effects. *Florida Journal of Educational Research*, 42, 1–21.
- Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, 98(2), 438-456.
- Tsukayama, E., Duckworth, A.L. & Kim, B. (2013). Domain-specific impulsivity in school-age children. *Developmental Science*, 16(6), 879–893.

- West, M., Kraft, M. A., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148–170.
- White M., & Rowan B. (2012). A user guide to the “core study” data files available to MET early career grantees. Ann Arbor: Inter-University Consortium for Political and Social Research, The University of Michigan
- Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100, 61-78.
- Zamarro, G., Engberg, J., Saavedra, J. E., & Steele, J. (2015). Disentangling Disadvantage: Can We Distinguish Good Teaching from Classroom Composition? *Journal of Research on Educational Effectiveness*, 8(1), 84–111.
- Zamarro, G., Cheng, A., Shakeel, M., & Hitt, C. (2016). Comparing and validating measures of character skills: Findings from a nationally representative sample. Paper presented at the Association for Education Finance and Policy 41st Annual Conference. Denver, CO.

## Appendix: Items on Student Effort Scale

### Items for Elementary School Students

1. In this class, I take it easy and do not try very hard to do my best. *(Reverse coded)*
2. In this class, I stop trying when the work gets hard. *(Reverse coded)*
3. When doing schoolwork for this class, I try to learn as much as I can and don't worry about how long it takes.
4. I have pushed myself hard to completely understand lessons in this class.

### Items for Secondary School Students

1. In this class, I take it easy and do not try very hard to do my best. *(Reverse coded)*
2. In this class, I stop trying when the work gets hard. *(Reverse coded)*
3. When doing schoolwork for this class, I try to learn as much as I can and don't worry about how long it takes.
4. I have pushed myself hard to completely understand lessons in this class.
5. Overall, between homework, reading, and other class assignments, I work hard.



#### **Chapter 4: Fostering School Mission Coherence for Noncognitive-Skill Development**

Public agencies and bureaucracies often face issues that may undermine their effectiveness, two of which are principal-agent problems and goal conflict (Downs, 1967; Rainey, 2014; Wilson, 1991). To address these problems, public administration research often emphasizes fostering mission coherence within the organization — that is, clarifying and building agreement over organizational goals. In doing so, workers would be less motivated to shirk and, instead, pursue the organization's mission. One possible way to build mission coherence is through the use of personnel policies. Administrators may be able to compose a staff sharing the norms, values, and goals of the organization, minimizing the incidence and prevalence of principal-agent problems (DiIulio, 1994; Downs, 1967; Meier & O'Toole, 2006; Wilson, 1991).

According to principal-agent theory, less restrictive personnel policies would give principals more autonomy to shape their organizations through the composition of their agents (Brehm & Gates, 1999; Moe, 1984). In particular, less restrictive personnel policies may influence organizational effectiveness directly through the selection of the most talented individuals. Alternatively, less restrictive possibilities may indirectly enhance organizational effectiveness by enabling managers to foster mission coherence. I test the plausibility of this latter proposition. Specifically, I test whether the (a) level of influence managers have over hiring or (b) presence of procedural and institutional barriers to dismissing agents is associated with weaker organizational mission coherence. The aim is simply to find descriptive, not causal, evidence of such a relationship.

On the one hand, it is useful to simply check whether a relationship between mission coherence and personnel policies even exists because the prospects of using personnel policies to

influence mission coherence have not received much prior attention in the research literature and in practice (Berman et al., 2012; Hess & Kelly, 2007). Beyond that, finding such a relationship would refine the understanding of how personnel policies, mission coherence, and organizational effectiveness interact. Claims about how and why personnel policies theoretically affect organizations remain unsubstantiated absent such evidence, whereas the existence of such evidence raises the possibility that personnel policies could indirectly improve organizations by enabling managers to foster mission coherence. This conclusion would carry implications for practice and raise additional research questions such as evaluating the efficacy of various personnel policies on mission coherence.

The remainder of the paper is divided into five sections. The first section is a review of the literature regarding the relationship between personnel policies and mission coherence, with some particular attention to the field of educational administration. The second section describes the data and methods used in the analysis. I conduct this analysis in the context of the US public schools system. Because it operates much like other typical public agencies, the US traditional public school system provides an appropriate context (Chubb & Moe, 1988). I then present and discuss the results of the analysis in the third and fourth sections, respectively. The fifth section concludes.

## **Literature Review and Theoretical Framework**

### **The Importance of Organizational Mission Coherence**

Public administration research has often emphasized the importance of fostering mission coherence within an organization. Much of the literature has argued that the success and productivity of various public agencies is due, in large part, to having workers who are dedicated to the organizational mission (Bright, 2007; Chun & Rainey, 2005; Goodsell, 2011; Kim, 2005; Meier & O'Toole, 2006; Wolf, 1993). Mission coherence enhances organizational effectiveness

primarily by addressing principal-agent problems. It reduces the incidence and prevalence of shirking by public-sector workers. For instance, scholars have shown that having clear organizational goals and highlighting their importance motivates workers to prioritize organizational goals over personal goals (Paarlberg & Lavigna, 2010; Wright, Moynihan, & Pandey, 2012). Other researchers have shown that when worker and organizational goals align, workers tend to exhibit more job commitment (Bright, 2007; Wright & Pandey, 2008). Sometimes the desire to conform and follow organizational norms motivates workers to do their jobs well. Brehm and Gates (1999) argue that many public-sector workers refrain from shirking for solidary reasons. Workers do not want to be held in disdain by their peers for not following organizational norms and being less effective. This dynamic could be stronger in organizations that have well-defined ends and consensus over the means to achieve those ends. In other words, mission coherence may raise organizational effectiveness because it raises the salience of organizational norms that are conducive to success (see also Scott, 2014).

Fostering mission coherence primarily attempts to address principal-agent problems by appealing to the intrinsic motivation of public-sector workers, namely, their desire to fulfill the organizational mission by doing public service in and of itself. Research on public-sector workers suggests that they particularly are motivated by these types of intrinsic, prosocial impulses (Buelens & Van den Broeck, 2007; Georgellis, Iossa, & Tabvuma, 2011; Goodsell, 2011). For this reason, DiIulio (1994) has encouraged managers who desire to raise organizational effectiveness to seek out *principled agents*. Principled agents are

workers who do not shirk, subvert, or steal on the job even when the pecuniary and other tangible incentives to refrain from these behaviors are weak or nonexistent, and who often perform “thankless tasks” and make virtual “gifts” of their labor (p. 282).

This attitude among public-sector workers also has classically been referred to as public service motivation by Perry and Wise (1990). Notably, research suggests that public service motivation is tied to enhanced organizational performance (Bright, 2007; Kim, 2005; Naff & Crum, 1999). Given these findings, together with the fact that fostering mission coherence directly appeals to the intrinsic motivations of workers, it may not be surprising that mission coherence is also positively linked to organizational effectiveness.

Research in economics also suggests that worker productivity is increased when the worker's personal mission matches the organizational mission. Akerlof and Kranton (2005) model how worker motivations can be influenced by a sense of job identity. Workers may exert greater effort not only to earn greater material compensation (e.g., money, status) but also to live up to normative expectations based upon archetypal conceptions of the ideal worker. Citing an example of this dynamic from the US military, Akerlof and Kranton write:

[West Point] wish[es] to inculcate non-economic motives in the cadets so that they have the same goals as the U.S. Army. Alternatively stated, the goal of West Point is to change the identity of the cadets, so they will think of themselves, above all else, as officers in the U.S. army. They will feel bad about themselves—they will lose utility—if they fall short of the ideals of such an officer. This change in identity is a way to motivate employees, different than incentives from monetary compensation. (p. 9)

Similarly, Besley and Ghatak (2005) craft an economic theory explaining how matching between the worker's mission and his organizational mission mitigates principal-agent problems and enhances worker productivity. Using experimental economics, Carpenter and Gong (2016) have produced empirical evidence to support these economic theories.

Regarding schools, research in educational leadership suggests that administrators can improve school effectiveness through developing mission coherence. Schools that possess clearly defined and communicated goals typically have greater student achievement growth (Bryk et al., 2010; Hallinger & Heck, 1998; Teddlie & Reynolds, 2000). Likewise,

transformational leadership — an approach to school leadership that appeals to the intrinsic motivation of workers to serve the organizational mission — has been shown to improve school productivity (Leithwood & Jantzi, 2005; Leithwood & Sun, 2012). Evidence also suggests that schools with greater mission coherence tend to exhibit more characteristics that mediate school improvement, such as positive student behavior and higher student motivation (McShane, 2013). In summary, fostering mission coherence may be an important intermediate goal for raising organizational effectiveness for schools and other public agencies, improving public service.

### **Hiring and Dismissal as a Mission Building Mechanism**

The relevant task, then, is to determine how managers can foster mission coherence, even as they are faced with the need to address the potential for their workers to shirk (Moe, 1984; Rainey, 2014; Wilson, 1991). There are three useful strategies to mitigate this problem. First, managers could appeal to the extrinsic motivations of workers through rewards or sanctions. This approach might include regulating worker behaviors through monitoring or establishing procedural rules to restrain undesired behavior (Downs, 1967). Yet it is debatable how salient extrinsic sanctions and incentives are to public-sector workers (Alonso & Lewis, 2001; Buelens & Van den Broeck, 2007; Brehm & Gates, 1999; Georgellis, Iossa, & Tabvuma, 2011; Stazyk, 2013).

A second strategy for managing workers appeals to their intrinsic motivation. When hired, the goals of workers may not match the goals of the organization. In this case, managers could socialize their workers so that organizational and worker goals match. A basic training regimen for military recruits typically adopts this approach (Wilson, 1991). Transformational leadership is premised on this approach and also lessens the need for monitoring and the establishment of formal rules. After all, workers will already do what managers desire even absent those rules and socialization efforts. In fact, some researchers have demonstrated that

fostering shared values and commitments among agents is a more promising strategy than calling managers to exert more top-down control for improving the effectiveness of public agencies (Brehem & Gates, 1999; Meier & O'Toole, 2006).

As a third option for reducing shirking behaviors among their workers, managers could directly recruit principled agents who are loyal to and are already intrinsically motivated to achieve the organization's goals (DiIulio, 1994; Downs, 1967; Wilson, 1991). Selective recruitment and hiring of principled agents helps to establish goal consensus (i.e., mission coherence) throughout the larger organization (Kaufman, 1960). Indeed, this approach partially explains why law enforcement agencies often prefer to hire workers with prior military experience. Such workers presumably already share the values of their employers (Aamodt, 2004). Selecting workers whose values match those of the organization additionally lessens the need for managers to invoke the two previous strategies. That is to say, monitoring, establishing formal rules designed to solicit desired behavior, and socializing workers becomes less useful when organizational and worker goals already match (Brehem & Gates, 1999).

Despite the promise of using hiring or dismissal to build mission coherence, it is important to note some limits. Dismissal practices, in particular, have a crucial limit. Previous work demonstrates that dismissals can improve organizational performance, but only up to a certain point. Past some threshold, high rates of dismissals may have negative effects on organizations (Meier & Hicklin, 2008). Excessive dismissals might directly hinder organizational mission coherence by creating an uneasy work environment. Indeed, some research suggests that job insecurity is directly linked with lower organizational commitment (Battaglio, 2010; Sverke, Hellgren, & Haswall, 2002). Excessive dismissals may also indirectly harm organizational commitment and mission coherence through their effects on job dissatisfaction. Punitive

sanctions, such as the threat of dismissal, could lead to burnout, stress, low morale, and other indicators of job dissatisfaction, which are all tied to lower organizational commitment (Hargreaves & Fink, 2005; Kim & Lee, 2007; Wright & Pandey, 2008; 2011).

The use of personnel policies to build mission coherence and, ultimately, to improve organizational effectiveness is a crucial component to the New Public Management (NPM) or the Reinventing Government paradigms, which focus on results and productivity. Efforts to expand managerial discretion, to scale back procedural restrictions, to measure outcomes, and to use incentives to raise the effectiveness of public agencies are endemic to NPM and Reinventing Government (Hood 1991; Osborne & Gaebler, 1993). Many of these ideas are borrowed from private-sector management, and relaxing restrictions to managing personnel is one of them (Ban, Goldenberg, Marzotto, 1982). However, the NPM and Reinventing Government paradigms are not without reasonable criticisms (Wolf, 1997). Many formal rules governing personnel management of public agencies exist to protect against unwanted consequences such as political corruption and patronage. For instance, employment-at-will policies, where employers may dismiss employees for any reason, have gained popularity as a means to improve the efficiency of public agencies. Yet the gain in managerial discretion and expediency in personnel management comes at the loss of procedural protections for employees, which may also erode organizational commitment (Battaglio, 2010; Battaglio & Condrey, 2009; Kellough & Nigro, 2006; Selden, 2006).

### **Personnel Policies and Organizational Effectiveness: Mission Coherence as Mediator?**

Controversy over the NPM and Reinventing Government paradigms also presents itself within the public schools system. While some scholars have begun exploring how public school administrators can employ NPM ideas (Ritter, Maranto, Buck, 2009; Hess, 2013), others caution against the implementation of those ideas (Spillane, 2006; Hargreaves & Fink, 2005).

Nonetheless, reforms drawing upon the NPM and Reinventing government paradigms have been tried in some public school districts such as Washington DC and New York City. In addition to focusing on measuring teacher effectiveness with regular student testing and compensating teachers based on their performance, school administrators in these districts were given more flexibility to hire teacher candidates and to dismiss low-performing teachers. Interestingly, these school districts experienced dramatic improvements in student achievement after the adoption of these reforms.

Although it is unclear what exact factors caused the rapid improvement of student outcomes in New York City or Washington DC, there are two plausible explanations for how greater discretion over personnel policies may do so. First, greater discretion may enable administrators to hire more qualified teachers, as measured by their ability to raise student achievement. Simply changing the composition of teaching staff by replacing the lowest-performing teachers even with average teachers results in large student-achievement gains (Boyd et al., 2011; Hanushek, 2011; Jacob, 2011). This dynamic may have occurred in DC Public Schools. There, administrators implemented a new program, called IMPACT, where teachers could receive large bonuses for raising student achievement or face dismissal for failing to do so. Among other findings, analyses of IMPACT indicate that it improved the performance of teachers threatened with dismissal for low performance and induced other lower-performing teachers to voluntarily leave their jobs (Dee & Wyckoff, 2014). This evidence suggests that flexibility over personnel management, together with performance pay, could improve student outcomes by increasing average teacher quality. Greater managerial discretion over personnel policies in New York City also had a similar effect (Goertz, Loeb, & Wyckoff, 2011, Ouchi, 2009).



A second reason for why greater flexibility over personnel management improved Washington DC and New York City public schools is that it enabled administrators to recruit, hire, and retain more mission-driven, committed teachers (Maranto & Wolf, 2013). In other words, administrators were able to foster more mission coherence, which according to other research, improves organizational effectiveness (Bright, 2007; Chun & Rainey, 2005; Goodsell, 2011; Kim, 2005; Wright & Pandey, 2008). Bryk and Schneider (2003) have argued that hiring teachers who agree and fit with the school's mission, while dismissing those who do not, enables administrators to maintain a high level of mission coherence. This builds trust and commitment within schools and, ultimately, leads to improvement (see also Bryk et al., 2010; Egalite et al., 2014).

Public charter schools utilize personnel policies to foster mission coherence more commonly than traditional public schools. Charter schools are publically-financed schools that are granted a higher degree of autonomy than traditional public schools but held accountable by an authorizing agency to meet certain goals. Charter schools have typically used their autonomy to employ various strategies to recruit “mission-driven people that they [believe] would fit their schools and programs,” while dismissing those that do not (DeArmond, et al. 2012, p. 7; see also Merseth et al, 2009). Distinctive charter schools such as KIPP and YES Prep stay true to their unique missions and approaches to their everyday operations in part because of such personnel management practices (Furgeson et al., 2012; Maranto & Shuls, 2011). Notably, these types of charter schools are the most successful among other charter schools which are only slightly more effective than other public schools (Cheng et al., 2015, CREDO, 2013).

So, how do less restrictive personnel policies facilitate organizational improvement? Do such policies enable administrators to more easily compose a higher-quality teaching staff? Or

do they enable administrators to more easily foster mission coherence? Both possibilities are plausible and likely interact. And while the former possibility has empirical support (e.g., Jacob, 2011; Dee & Wyckoff, 2014), the latter question regarding the relationship between personnel policies and mission coherence is largely unexplored. Answering the aforementioned questions provides insight of theoretical and practical import as it may identify a mechanism by which to raise organizational effectiveness. From a theoretical standpoint, it is useful to not only test *if* personnel policies might improve organizational effectiveness, as suggested by the NPM paradigm, but also to know *why* those personnel policies would do so. This study fills this gap in the literature by examining whether a link between mission coherence and personnel policies even exists in the first place. If so, then the positive influence of personnel policies on organizational effectiveness, as experienced in Washington DC and New York City public schools, could be partially mediated by its influence on mission coherence.

From a more practical standpoint, school administrators may be interested in hiring based on mission-fit instead of ability to raise student test scores because it is often difficult to differentiate teachers based on the latter criteria. Research suggests that though school administrators, for example, are only able to identify the lowest-quality and highest-quality teachers (i.e., the top and bottom 10 to 20 percent), they cannot differentiate between the 60 to 80 percent of remaining teachers in the middle range of quality (Jacob & Lefgren, 2008). Hiring regulations also preclude administrators from identifying high-quality teachers. Signals intended to differentiate teacher quality, such as credentials, master's degrees, and years of experience, are all generally uncorrelated with teacher quality (Hanushek & Rivkin, 2006). Of course, these considerations assume that effectiveness can be easily measured. Wilson (1991) classifies schools as *coping agencies*, where outcomes are difficult to measure. Given concerns over the

validity and reliability of measures for teacher effectiveness, personnel decisions based upon effectiveness could be difficult to make and error-prone (Rothstein, 2009; but see Kane, 2014). Hence, exploring the link between personnel policies and the intermediate outcome of mission coherence, rather than the ultimate outcome of organizational effectiveness, would be useful for practical reasons, especially if managers of public agencies are consigned to hire based upon person-organizational fit due to poor alternative measures of quality.

Much can be learned even from a straightforward test for a relationship between personnel policies and mission coherence. More specifically, the following two hypotheses are tested:

H1: Organizations where managers have more influence over hiring workers exhibit higher levels of mission coherence.

H2: Organizations where managers face more barriers against dismissing workers exhibit lower levels of mission coherence.

In the next section, I describe the data and methods to test these hypotheses.

## **Research Methodology**

### **Data**

The Schools and Staffing Survey (SASS) consists of a series of questionnaires for school administrators and teachers. The general purpose of SASS is to provide a nationally-representative picture of US schools. Every four years, the US Department of Education surveys a cross-sectional, stratified random-sample of US primary and secondary schools. A wide range of contextual details are collected in SASS. On the teacher questionnaire, for instance, teachers are asked to describe their job duties, the way they run their classes, their educational background, the type of certification or professional training they have completed, their perceptions of school climate and working conditions, and their demographic characteristics.

Administrators report similar information on the administrator questionnaire. This information includes the way their schools are organized and governed; special programs, facilities, or services offered by the school; particular policies (e.g., teacher merit pay, admissions requirements) implemented within their school; their educational and professional history; and demographic characteristics of the student body.

Data from this study comes from the 2011-2012 version of SASS. Private and public schools are included in SASS, but this analysis is limited to public schools, including charter schools. In all, information from administrators and 37,000 teachers<sup>27</sup> belonging to over 6,400 public schools across the United States are used in the analysis.

### **Measure of Mission Coherence**

The school mission coherence index is constructed based upon a subset of items on the teacher questionnaire of SASS. Again, mission coherence is defined as the degree to which members in a public agency agree with and act in accordance with the organizational mission. Regarding schools, mission coherence among teachers entails the extent to which they agree over the purpose of schooling and work in unison to achieve that end. This requires collaboration, coordination of curriculum, and consistent enforcement of school rules (Bryk et al., 2010). Mission coherence is captured by the following four items: (a) Most of my colleagues share my beliefs and values about what the central mission of the school should be; (b) I make a conscious effort to coordinate the content of my courses with that of other teachers; (c) Rules for student behavior are consistently enforced by teachers in this school, even for students who are not in their classes; and (d) There is a great deal of cooperative effort among the staff members. These

---

<sup>27</sup>A random, stratified sample of teachers within each school is surveyed in SASS. The number of teachers sampled per school ranges from 1 to 18 and is proportional to the size of the school. An average of 6 teachers are represented in each school.

items mimic items used in other work to measure facets of mission coherence (Chun & Rainey, 2005; Wright & Pandey, 2011). Teachers respond to these items on a 4-point Likert scale.

Responses to each question are numerically coded and averaged for each teacher. Each of these averages is then averaged across teachers within a particular school. Higher values denote higher degrees of mission coherence. The mission coherence scale also possesses a Cronbach's alpha equal to 0.75, suggesting that the responses to four items on the scale are well-correlated with each other. A factor analysis also confirms that all items load onto a single factor.

### **Measure of Administrator Influence over Hiring Decisions**

The administrator questionnaire in SASS asks school administrators to rate the amount of influence they have over various organizational decisions and processes, such as setting academic curriculum or budget. One of these items asks administrators to rate, on a four-point scale, their influence over teacher-hiring decisions. This rating is used as the independent variable of interest to test H1. Higher values indicate greater administrator influence.

### **Measure of Barriers to Dismissal**

An index of formal barriers to dismissing teachers (henceforth referred to as the formal barriers index) is constructed based upon items contained in the administrator questionnaire. This questionnaire presents school administrators with a list of ten potential barriers to dismissing poor-performing teachers. For each potential barrier, the administrator indicates whether it is a barrier that he or she faces or does not face. Seven of the barriers are formal barriers (i.e., bureaucratic rules, organizational characteristics) that typically receive attention in educational research and policy debate. The items that capture formal barriers are: (a) personnel policies, (b) termination decisions are not upheld, (c) length of time required for termination process, (d) effort required for documentation, (e) tight deadlines for completing documentation, (f) tenure, and (g) teacher associations or unions. The three remaining items are excluded from the analysis

because they appear to capture non-formal barriers. They are (a) dismissal is too stressful and/or uncomfortable (b) difficulty in finding suitable replacements, and (c) resistance from parents.

Summary statistics for each item are displayed in the first two columns of Table 1.

Table 1

*Summary Statistics and Two-Factor Loadings of Barriers-to-Dismissal Items*

Item	Mean	Standard Deviation	Formal- Barriers Factor Loading	Non-Formal- Barriers Factor Loading
Personnel policies	0.519	0.500	0.675	0.048
Termination decisions are not upheld	0.199	0.400	0.443	0.233
Length of time required for termination process	0.611	0.487	0.712	0.087
Effort required for documentation	0.654	0.476	0.670	0.192
Tight deadlines for completing documentation	0.321	0.467	0.495	0.319
Tenure	0.696	0.460	0.633	-0.150
Teacher associations or unions.	0.614	0.487	0.648	-0.141
Dismissal is too stressful and/or uncomfortable	0.114	0.318	0.185	0.536
Difficulty in finding suitable replacements	0.135	0.342	-0.058	0.672
Resistance from parents	0.039	0.194	0.030	0.657

Notes: Means and standard deviations are based on unweighted data. All variables are dummy variables. A principal components factor analysis and varimax rotation were used to calculate these factor loadings. In general, a minimum factor loading of 0.32, as suggested by Tabachnick and Fidell (2001), was necessary to consider an item to load onto a particular factor. The item “tight deadlines for completing documentation” only marginally loaded onto the non-formal barriers factor. But given that it was hypothesized to be a formal barrier and significantly loaded onto the formal-barriers factor, it was ultimately considered a formal-barriers factor. Rerunning the analysis while omitting responses for “tight deadlines for completing documentation” and “termination decisions are not upheld,” which loaded less weakly onto the formal barriers factor, does not substantively change the results.

The Cronbach's alpha for the ten items is approximately 0.72, suggesting that administrators who indicate that they face a certain barrier also tend to indicate that they face other barriers associated with it. Furthermore, an exploratory factor analysis was used to confirm that the items loaded onto the expected factors. As shown in Table 1, the seven and three items load onto the formal and non-formal barriers factors, respectively. These results lend confidence that the seven items selected for a formal barriers index capture the extent of formal barriers that administrators face.

The formal barriers index is then constructed by computing a weighted average of the seven items that capture formal barriers.<sup>28</sup> Weights for each of the items are derived from the factor analysis. The seven items that capture formal barriers are used to estimate a factor model to estimate the underlying formal-barriers factor for each observation in the sample.<sup>29</sup> Higher values indicate a greater presence of barriers. The formal barriers index is the independent variable of interest in testing H2. Summary statistics for the formal barriers index as well as other key variables of interest (i.e., administrator influence over hiring, and mission coherence)

---

<sup>28</sup> This weighted average is computed to facilitate interpretation of the results. Unlike other variables in the analysis, the items used to construct the formal barriers index are yes/no questions. Had the weighted average not been created, then the formal barriers index would simply be a count of the number of barriers against dismissing teachers that administrators faced. Regression results would consequently communicate the change in mission coherence for each additional barrier faced. However, not all barriers have equal influence on mission coherence, obscuring the relationship between the two variables. On the other hand, using a weighted average based on the factor analysis creates a latent and more general variable, which captures the difficulty that formal barriers impose on administrators when they want to dismiss teachers.

<sup>29</sup> The equation used to estimate the underlying formal barriers factor is: Underlying Formal Barriers Factor =  $0.24068 \times (\text{Personnel policies}) + 0.014014 \times (\text{Termination decisions are not upheld}) + 0.026293 \times (\text{Length of time required for termination process}) + 0.22900 \times (\text{Effort required for documentation}) + 0.15364 \times (\text{Tight deadlines for completing documentation}) + 0.25274 \times (\text{Tenure}) + 0.25408 \times (\text{Teacher associations or unions})$ . Each barrier takes on the value equal to one if the administrator indicates that he faces that barrier. Else, the barrier equals 0.

are shown in Table 2. Table 2 also lists the control variables in the analysis, which I describe in the next section.

Table 2

*Summary Statistics for Independent Variables*

Variable	Sample Average	Standard Deviation	Minimum	Maximum
<i>Key Independent and Dependent Variables</i>				
Mission Coherence Index	3.10	0.39	1.25	4.00
Administrator Influence on Hiring	3.84	0.48	1.00	4.00
Formal Barriers Index	0.84	0.47	0.00	1.53
<i>Principal Characteristics</i>				
Years of Experience as Principal	7.41	6.51	0.00	45.00
Years of Experience as Principal at Current School	4.37	4.62	0.00	45.00
Male <sup>a</sup>	0.58	0.49		
Age	48.03	8.92	23.00	80.00
Principal Race/Ethnicity				
American Indian or Native Alaskan	0.01	0.08		
Asian	0.01	0.08		
Black	0.08	0.27		
Hispanic	0.05	0.21		
More than one Race	0.04	0.19		
White	0.83	0.38		
Educational Background				
No Graduate Degree	0.04	0.20		
Graduate Degree in Educational Field	0.93	0.17		
Graduate Degree outside of Educational Field	0.03	0.25		



Table 2 (Cont.)

Variable	Sample Average	Standard Deviation	Minimum	Maximum
<i>School Characteristics</i>				
Percent of Students Eligible for Free or Reduced-Price Lunch	45.39	28.36	0.00	100.00
Percent of English Language Learners	5.60	11.49	0.00	100.00
Percent of Students Classified as Special Education	14.00	13.88	0.00	100.00
Percent of Students from Racial/Ethnic Minority Backgrounds	30.27	30.17	0.00	100.00
School Enrollment (natural log)	6.23	0.83	0.69	9.21
Charter School	0.05	0.22		
Locale				
Urban	0.20	0.40		
Rural	0.38	0.48		
Town	0.17	0.37		
Suburb	0.25	0.44		
School Level				
Elementary School	0.30	0.46		
Middle School	0.26	0.44		
High School	0.33	0.47		
Combined Grade Configurations	0.11	0.31		
<i>Teacher Characteristics (Aggregated to School-Level)</i>				
Years of Experience	13.95	5.54	0.00	48.00
Male	0.29	0.26		
Age	42.74	6.75	22.00	79.00
Master's Degree	0.55	0.32		

Notes: Summary statistics are calculated without using sampling weights. Variables for which no minimum or maximum values are specified are dummy variables.

## Empirical Strategy

Ordinary least squares regression was used to test the hypotheses that schools where administrators face more barriers against dismissing teachers or have less influence over hiring decisions tend to exhibit lower levels of mission coherence. Sampling weights are used to ensure that the results generalize to a nationally-representative sample of US schools. Moreover, bootstrapping procedures are used estimate robust standard errors. The following empirical model is estimated:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 B_i + \beta_3 A_i + \beta_4 S_i + \mu_i.$$

In the model,  $M_i$  is the school mission coherence index for school  $i$ . The variables of interest are  $H_i$ , the influence that administrators have over teacher hiring, and  $B_i$ , the barriers-to-dismissal index. The random-error term is denoted by  $\mu_i$ .

$A_i$  and  $S_i$  are vectors of administrator and school characteristics, respectively. These variables serve as control variables. It is important to control for these variables as they might be correlated with the independent variables of interest. Failing to control for them would consequently bias the estimates. Although the aim of the study is not to uncover causal relationships, controlling for other variables provides a more detailed picture of the relationships between the variables of interest. Results would convey the link between the variables of interest, independent of other possible confounding factors.

For example, control variables for administrator characteristics include years of experience, gender, age, race, and educational background. It is possible that a particular type of administrator would tend to work in schools with particular hiring and dismissal restrictions. Likewise, school control variables include the percentage of students who are classified as English-language learners, require special education, come from racial minority backgrounds, and qualify for free- or reduced-priced lunch (a proxy for poverty). The school's student

enrollment size and indicator variables for whether the school is a charter school, the grade levels that it serves, and the urbanicity of its locale are also included. A final set of school control variables captures the characteristics of teachers at each school. These variables include average age and years of experience for teachers at the school. Also included are percentages of teachers at the school who are white, possess a master's degree, and are male.<sup>30</sup> Like the controls for administrator characteristics, the school characteristics are likely correlated with the independent variables of interest, and controlling for them more accurately captures the relationship between those independent variables of interest and mission coherence.

Before presenting the results of this analysis, it is worthwhile to note that the formal barriers index and the amount of influence that an administrator has over teacher hiring are generated from information supplied by the administrator, whereas the mission-coherence index is generated from information supplied by teachers. Because these two variables are not based upon responses provided by the same individual, these variables provide independent information about the school. This feature lends more confidence that the correlation between the mission coherence and formal barriers indices is due to the relationship between those two variables, instead of the response bias that would potentially occur if the same respondent provided information for these two indices. Controlling for administrator-, teacher-, and school-level characteristics additionally assuages this concern. The distinction between administrator and teacher responses also yields a more compelling case that the personnel policies faced by managers are linked to mission coherence among the workers in the organization.

---

<sup>30</sup> SASS does not report the raw values of these percentages. These percentages are estimated using a weighted average of these characteristics for teachers within each school who responded to SASS. Weights are based upon the teacher's probability of selection into the SASS sample.

## Results

To reiterate, this analysis explores the relationship between mission coherence and two distinctive features of personnel policies, namely, hiring and dismissal. The results indicate that schools in which administrators have greater influence over teacher-hiring decisions also have greater levels of mission coherence. As shown in column 1 of Table 3, increasing the amount of influence that administrators have over hiring decisions by one scale point is associated with an increase of mission coherence by 0.04 scale points, or 9 percent of a standard deviation ( $p < 0.05$ ).

Table 3

### *Coefficient Estimates of Empirical Model*

Dependent Variable: Mission Coherence Index	(1)	(2)
Administrator Influence on Hiring	0.035* (0.015)	0.037* (0.015)
Formal Barriers Index	-0.033* (0.014)	-0.024 (0.015)
Charter		0.007 (0.036)
Charter * Administrator Influence on Hiring		-0.048† (0.027)
Charter * Formal Barriers Index		-0.030 (0.049)
$R^2$	0.191	0.192

Notes: N=6,422. Model also control for all other covariates listed in Table 2. Standard errors in parenthesis. Sampling weights are used so that results generalize to a nationally-representative sample of US schools. Robust standard errors are estimated using bootstrapping and shown in parenthesis. State fixed effects are included in all models. \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.1$

Turning to the second feature of personnel policies — those that concern dismissal, results indicate that schools in which administrators face a greater number of barriers against dismissing mismatched teachers also tend to have less mission coherence. Increasing the formal barriers index by one scale point decreases mission coherence by about 0.03 scale points, or 8 percent of a standard deviation ( $p < 0.05$ ). Note too that the administrator influence over hiring and the extent of formal barriers against dismissal are estimated jointly. Thus, the link between

mission coherence and one of the two personnel policies is net of the relationship between mission coherence and the other personnel policy.

Column 2 presents results for a test of whether charter schools are systematically different than traditional public schools regarding the relationships between mission coherence, the formal barriers index, and administrator influence over hiring. Before discussing these results, it is worthwhile to point out that the data generally support the claim that charter school administrators have slightly more influence over hiring decisions and face drastically fewer barriers to dismissing ineffective teachers. Table 4 displays regression results where the formal barriers index and the administrator influence on hiring variable are used as dependent variables, while a charter dummy variable as well as other administrator and school control variables are used as independent variables. Charter school administrators report fewer barriers to dismissing teachers. The difference between charters and other public schools, with respect to the barriers-to-dismissal index, is slightly over one standard deviation or about one-half point ( $p < 0.001$ ). Administrators in charter schools also appear to have more flexibility over hiring decisions, but just beyond conventional levels of statistical significance ( $p = 0.101$ ).

Turning to Column 2 of Table 3, the analysis yields mixed evidence that greater influence over personnel decisions helps charter-school administrators build mission coherence. First, consider the ability of principals to influence hiring decisions. The sum of the coefficients for the administrator influence on hiring variable and the associated charter interaction term is not statistically different from zero ( $p = 0.90$ ). Greater influence over hiring among administrators of other public schools, however, remains linked to greater mission coherence, as indicated by the coefficient estimate of 0.37 ( $p < 0.05$ ). Turning to teacher dismissal, charter-school administrators who face fewer barriers to dismissing teachers also tend to have schools with more mission

coherence. Summing the coefficients for the barriers-to-dismissal index and the respective interaction term, a one-point increase in the barriers-to-dismissal index is associated with a 0.05-point decrease in the mission coherence index in public charter schools, or about 13 percent of a standard deviation ( $p < 0.05$ ). Among other public schools, the presence of barriers to dismissing teachers is negatively related to mission coherence, but the relationship is not significant at conventional levels ( $p = 0.12$ ).

Table 4

*Personnel Policy Differences between Charter and other Public Schools*

	Dependent Variable	
	Barriers-to-D dismissal Index	Administrator Influence on Hiring
Charter School	-0.494 (0.041) $p < 0.001$	0.062 (0.037) $p = 0.101$
Observations	6,475	6,422
R <sup>2</sup>	0.173	0.126

Notes. Coefficients depict differences between charter schools and other traditional public schools. Standard errors in parenthesis; p-values shown below. Ordinary least squares regression was used to estimate the coefficients. Models control for all administrator and school-level characteristics as in other models. State fixed effects included. Sampling weights are used so that results generalize to a nationally-representative sample of US schools. Robust standard errors are estimated using bootstrapping and shown in parenthesis.

### Discussion

The results generally lend credence to the two proposed hypotheses in this study. First, schools where administrators have more influence over hiring workers exhibit higher levels of mission coherence, confirming the first hypothesis (H1). This is not the case, however, for charter schools, where influence over hiring decisions is uncorrelated with mission coherence. This study also produced supporting evidence for the second hypothesis (H2): Schools where there are more barriers to dismissing teachers tend to have less mission coherence. The

magnitude of this relationship is more pronounced among charter schools than other public schools. Such evidence supports long-standing and mostly untested theories that personnel policies within an organization influence the degree of mission coherence within the organization. Specifically, the extent to which a manager is able to hire and to dismiss workers could possibly influence his or her potential to craft staff with a mission that matches the organizational mission (Brehm & Gates, 1999; Bryk & Schnieder, 2003; Chubb & Moe, 1988; DeArmond et al., 2012; Downs, 1967; Kaufman, 1960; Wilson, 1991).

One must be careful, though, not to overstate the findings. There are two reasons for this. First, the research design used in this study only allows for a descriptive interpretation. Other unobservable characteristics about the administrator, school, or teachers that are associated with having more barriers in dismissing teachers or the administrator's influence over hiring decisions may also be associated with or cause changes in mission coherence. It is possible that administrators who are adept at fostering mission coherence are also adept at sidestepping formal barriers to dismissing teachers or have political savvy to influence hiring decisions.

Second, the magnitude of the relationships between mission coherence and flexibility in hiring or dismissing teachers is relatively small. Findings indicate that a one-standard deviation change in administrator discretion over hiring or dismissing teachers amounts to about a four-percent change in mission coherence. Nevertheless, this relationship of a small magnitude should not be overlooked. The claim that personnel policies can be used to foster mission coherence within an organization has received little empirical attention. Detecting this relationship is important. Up until now, some research has established that organizational effectiveness is tied to mission coherence (Bright, 2007; Chun & Rainey, 2005; Downs, 1967; Goodsell, 2011; Kim, 2005; Wilson, 1991; Wolf, 1993). However, less is known about how administrators can foster

mission coherence in practice. Socializing workers, as advocated by transformational leadership, or supervising them may be useful (Downs, 1967; Leithwood & Jantzi, 2005; Leithwood & Sun, 2012; Wilson, 1991). Yet a third possibility of recruiting, hiring, and recruiting workers who fit with the organizational mission may be even more promising (e.g. Egalite et al., 2014). After all, such workers do not need to be supervised or socialized as much. Findings from this study suggest that this third way is a viable approach. Although the magnitude of the relationship between personnel policies and mission coherence is relatively small, the very existence of the relationship, at minimum, warrants more research attention.

Likewise, scholars have argued that less restrictive personnel policies explain much of the success experienced in New York and Washington DC public schools. The mechanisms behind this relationship, however, are less clear. Although it has been established that administrators were enabled to dismiss lower-quality teachers and retain higher-quality ones, school improvement may also have stemmed from enabling administrators to foster greater mission coherence (Dee & Wyckoff, 2014; Goertz, Loeb, & Wyckoff, 2011; Maranto & Wolf, 2013). Specifically, less-restrictive personnel policies may have enabled administrators to compose a teaching staff of workers whose missions matched that of the school. The results of this present study, at the very least, provide empirical support for this possibility, even if the coefficient estimates are relatively small. In other words, less restrictive personnel policies not only directly impact organizational effectiveness by enabling managers to hire more talented workers. Less restrictive personnel policies may also indirectly impact organizational effectiveness by enabling managers to foster mission coherence. Making these distinctions refines the theoretical mechanisms by which organizations can improve.



Thus, additional examination of the relationship between personnel policies, mission coherence, and organizational effectiveness are in order. Utilizing experimental design would be useful for establishing causal relationships. But even when experimental design is not feasible, using better data to test the relationship between personnel policies, mission coherence, and organizational effectiveness would be worthwhile. SASS, though it provides a rich data set of a nationally-representative sample of US schools, has not been designed with the intention of conducting such a test. Student test scores or other measures of educational outcomes are not included in SASS. More detailed measures of, for example, the administrator's influence over hiring decisions would also be useful. It is not unreasonable to posit that an analysis that captures additional aspects of the personnel management process (e.g., recruitment, professional training, promotions, compensation arrangements) may further disentangle the relationship between mission coherence and the restrictiveness of personnel policies. A more comprehensive and nuanced measure of mission coherence would be useful as well. In other research, measures of value congruence (Wright & Pandey, 2008) or goal clarity (Jung & Rainey, 2011; Wright & Pandey, 2011) are used to capture distinct aspects of organizational mission.

Measures of mission coherence should also incorporate a wider range of stakeholders. SASS, in particular, is unable to measure the alignment of school mission between administrators, teachers, and families. Yet despite data limitations in SASS and resulting models with relatively lower  $R^2$  statistics, it is remarkable that theoretically predicted relationships between personnel policies and mission coherence are detected.

Finally, it is important to explore potential negative effects of less restrictive personnel policies. Excessive dismissal and fewer worker protections may impose other costs, which need to be weighed against realized benefits. Indeed, NPM and the Reinventing Government

paradigms are not without concerns and limitations (Battaglio, 2010; Battaglio & Condrey, 2009; Kellough & Nigro, 2006; Meier & Hicklin, 2008; Selden, 2006). Such inquiry cannot be overlooked as they carry nontrivial implications for policy and practice.

### **Conclusion**

Public administration research has argued that building mission coherence facilitates organizational effectiveness. The important question, therefore, is how managers of public agencies can be best equipped or empowered to foster mission coherence in their organizations. This study suggests one of the potential ways forward, namely, through personnel policies. In particular, providing managers with the wherewithal to compose a particular staff may help to foster mission coherence in the organization (Hess, 2013; Hess & Kelly, 2007). As this study demonstrates with a nationally representative sample of US public schools, schools in which administrators possess less influence over hiring decisions or face more barriers to dismissing mismatched teachers also tend to exhibit lower levels of mission coherence. Such inquiry is lacking in the field of public administration, despite its important implications for researchers and practitioners who wish to improve public agencies.

## References

- Aaomodt, M.G. (2004). *Research in law enforcement selection*. Boac Raton, FL: Brown Walker Press.
- Akerlof, G.A., & Kranton, R.E. (2005). Identity and the economics of organizations. *The Journal of Economic Perspectives*, 19(1), 9-32.
- Alonso, P., & Lewis, G. (2001). Public service motivation and job performance: Evidence from the federal sector. *American Review of Public Administration*, 31(4), 363–380.
- Ban, C., Goldenberg, E.N., & Marzotto, T. (1982). Firing the unproductive employee: Will civil service reform make a difference? *Review of Public Personnel Administration*, 2(2), 87-100.
- Battaglio, R.P. (2010). Public service reform and motivation: Evidence from an employment at-will environment. *Review of Public Personnel Administration*, 30(3), 341-363.
- Battaglio, R.P., & Condrey, S.E. (2009). Reforming public management: Analyzing the impact of public service reform on organizational and managerial trust. *Journal of Public Administration Research and Theory*, 19(4), 689-707.
- Berman, E.M., Bowman, J.S., West, J.P., Van Wart, M.R. (2012). *Human resource management in public service: Paradoxes, processes, and problems (4th Ed.)* Los Angeles, CA: SAGE Publications.
- Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, 95(3), 616-636.
- Boyd, D., Lankford, H., Loeb, S., Wyckoff, J. (2010). Teacher layoffs: An empirical illustration of seniority versus measures of effectiveness. *Education Finance and Policy*, 6(3), 439-454.
- Brehm J., & Gates, S. (1999). *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, MI: University of Michigan Press.
- Bright, L. (2007). Does person–organization fit mediate the relationship between public service motivation and the job performance of public employees. *Review of Public Personnel Administration*, 27(4), 361–379.
- Bryk A.S., & Schneider B. (2003). Trust in Schools: A Core Resource for School Reform. *Educational Leadership*, 60(6): 40-45.
- Bryk A.S., Sebring P.B., Allensworth E., Luppescu S., & Easton J.Q. (2010). *Organizing Schools for Improvement: Lessons from Chicago*. Chicago, IL: The University of Chicago Press.

- Buelens, M., & Van den Broeck, H. (2007). An analysis of differences in work motivation between public and private sector organizations. *Public Administration Review*, 67(1), 67–74.
- Carpenter J., & Gong, E. (2016). Motivating agents: How much does mission matter? *The Journal of Labor Economics*, 34(1), 211-236.
- Cheng, A., Hitt, C., Kisida, B., Mills, J.N. (2015). *The impact of "No Excuses" charter schools on academic achievement*. National Center for the Study of Privatization in education. New York: Teachers College, Columbia University.
- Chubb, J., & Moe, T. (1988). Politics, Markets, and the Organization of Schools. *The American Political Science Review*, 82(4), 1065-1087.
- Chun, Y.H., & Rainey, H.G. (2005). Goal ambiguity and organizational performance in U.S. federal agencies. *Journal of Public Administration Research and Theory*, 24(2), 529-557.
- CREDO. (2013). National Charter School Study 2013. Stanford, CA. Stanford University.
- DeArmond M., Gross B., Bowen M., Demeritt A., & Lake R. (2012). *Managing Talent for School Coherence: Learning from Charter Management Organizations*. Seattle, WA: Center on Reinventing Public Education.
- Dee, T., & Wyckoff, J. (2013). *Incentives, Selection, and Teacher Performance: Evidence from Impact* (NBER Working Paper No. 19529). Cambridge, MA: National Bureau of Economic Research.
- DiIulio, J.D. (1994). Principled Agents: The cultural Bases of Behavior in a Federal Government bureaucracy. *Journal of Public Administration Research and Theory*, 4(3), 277-318.
- Downs, A. (1967). *Inside Bureaucracy*. Boston, MA: Little, Brown and Company.
- Egalite, A.J., Jensen, L.I., Stewart, T., Wolf, P.J. (2014). Finding the right fit: Recruiting and retraining teachers in Milwaukee Choice Schools. *Journal of School Choice*, 8(1), 113-140.
- Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., Teh, B., Veritsky-Savitz, N., Bowen, M., Demeritt A., Hill, P., & Lake, R. (2012). *Charter School Management Organizations: Diverse Strategies and Diverse Student Impacts*. Washington, DC. Mathematica Policy Institute
- Georgellis, Y., Iossa, E., & Tabvuma, V. (2011). Crowding out intrinsic motivation in the public sector. *Journal of Public Administration Research and Theory*, 21(3), 473–493.
- Goodsell, C. T. (2011). *Mission mystique: belief systems in public agencies*. Washington, DC: CQ Press.

- Goertz M., Loeb S., & Wyckoff J. (2011). Recruiting, Evaluating, and Retaining Teachers. In J.A. O'Day, C.S. Bitter, & L.M. Gomez (Eds.), *Education reform in New York City: Ambitious Change in the Nation's Most Complex School System* (pp. 157-180). Cambridge, MA: Harvard Education Press.
- Hallinger P., & Heck R.H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hanushek, E. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30, 466-479.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher Quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1052-1078). Amsterdam, The Netherlands: Elsevier B.V.
- Hargreaves, A., & Fink, D. (2005). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- Hess, F.M. (2013). *Cage-busting leadership*. Cambridge, MA: Harvard Education Press.
- Hess F.M., & Kelly, A. (2007). Learning to Lead: What Gets Taught in Principal-Preparation Programs. *Teachers College Record*, 109(1): 244-274.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3-19.
- Jacob, B. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, 33(4), 403-434.
- Jung, C.S., & Rainey, H.G. (2011). Organizational goal characteristics and public duty motivation in U.S. federal agencies. *Review of Public Personnel Administration*, 31(1), 28-47.
- Kane, T. (2014). *Do value-added estimates identify causal effects of teachers and schools?* Washington, DC: Brookings Institution.
- Kaufman, H. (1960). *The forest ranger: A study in administrative behavior*. Baltimore, MD: Johns Hopkins Press.
- Kellough, J.E., & Nigro, L.G. (2006). Dramatic reform in the public service: At-will employment and the creation of a new public workforce. *Journal of Public Administration Research and Theory*, 25(1), 447-466.
- Kim, S. (2005). Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research and Theory*, 15(2), 245-261.
- Kim, S.E., & Lee, J.W. (2007). Is mission attachment an effective management tool for employee retention? An empirical analysis of nonprofit human services agency. *Review of Public Personnel Administration*, 27(3), 227-248

- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, 4(3), 177–199.
- Leithwood, K., & Sun, J. (2012). The nature and effects of transformational school leadership: A meta-analytic review of unpublished research. *Educational Administration Quarterly*, 48(3), 387-423.
- Maranto, R., & Shuls J. (2011). Lessons from KIPP Delta. *The Phi Delta Kappan*, 93(3), 52-56.
- Maranto, R., & Wolf, P.J. (2013). Cops, teachers, and the art of the impossible: Explaining the lack of diffusion of innovations that make impossible jobs possible. *Public Administration Review*, 73(2), 230-240.
- McShane, M.Q. (2013). Does mission matter? Examining the Role of Organizational Mission and Religious Identity in Schools Participating in the Milwaukee Parental Choice Program. (Unpublished doctoral dissertation). University of Arkansas, Fayetteville, AR.
- Meier, K.J., & O'Toole, L.J. (2006). *Bureaucracy in a Democratic State: A Governance Perspective*. Baltimore, MD: Johns Hopkins University Press.
- Meier, K.J., & Hicklin, A. (2008). Employee Turnover and Organizational Performance: Testing a Hypothesis from Classical Public Administration. *Journal of Public Administration Research and Theory*, 18(4): 573-590.
- Merseth, K., Cooper, K., Roberts, J., Tieken, M.C., Valant, J., & Wynne, C. (2009). *Inside Urban Charter Schools: Promising Practices and Strategies in Five High-Performing Schools*. Cambridge, MA: Harvard Education Press.
- Moe, T.M. (1984). The new economics of organization. *American Journal of Political Science*, 28(4), 739-777.
- Naff, K.C., & Crum, J. (1999). Working for America: Does public service motivation make a difference. *Review of Public Personnel Administration*, 19(4), 5-16.
- Osborne, D. & Gaebler, T. (1993). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York, NY: Plume.
- Ouchi, W.G. (2009). The secret of TSL: *The revolutionary discovery that raises school performance*. New York, NY: Simon and Schuster.
- Paarlberg, L. E., & Lavigna, B. (2010). Transformational leadership and public service motivation: Driving individual and organizational performance. *Public Administration Review*, 70(5), 710–718.
- Perry, J. L., & Wise L.R. (1990). The motivational bases of public service. *Public Administration Review* 50(3), 367-373.

- Rainey, H.G. (2014). *Understanding and Managing Public Organizations, 5th ed.* San Francisco, CA: Jossey-Bass.
- Ritter, G.W., Maranto, R., & Buck, S. (2009). Harnessing private incentives in public education. *Review of Public Personnel Administration, 29*(3), 249-269.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy 4*(4), 537-571.
- Scott, W. R. (2014). *Institutions and Organizations: Ideas, Interests, and Identities.* Los Angeles, CA: Sage Publications, Inc.
- Selden, S.C. (2006). The impact of discipline on the use and rapidity of dismissal in state governments. *Review of Public Personnel Administration, 26*(4), 335-355.
- Spillane, J. P. (2006). *Distributed Leadership.* San Francisco, CA: Jossey-Bass.
- Stazyk, E.C. (2012). Crowding out public service motivation? Comparing theoretical expectations with empirical findings on the influence of performance-related pay. *Review of Public Personnel Administration, 33*(3), 252-274.
- Sverke, M., Hellgren, J., & Naswall, K. (2002). No security: A meta-analysis and review of job insecurity and its consequences. *Journal of Occupational Health Psychology, 7*(3), 242-264.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using Multivariate Statistics, 5th Edition.* Boston, MA: Allyn and Bacon.
- Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research.* New York, NY: Routledge.
- Wilson, J.Q. (1991). *Bureaucracy: What government agencies do and why they do it.* New York, NY: Basic Books.
- Wolf, P.J. (1993). A case survey of bureaucratic effectiveness in U.S. Cabinet Agencies: Preliminary Results. *Journal of Public Administration Research and Theory, 3*(2), 161-181.
- Wolf, P.J. (1997). Why must we reinvent the federal government? Putting historical developmental claims to the test. *Journal of Public Administration Research and Theory, 7*(3), 353-388.
- Wright, B. E., Moynihan, D. P., & Pandey, S. K. (2012). Pulling the levers: Transformational leadership, public service motivation, and mission valence. *Public Administration Review, 72*(2), 206–215.

- Wright, B. E., & Pandey, S. K. (2008). Public service motivation and the assumption of person-organization fit: Testing the mediating effect of value congruence. *Administration and Society*, 40(5), 502–521.
- Wright, B. E., & Pandey, S. K. (2011). Public organizations and mission valence: When does mission matter? *Administration and Society*, 43(1), 22–44.



## Chapter 5: Conclusion

In recent decades, improving student achievement on standardized tests of cognitive ability has been the primary goal in US education policy. Schools and other educational interventions have been assessed based upon their impacts on student cognitive ability.

On one hand, the focus on improving student cognitive ability is not wholly misplaced. Human capital research underscores the important role that cognitive skills play in influencing educational attainment, labor-market outcomes, and other indicators of well-being in life (Becker, 1964; Hanushek, 2011). However, scholars and policymakers have arguably overemphasized this outcome measure. More recent research from psychology and economics demonstrates that noncognitive skills play an independent role in determining the long-run well-being of individuals (Almlund et al, 2011; Borghans et al. 2008; Borghans, ter Weel & Weinberg, 2008; Bowles, Gintis, & Osborne, 2001; Deke & Haimson 2006; Heckman, 2000; Heckman & Rubinstein 2001; Heckman & Kautz, 2012; Heckman, Stixrud, & Urza 2006; Kaestner & Callison 2011; Lindqvist & Vestman 2011; Lundborg, Nystedt, & Rooth, 2014; Mueller & Plug, 2006).

It is, therefore, essential to understand how schools and other education interventions affect students in a variety of ways beyond achievement. If policymakers desire to improve long-run wellbeing for students, it also behooves them to better understand how noncognitive skills are developed and formed. It is for these reasons that in this dissertation I have investigated the role that teachers play in affecting student noncognitive skills.

Psychologists have maintained that learning is social. Individuals learn behavioral norms and values by observing role models whom they seek to emulate or not to emulate (Bandura, 1977; Bandura & Walters, 1963). Philosophers since Aristotle have postulated that the development of virtue and character occurs in the same manner (MacIntyre, 2007). In chapters 2

and 3 of this dissertation I empirically test this theory using novel behavioral measures of conscientiousness that are based upon survey effort.

In chapter 2, I use panel data and techniques to show that students experience declines in conscientiousness in years when they are taught by teachers who exhibit less conscientiousness. Assuming that students are not systematically sorted to teachers with varying degrees of conscientiousness based upon time-varying student characteristics, the relationship between student and teacher conscientiousness can be interpreted as causal. In chapter 3, I corroborate these findings in an analysis that possesses greater internal validity. Using data in which teachers were randomly assigned to classrooms, I find that students become more conscientious when they are taught by more conscientious teachers.

If teacher noncognitive skills are transmitted to students, particularly by role modeling, it may be desirable for school leaders to ensure that their school communities exhibit coherence over a set of values that they desire students to embrace. For instance, if school leaders may wish to instill conscientiousness among their students, then creating a school culture where conscientiousness is embodied by teachers and other staff may yield significant dividends towards that aim. Indeed, in chapter 4, I show descriptively that schools where administrators have more leeway over personnel decisions tend to exhibit more values coherence. Presumably, administrators are able to use their flexibility in hiring and dismissing teachers to build a staff that embodies a core set of values.

That being said, the extent to which teachers, schools, and educational interventions ought to focus on student cognitive and noncognitive skills is a philosophical question that cannot be adjudicated solely by methods of empirical science. In chapters 2 and 3, I find that more conscientious teachers do not necessarily have commensurate impacts on student

achievement. This result suggests that different teachers benefit their students in different ways. Some teachers are more effective at improving student achievement while others are more effective at improving student noncognitive skills. The range of ways in which teachers and, by extension, educational organizations or institutions shape students, at the very least, underscores the importance of clearly articulating the means and ends of schooling. The emphasis on student achievement that currently dominates the discourse among policymakers and scholars, though useful, appears to be too narrow. There likely needs to be a broadening and enriching of such discourse to consider the plurality of existing conceptions of the purpose of schooling. Many different student outcomes are important to foster. It will ultimately be up to families, school leaders, policymakers, and other stakeholders to adjudicate, discern, and decide what educational goals to emphasize when tradeoffs must be made. If anything, it is my hope that this dissertation will remind the reader that visions of student success and flourishing are plural. Education is a values-laden and moral enterprise. Perhaps recognition of these facts will assist efforts that seek to improve, reform, and even reimagine educational institutions.

## References

- Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. D. (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 1–181). Amsterdam: Elsevier.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A., & Walters, R.H. (1963). *Social learning and personality development*. New York: Holt, Rinehart and Winston.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York, NY: National Bureau of Economic Research.
- Borghans, L., Angela L. Duckworth, Heckman, J., & Weel, B. ter. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4), 972–1059.
- Borghans, L., ter Weel, B., & Weinberg, B.A. (2008). Interpersonal styles and labor market outcomes. *Journal of Human Resources*, 43(4), 815–858.
- Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, 1137–1176.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679.
- Deke, J., & Haimson, J. (2006). *Valuing Student Competencies: Which Ones Predict Postsecondary Educational Attainment and Earnings, and for Whom?* Princeton, NJ: Mathematica Policy Research.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466–479.
- Heckman, J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3–56.
- Heckman, J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), 145–149.
- Heckman, J.J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464.
- Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Kaestner, R., & Callison, K. (2011). Adolescent Cognitive and Non-cognitive Correlates of Adult Health. *Journal of Human Capital*, 5(1), 29–69.

- Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 101–128.
- Lundborg, P., Nystedt, P., & Rooth, D. (2014). Height and earnings: The role of cognitive and noncognitive skills. *Journal of Human Resources*, 49(1), 141–166.
- Macintyre, A. (2007). *After Virtue: A Study in Moral Theory* (3rd Ed.). Notre Dame, IN: University of Notre Dame Press.
- Mueller, G., & Plug, E. (2006). Estimating the effect of personality on male and female earnings. *Industrial and Labor Relations Review*, 60(1), 3–22.