12-2013

Risky Business? An Analysis of Teacher Characteristics and Compensation Preferences

Daniel Henry Bowen
University of Arkansas, Fayetteville

Risky Business? An Analysis of Teacher Characteristics and Compensation Preferences

Risky Business? An Analysis of Teacher Characteristics and Compensation Preferences

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

by

Daniel H. Bowen
University of Notre Dame
Bachelor of Arts in Political Science, 2006
University of Notre Dame
Master of Education, 2008

December 2013
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Dr. Jay P. Greene
Dissertation Chair

_____
Dr. Patrick Wolf
Committee Member

_____
Dr. Scott Eidelman
Committee Member

Abstract

Teacher quality has a significant impact on both student learning gains and later life outcomes. With this is mind, policymakers implement reforms to attract and retain more effective educators. A major obstacle for designing these policies is that the ingredients for training, as well as initially identifying, effective teachers remain largely a mystery. However, there are strong theoretical arguments for certain education policy reforms producing improvements in the quality of the teacher workforce. One increasingly popular example is performance-based pay. Performance pay has the potential to better align teachers' incentives to produce increases in student achievement. Paying teachers based on students' learning gains, rather than years of experience, increases lifetime earnings for effective teachers which could attract and retain higher-caliber educators. However, since performance-based pay programs rarely last longer than a few years and tend to be small in scale, researchers have not been able to evaluate how changing the compensation structure affects the composition of the teacher workforce.

I provide preliminary evidence for how implementing a performance-based pay program could impact the teacher workforce. Based on analyses from three studies, I conclude that individuals who enter the teaching profession are significantly more risk averse than individuals entering other professions. This finding supports a common stereotype about teachers and possibly provides an explanation for their resistances to education reforms such as merit pay, even when such policies are fairly popular with the rest of the general public. In a follow-up study, the preferences of teachers with a preference for performance-based pay are compared to those of other teachers. The hypothesis of this study is that changing a central component of teacher compensation, the step and lane pay scale, is likely to attract less risk-averse employees.

The evidence from this study is inconclusive but suggests that performance pay might alter the composition of the teacher workforce by either attracting more risk-loving teachers to the classroom or deterring relatively risk-averse individuals from entering the profession. Finally, I link teachers' risk and performance pay preferences to measures of teacher quality. I find that teachers who are the least supportive of performance pay are actually more effective in the classroom. This result contradicts the argument that a compositional shift from performance-based pay will necessarily improve teacher quality. This finding is somewhat counterintuitive but is possibly explained by prior research on the negative relationship between intrinsic or mission-driven motivations and the introduction of performance-based financial incentives.

Acknowledgements

I would first like to acknowledge the mentoring, guidance, and feedback that the members of my dissertation committee have provided me over the years; their time and efforts go above and beyond those spent serving on my dissertation committee. Jay Greene, Patrick Wolf, and Scott Eidelman have each played vital roles in my development as an academic. Jay has been a fantastic advisor throughout my years in the program. He has always been encouraging of my research interests, no matter how crazy they might sound. Patrick has been an outstanding teacher and role model. The strides I have made as an education researcher are largely attributed to his tutelage. Scott has been incredibly generous and gracious with sharing his knowledge and expertise on conducting experimental research. With his incredibly helpful, constructive feedback, he has been a major inspiration for the research agenda I hope to pursue as an academic.

Although they did not officially serve on my dissertation committee, Gary Ritter and Cary Deck deserve recognition for the significant roles they have had in the progress I have made throughout my graduate studies. Gary Ritter has always been incredibly supportive of my efforts and has played a very critical role in many of my achievements over the past four years. Cary Deck has been one of the best teachers I have ever had. His instruction has unquestionably influenced my research interests and abilities with regard to my interests and methodological approaches.

I also want to acknowledge everyone who had a direct role in the work that has gone into this dissertation. Without their help, this research would not have been possible. I thank my coauthors, Jonathan Mills, Cary Deck, James Shuls, and Stuart Buck. It has been a great pleasure to work with them throughout these projects, and I hope that I get the opportunity to work with

them again. I also owe a great debt of gratitude to Maria Perez, Corey Koedel, Jonah Rockoff, Seth Gershenson, and the panel participants at the 2012 and 2013 Annual Education Finance and Policy and the 2012 Association for Public Policy Analysis and Management Conferences for their helpful comments and feedback. I am grateful to Gary Ritter, Jay Greene, Gary Ferrier, and Moez Limayem for providing the funding necessary to conduct this research. Finally, I appreciate the assistance and access that Anna Egalite, Michael McShane, James Woodworth, Albert Cheng, Caleb Rose, Josiah Foster, Jennifer Beasley, Freddie Bowles, Joshua Foster, Christian Goering, John Bacon, Darin Beckwith, and Kristi Campbell provided throughout the data collection process.

This dissertation represents the conclusion of more than two decades of my formal education. Therefore, I would like to acknowledge some of the incredible teachers, professors, and coaches who went to great lengths to provide me with the guidance, foundation, and challenges I needed to make it this far: Patrice Majewski, Paula Kratz, Jane Ketteringham, David Smith, Don Lee, Bill Soule, Nick Mead, Jamie Ferbezar, Tim Meier, Annette Goode, Jennifer Garner, Cynthia Murphy, Peggy Soule, Greg Seibert, Steven Noll, Alvin Tillery, Anthony Messina, Ryan Clark, and Anne Ceasar. While I have spent a significant portion of my years as a graduate student studying teacher quality, I have truly learned most of what I know about great teaching from these remarkable educators.

Finally, I wish to thank all of my extended family and friends who have provided me with so much love and encouragement over the years. There are truly too many people to list out by name, and I do not wish to demean their importance or offend anyone as a result of any accidental omissions. The fact that this list is so extensive is incredibly humbling and reflects how truly fortunate I have been.

Dedication

I dedicate this dissertation to my primary and principal educators, my parents, Dan and Teri Bowen. I would not have made it this far without your constant love and support.

Table of Contents

Introduction

Arguably the most important issue in public education in the United States today is teacher quality. While policymakers and school administrators have tried everything from increasing funds to reducing class sizes to improving classroom technology, researchers consistently find that the most critical school input for students' success is the quality of their teacher. Chetty, Friedman, and Rockoff (2011) conclude that the value of having teachers who are one standard deviation above the median in terms of value-added scores brings about an additional $25,000 to a child's lifetime earnings. Hanushek (2002) argues that very effective teachers produce as much as an additional year of learning for their students compared to ineffective educators. U.S. Secretary of Education Arne Duncan has been very vocal in emphasizing the importance of teachers; "Teachers are central to every single community in America. Each day, teachers come to school ready to tackle a job that is critically important, extraordinarily complex, often joyful, and, at times, heartbreaking. In essence, teachers help others to become their best selves and I can think of no more important work" (2013). Therefore, it is not surprising that the issue of attracting and retaining high-caliber individuals to the teacher workforce has captivated education stakeholders.

Even though the importance of teacher quality is widely accepted, researchers and policymakers have struggled to find answers for how to identify and attract more effective educators. Some of the more commonly used teacher quality screens tend to be inadequate. Kane, Rockoff, and Staiger (2008) demonstrate that teacher certification has a very small impact on teacher quality in terms of students' achievement gains. Hanushek, Rivkin, Rothstein, and Podgursky (2004) find the same to be true for a master's degree in education. In his overview of the empirical research on the effectiveness of various school resources, Hanushek (1997)

concludes that teacher education and experience fail to consistently predict effectiveness for teachers with a few years of experience. One notable exception from Hanushek's (1997) findings is the positive relationship between teacher IQ and test score gains, but screen or retention policies explicitly focused on teacher intelligence are relatively nonexistent beyond the requirement of earning a bachelor's degree.

An initial step for addressing this issue, the attempt to build fair and consistent measures of teacher effectiveness, has proven to be a major endeavor. The Bill and Melinda Gates Foundation has invested over $45 million in the Measures of Effective Teaching Project to find predictors of student learning. While there is considerable debate about whether this study's findings have produced a valid means for identifying more effective teachers (e.g. Greene, 2013; West, 2013), there is less controversy with regard to the argument that these findings do not provide much guidance for *how* to actually attract and train more effective teachers (Hyslop, 2013).

Despite the challenges associated with identifying effective teachers, education policymakers have designed and implemented reforms that aim to improve the caliber of the teacher workforce. One example of this type of reform is a proposal to shift away from a compensation format that rewards teachers for years of experience (i.e. step and lane pay scales) and move towards basing salaries more on student achievement. The theoretical argument for this reform is based on the position that paying for outputs rather than inputs should better align teachers' incentives to produce increases in student achievement (Dee & Keys, 2004; Springer et al., 2010). The primary objective of performance pay is to strengthen this alignment of incentives by increasing the motivations for teachers to work harder to improve student learning (Besley & Ghatak, 2006). Implementing a performance-based pay program over an extended period of time

should also facilitate improvements in the composition of the workforce (Ballou & Podgursky, 1995). If effectiveness is rewarded over experience, then effective teachers have the potential for higher lifetime earnings. The potential for higher lifetime earnings likely attracts individuals with higher wage-earning skillsets which should improve the overall composition of the teacher workforce (Lazear, 2000).

The literature on performance pay policies has generally focused almost entirely on whether financial incentives lead to significant increases in current teachers' motivations that then lead to improved student outcomes (Podgursky & Springer, 2007). The major obstacle that prevents studies on such a policy's impact on the workforce composition is the fact that teacher performance-based pay programs are often too small in scale or too short in duration to significantly alter the teacher labor pool. One exception is Woessmann's (2011) examination of the relationships between countries' international test score performances and the formats of their teacher compensation policies. Controlling for demographics, Woessmann finds that there is a positive, significant relationship between the degree to which a more performance-based pay structure is used and student performance in both math and reading. However, Woessmann concludes that although this finding is significant and substantial, it does not demonstrate a causal relationship.

I hope to add to this literature by examining the relationships between teachers' characteristics and pay preferences and analyzing how these attributes could affect teachers' responses to the implementation of a performance-based pay program. I investigate these relationships by addressing three separate research questions.

The first article evaluates whether individuals who select into the teaching profession have fundamentally different risk preferences than individuals going into other professions. The

primary goal of this article is to assess the validity of the stereotype that teachers are relatively risk averse. Finding evidence that individuals who select into teaching are more risk averse could have significant implications with regard to the implementation of a performance-based pay program. First, if more risk-averse individuals are not amenable to performance-based pay policies, then the pursuits of radically restructuring the format for teacher compensation could have substantial political costs. Second, greater financial incentives are typically needed to get employee buy-in and cooperation for more risk-averse individuals. In other words, to offset the strong, negative fears of ambiguity that come with performance-based pay, programs would almost certainly have to provide bigger rewards in order to achieve their desired outcomes than would be necessary for a more risk-neutral population. Finally, if individuals *entering* the profession are generally more risk averse, then this would suggest that teacher risk aversion is not solely a product of socialization.

The next article attempts to determine whether the shift from the more traditional step and lane pay schedules to performance-based pay attracts more risk-loving teachers. In this study, the risk preferences of teachers who chose to work at schools during the implementation of a performance-based pay program are compared to the other teachers at these schools (including those hired before the program's implementation as well as those hired after the program was discontinued). The goal of this article is to build upon the findings from the first article and see whether pay and risk preferences are related. If performance-based pay attracts individuals with significantly different risk preferences, then the implication might be that changing the format of teacher pay would alter the composition of the workforce. The change in risk preferences might not be better or worse for students, but the fact that a new format attracts

individuals with significantly different characteristics reflects that compensation format can play a major role in who is ultimately attracted to entering the field.

The last article examines whether pay and risk preferences are significantly related to measures of teacher quality. Teachers' responses from the second study are linked to their students' test score gains in addition to the performance evaluations that their principals provide at the end of each school year. The primary objective of this article is to provide evidence for whether there might be cause for hope or concern with regard to the type of compositional shift that could take place in the event of a substantial change to the format of teacher compensation. If there is a positive relationship between teachers' preferences with regard to performance pay and their effectiveness as educators, then this study would provide evidence for this reform plausibly improving teacher quality. However, if there is a negative or even null relationship, then performance-based pay policies might lead to greater costs than benefits at least with regard to their compositional effects. Moreover, a negative relationship between teacher effectiveness and preference for performance-based pay might signify a conflict between motivations for teaching and the influences that financial incentives have on teachers.

With performance-based pay programs being relatively small in scale and duration at this time, circumstances make it difficult to address these questions with research in the field. Therefore, the data collection for these studies took place in a more laboratory-like setting. This type of approach is fairly uncommon in current education policy research. However, the increasing influence of experimental and behavioral economics in other public policy arenas (e.g. Birol, Karousakis, & Koundouri, 2006; Gowdy, 2008; Lancsar & Louviere, 2008) might serve as precursors for incorporating these types of methods into education policy research. Moreover,

Jabbar (2011) argues that controlled laboratory settings provide a useful means for developing new hypotheses that can be later tested in the field.

In order to measure and compare individuals' risk preferences, 132 teacher candidates and then 120 different teachers from two different school districts participated in the Holt and Laury (2002) experimental task. The Holt and Laury (2002) risk-elicitation task is commonly used in the field of experimental economics and has the benefit of providing means for conducting both absolute and relative measures of individual risk preferences. All of the participants also completed surveys in order to provide demographic information that is used to analyze how participants' characteristics and backgrounds potentially influence risk and pay preferences. I also received permission from administrators of the two school districts to obtain the names of the participants from the later study. I link their responses to the districts' teacher performance data, including teachers' student test score gains and the ratings on their end of the year performance evaluations.

Overall, I find that individuals who want to go into teaching are significantly more risk averse than individuals with similar educational backgrounds whom are entering other, non-teaching professions. Prospective teachers do not have a greater aversion or propensity for performance-based pay; however, this finding is possibly attributed to these individuals not having options to distinguish their preferences for certain types of pay. More specifically, compensation formats were not presented in this study as tradeoffs.

I also find that teachers who "truly seek out performance-based pay" are significantly more risk loving than other teachers. I identify teachers as seeking out performance-based pay as those teachers employed at a time when the merit pay system was in place. When analyzing differences in teachers who opt in versus those who opt out, I do not find a significant difference

in risk preference. However, this finding could be attributed to the fact that the merit pay aspect was an afterthought. Therefore, I run additional analyses where true performance-pay seekers are identified as those having opted in as well as indicating that the merit pay program was "a deciding factor" in their employment decision. This finding however is not statistically significant at conventional levels due to such a small sample of teachers designated as being truly drawn to performance-based pay.

Finally, I find that a teacher's inclinations towards a more performance-based pay system as well as the degree to which they are risk loving are both negatively related to teacher effectiveness in terms of value-added measures. This finding is especially perplexing given the fact that the most effective teachers would seemingly benefit from having pay based on student performance. There is no significant relationship between teachers' preferences and their ratings on the end of the year performance evaluations. This evidence could serve as an important caveat for policymakers. Changing the structure of teacher compensation may have unintended consequences. If the most effective teachers are opposed to performance-based pay, then implementing this type of policy could make it more difficult to retain these teachers. Moreover, this finding could assert the possibility that mission orientation or intrinsic motivation is positively related to teacher quality. In other words, high quality teachers could be more likely to find that financial incentives crowd out or conflict with their motives for teaching, and even though they benefit the most financially from this compensation format, they might grow to resent performance pay the most. This explanation is purely hypothetical but perhaps the relationships between teachers' motivations, efficacy, and perspectives, should be further examined in future research.

The remainder of this dissertation is divided up into four sections, consisting of three separate articles and a conclusion. The second through fourth sections are the separate articles that examine the three research questions that guide this dissertation. The second section examines the question of whether individuals who go into the teaching profession are significantly more risk averse than individuals with similar educational backgrounds selecting into other professions. The third section investigates whether performance-based pay could alter the composition of the teacher workforce. The fourth section links teachers' risk and survey responses to their value-added scores and year-end performance evaluations to analyze whether pay and risk preferences predict measures of teacher quality. Finally, the last section summarizes the findings of these three articles, provides some policy implications, describes some of the limitations of this research, and proposes questions to hopefully guide future research.

References

Ballou, D., & Podgursky, M. (1995). Recruiting smarter teachers. *The Journal of Human Resources, 30* (2), 326-338.

Besley, T. & Ghatak, M. (2006). Sorting with motivated agents: Implications for school competition and teacher incentives. *Journal of the European Economic Association, 4* (2-3), 404-414.

Birol, E., Karousakis, K., & Koundouri, P. (2006). Using economic valuation techniques to inform water resources management: A survey and critical appraisal of available techniques and an application. *Science of the Total Environment, 365*(1), 105-122.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (NBER Working Paper No. 17699). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w17699

Dee, T., & Keys, B.J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management, 23*, 471-488.

Duncan, A. (2013, May, 6). *Statement by U.S. Secretary of Education Arne Duncan recognizing Teacher Appreciation Week, May 6-10, 2013*. Retrieved from https://www.ed.gov/news/press-releases/statement-us-secretary-education-arne-duncan-recognizing-teacher-appreciation-we

Gowdy, J. M. (2008). Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization*, *68*(3), 632-644.

Greene, J.P. (2013, January 9). Understanding the Gates Foundation's Measuring Effective Teachers Project. Retrieved from http://jaypgreene.com/2013/01/09/understanding-the-gates-foundations-measuring-effective-teachers-project/

Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.

Hanushek, E.A. (2002). Teacher quality. In L.T. Izumi and W.M. Evers (Eds.), *Teacher quality* (pp. 1-12). Stanford, CA: Hoover Press.

Hanushek, E. A., Rivkin, S. G., Rothstein, R., & Podgursky, M. (2004). How to improve the supply of high-quality teachers. *Brookings papers on education policy*, (7), 7-44.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655.

Hyslop, A. (2013, January 17). What to think about the MET Project results. Retrieved from http://edmoney.newamerica.net/blogposts/2013/what_to_think_about_the_met_project_results-77643

Jabbar, H. (2011). The behavioral economics of education: New directions for research. *Educational Researcher, 40*(9), 446-453.

Kachelmeier, S. J., & King, R. R. (2002). Using laboratory experiments to evaluate accounting policy issues. *Accounting Horizons*, *16*(3), 219-232.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631.

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review, 90*, 1346-1361.

Podgursky, M., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26* (4), 909-949.

Springer, M.G., Ballou, D., Hamilton, L., Le, V., Lockwood, J.R., McCaffrey, D., Pepper, M., & Stecher, B. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

West, M.R. (2013, January 10). Misunderstanding the Gates Foundation's Measures of Effective Teaching Project. Retrieved from http://educationnext.org/misunderstanding-the-gates-foundation%E2%80%99s-measures-of-effective-teaching-project/

Woessmann, L. (2011). Merit pay international. *Education Next, 11*(2), 73-77.

Risky Business: An Analysis of Teacher Risk Preferences

By: Daniel H. Bowen, Stuart Buck, Cary Deck, Jonathan N. Mills, & James V. Shuls

Abstract

Teachers tend to oppose policy reforms aimed at improving teacher effectiveness. One potential explanation for their objections is that teachers, as a group, are relatively risk averse. This paper addresses this explanation through the use of a risk preference elicitation task commonly used in experimental economics. Comparing the risk preferences of future teachers with those entering other professions, we find that individuals selecting into teaching careers are significantly more risk averse. At the same time, we do not find evidence that compensation format preferences are independently responsible for attracting more risk-averse individuals into teaching, as risk aversion does not appear to correlate with a stated preference for a particular compensation format. We believe that policymakers need to take into account teachers' risk characteristics when considering policy changes that may clash with teacher preferences.

Introduction

Research has demonstrated that an effective teacher can significantly increase how much students learn in the classroom (Chetty, Friedman, & Rockoff, 2011; Hanushek, 1992). In fact, the difference between an effective and ineffective teacher could be as much as an extra year of learning for students (Hanushek, 2002). However, while private sector workers are more commonly paid according to on-the-job performance, most American teacher salaries are determined by a single salary schedule, often referred to as a "step and lane" pay scale. First introduced in 1921, by 1950 roughly 97% of all schools in the United States adopted a step and lane pay schedule (Prostik, 1996). Currently, the vast majority of K-12 teachers in a school district start at the same base salary regardless of teaching subject or grade level. Rather than

rewarding teachers for student performance, step and lane pay schedules offer salary increases for each year of service and for additional coursework, typically college credit hours or advanced degrees. Therefore, under this pay scheme, a highly effective teacher and a highly ineffective teacher, who begin teaching in the same school at the same time, will earn the exact same salary.

While step and lane salary schedules have been the norm in public education for nearly six decades, there has been a recent push to reform teacher pay. Secretary of Education Arne Duncan and President Obama have publicly supported the development of merit pay systems for teachers. The Department of Education has established a grant program, the Teacher Incentive Fund, that supports the development of performance-based compensation systems. Paying teachers based on their performances has also grown in popularity with the general public. In a recent public opinion poll, 47% of those surveyed favor "paying teachers, in part, based on the academic progress of their students on state tests," while only 27% were opposed to merit pay (Howell, West, & Peterson, 2011, p. 14). However, the same poll finds that 72% of teachers oppose performance-based pay.

There are many reasons why teachers might oppose a more performance-based incentive scheme. A step and lane scale may contribute to a steady workforce by encouraging individuals to remain in the profession throughout their career. A stable pay scale might also enable teachers to experiment with new and potentially innovative teaching techniques that they might not try otherwise. Teachers might fear that merit pay could bring about a more acute focus on a narrow academic element of schooling to the detriment of other valuable areas (Wilms & Chapleau, 1999). Teachers may also worry about the reliability of tests to accurately measure student performance (Koedel, 2009). Additionally, teachers may fear that rewarding individual

performance may reduce the incentives for teachers to collaborate with one another and could even cause conflict (Ritter & Jensen, 2010).

Another plausible, yet relatively unexplored, explanation for teacher opposition is that individuals who go into teaching may be relatively more risk averse than the general population. Studies in a variety of fields have noted that compensation schemes impact workforce composition (e.g. Dohmen & Falk, 2011; Lazear, 2000). Relative to private sector employees, teachers receive lower average pay but have less pay variability and more generous health benefits, pensions, and job security (Podgursky, 2003). These features of the compensation system could entice relatively risk-averse individuals into the profession. If teachers are in fact more risk averse than other workers, teacher opposition to performance pay systems is unsurprising.

In this paper, we explore teachers' risk preferences using a lottery-based risk elicitation tool common to the experimental economics literature. Using responses on this task, we compare the risk attitudes of future teachers with the risk preferences of students in master of business administration (MBA) and juris doctorate (JD) programs. By focusing on graduate students preparing to enter teaching and other professional careers, we can focus on the issue of self-selection into a career while eliminating affects that experiences in these professions might have on preferences. We also survey participants on key demographic information, as well as their likelihoods of going into teaching, and their preferences for four different types of pay systems: pay based on individual performance, developing knowledge and skills, team performance, and experience.

Literature Review

Public sector jobs often come with tenure or civil service protections that are far more protective than the employment-at-will relationships pervasive in the private sector. Unsurprisingly, economists have long found that public sector workers tend to be more risk averse than private sector workers (e.g. Dohmen et al., 2005; Hartog, Ferrer-i-Carbonell, & Jonker, 2002; Masclet, Colombier, Denant-Boemont, & Loheac, 2009). Bellante and Link (1981) find that risk aversion, measured from survey responses, is a significant predictor of public sector employment. Buurman, Delfgaauw, Dur, and Van den Bossche (2012) find that public sector workers' odds of choosing a riskier reward for filling out a survey are only slightly more than half of that of a private sector worker.

As for the risk aversion of teachers in particular, previous work has been anecdotal (e.g. Wagner, 2001) or based on survey data lacking salient rewards (e.g. Davis, 1994). One exception is Perez (2011) who compares the risk preferences, attitudes towards pay inequity, and preferences toward competition of female teacher candidates to female law students. In the study, subjects were asked to complete a series of 10-minute rounds of solving mazes with different pay schemes. In the final round, some participants were randomly asked to choose which pay scheme they would prefer. The other participants were asked to make the same decision, except that they were told that one of their decisions (randomly determined) would apply to the rest of the group.

Perez finds that teachers do not exhibit higher levels of risk aversion than the lawyers because they are no more or less likely to choose performance-based compensation when determining the pay format that only applies to them. However, teachers were significantly more likely to be averse to pay inequity because when the decision potentially applies to everyone

else, teachers are much less likely to opt for performance-based pay. Perez claims that these results suggest that teachers have fundamental inclinations towards greater pay equity and that policymakers should take this inclination into consideration.

While our study is closely related to Perez's, it differs in a few ways. First, we include males in our sample. Even though the majority of teachers are women, we want to expand our analyses to also include men drawn to the teaching profession. Second, we use individuals' choices with regard to different lotteries in order to elicit risk preferences. Finally, we examine whether there is a significant relationship between individuals' risk preferences and their preferences for different forms of compensation. In the next section, we introduce the risk-elicitation instrument and outline our experimental procedure. Then, we present our results with regard to differences in risk preferences and whether these related to surveyed preferences for different compensation formats. Finally, we conclude with a discussion of our results and their potential implications.

Methods

Risk Elicitation Task

To measure the risk attitudes of future teachers relative to the participants in the MBA and JD programs, we use the Holt and Laury (2002) risk preference task. This tool is a well-known, commonly used risk-elicitation procedure in the experimental economics literature (e.g. Anderson, Harrison, Lau, & Rutstrom, 2008; Dohmen, Falk, Huffman, & Sunde, 2010; Eckel & Wilson, 2004). While there are alternative controlled procedures for measuring risk, this procedure has become a generally accepted standard. In part this is due to the ease with which it can be implemented and explained to subjects. Further, Harrison, Johnson, McInnes, and

Rutstrom (2005) have found this procedure to have high retest reliability over an extended time frame.

The procedure, shown in Table 1 as it was presented to participants, estimates risk preferences by examining subject choices between lotteries with different real dollar payouts. In particular, subjects are asked to choose between two options (A and B) for each of 10 lotteries with the understanding that they will receive the outcome from one particular lottery chosen at random. The possible payouts are held fixed for each option, with Option A payoffs ($4.80 or $6.00) having less variability than Option B payoffs ($0.30 or $11.55). The chance of receiving the higher of the two payouts for either option increases by 10 percentage points with each lottery. In Lottery 10, a participant is guaranteed to receive the larger payoff amount from the selected option and thus should strictly prefer the $11.55 from Option B to the $6.00 from Option A.

Risk preferences are modeled using a constant relative risk aversion functional form of utility, $u(x) = x^{1-r}$, where individual utility is a function of payout x and their risk preference parameter r. In this framework, individuals are classed into one of three groups based on the value of r: risk averse ($r>0$), risk neutral ($r=0$), and risk preferring($r<0$). Individuals with $r=0$ are considered risk neutral because their utility is solely determined by their expected payout. In contrast, individuals with $r>0$ are risk averse as their utility down-weights the expected payout because $(1-r) <1$ when $r>0$. Finally, $r<0$ is associated with risk loving preferences as such individuals prefer increased uncertainty.

The switching point from Option A to Option B identifies a range of risk parameters that are consistent with the observed choice (Holt & Laury, 2002). It is conceivable that an extremely risk-loving individual would select Option B for every choice, but monotonicity is sufficient to

cause people to select Option B for choice 10 because one is assured of receiving the higher payoff. A risk-neutral individual is concerned with the expected payouts from the two options in a given lottery and will choose Option A for the first four lotteries and Option B for the remaining lotteries. Risk-averse individuals are willing to forgo expected value in exchange for reduced uncertainty and therefore will continue choosing Option A even after the fourth lottery despite the higher expected payout from Option B. Risk-loving individuals will make the switch to Option B before the fifth lottery despite the higher expected payout from Option A.

The Holt and Laury (2002) procedure not only provides rich information on individual preferences, it also allows us to check the extent to which subject confusion exists in our data: individuals should never select Option A on Lottery X after having selected Option B for Lottery Y if X>Y. Therefore multiple switches serve as one indicator of participant confusion.

Risk Task Procedures

A total of 132 subjects completed the study, all of whom were graduate students at a major, public university in the southeast region of the United States. Our group of interest consisted of 65 prospective teachers recruited from the university's Master's in the Art of Teaching (MAT) Program. Of these, 32% were studying to teach at the elementary level while 68% were studying to teach at the middle or secondary level. The students spend an academic year taking coursework while student teaching at traditional public schools. Upon completion of the program, students become certified teachers in the state. We chose MAT students, as opposed to undergraduates with declared education majors, because of the greater certainty that these students will ultimately end up teaching in the near future. In fact, according to a conservative estimate provided by the MAT program coordinator, more than 90% of the MAT

students accept teaching positions in the academic year immediately following the completion of the program.

The MAT program is the only route a student at this particular university can take to earn a traditional teaching license. Future elementary teachers can earn a bachelor's degree in education, while middle and secondary teachers typically first earn a bachelor's degree in their content areas. Therefore, while education undergraduates are almost entirely elementary education majors, MAT students represent a more even distribution of elementary, middle, and high school teachers.

The comparison group for this study is composed of graduate students in other fields. In particular, future teachers are compared to 43 students in the university's MBA program in addition to 24 JD students. We chose these students to serve as a comparison group because they are similarly pursuing professional graduate degrees.

To conduct the study, we obtained permission from university faculty to offer their students the opportunity to participate in a paid research study on economic decision making. The experiments were conducted during the last fifteen minutes of class. Instructors were asked to leave the room prior to the study and participants were aware that their responses would be kept anonymous. The students were not made aware of the opportunity to participate in the study prior to our arrival. After the instructor left, we briefly described the experiment and reviewed the participation consent form that subjects would be required to sign in order to receive payment. Students were free to leave at any point as participation was voluntary.[1]

Participants were then given a single, two-sided sheet of paper with instructions for the experiment. This document contained a sample question to verify if the participant understood

[1] Only two potential subjects opted not to participate; one male and one female in the JD program.

the experiment, the risk preference elicitation instrument, and a survey of compensation preferences as well as demographic questions (see Appendix). Participants were told to read the directions, complete a comprehension question, and then raise their hands so that one of the monitors could verify that the subject understood how the elicitation procedure worked before completing the task. If the subject completed the sample question correctly, the monitor discreetly marked the form in a particular location on the paper. If the subject had answered incorrectly and needed additional explanation, the monitor discreetly marked the form in a different location. This process allowed us to track which subjects experienced at least some initial confusion about the task.

After an experimenter checked the comprehension question and verified that the participant understood the instructions, the participant was asked to complete the risk elicitation task and then complete a brief survey on the other side (see Appendix). Participants were aware the survey existed when completing the risk-elicitation tool, but they were unaware of the survey's contents. The survey was completed after the lottery in order to avoid the possibility that these questions could frame or influence subjects' behaviors while listing their lottery preferences.

The first half of the survey asked participants about their preferences for four different pay system scenarios. We used the same scenarios that Milanowski (2007) used to survey students who were preparing to become teachers. We asked participants, on a scale of -4 (highly undesirable) to +4 (highly desirable), to indicate how desirable or undesirable they would find each particular pay format for their first chosen occupation. The four formats were pay for individual performance, pay for the development of knowledge and skills, pay for team performance, and pay not based on performance.

The last part of the survey included questions about background information to control for other potentially salient characteristics. Specifically, we collected information on each participant's age, gender, proxies for income level, mother's level of education, and the individual's estimated likelihood that they would ever enter the teaching force.[2] Our proxies for wealth include the make and model of a participant's car and how many times they eat at a restaurant in a given week. We used these proxies rather than asking for current income level, both to try and elicit honest answers and to account for the possibility of inaccurate reports of graduate student incomes.[3]

Upon completion of the study, participants brought their response forms to the front of the room. Monitors verified that subjects had completed all of the questions and then rolled a ten-sided die to determine which lottery choice would be used to determine the subject's payment.[4] Next, the ten-sided die was rolled again to determine the actual payment according to the subject's choice of Option A or B for the randomly chosen lottery. On average, participants in the experiment received a payment of $7.00. After subjects received their payments, they were dismissed from the study and then exited the room. There was no identifying link between the payment record and the participants' response forms in order to preserve their anonymity.

[2] The latter variable was collected to verify that one's degree program was a reliable proxy of career intentions.

[3] In other words, these approaches seem to more accurate measure of a participant's wealth than having them self-report income levels. A graduate student could potentially report no salary due to their student status, but others might include family and/or parents' income levels in their responses without us being able to distinguish how the participant approached this question.

[4] Randomly selecting one task for payment is a common approach in experiments where the researcher wants to control for potential wealth effects (see e.g. Holt & Laury, 2002).

Results

In this section we present both the risk task and survey results. Before examining differences in behavior between prospective teachers and other students, we compare the composition of the two groups. To make sure that the two groups accurately distinguish individuals who are most likely to teach from those who will not, we asked how likely the participants were to teach in a K-12 setting. More than 97% of MAT students indicated there was greater than a 75% chance they would enter teaching with 60% stating the likelihood was 100%. In comparison, only 3% of non-MAT students indicated there was greater than a 75% chance that they would teach, while more than 89% indicated there was less than a 25% chance. These responses give us confidence that the participants in our study accurately capture future teachers and other professionals not likely to work in education.

Table 2 provides summary demographic information for the MAT and non-MAT samples. The overall sample consists of mostly female participants, but males represent a majority among non-MAT students. White participants are the overwhelming majority in the MAT sample, and the non-MAT sample is more diverse. In addition to simple demographics, we examined the extent to which the MAT and non-MAT samples varied on wealth proxies. In general, the two samples did not significantly differ on level of mother's education, number of days they ate out in a given week, or with the Blue Book value of their personal vehicles.

Our experimental procedures provided us with two opportunities to ascertain individual confusion with the risk elicitation tool. First, after having participants read through the instructions, we had them answer a question on a hypothetical payout. The second test of comprehension is whether the individual made consistent responses (i.e. as opposed to making multiple switches between Lotteries A and B). In Table 3, we present the percentages of correct

answers to the comprehension question and percentages of students with consistent responses on the risk elicitation tool. When comparing the overall percentages of confusion in the MAT and non-MAT groups, we find no statistically significant difference between the groups.

We present the results of our study in terms of risk preferences and pay preferences. These results include all individuals from our study; we do not limit our analyses to only those participants who lacked any difficulty comprehending the risk task. This allows us to utilize our full data set. We have also conducted the same analyses both controlling for and excluding individuals that exhibited confusion. While point estimates differ between the two sets of analyses, estimate signs and significance generally remain unchanged.[5]

Risk Preferences

In Table 4, we present a comparison of the average number of times an individual went with the "safer" (i.e. lower payout variance) choice, Option A. On average, participants make 4.7 safe choices. MAT students, on average, make 5 safe choices while non-MAT students make 4.3. Figure 1 presents the distribution of the proportion of individuals making safe choices in each lottery for MAT and non-MAT students. The dotted line serves as a reference, indicating how a perfectly risk-neutral group would respond to the risk-elicitation tool. Consistent with the comparison of averages in Table 4, Figure 1 indicates non-MAT students are more willing to take risks. For example, in lottery 5 (where a risk-neutral individual would first make a switch away from the safer choice), 39% of non-MAT, as opposed to 60% of MAT students, chose the

[5] The only changes in significance that take place are in 2 of the 12 analyses. The coefficient for MAT is always positive and statistically significant at conventional levels in 10 of these 12 analyses and always for models 1, 2, and 3. However, when removing participants that did not initially get the initial item check correct, the coefficient on MAT is 0.90 ($p = 0.11$) in the 4th model when controlling for the number of times that a participant eats out in a given week (wealth proxy 2). When removing any participants that demonstrated confusion in terms of switching options more than once, the coefficient on MAT is 0.91 ($p = 0.12$) in the 3rd model when controlling for the number of times that a participant eats out in a given week.

safer option. The greater willingness of non-MAT students to take risks is supported by a Kolmogorov-Smirnov test that compares these two distributions (p-value < 0.01). The central tendencies of the two distributions also differ based upon a Mann-Whitney test (p-value < 0.01).

Since our sample of future teachers is mostly female, a gender effect may be driving our main result. Figure 2 presents a comparison of outcomes for females and males. The strong similarities between the male and female distributions suggest a lack of a gender effect. The Mann-Whitney and Kolmogorov-Smirnov tests both support this conclusion as they fail to find a significant gender effect (p-values = 0.79 and 0.67, respectively). This finding may seem counterintuitive as prior research has found women to be more risk averse than men (e.g. Eckel & Grossman, 2008; Schubert, Gysler, Brown, & Brachinger, 1999). Nevertheless, there are also experimental studies that have found no significant gender differences in risk behavior (e.g. Harbaugh, Krause, & Vesterlund, 2002; Moore & Eckel, 2003).

We also examine results of gender differences within degree program in Figure 3. While non-MAT males do not appear to strongly differ from non-MAT females, there appears to be a gap between MAT males and females, with males significantly choosing the safer choice more often on average. Among males in the MAT program, 63% made 6 safe choices or more indicating a high degree of risk aversion. Mann-Whitney and Kolmogorov-Smirnov tests reject the null hypothesis of similar distributions between male MAT and both female MAT and male non-MAT participants (see Table 5).

In addition to comparing the number of safe choices by each group, we conduct regression analyses to examine the risk attitudes between those in the two degree programs while controlling for other demographic variables. Table 6 reports the results of four different specifications where the dependent variable is the number of safe choices. The first specification

examines only the relationship between degree program and risk aversion. While the second through fourth specifications include additional covariates such as gender, race, age, income, and confusion, MAT program enrollment is the only variable that predicts risk attitude. In the first specification, MAT program enrollment is associated with participants choosing an average of 0.70 more safe choices (Option A). This result holds when we additionally control for gender in the second specification: MAT program enrollment is associated with significantly more safe choices; however gender has no significant impact on the number of safe choices.

In the third specification, we include gender interacted with the MAT program as well as indicators for race, age, and whether a subject experienced confusion in interpreting the lottery task. Males in the MAT program make 1.03 more safe choices than non-MAT males ($p = 0.05$). Additionally, we compare female MAT and non-MAT student risk preferences by examining the statistical significance of the sum of the MAT and Female*MAT coefficients. A joint F-test rejects the hypothesis that these coefficients sum to zero ($p = 0.08$). Finally, we can examine if there is a gender difference in the number of safe choices among MAT students by testing the statistical significance of the sum of the Female and Female*MAT coefficients. A joint F-test fails to reject the hypothesis that these coefficients sum to zero ($p = 0.57$). Thus, MAT status is significantly related to making more safe choices when examining different subgroups and controlling for demographics. While MAT males still appear to be the most risk averse subgroup, they are not significantly more risk averse than MAT females.

In the fourth specification, we include a host of variables that attempt to control for student wealth. In particular, we include three wealth proxies: the log of participant car Blue Book values; the number of times the participant reports to eat at a restaurant in a given week; and reported level of mother's education. We ran this analysis with each of these proxies for

wealth separately as these measures are likely highly correlated. The findings from the other model specifications remain unchanged. Specifically, we find no significant differences in risk preferences based on an individual's wealth as measured by our proxies. The coefficients for MAT, gender, age, etc. marginally change. However, each coefficient does not change with regard to direction or significance.

In summary, the results from our analyses of individual risk-preferences indicate that future teachers in our sample are significantly more risk averse than other graduate students. This result holds when we disaggregate by gender: both male and female MAT students are significantly more risk averse than respective non-MAT counterparts. At the same time, we do not find significant differences in risk preferences among male and female MAT students. These results hold when we control for additional individual demographics. In the next section, we examine the extent to which these estimated differences in risk preferences correspond to differences in preferences for performance-based compensation schemes.

Pay Preferences

Table 7 describes the average ratings given to each pay system, broken out by group and gender. Overall, there was relatively strong support for individual merit pay based on objective performance as well as based on one's development of knowledge and skills. MAT students were significantly less supportive of team-based pay (K-S p-value = 0.03; M-W p-value < 0.01). We also have dichotomized participants' preferences in order to examine the percentage within each group who generally support the different types of pay systems. MAT students are still found to be significantly less likely to support pay for team performance. MAT students are also less likely to support pay for individual performance and pay for developing their knowledge and skills, and they are more likely to support pay raises that are not based on performance.

However, these differences fail to achieve traditionally accepted levels of statistical significance. Finally, there is little to no correlation between these participants' risk and pay system preferences.

The strong overall support for individual merit pay might seem anomalous, especially in the case of risk-averse individuals. This might be because our survey question described individual merit pay as the potential to earn a pay increase (i.e. a bonus). It is plausible then that even risk-averse individuals would be supportive of the chance to earn a pay raise when there is no risk of wage loss. Moreover, the prospect of pay for performance for a novice teacher is possibly not perceived as being all that risky. Moreover, one potential shortcoming of the design for this section of the survey is that we are not able to observe how these participants necessarily perceive these different pay formats in relation to one another. Perhaps having the participants choose the extent to which they would prefer individual versus team pay or performance-based versus automatic raises would have provided more useful information at least for comparing these groups' pay preferences. In the following section we summarize our findings and discuss their implications.

Discussion & Conclusion

We conduct an experiment that allows us to ascertain relative risk preferences for individuals in a MAT program and a comparison group of MBA and law students. Our results show clear evidence that those who opt to pursue teaching careers are more risk averse than those pursuing careers in business or law. However, based on the survey responses of our participants, it does not appear as though an early preference for step and lane pay (or an aversion to performance-based pay) systems can independently explain why more risk-averse individuals choose to go into teaching.

Support for the notion that teachers tend to be risk averse could have important implications for education policymakers. Buurman et al. (2012) note that if public employees are more risk averse, "pay-for-performance is a more costly instrument to induce effort in the public sector than in the private sector" (p. 4). Similarly, implementing performance pay may increase the levels of tension and worker dissatisfaction (Dohmen & Falk, 2010; Perez, 2011). Specifically, this disapproval could be attributed to the possibility that the profession has attracted individuals who are relatively risk averse.

In our sample, both future teachers and the comparison group preferred merit pay systems more than they preferred a uniform pay raise system which is commonly used in education. There are at least a couple of reasons why this might be the case for the participants of this study. First, if these future teachers are comparing their possible wages under a performance-based pay system to what they will receive under the step and lane system, then the former system possibly looks more appealing. Early career teachers have the lowest wages in a step and lane pay system and may therefore be more likely to prefer performance-based pay because it offers the chance of making higher wages with essentially no risk. Milanowski's (2007) research supports this explanation, finding that support for merit pay appears to be more common for less experienced than veteran teachers. Future teachers may also overestimate their abilities to be high performers. In other words, without evidence to the contrary, MAT students may not find performance-based pay to be all that risky if they see themselves as being highly effective teachers. Finally, the result may simply be driven by the wording of the survey question, as participants may prefer performance pay when it takes the form of a bonus (i.e. no risk of loss).

There was a significant difference in preference for team-based merit pay plans, with non-MAT students preferring this system more than MAT students. We have no firm hypothesis

for why this might be the case, but we speculate that it could be tied to the nature of their prospective professions. The MAT students' opposition to team performance pay could reflect how these future teachers assess their abilities relative to other teachers. In other words, if these future teachers believe that they will be better than average, they may view team performance pay as undesirable (at least relative to their expected earnings under an individual performance pay program). Another possibility is that MBA and law students might be more likely to work together on a single project or case. In which case, rewarding workers for the success of the team may closely align with the type of work these future employees aspire to do upon entering the workforce. Teachers, on the other hand, have just their classroom of students and may have little influence over how other teachers perform.

Limitations

In addition to our small sample size, our use of a laboratory environment limits the extent to which we can make broader, more generalizable claims about all teachers' risk characteristics. Moreover, our sample is restricted to prospective teachers coming from one university. Despite these limitations, we hope that our findings build upon as well as facilitate the research on teacher characteristics. We believe that incorporating behavioral measures into this body of research can help increase our knowledge about the types of individuals who are attracted to the teaching profession in addition to a better understanding of the specific aspects of the profession that have fostered this attraction. We do not believe that our findings provide any specific policy prescriptions. Nevertheless, this research can help inform and caution policymakers going forward as they consider ways to attract and retain high quality teachers in addition to positively influencing the performances of those currently in the workforce.

Conclusion

Our findings provide suggestive evidence that future teachers are indeed more risk averse than individuals in other professions. At the same time, our survey data do not provide evidence supporting the notion that uniform salary schedules or even aversion to performance-based pay is solely responsible for attracting these relatively risk-averse individuals to the teaching profession. It could be the case that other aspects of teaching are attracting risk-averse individuals to the profession. Another possibility is that these future teachers do not view their participation in performance-based pay programs as very risky. These views could be due to the fact that incoming teachers have relatively little to lose in a step and lane pay system. It could also be the case that these future teachers have a high estimate of their likely effectiveness in the classroom and therefore see little risk in compensation based on their abilities.

Although there are many other issues that need to be taken into consideration in terms of how these results can shape and influence education policy, this evidence can at least inform policy discussions on teacher quality in at least a couple of different ways. For example, our results may suggest that policymakers need to take into greater account teachers' risk preferences when designing performance incentives (e.g. performance pay programs). Moreover, these findings may suggest that the efficacy of policies could depend more on facilitating systemic changes that attract a different pool of individuals into the profession. However, while implementing certain reforms could entice less risk-averse individuals into the teaching profession as well as deter more risk-averse individuals from entering the profession, the impact that this systemic shift would have on student achievement is unknown.

Future research could compare prospective teachers' preferences, characteristics, and attitudes to more veteran teachers. Perhaps there is a selection effect over time: people who

especially like certain aspects of the job might stay in the teaching profession, and those who do not care for these aspects eventually leave the profession. In addition, people who teach for several years might dislike certain characteristics of the job initially, but they might grow to prefer these characteristics over time because of social conformity or because it is a system that fundamentally advantages them as they transition to veteran status.

With regard to systemic changes, future analyses could further explore the nature of personnel changes that might occur with the implementation of certain reforms (e.g. examining whether performance-based pay or the removal of tenure attracts less risk-averse teacher candidates) and assessing if these changes benefit student outcomes. These questions, as well as the ones addressed in this research, will hopefully better inform policymakers on important considerations when it comes to implementing changes that potentially alter the nature of the teaching profession and, subsequently, the composition of the teacher workforce.

References

Anderson, S., Harrison, G. W., Lau, M. I., & Rutstrom, E. E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583-618.

Bellante, D. & Link, A.N. (1981). Are public sector workers more risk averse than private sector workers? *Industrial and Labor Relations Review, 34*(3), 408-12.

Buurman, M., Delfgaauw, J., Dur, R., & Van den Bossche, S. (2012). Public sector employees: Risk averse and altruistic? *Journal of Economic Behavior & Organization 83*(3), 279-291.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (No. w17699). National Bureau of Economic Research.

Davis, J.B. (1994). A closer look at those who have decided to teach. *High School Journal, 77* (4), 274–79.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review, 101*(2), 556-590.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*(3), 1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G.G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey.

Eckel, C.C. & Grossman, P.J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization, 68*(1), 1-17.

Eckel, C. C. & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization, 55*, 447-465.

Hanushek, E. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Hanushek, E. (2002). Teacher quality. In L.T. Izumi and W.M. Evers (Eds.), *Teacher quality* (pp. 1-12). Stanford, CA: Hoover Press.

Harbaugh, W.T., Krause, K., & Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics, 5*(1), 53-84.

Harrison, G. W., Johnson, E., McInnes, M. M., & Rutstrom, E. E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1(1), 31-35.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking measured risk aversion to individual characteristics. *Kyklos, 55,* 3-26.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655.

Howell, W. G., West, M. R., & Peterson, P. E. (2011). The public weighs in on school reform. *Education Next, 11*(4), 10-22.

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review, 28*(6), 682-692.

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review, 90,* 1346-1361.

Masclet, D., Colombier, N., Denant-Boemont, L., & Loheac. Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization, 70,* 470-84.

Milanowski, A. (2007). Performance pay system preferences of students preparing to be teachers. *Education Finance and Policy, 2*(2), 111-132.

Moore, E. & Eckel, C.C. (2003). Measuring ambiguity aversion. Unpublished manuscript, Department of Economics, *Virginia Polytechnic Institute and State University*.

Perez, M. (2011). Inequity aversion differences: Experimental evidence among prospective teachers and lawyers. *Available at SSRN 2018697*.

Podgursky, M. (2003). Fringe benefits: There is more to compensation than a teacher's salary. *Education Next, 3*(3), 71-76.

Prostik, J. (1996). History of teacher pay and incentive reforms. *Journal of School Leadership, 6*(3), 265-89.

Ritter, G. W. & Jensen, N. C. (2010). The delicate task of developing an attractive merit pay plan for teachers. *Phi Delta Kappan, 37,* 32-37.

Schubert, R., Gysler, M., Brown, M., & Brachinger, H.W. (1999). Financial decision-making: Are women really more risk averse? *American Economic Review Papers and Proceedings, 89,* 381-385.

Wagner, T. (2001). Leadership for learning: An action theory of school change. *Phi Delta Kappan, 82,* 378–83.

Wilms, W. W. & Chapleau, R. R. (1999, November 3). The illusion of paying teachers for student performance. *Education Week.* Retrieved from http://www.edweek.org/ew/articles/1999/11/03/10wilms.h19.html.

Table 1: Lottery Choices

Lottery	Option A		Possible Roll(s) for Payout	Option B	
	☑	Payout		Payout	☑
1.		$6.00	1	$11.55	
		$4.80	2,3,4,5,6,7,8,9,10	$0.30	
2.		$6.00	1,2	$11.55	
		$4.80	3,4,5,6,7,8,9,10	$0.30	
3.		$6.00	1,2,3	$11.55	
		$4.80	4,5,6,7,8,9,10	$0.30	
4.		$6.00	1,2,3,4	$11.55	
		$4.80	5,6,7,8,9,10	$0.30	
5.		$6.00	1,2,3,4,5	$11.55	
		$4.80	6,7,8,9,10	$0.30	
6.		$6.00	1,2,3,4,5,6	$11.55	
		$4.80	7,8,9,10	$0.30	
7.		$6.00	1,2,3,4,5,6,7	$11.55	
		$4.80	8,9,10	$0.30	
8.		$6.00	1,2,3,4,5,6,7,8	$11.55	
		$4.80	9,10	$0.30	
9.		$6.00	1,2,3,4,5,6,7,8,9	$11.55	
		$4.80	10	$0.30	
10.		$6.00	1,2,3,4,5,6,7,8,9,10	$11.55	
		$4.80	---	$0.30	

Table 2: Descriptive Statistics

Item		non-MAT		MAT		Total	
		N	%	N	%	N	%
Gender							
	Male	41	61.2	16	24.6	57	43.2
	Female	26	38.8	49	75.4	75	56.8
Race/Ethnicity							
	White	46	68.7	62	95.4	108	81.8
	Black	5	7.5	---	---	5	3.8
	Asian	12	17.9	---	---	12	9.1
	Hispanic	2	3.0	1	1.5	3	2.3
	Other	2	3.0	2	3.1	4	3.0
Mother's Education							
	< H.S.	2	3.0	---	---	2	1.5
	H.S.	9	13.4	15	23.1	24	18.2
	Some College	13	19.4	16	24.6	29	22.0
	Bachelors	28	41.8	25	38.5	53	40.2
	Grad. Degree	15	22.4	9	13.4	24	18.2
Eating Out (per week)							
	Never	5	7.5	1	1.5	6	4.6
	1-2	37	55.2	37	56.9	74	56.1
	3-4	20	29.9	22	33.9	42	31.8
	5-6	4	6.0	1	1.5	5	3.8
	Daily	1	1.5	4	6.2	5	3.8
Car Value							
	< $1k	12	17.9	5	7.7	17	12.9
	$1k-$5k	23	34.3	23	35.4	46	34.8
	$5k-$10k	17	25.4	23	35.4	40	30.3
	$10k-$15k	8	11.9	6	9.2	14	10.6
	> $15k	7	10.4	8	12.3	15	11.3

Note: We test for gender and race/ethnicity differences using a chi-squared test. Participants in the MAT group are significantly more likely to be female and white (p-values < 0.01). Using a Mann-Whitney two-sample rank-sum test for both mother's education and number of times eating out and a t-test for car values, we find no statistically significant differences between the MAT and non-MAT groups on these items.

Table 3: Percentages of Confusion with the Risk Elicitation Tool

Cohort	Correct Comprehension Question	Consistent Responses	Correct and Consistent
MBA	76.7%	88.4%	74.4%
Law	91.7%	100%	91.7%
Elementary MAT	76.2%	85.7%	71.4%
Secondary MAT	81.4%	90.9%	76.7%
Non-MAT	82.1%	92.5%	80.6%
MAT	76.7%	89.2%	75.0%

Table 4: Average Number of Safe Choices by Category

Degree Type	All		Gender			
			Male		Female	
	N	Average	N	Average	N	Average
non-MAT	67	4.3 (1.4)	41	4.3 (1.6)	26	4.3 (1.3)
MAT	65	5.0 (1.9)	16	5.4 (2.1)	49	4.9 (1.8)

Note: Standard errors in parentheses.

Table 5: Statistical Comparison between Subgroups

	non-MAT Males	MAT Females
MAT Males	0.018, M-W 0.024, K-S	0.095, M-W 0.022, K-S
non-MAT Females	0.774, M-W >0.999, K-S	0.122, M-W 0.257, K-S

Note: Reported values are estimated p-values for the Mann-Whitney (M-W) and Kolmogorov-Smirnov (K-S) tests for differences in distributions.

Table 6: Regression Results for Number of Safe Choices

Explanatory Variable	(1)	(2)	(3)	(4)
MAT	0.70**	0.77**	1.03**	Positive &
	(0.29)	(0.32)	(0.51)	Significant
Female		-0.19	-0.01	Not
		(0.32)	(0.43)	Significant
Female*MAT Degree			-0.27	Not
			(0.66)	Significant
Minority			0.37	Not
			(0.42)	Significant
Age			0.02	Not
			(0.02)	Significant
Confusion			0.40	Not
			(0.37)	Significant
Constant	4.34***	4.42***	3.23***	Positive &
	(0.20)	(0.24)	(0.79)	Significant
Log Car Value (Proxy 1)				-0.04
				(0.06)
Never Eat Out (Proxy 2)				-0.57
				(0.76)
Eat Out 1-2 times (Proxy 2)				-0.18
				(0.33)
Eat Out 5-6 times (Proxy 2)				-0.42
				(0.84)
Eat Out Daily (Proxy 2)				-0.83
				(0.83)
Mother Edu - Less Than High School (Proxy 3)				0.93
				(1.32)
Mother Edu - Some College (Proxy 3)				-0.18
				(0.48)
Mother Edu - Bachelor's (Proxy 3)				-0.13
				(0.43)
Mother Edu - Grad. Degree (Proxy 3)				0.15
				(0.50)
N	132	132	132	132
R-Squared	0.04	0.05	0.07	---

Note: Numbers in parentheses below coefficient estimates are standard errors. ** and *** indicate significance at the 5%, and 1% p-value levels, respectively, in a two-sided alternative to the null that the coefficient value is 0.

Table 7: Average Ratings for Pay Systems

Scenario	MAT Students			non-MAT Students		
	Females	Males	All	Females	Males	All
Pay for individual performance	2.78 (1.65)	2.69 (1.54)	2.75 (1.61)	2.44 (1.87)	3.17 (0.97)	2.89 (1.42)
Pay for developing your knowledge and skills	2.37 (1.39)	2.44 (1.75)	2.38 (1.48)	2.24 (1.36)	2.41 (1.80)	2.35 (1.64)
Pay for team performance	-0.06 (2.59)	-1.13 (2.42)	-0.32 (2.57)	0.68 (2.56)	1.34 (2.09)	1.09 (2.29)
Pay not based on performance	0.18 (2.88)	-0.94 (2.70)	-0.09 (2.85)	-0.48 (2.42)	-0.49 (2.64)	-0.48 (2.54)

Note: Standard deviations in parentheses

Figure 1: Percentage of Safe Choices in Each Lottery by Degree Program

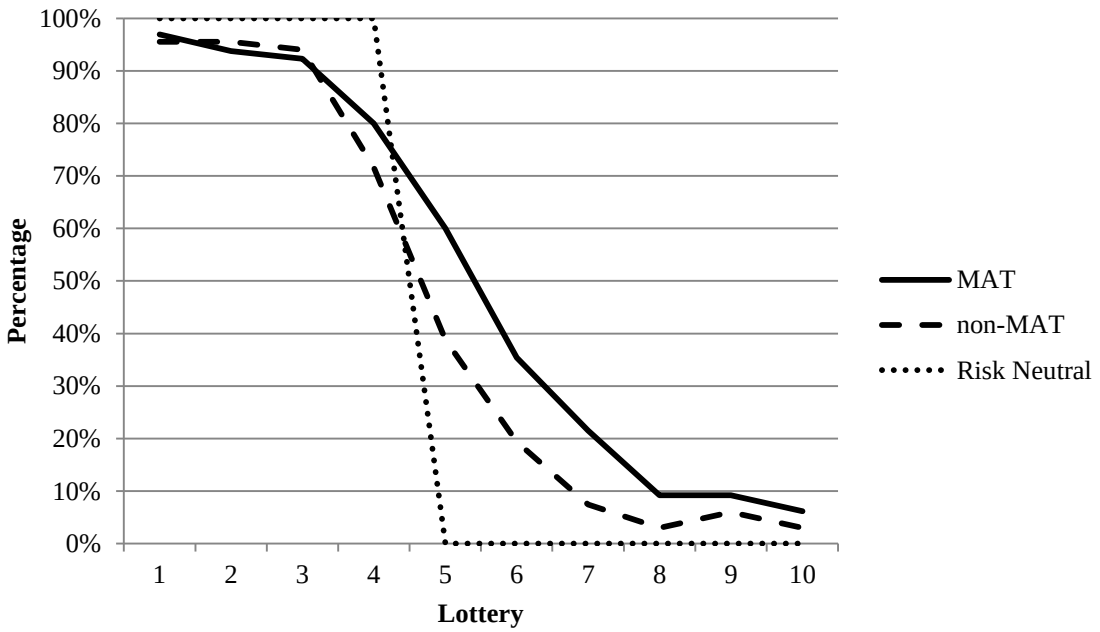


Figure 2: Percentage of Safe Choices in Each Lottery by Gender

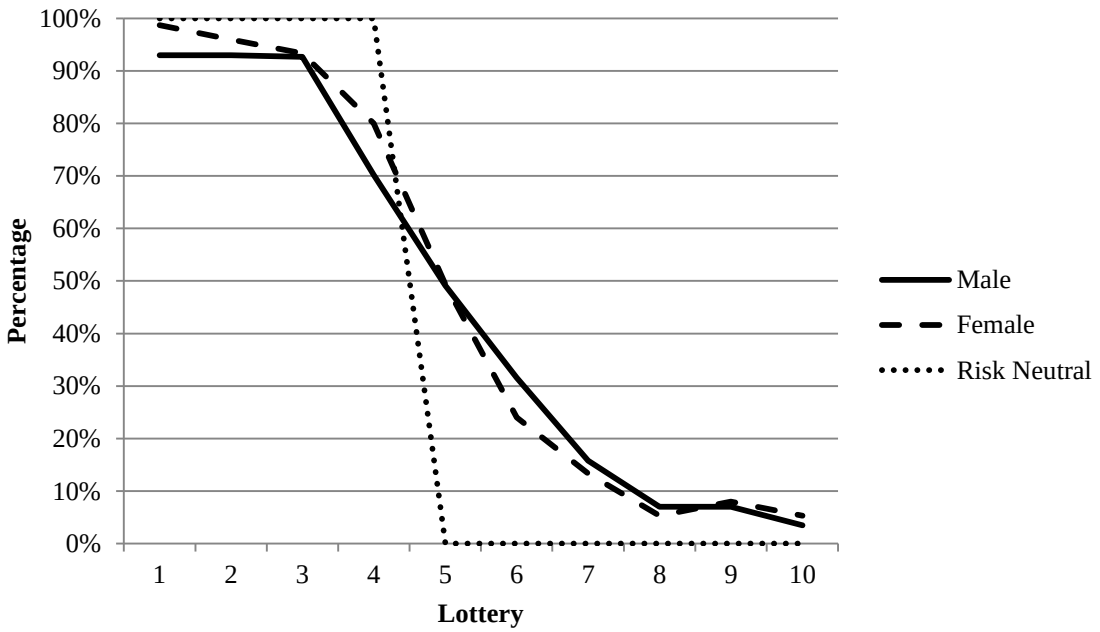
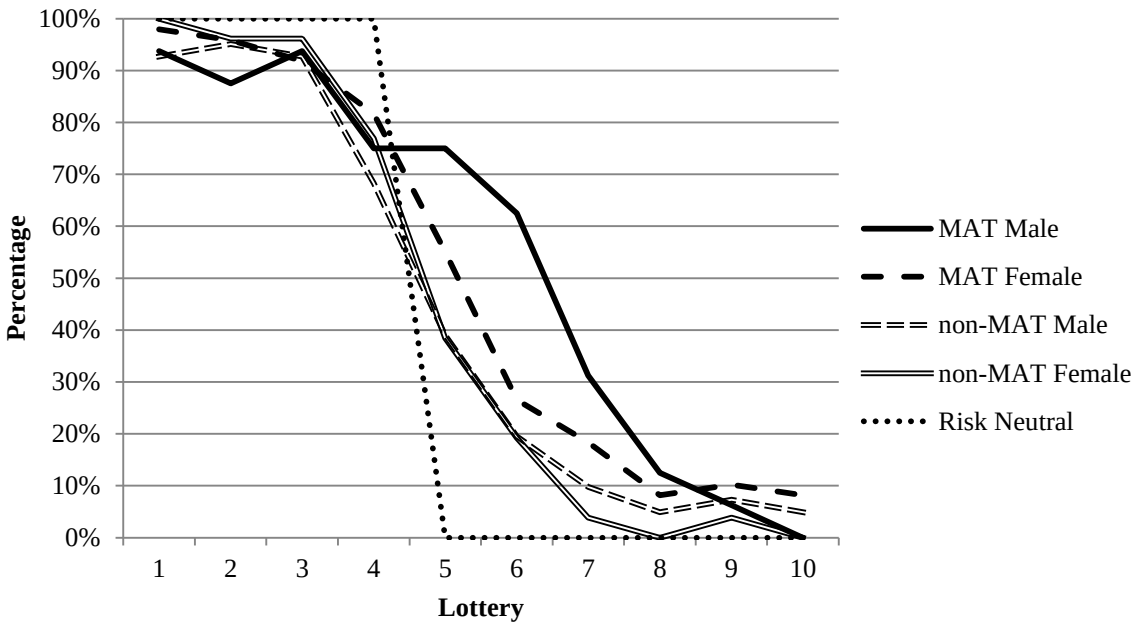


Figure 3: Percentage of Safe Choices in Each Lottery by Degree Program and Sex



Appendices

Note: Actual document presented to the participants was formatted to take up the front and back of only one page.

Please complete this page first, then proceed to the back to complete the study. Upon completion you will be paid in cash for your participation as described below.

For each of the 10 lottery pairs listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option, either A or B, for each lottery.

In each lottery pair you will be selecting between a lottery that will pay either $6.00 or $4.80 (Option A) and a lottery that will pay either $11.55 or $0.30 (Option B). In the first lottery there is a 10% chance of receiving the larger payout for lottery 1 and a 90% chance of receiving the smaller amount. In each subsequent lottery pair the chance of earning the higher payout increases by 10%.

After you complete this study the experimenter will roll a 10-sided die to randomly select which lottery will be used.

Next, the experimenter will roll the same die a second time to determine your actual payoff based on the option you chose for that particular lottery.

The number on the die for the second roll will determine whether you receive Payout 1 or Payout 2.

Example: If the experimenter's first roll is "3," then your payoff will based on Lottery 3. If you chose Option B for Lottery 3 and the second roll is "1", “2”, or “3”, you will receive a Payout of $11.55, but if the second roll is “4”, “5”, “6”, “7”, “8”, “9” or “10”, you will receive a Payout of $0.30.

Please answer the following question which will not impact your payoff but is intended to ensure you understand this task. Suppose the experimenter rolls a "2" first and then rolls a “9”. If you have selected Option A for Lottery 2, what will your payout be? _____

Please notify the experimenter when you have answered this question before you continue with the experiment.

Lotteries. For each of the 10 lotteries listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option per lottery.

Lottery	Option A		Possible Roll(s) for Payout	Option B	
	☑	Payout		Payout	☑
1.		$6.00	1	$11.55	
		$4.80	2,3,4,5,6,7,8,9,10	$0.30	
2.		$6.00	1,2	$11.55	
		$4.80	3,4,5,6,7,8,9,10	$0.30	
3.		$6.00	1,2,3	$11.55	
		$4.80	4,5,6,7,8,9,10	$0.30	
4.		$6.00	1,2,3,4	$11.55	
		$4.80	5,6,7,8,9,10	$0.30	
5.		$6.00	1,2,3,4,5	$11.55	
		$4.80	6,7,8,9,10	$0.30	
6.		$6.00	1,2,3,4,5,6	$11.55	
		$4.80	7,8,9,10	$0.30	
7.		$6.00	1,2,3,4,5,6,7	$11.55	
		$4.80	8,9,10	$0.30	
8.		$6.00	1,2,3,4,5,6,7,8	$11.55	
		$4.80	9,10	$0.30	
9.		$6.00	1,2,3,4,5,6,7,8,9	$11.55	
		$4.80	10	$0.30	
10.		$6.00	1,2,3,4,5,6,7,8,9,10	$11.55	
		$4.80	---	$0.30	

Pay Scenarios: Below are descriptions of four different methods of providing pay increases to workers. Please read each description, then circle the number that indicates how desirable or undesirable you would find that pay increase method for your first job in your chosen occupation or career field.

1. Pay for individual performance - In this system, you could get up to a 10 percent pay increase each year, depending on your individual job performance, as measured by objective factors. If your job performance was excellent, you would get a 10 percent increase; if very good, a 6 percent increase; if minimally acceptable, a 3 percent increase; and if poor no increase.

Highly Undesirable										Highly Desirable
	-4	-3	-2	-1	0	1	2	3	4	

2. Pay for developing your knowledge and skills - In this system, you could get up to a 10 percent pay increase each year, depending on how well you develop a specified body of knowledge and skills, as judged by your supervisor and a group of more experienced peers. If you developed all of the specified skills to a high level, you would receive a 10 percent pay increase. If you developed all of the skills to a satisfactory level, you would receive a 6 percent increase. If you developed some but not all of the skills, you would receive a 3 percent increase. If you did not develop any of the skills, you would get no increase.

Highly Undesirable										Highly Desirable
	-4	-3	-2	-1	0	1	2	3	4	

3. Pay for team performance - In this system, you would get up to a 10 percent pay increase each year depending on your team's performance. Performance would be measured by objective factors. If your team's performance was excellent, you would get a 10 percent increase; if very good, a 6 percent increase; if about average, a 3 percent increase; and if poor, no increase.

Highly Undesirable										Highly Desirable
	-4	-3	-2	-1	0	1	2	3	4	

4. Pay not based on Performance - In this system, you and all other workers would get a 5 percent pay increase each year regardless of how well or poorly you or your team performed, or how well you developed your skills, as long as you performed well enough to keep your job.

Highly Undesirable										Highly Desirable
	-4	-3	-2	-1	0	1	2	3	4	

Background Information

1. Year of Birth: _____ 2. Gender (circle please): Male Female

3. Race/Ethnicity (circle please): White Black Asian Hispanic Other (specify): _____

4. What graduate degree are you currently working towards? (circle please)
Ed.D. M.A.T. M.B.A M.Ed. Other (specify field and degree):_____

5. What is your mother's highest level of education? (circle please)
Less than H.S. H. S. Some College Bachelor's Degree Graduate Degree

6. On average, how many times a week do you eat at a restaurant off campus? (circle please)
None 1-2 3-4 5-6 Daily

7. What is the year, make, and model of your car? _____

8. What do you think the likelihood is that you go into teaching in a K-12 setting at some point in your life? (circle please)

No Chance (0%) Very Slim (1-25%) Modest (26-75%) Very Likely (76-99%) Guaranteed (100%)

College of Education and Health Professions
Department of Education Reform

201 Graduate Education Building
Fayetteville, Arkansas 72701
(479) 575-3172
(479) 575-3196 (FAX)

August 14, 2013

To: The Graduate School

Re: Confirmation of Authorship

This letter is to confirm that Daniel H. Bowen is the lead author of the article "Risky Business: An Analysis of Teacher Risk Preferences." He completed more than 51% of the work that went into this paper.

_____

Jay P. Greene, Ph.D.
Endowed Chair and Head of the Department of Education Reform

November 22, 2011

MEMORANDUM

TO:	Daniel Bowen Jonathan Mills James Shuls Jeremy Buck Cary Deck
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	11-11-239
Protocol Title:	*Risk Attitudes of Teachers in Training*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 11/22/2011 Expiration Date: 11/21/2012

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (http://vpred.uark.edu/210.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 200 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 210 Administration Building, 5-2208, or irb@uark.edu.

Does Merit Pay Attract Teachers with Different Risk Preferences?

By: Daniel H. Bowen & Jonathan N. Mills

Abstract

Prior research suggests that teachers are more risk averse than individuals with comparable levels of education. Given the relatively high job security and step and lane salary schedules prevalent in the teaching profession, this finding may not be particularly surprising. There is reason to believe that a more ambiguous pay system, based more on student performance, would attract more relatively risk-loving individuals while also deterring risk-averse teachers to the profession. This paper examines this hypothesis through the use of a risk-elicitation tool common in experimental economics. We compare the risk preferences of teachers who choose to work at schools with performance-based pay to teachers at the same schools that were hired either before the implementation of this pay program or claimed that performance pay was not a deciding factor in their employment decisions.

We find that teachers who actively choose schools with performance-pay programs are not significantly more risk loving; however, teachers who indicate that performance pay was a deciding factor in their employment decisions are significantly more risk loving. In addition to this relationship, we also find that number of years of teaching experience is a significant predictor of risk aversion. This finding is likely explained by the fact that individuals who remain in the same line of work for more extensive periods of time, regardless of profession, are probably less willing to take risks. While our finding does not indicate whether performance-based pay improves overall teacher quality, this finding provides some affirmation for the argument that reforming teachers' compensation format will alter the composition of the teacher workforce.

Introduction

Teacher quality is arguably the most important school input with regard to student achievement (Chetty, Friedman, & Rockoff, 2011; Hanushek & Rivkin, 2004; Rockoff, 2004). The importance of teacher quality has led some education researchers to question how schools can attract the best and the brightest to come and work in the classroom (e.g. Ballou & Podgursky, 1995; Finn, 2001; Fox, 1984; Wallis, 2008). Teacher performance-based pay (i.e. merit pay) is one potential policy solution that continues to attract attention and gain public support (Howell, Peterson, & West, 2011; Meckler, 2009; Resmovits, 2011). While performance-based pay could incentivize current teachers to increase their efforts, it also has the potential to attract a new, better crop of teachers to the workforce (Podgursky & Springer, 2007).

Career choice, by nature, is almost an entirely endogenous decision. Individuals want careers that maximize their utilities, and employers structure jobs and compensation packages in hopes of attracting the best possible employees (given their finite resources). Labor economists have provided empirical research demonstrating that particular parameters of a job's compensation packages can highly influence the matching of worker to employers (Salop & Salop, 1976). There is good reason to believe that teaching, as a profession, is not much different than other careers with regard to the importance of job and compensation structure as well as employee preferences and characteristics.

Compensation for teachers has traditionally been through a step and lane pay scale that weights salary according to years of experience and the earning of additional credentials and certifications such as master's degrees and a "Master Teacher Certification" (Podgursky, 2003; Sharpes, 1987). While an inherent advantage of this pay scale is that it can encourage those interested in the teaching profession to remain in the profession for longer, it comes at the

expense of failing to differentiate compensation according to a teacher's effectiveness in the classroom (National Research Council Committee on Performance Appraisal for Merit Pay, 1991). As a result, it is plausible that the structure of teacher compensation attracts individuals with a set of characteristics that differ from professions with different compensation packages.

Advocates for performance-based pay argue that compensation based on students' performances would better align teacher incentives with more desirable student outcomes (Dee & Keys, 2004; Podgursky & Springer, 2007). Moreover, some advocates also contend that changing pay formats will likely change the composition of the teacher workforce in hopes of this change bringing about improvements in the overall quality of the teacher labor pool (Woessman, 2011).

The research on performance-based pay has almost entirely focused on whether changing pay structure facilitates increases in the effort levels of current teachers, typically measured in terms of student growth scores on standardized assessments. While improving the productivity levels of the current teacher labor force is an important and potentially valuable outcome of performance pay, this line of research does not comprehensively examine the total impact of a performance-based pay system. The other important aspect is whether changing the structure of teacher compensation can transform the composition of the teacher workforce.

Beyond theoretical assumptions, little is known as to whether or not this compositional change would in fact take place within the teaching profession. An underlying issue with conducting such an evaluation is the fact that performance-based pay policies rarely exist long enough or are too small in scale to have a lasting impact on the workforce (e.g. the performance pay policies in Colorado, Tennessee, and Texas). However, when field studies are not possible,

laboratory-type research can provide a more controlled setting for testing certain underlying hypotheses of a policy.

In a prior study Bowen, Buck, Deck, Mills, & Shuls (2013) find evidence that prospective teachers are indeed more risk averse than individuals with comparable levels of education and backgrounds. They argue that this finding suggests that the structure of the K-12 teaching profession attracts individuals with certain characteristics, specifically those with lower preferences for risk. Therefore, it is likely that changing certain aspects of the profession would change the composition of the teacher workforce. Bowen et al. (2013) also do not find conclusive evidence for whether choice of profession or risk preference predict the desire for one particular compensation format over another. They conclude that this finding could be the result of a lack of real-life experience with different salary types. They also suggest that this finding could be attributed to the relatively low risk that a performance-based pay program poses to an entry-level employee.

In this study, we hope to further explore the relationship of how teacher quality reform policies might alter the composition of the teacher labor force. More specifically, we evaluate whether individuals drawn to teacher performance-based pay have different risk preferences than more traditional teachers. We use a risk-elicitation tool that is commonly used in experimental economics to directly measure teachers' risk preferences. We examine whether or not teachers who are explicitly drawn to schools that offer performance-based pay have different risk preferences than other teachers at the same school. By focusing on teachers at the same school, we eliminate many factors that might influence characteristics that go beyond the explicit desire to be compensated through performance-based pay. We also survey participants on key demographic information in order to control for characteristics such as teacher age, experience,

and subject area to try and determine the relationship that these characteristics have with regard to teacher preferences.

Literature Review

The majority of the research on the effectiveness of performance pay, to this point, has not addressed the issue of whether performance pay programs can alter the composition of the labor force (Lazear, 2004; Sawchuck, 2010). Dohmen and Falk (2010) refer to this alternative aspect of performance pay as the sorting component: "Agents with different individual characteristics and personality feel attracted by different types of incentives. In this sense providing incentives in firms or organisations has two important effects, an incentive effect *per se* and a selection effect. Importantly, these effects need not be complementary" (p. F256). In other words, performance pay programs have the potential to incentivize the current workforce to be more productive. However, reforming the structure of compensation can also incentivize changes in outcomes that come about as a result of a significant shift in the characteristics for the types of individuals drawn to a particular profession.

There substantial literature on the research that examines the effectiveness of performance pay as a motivator for the current teacher workforce (e.g. Dee & Keys, 2004; Springer et al., 2008; Springer et al., 2010). Based on the current body of evidence, performance pay tends to range from having statistically significant effects in both directions as well as no impact with regard to increasing students' test scores (e.g. Figlio & Kenny, 2007; Podgursky & Springer, 2007; Sawchuck, 2010; Springer et al., 2008; Springer et al., 2010). There are several plausible explanations for why performance pay programs have not produced the desired results (e.g. unions co-opting reward design; rewards not large enough to induce significantly higher levels of effort; etc.). Yuan et al. (2013) provide empirical evidence for one commonly held

belief for the lack of substantial performance pay results. They find that teachers do not adapt their practices or increase efforts when a performance pay program is implemented. While this finding could be interpreted as meaning that teachers are stuck in their ways and unwilling to change how they teach, an alternative explanation is that teachers already maximize their efforts; therefore, it is difficult for them to increase their efforts due to ceiling effect (Gratz, 2009). If performance-based pay programs are found to be unsuccessful when it comes to increasing student achievement, then the success of performance pay programs would entirely depend on whether they can significantly change the composition of the teacher workforce.

Different types of pay or incentive schemes will attract potential employees according to characteristics and skills (Salop & Salop, 1976). Oyer and Schaefer (2005) provide the example of a private sector firm offering lower-level employees stock options. New businesses will often choose to compensate employees with stock options because they tend to lack the resources necessary for offering higher wages. While cost convenient for a new firm, such a policy can also induce a selection process that screens on characteristics that a particular business finds desirable. Employees that select jobs where employers provide stock options over a more fixed salary are more likely to feel optimistic about the prospects of the business, develop a sense that they have more "skin in the game," and demonstrate less risk-averse tendencies. Assuming that an employer desires these traits, providing an alternative form of compensation provides a desirable sorting mechanism. Lazear (2000) offers empirical evidence supporting personnel economics theory with data from an auto glass company. He finds that, after a shift to a piece-rate compensation system, workers become significantly more productive and that, over time, this change begins to positively influence the composition of the firm's applicant pool.

Public sector jobs often come with tenure or civil service protections that are more protective than the private sector. Economists often find that these public sector workers tend to be more risk averse than those in the private sector (Dohmen et al., 2005; Hartog, Ferrer-i-Carbonell, & Jonker, 2002; Masclet, Colombier, Denant-Boemont, & Loheac, 2009). Using Panel Study of Income Dynamics (PSID) data, Bellante and Link (1981) conduct surveys and find that risk aversion was a significant predictor of employment in the public sector. Buurman, Delfgaauw, Dur, and Van den Bossche (2012) analyze the survey data of more than 3,000 employees in the Netherlands; employees were offered a reward for completing the survey and offered a choice between a gift certificate, a national lottery ticket, or donating the money to charity. After controlling for some key demographics, they find that public sector workers' odds of choosing the lottery ticket were 0.68 times the odds of a private sector worker.

Dohmen, Falk, Huffman, and Sunde (2010) provide suggestive evidence regarding teacher risk attitudes. In one experiment, they asked subjects to complete as many multiplication problems as they could in a 10-minute period. After this task, subjects were asked to choose either a fixed compensation or a payment that varied with individual success at getting a subsequent round of multiplication problems correct. Dohmen et al. find that higher-performing subjects are more likely to choose a variable compensation method. They also find that, even controlling for performance on the task, women were 20% less likely to choose the variable pay scheme. This finding suggests that women tend to not prefer competition or greater uncertainty when it comes to compensation. This finding is potentially relevant to teaching, which is primarily composed of females.

Perez (2011) conducts an experiment, examining risk preferences, attitudes towards pay inequity, and preferences toward competition to assess whether changing teacher pay might

change the composition of the teacher labor force. In the study, she compares female candidates in the Stanford Teaching Education Program with female law students. Participants completed a series of 10-minute rounds of solving mazes and made decisions with regard to compensation for subsequent rounds of puzzle solving. Noting that similar numbers of prospective teachers and lawyers chose a tournament pay scheme, Perez concludes "teachers do not exhibit a higher level of risk aversion [or] a dislike for competition that is higher than female lawyers" (p. 18-19). However, teachers were significantly more likely to be averse to pay inequity, meaning they tended to like differences in pay between individuals less than women in law school. Perez claims that this finding could show that teachers have fundamental inclinations towards greater pay equity and that policymakers should take this inclination into consideration.

Finally, Bowen et al. (2013) conduct an experiment comparing the risk preferences of teacher candidates to law school students and students pursuing a Masters of Business Administration. They find that these teacher candidates are significantly more risk averse than the comparison group. The authors argue that this finding is likely due to a sorting effect that could be attributed to the general structure of the teaching profession. Bowen et al. conclude that policymakers should take this finding into account when designing education policies that may be at odds with the characteristics and preferences of the general teacher workforce, especially when these policies aim to influence the behaviors of the current workforce.

Methods

Risk Task

To directly measure the risk attitudes of those teachers explicitly choosing performance-based pay as opposed to other teachers in the same schools, we use a procedure developed by Holt and Laury (2002). This tool is a well-known risk-elicitation procedure used in the

experimental economics literature (e.g. Anderson, Harrison, Lau, & Rutstrom, 2008; Dohmen et al., 2010; Eckel & Wilson, 2004). While several experimental risk-elicitation tools are available (e.g. Eckel & Grossman, 2002), this procedure has become a generally accepted standard. This status is due in part to the ease with which it can be implemented and explained to study participants. Harrison, Johnson, McInnes, and Rutstrom (2005) have found this procedure to have good retest reliability over time. Anderson and Mellor (2008) have also found that higher levels of risk aversion on the Holt and Laury task are negatively and significantly related to failure to use a seatbelt, cigarette smoking, being overweight or obese, and heavy drinking.

The task presented in Table 1 is our version of the Holt and Laury (2002) task that was provided to the participants of this study. Participants were asked to choose between options A and B for 10 different lotteries. There are two possible payouts for each option, and these payouts are fixed across all lotteries. The payouts for Option A are $4.80 or $6.00, and the payouts for Option B are $0.30 or $11.55. We refer to Option A as the "safe" choice because it has less disparity in the potential payouts than Option B. As the participant progresses from one lottery to the next the likelihood of receiving the higher of the two payouts for both options increases by 10 percentage points. Since the likelihood of receiving the higher payout consistently increases as one progresses through the different lotteries, a consistent individual will choose Option A for the first *i* lotteries and then after making the switch to Option B will stick with this option for the remaining lotteries. Participants were instructed that their payout would ultimately be determined by their choice for one of the lotteries and that this lottery would be determined by the roll of a ten-sided die.

The expected payout is determined by multiplying the two possible payouts within an option by the likelihood of receiving each payout and then summing these values. A risk-neutral

individual always chooses the option with the higher expected payout. Option A has a higher expected payout for the first four lotteries; Option B has a higher expected payout for the remaining six lotteries. The switching point from Option A to Option B identifies a range of risk parameters that are consistent with the observed choice (see Holt & Laury, 2002). Risk-loving individuals will choose Option B before the fifth lottery because the possibility of receiving the high payout is valued more than the difference between the expected payouts of Options A and B. Risk-averse individuals will not switch to Option B before the sixth lottery because their value of reducing uncertainty is greater than the differences in the expected payouts. An extremely risk-averse individual may choose to disproportionately stick with Option A to ensure at least a $4.80 payout, but all participants should choose Option B for lottery 10 because at this point the higher payout is guaranteed.

The Holt and Laury (2002) procedure not only provides rich information on individual preferences, it also allows us to check the extent to which subject confusion exists in our data: rational individuals should never select Option A on Lottery X after having selected Option B for Lottery Y if X>Y. Multiple switches could serve as one indicator of participant confusion. Rational individuals should also choose Option B for the tenth lottery because of the guaranteed higher payout. Therefore, multiple switches and choosing Option A for the tenth lottery serve as indicators of participant confusion.[6]

[6] It is worth noting that there is anecdotal evidence that some participants may have made seemingly irrational choices that are not the result of confusion. For example, one subject seemed to fully comprehend the Holt and Laury procedure but still made multiple switches. Participants were never asked to explain their choices, but this participant stated that the decision to switch over to Option B for two of the lotteries before switching back to Option A was attributed to those lotteries being her “lucky numbers”.

Risk Elicitation Procedures

A total of 120 teachers participated in this study. All participants were employed at two different school districts which have implemented a performance-based pay program. One district is a charter school that had a performance-based pay program in place between the 2008-09 and 2011-12 school years.[7] The other district is a traditional public school district that implemented performance pay in 2009-10 and still has the program in place as of the 2012-13 school year. Of the 120 participants, 48 (40%) are designated as having opted in to a performance-based pay system. These 48 participants decided to work in a school with a performance pay program in place. Of these 48 participants, 13 (27%) indicated that performance-based pay was a "deciding factor" in their decision to seek employment at their current school.[8]

To conduct the study, we first obtained permission from the school administrators to offer their teachers the opportunity to participate in a paid research study that examines how individuals make economic decisions. Teachers had the opportunity to participate during their planning periods, lunch breaks, or at the end of the day. Teachers were notified about the opportunity to participate in this study either the day before or the morning of the study.

When the teachers arrived, the researcher greeted the participant, briefly described the study as being about how different people make economic decisions, and then provided a handout of the participation consent form. Teachers were made aware that they were free to

[7] This school stopped performance-based pay because of lack of funding, not because of an administrative or faculty decision.

[8] Two participants that did not opt in to the performance pay systems also indicated that this policy was a deciding factor. While these teachers may have chosen to stay at their current schools, rather than seek employment elsewhere because of this policy, we do not include them in as making performance-based pay a "deciding factor".

leave at any point since participation was completely optional and voluntary.[9] If they agreed to still partake, participants were then given a single, two-sided sheet of paper with instructions. On the front side was a sample question to verify that the subject understood the Holt and Laury task in addition to the actual risk-elicitation instrument. Survey questions were located on the back side of the paper (see Appendix).

Participants were asked to read the directions, complete the comprehension question, and then notify the researcher upon completion in order to verify that the participant understood how the elicitation procedure worked before the rest of the survey was completed. If the subject completed the sample question correctly, the monitor marked the form in a particular location on the paper. If the subject got the sample question incorrect or requested further explanation, then the monitor marked the survey form in a different location to indicate a lack of initial comprehension. Marking the surveys in this way allowed us to discreetly track which subjects experienced at least some initial confusion about the procedure.

After the monitor verified that the participant understood the instructions, the participant was then asked to complete the risk-elicitation task before proceeding to the other side of the paper to complete a brief survey. Participants were aware the survey existed before completing the risk-elicitation task, but they did not know its contents. The survey was completed after the risk-elicitation task in order to prevent the possibility that these questions could frame or influence subjects' behaviors while indicating their lottery preferences. Researchers verified that participants completed the entire form and tried to ensure that subjects did not talk or otherwise communicate with each other while completing the study.

[9] Three teachers who showed initial interest in taking part in the study ultimately decided not to participate.

The survey included background information to control for and examine other potentially salient characteristics. On top of age, gender, income level, undergraduate institution, work history, and mother's level of education, we also asked participants about their number of years of teaching experience, the year they were hired to work at the school, and their opinions and preferences with regard to performance pay.

Upon completion of the study, participants notified the researcher. The researchers first verified that subjects completed all questions, and then rolled a ten-sided die to determine which lottery choice would be used for determining the participant's payment.[10] Then, the ten-sided die was rolled again to determine the actual payment according to the subject's choice of Option A or B for the randomly chosen lottery. Participants then signed a participant payment form to verify their payout amounts and were thanked for their participation in the study. On average, participants earned $6.29 for about fifteen minutes of their time.

Sample

Before examining teacher behavior, we first examine the descriptive statistics of our sample. Our sample consists of 120 teachers, 50 from a charter school and 70 from a traditional public school district. These two school campuses consisted of different school buildings for different school grade levels. Of the teachers in our sample, 43% taught at an elementary school, 23% at a middle school, and 34% at a high school. The average participant was 40 years old and had 11 years of teaching experience. The overwhelming majority of the participants were female (80%) and white (93%). The majority (68%) of the teachers were instructing in core subject areas (i.e. mathematics, science, language arts, and/or social studies). The other teachers either

[10] Randomly selecting one task for payment is a common approach in experiments where the researcher wants to control for potential wealth effects (see e.g. Holt & Laury, 2002).

were strictly involved in special education (9%) or taught in a non-core subject area (e.g. physical education, art, foreign language, etc.) (24%).

We also examine descriptive statistics for our groups of interest: teachers who were employed with the performance pay program in place and teachers who indicated that performance-based pay was a deciding factor in their employment decision. There are some significant differences in these groups worth noting. Teachers who opted in as well as those who indicated performance pay as a deciding factor were significantly younger and less experienced. These teachers were also much more likely to live in households that were making less than $50,000 a year and have mothers who had earned at least an undergraduate degree. Finally, these teachers were much more likely to be employed in the charter school district.

These differences are worth noting but are not all that surprising. Teachers employed during the relatively recent implementation of the performance pay program are almost certainly going to be younger and less experienced and therefore have lower household incomes. There are no significant differences in gender, race, subject area, and grade level. See Table 2 for the sample descriptive statistics.

Results

The primary research question for this study is whether teachers who are drawn to performance-based pay systems differ in their risk preferences from other teachers. Our data allow us to answer this question in two different ways: examining individuals who opted in to the performance pay program and examining individuals who indicated that performance-based pay was a deciding factor in their employment decision. The results include all the participants from our study; we do not exclude participants who showed initial comprehension issues or inconsistencies. This allows us to use our entire data set. We have also conducted all analyses

with these participants excluded from our data set. Point estimates differ between these sets of analyses; however, the estimate signs and significances generally remain unchanged.[11]

First, we focus on individuals who indicated that they opted in to the performance pay program. In our sample, 48 (40%) of the participants made this selection.[12] In order to test for differences in risk aversion, we use the number of safe choices (Option A) as the measure of a participant's level of risk preference. Looking strictly at averages, participants who chose to work in a school with performance-based pay were essentially no different than the other teachers. In fact, teachers that opted in to the performance pay program made 0.11 more safe choices than those that did not select into the program. Results from a t-test for differences in means fails to reject the hypothesis that the two groups are similar on average (p-value = 0.81).

We also compare the two groups by looking at the full distributions of their choices. Figure 1 presents a comparison of the distribution of safe choices between participants who opt in to performance pay programs and those who do not. The dotted line in Figure 1 represents the choice distribution for a perfectly risk neutral group. The further the line remains above the dotted line after the 5th lottery, the more risk averse the group. Figure 1 clearly indicates that teachers who opted into the performance pay program are not significantly different than those teachers who did not opt in to this pay program. Indeed, a Kolmogorov-Smirnov test for distributional differences (p-value = 0.77) as well as a Mann-Whitney test of rank-sum differences (p-value = 0.98) both fail to reject the hypothesis of no significant differences.

[11] The coefficients on opting in and performance pay being a determining factor generally get larger when removing these 24 participants (20% of the total sample). However, significances still generally remain unchanged. The exceptions are with the fourth through seventh specifications. When removing participants deemed inconsistent, the significances on determining factor go from being significant at 99% confidence level to significant at the 95% confidence level (p-values = 0.05, 0.03, 0.02, and 0.03, respectively).

[12] One of the "opt in" participants is excluded from analyses because she did not indicate a preference for all ten of the lotteries.

Next, we test for whether there are differences for participants who indicate that performance-based pay was a deciding factor in their employment decisions. It could be the case that teachers identified as having "opted in" to schools with these programs were not actively seeking a school that offered performance pay. Figure 2 provides a comparison of the distribution of safe choices for these two groups of teachers. Interestingly, the distribution for teachers who considered performance pay to be a deciding factor tends to be lower relative to other teachers, a finding that would indicate that these teachers are more risk preferring on average. Indeed, the 13 participants who indicated that performance pay was a deciding factor in their decision to work at their school made 1.33 fewer safe choices on average.

The difference in these averages is significant at the 90 percent confidence level (p-value = 0.06). A Mann-Whitney rank-sum test of differences also rejects the hypothesis of no significant differences between groups at the 95 percent confidence level (p-value = 0.03). Despite evidence of significantly different means and visual evidence indicating that these teachers are more risk averse than other teachers, we are unable to reject the null hypothesis that the two groups of teachers do not share the same distribution (Kolmogorov-Smirnov p-value = 0.48). However, these results may be driven by having such a small number of teachers indicating that the performance pay program was a deciding factor, making these results still inconclusive.

The analyses presented so far have focused on differences in groups without controlling for additional demographic factors. Table 3 presents the results of 7 ordinary least squares (OLS) regressions that account for such differences. Column (1) indicates that teachers who opt in are no different than those who do not; however those who have opted in and indicate that merit pay was a deciding factor chose 1.7 fewer safe options than their colleagues (p-value = 0.03).

The model specification for column (2) analyzes how other factors associated with risk preferences (gender, age, race, experience, and school district) are associated with the number of safe choices. The sign for gender and experience coincide with intuition and prior research: females are more likely to choose more safe options and individuals with more teaching experience are more likely to choose the safer option. Interestingly, gender is not statistically significant in any of the specifications in Table 3. The sign for age is also somewhat surprising: older individuals choose slightly fewer safe options on average. This is explained by the fact that this specification additionally controls for experience, so teachers who enter the profession at older ages are slightly more risk preferring than teachers who entered the profession at a younger age. Furthermore, the estimated coefficient is not statistically distinguishable from zero.

The third specification examines how these demographics are associated with risk preferences after accounting for individual opt-in status. The only variables that are significantly associated with the number of safe choices one makes are the deciding factor variable and experience. Column (4) includes controls for mother's education and a postsecondary quality measure; column (5) controls for current household income, and the sixth specification includes controls for the teacher's subject area, grade level, and whether they have always been a teacher as opposed to transferring from another, non-education profession. None of these additional variables are significantly related to risk preferences. The only variables consistently related to the number of safe choices are whether the performance pay program was a deciding factor and years of experience.

In our final specification, we examine how these variables are related to risk preferences while accounting for all of these characteristics. Again, we find that only the indication that performance pay is a deciding factor in one's employment decision and teaching experience are

significantly related to the number of safe choices one makes. In particular, teachers who indicate that performance pay was a deciding factor make significantly fewer safe choices than other teachers who opted in to the program, and more experienced teachers make significantly more safe choices than other teachers.

Thus far, we have found that teachers who opt into a performance pay program are not necessarily less risk averse than other teachers; however, those who considered the performance pay program to be a deciding factor are somewhat more risk preferring. We have additionally found evidence that, controlling for more experience is associated with making more safe choices. In particular, we find that for each additional 10 years of teaching experience, participants make 0.7 more safe choices. This result is intuitive since individuals that stick with their jobs for longer periods, regardless of the occupation, are likely more risk averse.

Discussion & Conclusion

In this paper, we investigate if there are differences in risk preferences between teachers who opt into performance pay programs and those who do not. Given the relatively risky nature of a performance-based pay program compared to the traditional salary schedule, we expected to find teachers who voluntarily opted into such programs to be more risk preferring than those who did not. Surprisingly, our data do not support this conclusion: teachers who opted into the performance pay programs were not significantly more risk preferring than other teachers.

On the other hand, we do find evidence that teachers who indicate that the performance pay program was a deciding factor in their decision to work at the school are significantly more risk preferring than other teachers. It could be argued that these performance-preferring teachers are indeed those that would be drawn to a long-term performance pay program as opposed to the teachers who decide to work at a school that happens to offer performance pay. Nevertheless,

these teachers represent a relatively small subset of our entire sample (11%), so we believe this finding requires further exploration in order to be more conclusive.

Our other finding, that experience is positively related to risk aversion, is probably not that surprising. Intuitively, individuals who are more likely to remain with the same job for a longer period are going to be more risk averse. The fact that experience is a predictor of risk aversion that is independent of age makes it difficult to determine whether teacher risk aversion is something that is developed as a result of socialization. Based on the results from Bowen et al. (2013), we do not believe that socialization explains teachers' risk aversion. It is likely that job experience, regardless of profession, is positively associated with risk aversion.

While the size and nature of this study raise concerns with regard to external validity issues, our findings provide some insights into the issue of whether to expect compositional workforce changes as a result of implementing a performance-based pay program in the short term. It appears that some teachers have strong preferences for performance pay programs, and we further find evidence that these teachers may be more risk loving than other teachers. However, even though this study attempts to examine a causal relationship by taking advantage of the relatively exogenous implementation of the performance pay program, it is difficult to determine the accuracy of participants' claims that the program was a "deciding factor". In other words, performance pay may have been much more of a deciding factor for some of these participants than was indicated. If a teacher was initially excited about the prospect of performance pay and then had a bad experience, hindsight bias might influence how they answer that survey question. The same could be true for participants who indicated that performance pay was a deciding factor. These participants may not have been seeking performance pay, but positive experiences over the years might bias how they look back on their employment decision.

These results provide some preliminary insights into how changing teacher compensation structures can impact the composition of the teaching workforce. However, these findings do not speak to the differences in the quality between teachers who select into performance pay programs and those who do not as well as whether risk aversion predicts teacher effectiveness. This question is especially important because finding that performance pay alters the composition of the teacher workforce only tells half of the story. The other important question with regard to the shift in the composition of the labor pool is whether performance pay can attract higher quality teachers without producing a net loss in the quality of teachers currently drawn to the profession. This study supports the argument that performance-based pay would likely attract a fundamentally different group of individuals to the profession; in a subsequent study, I attempt to examine whether these teachers are more effective in the classroom.

References

Anderson, S., Harrison, G. W., Lau, M. I., & Rutstrom, E. E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583-618.

Anderson, L.R., & Mellor, J.M. (2009). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260-1274.

Ballou, D., & Podgursky, M. (1995). Recruiting smarter teachers. *The Journal of Human Resources, 30* (2), 326-338.

Bellante, D. & Link, A.N. (1981). Are public sector workers more risk averse than private sector workers? *Industrial and Labor Relations Review, 34*(3), 408-12.

Bowen, D. H., Buck, S., Deck, C., Mills, J. N., & Shuls, J. V. (2013). Risky Business: An Experimental Analysis of Teacher Risk Preferences (Working Paper No. 2013-01). Retrieved from University of Arkansas Department of Education Reform Working Papers site: http://www.uaedreform.org/working-papers/

Buurman, M., Delfgaauw, J., Dur, R., & Van den Bossche, S. (2012). Public sector employees: Risk averse and altruistic? *Journal of Economic Behavior & Organization 83*(3), 279-291.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (NBER Working Paper No. 17699). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w17699

Dee, T., & Keys, B.J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management, 23*, 471-488.

Dohmen, T. & Falk, A. (2010). You get what you pay for: Incentives and selection in the education system. *The Economic Journal, 120*(546), F256-F271.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G.G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey (IZA Discussion Paper No. 1730). Retrieved from the Institute for the Study of Labor (IZA) website: http://ftp.iza.org/dp1730.pdf

Eckel, C. C. & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23,* 281-295.

Eckel, C. C. & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization, 55,* 447-465.

Figlio, D., & Kenny, L. (2007). Individual teacher incentives and student performance. *Journal of Public Economics, 91*(5-6), 901-914.

Finn, Jr. C.E. (2001). Getting better teachers—and treating them right. In T. M. Moe (Ed.) *A primer on America's schools* (pp. 127-150). Stanford, CA: Hoover Institute.

Fox, J.N. (1984). Restructuring the teacher work force to attract the best and the brightest. *Journal of Education Finance, 10* (2), 214-237.

Gratz, D.B. (2009). The problem with performance pay. *Educational Leadership, 67*(3), 76-79.

Hanushek, E. A., & Rivkin, S.G. (2004). How to improve the supply of high quality teachers. Prepared for the Brookings Papers on Education Policy. Retrieved from Stanford website: http://hanushek.stanford.edu/sites/default/files/publications/hanushek%2Brivkin%202004%20BroPapEdu.pdf

Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1(1), 31-35.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking measured risk aversion to individual characteristics. *Kyklos, 55*, 3-26.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655.

Howell, W. G., West, M. R., & Peterson, P. E. (2011). The public weighs in on school reform. *Education Next, 11*(4), 10-22.

Lazear, E.P. (2004). Output-based pay: Incentives, retention or sorting? *Research in Labor Economics, 23*, 1-25.

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review, 90*, 1346-1361.

Masclet, D., Colombier, N., Denant-Boemont, L., & Loheac. Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization, 70,* 470-84.

Meckler, L. (2009, March 11). Education push includes merit pay. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB123668036405881929.html

National Research Council (U.S.) Committee on Performance Appraisal for Merit Pay (1991). *Pay for performance: Evaluating performance appraisal and merit pay*. Washington, DC: National Academy Press

Oyer, P., & Schaefer, S. (2005). Why do some firms give stock options to all employees? An empirical examination of alternative theories. *Journal of Financial Economics, 76*, 99-133.

Perez, M. (2011). Inequity aversion differences: Experimental evidence among prospective teachers and lawyers (July 8, 2011). Available at Social Science Research Network: http://ssrn.com/abstract=2018697

Podgursky, M. (2003). Fringe benefits: There is more to compensation than a teacher's salary. *Education Next, 3*(3), 71-76.

Podgursky, M., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26* (4), 909-949.

Resmovits, J. (2011, July 29). Arne Duncan boosts merit pay at teaching conference. *The Huffington Post*. Retrieved from: http://www.huffingtonpost.com/2011/07/29/duncan-boosts-merit-pay-a_n_913608.html

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*, 247-252.

Salop, J., & Salop, S. (1976). Self-selection and turnover in the labor market. *Quarterly Journal of Economics, 90*, 619-627.

Sawchuck, S. (2010, September 9). Merit pay found to have little effect on achievement. *Education Week*. Retrieved from: http://www.edweek.org/ew/articles/2010/09/21/05pay_ep.h30.html

Sharpes, D. (1987). Incentive pay and the promotion of teaching proficiencies. *The Clearing House 60*, 406-408.

Springer, M.G., Ballou, D., Hamilton, L., Le, V., Lockwood, J.R., McCaffrey, D., Pepper, M., & Stecher, B. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, M.G., Podgursky, M., Lewis, J.L., Ehlert, M.W., Gosh-Dastidar, B., Gronberg, T.J., Hamilton, L.S., Jansen, D., Lopez, O., Patterson, C.H., Stecher, B.M., & Taylor, L.L. (2008). Governor Educator Excellence Grant (GEEG) Program: Year one evaluation report. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Wallis, C. (2008, February 13). How to make great teachers. *TIME*. Retrieved from http://www.time.com/time/magazine/article/0,9171,1713473,00.html

Woessman, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review, 30*(3), 404-418.

Yuan, K., Vi-Nhuan, L., McCaffrey, D.F., Marsh, J.A., Hamilton, L.S., Stecher, B.M., & Springer, M.G. (2013). Incentive pay programs do not affect teacher motivation or reported practices: Results from three randomized studies. *Educational Evaluation and Policy Analysis, 35*(1), 3-22.

Figure 1: Percentage of Safe Choices in Each Lottery by Opt In Status

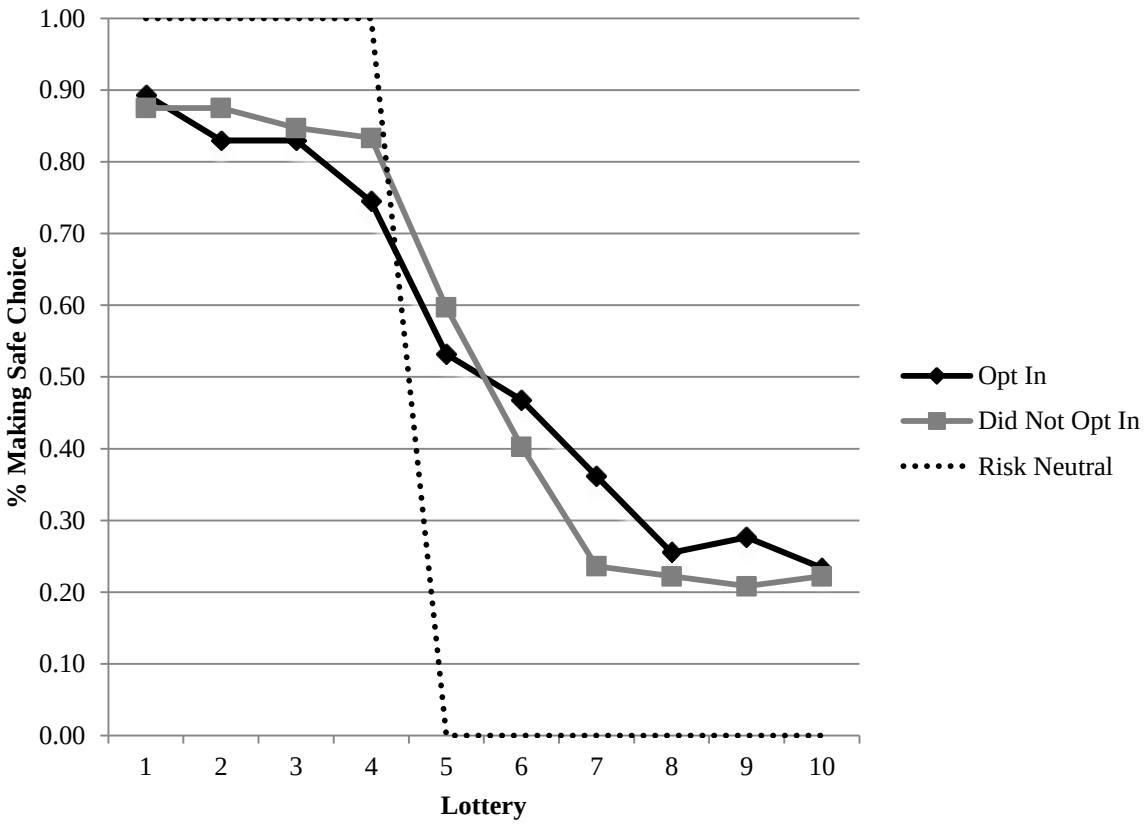


Figure 2: Percentage of Safe Choices in Each Lottery, Performance Pay as a Deciding Factor

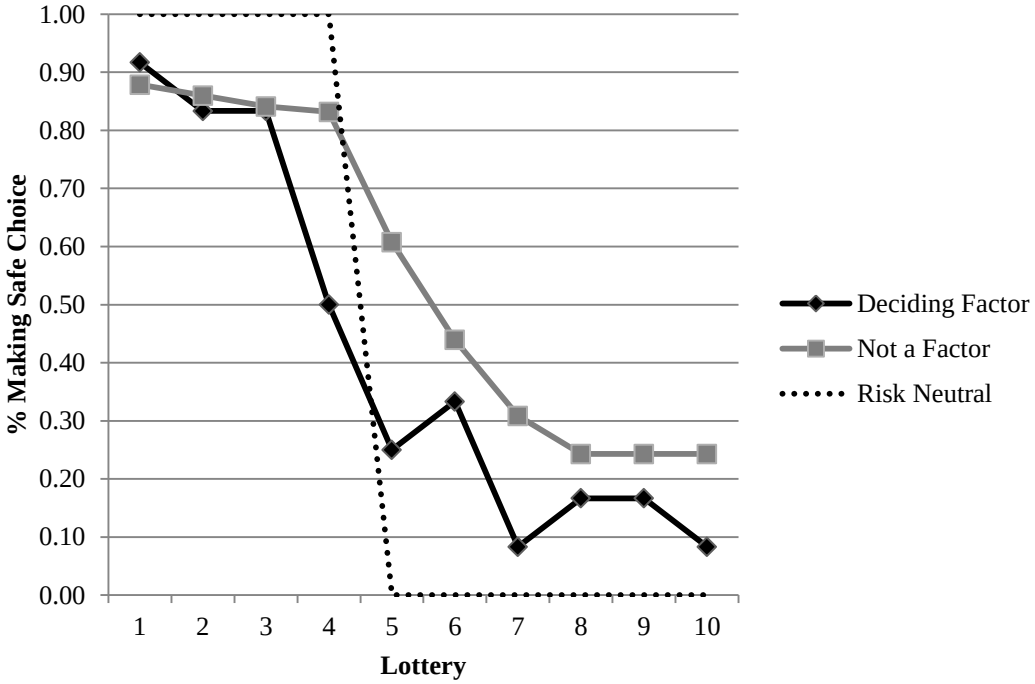


Table 1: Lottery Choices

Lottery	Option A		Possible Roll(s) for Payout	Option B	
	☑	Payout		Payout	☑
1.		$6.00	1	$11.55	
		$4.80	2,3,4,5,6,7,8,9,10	$0.30	
2.		$6.00	1,2	$11.55	
		$4.80	3,4,5,6,7,8,9,10	$0.30	
3.		$6.00	1,2,3	$11.55	
		$4.80	4,5,6,7,8,9,10	$0.30	
4.		$6.00	1,2,3,4	$11.55	
		$4.80	5,6,7,8,9,10	$0.30	
5.		$6.00	1,2,3,4,5	$11.55	
		$4.80	6,7,8,9,10	$0.30	
6.		$6.00	1,2,3,4,5,6	$11.55	
		$4.80	7,8,9,10	$0.30	
7.		$6.00	1,2,3,4,5,6,7	$11.55	
		$4.80	8,9,10	$0.30	
8.		$6.00	1,2,3,4,5,6,7,8	$11.55	
		$4.80	9,10	$0.30	
9.		$6.00	1,2,3,4,5,6,7,8,9	$11.55	
		$4.80	10	$0.30	
10.		$6.00	1,2,3,4,5,6,7,8,9,10	$11.55	
		$4.80	---	$0.30	

Table 2: Sample Descriptive Statistics

Variable	Overall Mean	Opted In	Opted In + Deciding Factor
N	120	48	13
Age	40.2	35.2	32.9
Experience	11.0	8.0	6.9
Female	80.0%	81.2%	84.6%
White	92.5%	91.7%	92.3%
Postsec. Avg. ACT	22.7	22.8	22.9
Mother's Education			
Less Than HS	5.8%	2.1%	---
HS Degree	29.2%	16.7%	7.7%
Some College	23.3%	22.9%	---
Undergraduate Degree	26.7%	41.7%	61.5%
Graduate Degree	15.0%	16.7%	30.8%
Household Income			
Less Than $50k	30.0%	37.5%	53.8%
$51k-$80k	27.5%	31.3%	23.1%
$81k-$110k	28.3%	22.9%	23.1%
More Than $110k	12.5%	8.3%	---
Teaching Subject Area*			
General Education	14.5%	19.1%	23.1%
Core Subject	68.3%	76.6%	84.6%
Math/Science	32.5%	31.9%	30.8%
Humanities	23.1%	27.7%	38.5%
Special Education	8.5%	8.5%	7.7%
Other	12.7%	8.4%	7.7%
School			
Charter	41.7%	72.9%	92.3%
Traditional Public	58.3%	27.1%	7.7%
Teaching Level			
Elementary	42.5%	50.0%	61.5%
Middle	23.3%	25.0%	30.8%
High	34.2%	25.0%	7.7%

Table 3: Regression Results for Number of Safe Choices

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Opted In	0.54		0.46	0.48	0.50	0.69	0.76
	(0.48)		(0.50)	(0.55)	(0.53)	(0.56)	(0.56)
Perf. Pay Deciding Factor x Opted In	-1.69**		-1.96**	-1.71**	-2.22***	-2.22***	-2.31***
	(0.77)		(0.79)	(0.79)	(0.79)	(0.80)	(0.83)
Female		0.28	0.34	-0.04	0.44	0.53	0.30
		(0.54)	(0.53)	(0.56)	(0.54)	(0.55)	(0.59)
Minority		-1.00	-1.30	-0.66	-1.66*	-1.62	-1.59
		(0.88)	(0.88)	(0.93)	(0.89)	(0.97)	(0.99)
Age		-0.01	-0.02	-0.01	-0.01	-0.02	-0.00
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Experience		0.07**	0.07**	0.08**	0.07**	0.07*	0.09**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Charter		0.45	0.66	0.54	0.68	0.78	0.81
		(0.49)	(0.55)	(0.56)	(0.57)	(0.60)	(0.65)
Additional Controls							
College Quality				X			X
Mother's Education				X			X
Household Income					X		X
Teacher Subject Area						X	X
School Grade Level						X	X
Teacher Only						X	X
Constant	5.32***	5.27***	4.83***	1.73	3.94***	4.31***	1.69
	(0.27)	(1.03)	(1.15)	(3.02)	(1.23)	(1.49)	(3.20)
N	119	119	119	111	119	112	105
R-Squared	0.02	0.02	0.06	0.11	0.08	0.07	0.12

Note: Numbers in parentheses below coefficient estimates are standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% p-value levels, respectively, in a two-sided alternative to the null that the coefficient value is 0.

Appendices

Note: Actual document was formatted to take up the front and back of only one page.

Please complete this page first, then proceed to the back to complete the study. Upon completion you will be paid in cash for your participation as described below.

For each of the 10 lottery pairs listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option, either A or B, for each lottery.

In each lottery pair you will be selecting between a lottery that will pay either $6.00 or $4.80 (Option A) and a lottery that will pay either $11.55 or $0.30 (Option B). In the first lottery there is a 10% chance of receiving the larger payout for lottery 1 and a 90% chance of receiving the smaller amount. In each subsequent lottery pair the chance of earning the higher payout increases by 10%.

After you complete this study, the experimenter will roll a 10-sided die to randomly select which lottery will be used.

Next, the experimenter will roll the same die a second time to determine your actual payoff based on the option you chose for that particular lottery.

The number on the die for the second roll will determine whether you receive Payout 1 or Payout 2.

Example: If the experimenter's first roll is "3," then your payoff will based on Lottery 3. If you chose Option B for Lottery 3 and the second roll is "1", “2”, or “3”, you will receive a Payout of $11.55, but if the second roll is “4”, “5”, “6”, “7”, “8”, “9” or “10”, you will receive a Payout of $0.30.

Please answer the following question which will not impact your payoff but is intended to ensure you understand this task. Suppose the experimenter rolls a "2" first and then rolls a “9”. If you have selected Option A for Lottery 2, what will your payout be? _____

Please notify the experimenter when you have answered this question before you continue with the experiment.

Lotteries. For each of the 10 lotteries listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option per lottery.

Lottery	Option A		Possible Roll(s) for Payout	Option B	
	☑	Payout		Payout	☑
1.		$6.00	1	$11.55	
		$4.80	2,3,4,5,6,7,8,9,10	$0.30	
2.		$6.00	1,2	$11.55	
		$4.80	3,4,5,6,7,8,9,10	$0.30	
3.		$6.00	1,2,3	$11.55	
		$4.80	4,5,6,7,8,9,10	$0.30	
4.		$6.00	1,2,3,4	$11.55	
		$4.80	5,6,7,8,9,10	$0.30	
5.		$6.00	1,2,3,4,5	$11.55	
		$4.80	6,7,8,9,10	$0.30	
6.		$6.00	1,2,3,4,5,6	$11.55	
		$4.80	7,8,9,10	$0.30	
7.		$6.00	1,2,3,4,5,6,7	$11.55	
		$4.80	8,9,10	$0.30	
8.		$6.00	1,2,3,4,5,6,7,8	$11.55	
		$4.80	9,10	$0.30	
9.		$6.00	1,2,3,4,5,6,7,8,9	$11.55	
		$4.80	10	$0.30	
10.		$6.00	1,2,3,4,5,6,7,8,9,10	$11.55	
		$4.80	---	$0.30	

Background Information

1. Year of Birth: _____

2. Gender (please circle): Female Male

3. Race/Ethnicity (please circle): White Black Asian Hispanic

Other (please specify): _____

4. Undergraduate Degree (please specify field and degree):_____

5. Undergraduate Institution:_____

6. What is your mother's highest level of education? (please circle)

Less than H.S. H. S. Some College Bachelor's Degree Graduate Degree

7. What is your estimated annual household income? (please circle)

Less than $35k $36K-$50k $51k-65k $66k-80k $81k-$95k $96k-$110k

$111k-$125k Greater than $125k

8. How many years total have you taught prior to this school year? _____

9. What is your primary teaching subject? _____

10. Did you participate in this school's performance pay program? (please circle) Yes No

11. If you participated in the performance pay program, were you employed at this school ***before it began***? (please circle) Yes No

For #12-15, please indicate how you feel about the following statements: (please circle)

12. The performance pay program was the deciding factor for me working at this school.

Strongly Disagree Disagree Neutral Agree Strongly Agree N/A

13. If I were looking to work in a different school, I would want to teach in a school that offered teacher performance pay.

Strongly Disagree Disagree Neutral Agree Strongly Agree

14. I believe that teachers should be paid based on student performance.

Strongly Disagree Disagree Neutral Agree Strongly Agree

15. I plan to spend the rest of my career as a teacher.

Strongly Disagree Disagree Neutral Agree Strongly Agree

On the scales below, please indicate how you, as a teacher, would prefer to be compensated.

16. ○ ○ ○ ○ ○ ○ ○

100% Student Performance 50/50 100% Teacher Experience

17. ○ ○ ○ ○ ○ ○ ○

100% Just My Performance 50/50 100% Team Performance

18. Please briefly list/describe any jobs that you had **after** college but before becoming a teacher:

_____

_____

Name Last: _____ Name First: _____

College of Education and Health Professions
Department of Education Reform

201 Graduate Education Building
Fayetteville, Arkansas 72701
(479) 575-3172
(479) 575-3196 (FAX)

August 14, 2013

To: The Graduate School

Re: Confirmation of Authorship

This letter is to confirm that Daniel H. Bowen is the lead author of the article "Does Merit Pay Attract Teachers with Different Risk Preferences?" He completed more than 51% of the work that went into this paper.

_____

Jay P. Greene, Ph.D.
Endowed Chair and Head of the Department of Education Reform

October 15, 2012

MEMORANDUM

TO:	Daniel Bowen Jonathan Mills Jay Greene
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	12-10-142
Protocol Title:	*Does Merit Pay Attract Teachers with Fundamentally Different Characteristics?*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 10/15/2012 Expiration Date: 10/14/2013

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (http://vpred.uark.edu/210.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 150 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 210 Administration Building, 5-2208, or irb@uark.edu.

Do Risk and Pay Preferences Predict Teacher Effectiveness?

By: Daniel H. Bowen

Abstract

School districts and charter schools implement performance-based pay programs with one of their goals being to attract and reward higher quality teachers. Prospective teachers who want to be compensated based more on their students' performances will likely envision themselves as being more effective in the classroom. However, there are aspects of performance pay that might deter individuals who are (or would be) effective teachers from remaining in (or entering) the profession. For example, introducing pay ambiguity and emphasizing student test performance could bring about unacceptable levels of stress for those currently interested in the teaching profession.

In this paper I attempt to provide some preliminary evidence for how implementing a transfer to performance-based pay might influence teacher quality. I examine the relationships between two teacher quality measures and individuals' preferences for performance-based pay as well as their risk preferences. Based on the evidence from this study, there appears to be a strong, negative relationship between the extent to which a teacher prefers pay based on performance to their value-added scores ($p = 0.09$). While there is also a strong, positive relationship between the extent to which a teacher is risk averse and value-added scores, this relationship fails to reach traditional levels of statistical significance. There are no significant associations between teachers' pay or risk preferences and their year-end performance evaluations. These findings suggest that the attraction of performance pay-preferring and risk-loving individuals will not necessarily improve student achievement in terms of test score gains. Moreover, these results

should concern policymakers looking to implement performance pay policies because this reform could potentially deter anti-merit pay and more risk-averse, though effective, teachers from entering or remaining in the profession.

Introduction

The importance of teacher quality is reflected both in the education research literature as well as the amount of attention it receives in the media as well as policy circles (e.g. Chetty, Friedman, & Rockoff, 2011; Darling-Hammond, 1999; Green, 2010; Kristof, 2012; Rice, 2003). Despite the emphasis on teacher quality, what makes for or how to attract better teachers remains, to a great extent, unknown (Hanushek & Rivkin, 2010; Kane, Rockoff, & Staiger, 2008; Rockoff, Jacob, Kane, & Staiger, 2011). Education policymakers develop and implement policies in hopes of attracting and retaining individuals who will improve student outcomes (e.g. Auguste, Kihn, & Miller, 2010; Finn, 2001). One of the increasingly popular policies, that is at least partly aimed at improving teacher quality, is performance-based (i.e. merit) pay for teachers (Ballou & Podgursky, 1995; Fox, 1984; Goldhaber, DeArmond, Player, & Choi, 2008). However, while one of the main goals of performance-based pay is to attract higher-quality individuals to the profession, introducing more ambiguity to the compensation format could have the unintended consequence of deterring individuals who prefer to avoid risk, yet would prove effective in the classroom.

Prior studies suggest that individuals who choose to go into K-12 teaching are more risk averse than other professionals with comparable educational backgrounds (e.g. Bowen, Buck, Deck, Mills, & Shuls, 2013; Davis, 1994). There is also evidence that teachers who choose to work at schools that offer performance-based pay are more risk loving (Bowen & Mills, 2013). These findings, taken together, help put forward the argument that significantly changing the

structure of teacher compensation to a more performance-based format could alter the composition of the workforce. However, the conclusions of these studies do not address the issue of whether such a change improves the overall quality of the workforce in terms of student achievement.

In this paper I hope to provide some preliminary evidence for how the introduction of performance pay might influence the quality of the teacher workforce. Using an instrument common to experimental economics as well as survey responses, I analyze 120 teachers' risk preferences and desires for performance pay and attempt to determine how these preferences relate to their effectiveness as educators. Examining individuals' desire for performance pay as well as risk preferences and relating them to measures of effectiveness should provide a glimpse into whether this reform will likely appeal to individuals with characteristics that relate to better teaching. Additionally, these analyses provide means for investigating the effectiveness of teachers whom are the least likely to support this type of reform and could be the most likely to leave a school or district in response to the implementation of such a policy.

Intuitively, the relationship between risk aversion and both measures of teacher quality could go in either direction. Teachers who prefer taking risks might be more willing to incorporate new instructional methods that increase students' achievement scores. Conversely, risk taking might not always benefit students and could even prove detrimental. Risk takers might also be prone to making rash decisions that have negative consequences with regard to student learning. The same might be true for year-end principal ratings. A reform-oriented principal might prefer a teacher who is more willing to take risks and is willing to adapt to new teaching styles and curricula. However, less reform-oriented principals may favor teachers who "don't rock the boat".

The connection between the stated the desire for basing compensation on performance and measures of teacher quality should be less ambiguous. Teachers who would rather work under performance-based compensation should also be more likely to financially benefit from this compensation format. Teachers should prefer merit pay when they view themselves as being highly effective, both in terms of students' test score gains and principal evaluations. However, the evaluation systems that come with performance-based pay might not be well-received with teachers with a mission focus or greater intrinsic motivation. Moreover, performance-based pay might simply stress out highly effective teachers who already maximize their teaching efforts.

I find that there are substantive, negative relationships between teachers' value-added scores and both the extent to which they prefer pay for performance as well as propensity for risk. A standard deviation increase on a preference for performance-based pay compensation scale is associated with a 0.20 standard deviation decrease in average student value-added score (p-value = 0.09). The decision to make one additional safe choice on the Holt and Laury (2002) risk-elicitation task relates to a 0.06 standard deviation increase in average student value-added score; however, this effect fails to achieve a traditional level of statistical significance (p-value = 0.15). Performance pay and risk preferences have no significant relationship to a teacher's year-end principal rating. The extent to which a teacher would prefer to have pay based on individual versus team performance also has no significant relationship to either teacher quality measure.

These findings suggest that if policies such as performance pay were to attract individuals more supportive of basing pay on student performance as well as more risk-loving individuals, then this compositional shift would not bring about higher student achievement. Moreover, such a policy could potentially deter effective teachers from entering or remaining in the profession. While these findings are preliminary and do not provide conclusions with regard to the causal

relationship between teachers' preferences and effectiveness, they do provide an early insight into whether performance pay would in fact have the desired outcome of attracting individuals who would prove to be more effective than the current workforce.

In the next section I provide a review of the literature focusing on the relationships between risk preferences and other characteristics related to teacher effectiveness in addition to relevant studies on the personalities of individuals who choose to become K-12 teachers. Then, I discuss the survey instrument and procedures. After that, I describe the sample used for this study. In the final sections, I present findings, discuss results, and conclude with potential policy implications as well as proposals for future studies that would help to further examine this issue.

Literature Review

Prior studies have concluded that K-12 teachers are generally more risk averse than professionals with similar educational backgrounds (Bowen et al., 2013; Davis, 1994; Wagner, 2001). While not conclusive, Bowen and Mills (2013) administer the Holt and Laury (2002) risk-elicitation task as well as a survey and find evidence that individuals who actively seek and opt-in to schools that offer performance-based pay are not as risk averse as their colleagues. Together, these studies provide an argument for the potential of performance pay policies to significantly change the composition of the teacher workforce, at least with regard to risk preferences.

There is a shared belief that the risk averseness of teachers is a source of concern when it comes to the quality of the workforce. Hannaway (2009), explains how performance pay can attract more effective teachers to the classroom; "If you reward performance, applicants who are confident in what they can do and are less risk-averse may be more attracted to the profession because they'll feel their contributions will be recognized." Di Carlo (2011) argues that this

sentiment is common among performance pay advocates; "The major benefit of merit pay, according to proponents, is that it could attract a 'different type' of candidate to the profession, one who is less risk-averse and more drawn to a job that rewards them according to results."

One reason for the desire to attract less risk-averse individuals to the profession is that this transition will lead to a broader base of people wanting to enter the profession. If certain aspects of the profession, unrelated to effectiveness in the classroom, are limiting who ultimately enters the profession, then changing or eliminating these components should increase the labor supply which should in turn boost the overall number of high quality teachers (Lavy, 2007). According to a study by Borghans, Golsteyn, Heckman, and Meijers (2009), less risk-averse individuals tend to be more ambitious and less neurotic, characteristics that could translate into greater teacher quality. However, when making changes to a particular aspect of the profession, it is important to make sure that these changes do not result in overall decreases in the quality of the composition of the workforce by dissuading more effective individuals currently drawn to the field than would be ultimately attracted (Oyer & Schaefer, 2008).

Another reason teacher risk preferences matter is attributed to findings that individuals' personalities are tied to the instructional styles that they adopt which could affect student learning. Fairhurst and Fairhurst (1995) find that Myers-Briggs personality profiles significantly predict teachers' preferred learning environments. For example, they find that introverts prefer quieter, controlled classroom settings. Studies have also shown that teachers' risk preferences significantly influence their instructional approaches. Individuals who are less risk averse are more likely to be innovative and open to trying new things (Chen, Greene, & Crick, 1998). Hills, Stroup, and Wilensky (2005) conclude that more risk-loving teachers are more likely to engage in constructivist-based learning. There is some evidence that constructivist-based environments

are more conducive to student learning (Calik, Ayas, Coll, Unal, & Costu, 2007; Keys & Bryan, 2001; Liang & Gabel, 2005), but there is also evidence to the contrary (Kirschner, Sweller, & Clark, 2006). However, regardless of whether constructivist-based instruction is beneficial, the willingness to incorporate constructivist-based learning signifies a greater inclination for differentiating instructional methods which tends to be related to more effective teaching (Subban, 2006).

Rushton, Morgan, and Richard (2007) provide the only study to my knowledge that explicitly examines the relationship between individuals' risk preferences and their effectiveness as teachers. They administered surveys to individuals who were nominated into the Florida League of Teachers, a prestigious group of educators nominated by their superintendents. Part of their survey included the Beiderman-Sensation Seeking Scale, a revised version of the Zuckerman Sensation Seeking Scale. Comparing these 58 teachers' responses to the general population, they conclude that outstanding teachers are "not generally extreme risk takers" (p. 436). However, when categorizing these teachers according to their responses and conducting a chi-square goodness of fit test, they do not find any significant differences between the distribution of risk preferences for this sample of effective teachers and that of the general population.

There is also almost no research on the relationship between teachers' perspectives of merit pay and how these views relate to their effectiveness as educators. Jensen (2012) interviews teachers who were participating in a newly implemented performance pay program, and asks them questions regarding their views of the new pay policy. One particularly interesting

finding from this study is that teachers who earn higher ratings, as well as those who earn higher bonuses, are no more likely to support merit pay than teachers with lower ratings or bonuses.[13]

I hope to add to this body of research by further investigating the relationship between pay and risk preferences to measures of teacher quality. Unlike the prior cited studies that examine teacher risk preferences through self-assessment surveys, I use the Holt and Laury (2002) experimental measure for eliciting risk preference. I also tie both pay and risk preferences to student achievement scores and principal evaluations as measures of teacher quality. These outcome measures have the advantage of being more continuous and tend to be more widely accepted measures of teacher quality (e.g. Chetty, Friedman, & Rockoff, 2011; Hanushek & Rivkin, 2010; Rubin, Stuart, & Zanutto, 2004).

Methods

Task Procedures and Data Collection

All the participants were employees at the time of this study at one of two school districts that had fairly recently implemented teacher performance pay programs. One district is a charter school district, and the other is a traditional public school district. In order to conduct the risk-elicitation task and issue the survey, permission was obtained from the schools' administrators to offer their teachers an opportunity to participate in a study about individuals' economic decision making. Teachers participated in the study during lunch breaks, planning periods, or after students were dismissed for the day. The respective schools' principals announced the opportunity to participate in the study shortly before the arrival of the research team and informed them of where to go if they were interested.

[13] This distinction is made because bonuses in this performance pay program were differentiated depending on teacher responsibilities. Core subject teachers in this program could make more money than higher rated, non-core subject teachers.

During the study, as prospective participants arrived, a researcher would greet the teacher, briefly describe the experiment, and then ask if he or she would care to participate. Teachers were made aware that their participation was entirely voluntary and that they were free to leave at any point during the study. They were also told that they would receive compensation in exchange for their participation but that the amount they earned would depend on their decisions. Potential participants were also informed that the task would take approximately 15 minutes. If the teacher still agreed to participate in the study,[14] then they were handed a single sheet of paper that had the Holt and Laury (2002) task with instructions on the front and 18 survey questions on the other side (see Appendix for the full set of instructions, the task itself, and the survey).

Teachers were instructed to first read through all of the directions. At the bottom of the risk task instructions, there was a comprehension question to verify that the participant understood how the task worked. The participants were instructed to notify the researcher after answering the comprehension question. The researcher would then ask the participant if there were any questions before proceeding. If the teacher was correct on the comprehension question, they were then asked to complete the task first and then answer the survey questions on the other side. If the teacher was incorrect, the researcher would ask the teacher to describe how they came to that answer and try to resolve any confusion they still might have even after eventually coming up with the correct answer to the comprehension question. The researcher would initial next to the each participant's answer to the comprehension question. If the participant understood the task without any additional explanation, then the researcher initialed to the right of the answer. If the researcher provided any assistance in helping the participant comprehend how the

[14] Only 2% of teachers who showed initial interest in participating in the study chose not to participate.

task worked, then the researcher would initial to the left of the answer to the comprehension question. Having the researcher indicate whether the participant needed any assistance makes it possible to document whether the participant may have been confused with the task while not bringing any unnecessary attention to the participant. After completing the survey, a researcher would verify that the participant did not leave any items blank and ensured that the teachers did not communicate with one another during the risk-elicitation task or survey.

After a researcher checked the participant's completed survey, a ten-sided die was rolled to determine by chance which of the ten lotteries would be used for determining the participant's payment. Randomly using only one of the ten lotteries is a common approach for these types of experiments to control for wealth effects (e.g. Holt & Laury, 2002). After the first roll of the die, the participants were notified of which of the lotteries would determine their payout. Then, a second roll of the die determined whether the teacher received the higher or lower payout for their option for that particular lottery. Participants were then given an envelope with their payment in cash and then signed a form to verify receiving payment and were thanked by the researcher. The average payout for each participant ended up being $6.29 for fifteen minutes of time.

Measuring Performance Pay Perspectives

Teachers were asked to provide their demographics and answer a few questions about their perspectives on performance-based pay. Three survey items were included specifically to elicit participants' preferences for performance-based pay. Teachers were requested to indicate on a five-point Likert scale the extent to which they agreed or disagreed about the following statement: "If I were looking to work in a different school, I would want to teach in a school that offered teacher performance pay". The two other performance pay items provided the teacher

with a spectrum with which they could reveal their preferences for an ideal teacher compensation format. The first of these items has the teacher indicate the extent to which she would prefer pay based on student performance versus years of experience. The second item has the participant choose the extent to which he would prefer compensation based on individual versus team performance.

In order to create a scale for teachers' preferences for performance-based pay I combine the responses to the items about working in a different school and the tradeoff between pay for student performance versus experience into a scale. I standardize the participants' responses to both items, take the average of the standardized responses, and then standardize this average to compare the extent to which a teacher supports performance-based pay in terms of standard deviations. The compensation preference item for preference of individual versus team pay is not included in this scale since responses do not necessarily reflect the desire for performance-based pay as much as they reflect how participants would prefer to have performance measured. However, since an individual's preference for pay based on individual performance could measure one's sense of self-efficacy, I include specifications in the analyses that incorporate a standardized measure for participants' responses to this survey item.[15]

Measuring Risk Preferences

Risk preferences are elicited from participants' decisions from the Holt and Laury (2002) experimental procedure. This risk measure is a commonly used procedure in experimental economics (e.g. Anderson, Harrison, Lau, & Rutstrom, 2008; Dohmen, Falk, Huffman, & Sunde, 2010; Eckel & Wilson, 2004). This risk-elicitation tool has the benefits of high test reliability

[15] The preference for individual over team pay does not definitively measure a sense of self-efficacy because teachers could still believe that they are very effective yet still prefer team-based pay. For example, a highly effective teacher might believe that team-based pay promotes more team teaching and collegiality amongst faculty and opt for this method of compensation.

(Anderson & Mellor, 2008). This task has also been externally validated with surveys of an individual's willingness to engage in risky behaviors (e.g. not using a seat belt, smoking, and heavy drinking) (Harrison, Johnson, McInnes, & Rustrom, 2005) as well as a psychometrically-validated survey of financial risk tolerance (Faff, Mulino, & Chai, 2008).

Study participants are asked to choose between two options, A and B, for 10 different lotteries. There are two payouts for each option. The payouts for each option are the same throughout the survey. Therefore, there were four total, possible payouts for the participants of this study: \$4.80 or \$6.00 for Option A; \$0.30 or \$11.55 for Option B.[16] I define Option A as the "safer" of the two because the difference in the possible payouts is substantially lower than it is for Option B.

As a participant proceeds from the first to the tenth lottery, the probability of being awarded the higher of the two payouts for a given option (\$6 for A and \$11.55 for B) increases at a consistent rate of 10 percentage points for each subsequent lottery. The participant has a 10% chance of obtaining the higher payout with the first lottery (a 90% chance of the lower payout), a 20% chance in the second lottery (an 80% chance of the lower payout), and so on until the tenth lottery where the higher payout is guaranteed. Most participants will start off choosing Option A because they wish to avoid the high probability of getting only a \$0.30 payout that comes with Option B in the earlier lotteries. The participant eventually switches to Option B at a point where the odds of getting the highest payout of \$11.55 make this option preferable to the safer choice, Option A.

[16] These payment amounts are not the exact same as the ones used in the original Holt and Laury (2002) study. The ratios are the same, but the payouts are three times that of the original study. The decision to triple the payouts was based on suggestions from colleagues with prior experiences with regard to the stakes that would typically be needed to efficiently maximize the rate of participation as well as get participants to seriously consider the lottery options.

The probability of getting the higher payout only increases as the participant goes through each of the 10 lottery choices. Therefore, after making the switch over to Option B, a participant who is consistent in their risk preferences should keep choosing B for the rest of the lotteries. Participants are told in advance that they will ultimately only be paid for the result of one of the ten lotteries and that this lottery would be determined randomly with the use of a ten-sided die at the conclusion of the study. Therefore, while they only have the chance to be paid for their choice in one of the lotteries, they should treat each lottery as if it will ultimately end up being the one that determines their payout.

Expected payouts for both options of each lottery can be calculated by multiplying each payout by the likelihood of obtaining that outcome and then summing these values. Comparing the expected payout value for each option within a lottery to participants' lottery choices indicate their risk preferences. Option A has a higher expected payout than Option B for the first four lotteries, and Option B has the higher expected payout for the remaining lotteries. A risk-neutral participant makes choices based entirely on the expected payout for each lottery. Therefore, a risk-neutral participant chooses Option A for lotteries 1-4 and Option B for lotteries 5-10. Someone who is risk averse will favor the guarantee of getting at least $4.80 (or depending on how it is viewed, the assurance of avoiding the $0.30) beyond the fourth lottery even though the expected payout for Option B is greater than Option A. A risk-loving individual will switch to Option B before the fourth lottery because they find that the potential benefits of choosing the riskier option outweigh the difference in expected value plus any concerns that come with taking this risk.

Participants' responses on the Holt and Laury (2002) risk-elicitation task also provide information for whether or not their preferences remain consistent throughout the procedure. The

likelihood of getting the higher payout only increases as the participant proceeds through the lotteries. Therefore, once a participant makes the switch from the safer option (A) to the riskier option B, they should not switch back and make the safer option in subsequent lotteries.[17]

Measuring Teacher Quality

I examine two available measures of teacher quality for this study. I obtained permission to collect the names of the teachers from the Bowen and Mills (2013) study and link their risk and survey responses to four years' worth of two measures of teacher quality.

The first measure is a teacher's average value-added score over a four year period. Student achievement scores on the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) are used to compute these value-added scores. An advantage of the NWEA test scores is that students are tracked longitudinally and given growth targets based on prior growth and trajectories computed from normative data. Teachers' value-added scores are determined by taking the average difference in students' scores over the course of the year and dividing by their average projected score target to get a ratio of how well their students did relative to goal growth. Therefore, a teacher earns an NWEA value-added score of 1 if their students' growth scores perfectly match their target scores. When included in analyses these scores were then standardized so that estimates for coefficients are presented in terms of standard deviation effect sizes.

The second measure of quality is the teachers' scores on their year-end performance ratings. This measure is almost entirely subjective because it is based on how the teacher's principal rates them on matters such as content knowledge, level of performance, collegiality,

[17] I have run all analyses without any of the inconsistent participants in the sample and find that this exclusion does not change the qualitative interpretations of the statistical significance for any estimates for the variables of interest.

and how well they adhere to school policies and culture. The one objective element of the performance evaluation is the teacher's attendance and punctuality. The score rating that the teacher receives is divided by the total number of possible points and then averaged across all years where the teacher has an evaluation on record during the performance pay program. These scores are also standardized so that estimates for coefficients are presented in terms of standard deviation effect sizes.

Sample

This study analyzes teacher quality measures for 93 of the 120 study participants from the Bowen and Mills (2013) study with either NWEA test score or principal evaluation data. Of the 120 participants from the study, I am able to link 70 teachers to student test score data and 79 teachers to principal evaluation scores.[18] Almost half (47%) of the 120 teachers have both test score and principal evaluation data.

An overview of the descriptive data for the sample is provided in Table 1. The majority of the teachers in the sample are employed at the traditional public school district (65%). The sample also consists of 43% teachers at the elementary level, 29% at the middle school level, and 28% at the high school level. The average teacher has 12 years of experience, and experience ranges from 0 to 39 years. The vast majority of teachers are female (78%), as is common to the profession.

The sample for this study consists of teachers who, on average, are probably considered effective based on their value-added scores. The average teacher's value-added score in this sample is a 1.17, meaning that their average student scores 17% above their NWEA target score

[18] The reason for study participants not having student test scores was that they taught untested subjects (e.g. social studies, foreign language, music, etc.). The primary reason for participants not having performance evaluation scores is that they had been employed at the school for less than a full academic year.

for the school year. However, there is significant variation in this measure. The average student achievement at the teacher level relative to projected growth has a standard deviation of 0.57 with ranges from -0.92 to 2.17. The measure for teacher performance is also generally positive and does not vary as much as the value-added score measure. The average teacher in this sample obtained 85% of the possible points available on the end of year evaluation and average teacher performance scores range from obtaining 55% to 100% of the points available.

On average, teachers in this sample prefer to have their wages determined more by experience than student performance. However, these teachers are willing to have close to 50% of their salaries based on performance. Preferences for the weighting of student achievement varies, ranging from the desire to have pay based solely on teacher performance to pay based entirely on experience. Teachers from this sample, on average, are neutral on whether they would seek employment at another school offering merit pay. These teachers also tend to prefer performance being more dependent on individual rather than team effectiveness. However, they are willing to have close to 50% of pay determined by team performance. Objectively, this sample is considered risk averse, making 0.44 more safe choices than a risk-neutral group would make. Risk preferences range from individuals making 0 to 10 out of 10 possible safe choices.

Empirical Strategy

I run multiple regressions using ordinary least squares (OLS) in order to control for observables that influence both measures of teacher quality. The controls used in all the regressions are a dichotomous variable for the teacher's school district (charter or traditional public) and school grade level (elementary, middle, or high school). Controlling for district is important because of the charter versus public distinction that can influence student achievement scores. Controlling for the different school grade levels helps account for differences that take

place with students and teachers at these different levels in addition to the fact that there are different principals providing performance evaluations at the different school levels. I also include model specifications that incorporate teacher experience in terms of years and gender in addition to a dichotomous variable for whether the participant was consistent in lottery responses.

Results

I find a significant, negative relationship between a teacher's stated preferences for pay based on student achievement and their value-added scores (see Table 2). Teachers scoring one standard deviation higher on the performance pay preference scale perform a fifth of a standard deviation lower on average in terms of students' gain scores relative to projected growth (p-value = 0.09). There is also a positive relationship between the number of safe choices that a teacher makes (i.e. risk aversion) and students' test score gains. One additional safe choice translates into a 0.06 standard deviation increase in students' score gains relative to target growth. However, this relationship fails to reach a traditionally accepted level of statistical significance (p = 0.15).

For the 79 teachers in the sample with year-end principal evaluations, both performance pay and risk preferences have no significant relationship to a teacher's principal ratings (see Table 3). A standard deviation increase on the support for merit pay scale is related to a 0.13 standard deviation increase in a teacher's year-end performance evaluation (p = 0.21). Moreover, making one additional safe choice is associated with a 0.04 standard deviation increase on the teacher's year-end performance evaluation (p = 0.29). However, neither of these estimates is statistically different from a null effect.

The preference for individual, as opposed to team, performance-based pay has no significant relationship to either teachers' value-added or year-end performance ratings. The

same is true for the number of years of teaching experience and whether the participant was consistent throughout the Holt and Laury (2002) task. Teaching at the high school is significantly, negatively related to students' gain scores relative to targeted growth. This result for high school teachers is likely attributed to the fact that there are curvilinear growth patterns; high school students' achievement scores are typically much more static than is the case for elementary students (Lee, 2010). While NWEA targets account for this with lower projected growth scores for high school students, the gains still appear to more regularly fall short of expectations.

Another interesting result is that females, on average, receive significantly higher performance evaluation scores. The male teachers in this sample receive 76% of the points possible in their year-end evaluations, and female teachers receive 83% of the year-end evaluation points. This result has been found in other research. Rinehart and Young (1996) examine whether there is a gender bias in teacher evaluations. They conclude that female teachers receive significantly higher ratings on professionalism and instruction. This effect could be attributed to female teachers simply performing better on these metrics or due to principals' perceptions.

Discussion & Conclusion

Finding evidence that teachers' desires for performance-based pay is negatively related to their effectiveness in terms of how well their students perform is intriguing. Teachers who are more effective in terms of value-added scores would seemingly have the most to benefit from a merit pay system. It is unlikely for the teachers in this sample to not have a good understanding of how well they perform on value-added measures because almost all of them had already

received student data and performance bonuses that reflected their effectiveness in the classroom.

One possible explanation for this relationship might have to do with a teacher's motivation for entering the profession. Teachers who do not want their salaries tied as much to student performance could be more intrinsically and mission-oriented, and individuals with this type of motivation probably have positive influences on student outcomes. Conversely, teachers who support performance-based pay might do so because they view it simply as an opportunity to make higher salaries, and teachers with this view are plausibly less effective than the more intrinsically and/or mission-oriented teachers.[19]

The behavioral economics research on the relationship between external reward and motivation provides evidence for this explanation. In a widely cited study, Deci (1971) concludes that "when money is used as an external reward for some activity, the subjects lose intrinsic motivation for the activity" (p. 114). Moreover, Deci, Koestner, and Ryan (1999) in a meta-analysis on the influence of incentives on motivation, conclude that the general finding on the use of rewards for signaling degrees of competence is that they tend to have very negative impacts on intrinsic motivation. Individuals often become resentful of these types of reward structures, finding them to be overbearing means for controlling employee behavior in addition to being negatively, rather than constructively, construed forms of feedback (p. 657).

Teachers are still financially compensated as educators under the more traditionally used step and lane pay format. However, the format of this compensation does not explicitly attempt

[19] It is worth noting that the average teacher in this sample would likely be deemed "effective" with regard to student growth scores. The average teacher in this sample produces student gains that are 17% higher than their targeted growth. Therefore, I am not claiming that teachers who support performance-based pay are objectively ineffective, just that they appear to be less effective than those more averse to student performance-based pay.

to motivate teachers to work harder to improve student outcomes. Educators who have more of a mission focus or are intrinsically motivated already maximize their efforts in the classroom. These teachers could conceivably acquire distaste for performance-based pay because the financial incentives are somehow perceived to diminish the value of their efforts. Weiber, Rost, and Osterloh (2010) conclude that pay for performance appears to have a crowding-out effect for individuals who more likely find their work to be more intrinsically rewarding.

Another explanation for this finding is possibly understood by considering the inverse of this result: teachers who prefer the more traditional step and lane pay are more effective when it comes to students' growth scores. A logical explanation for why this might be the case is that teachers who believe in rewarding experience see themselves as staying in the profession longer, and this motivation is positively related to student achievement. This finding corroborates with earlier findings that more effective teachers, in terms of value-added score measures, are actually less likely to leave the profession (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2007; Goldhaber, Gross, & Player, 2010; Hanushek, Kain, O'Brien, & Rivkin, 2005; Krieg, 2006).[20]

The fact that risk aversion and student performance are positively related could also support this finding. Bowen and Mills (2013) find that teachers with more teaching experience tend to be more risk averse. This finding makes sense because individuals who likely remain in the same career for a longer period of time are less likely to be risk takers. While risk aversion in and of itself does not likely explain teacher effectiveness, risk preference might relate to characteristics that increase the likelihood of an individual wanting to stay in the profession longer.

[20] Feng and Sass (2008) and West and Chingos (2009) however add an important caveat that even though the teachers with, on average, higher value-added scores are more likely to remain in the profession, exit rates for teachers on both the high and low end of the distribution are much higher than those in the middle.

Other aspects of teacher motivation that might influence perceptions of performance-based pay could have to do with a concern for the welfare of students. The explanation here is still that exceptional teachers are more mission oriented and/or intrinsically motivated, but this interpretation just comes from a different angle. More effective teachers might believe that even if performance pay provides them with greater financial rewards, such a policy could still ultimately have negative effects for students. These teachers might associate performance pay with something such as the negative aspects of "teaching to the test". Jensen (2012) supports this explanation, concluding that teachers employed at schools with performance-based pay become increasingly worried about the program's impact on students' outcomes; "There was a notable decrease in the percentage of teachers who thought the CPIP [performance pay program] improved educational outcomes for students, from 49% agreement in March to 37% agreement in October" (p. 116). These concerns are reflected in a decline in the belief that performance pay would increase student learning and an increase in the view that the program negatively affected students' levels of stress and anxiety.

These findings, however, are still very much preliminary. In order to get a better sense of the causal relationship between teachers' preferences and effectiveness, there needs to be larger-scale, more rigorous studies on this issue. One question worth further exploring is a more direct evaluation of the relationship between teachers' motives for entering the profession and how these motives affect student achievement. A similar, related issue is whether certain schools more effectively attract individuals with motivations that essentially translate into better student outcomes. Policymakers and researchers should also test whether adopting or eliminating particular education policies have long-term impacts on the composition of the teacher workforce.

Hopefully these analyses provide at least a glimpse into how pay and risk preferences might influence teacher quality and produce testable hypotheses for future research. This line of research could have major implications for the design and implementation of policies that aim to improve teacher quality. If individuals with more of a mission focus are found to be more effective educators, then policymakers should find ways to best attract and retain individuals with a genuine passion for teaching. Conversely, there could be cause for concern with regard to policies that dissuade more mission-focused and intrinsically-motivated individuals from entering or remaining in the classroom.

References

Anderson, S., Harrison, G. W., Lau, M. I., & Rutstrom, E. E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583-618.

Anderson, L.R., & Mellor, J.M. (2009). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260-1274.

Auguste, B. G., Kihn, P., & Miller, M. (2010). *Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research-based perspective*. McKinsey.

Ballou, D., & Podgursky, M. (1995). Recruiting smarter teachers. *The Journal of Human Resources, 30* (2), 326-338.

Borghans, L., Heckman, J. J., Golsteyn, B. H., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association, 7*(2-3), 649-658.

Bowen, D. H., Buck, S., Deck, C., Mills, J. N., & Shuls, J. V. (2013). Risky Business: An Experimental Analysis of Teacher Risk Preferences (Working Paper No. 2013-01). Retrieved from University of Arkansas Department of Education Reform Working Papers site: http://www.uaedreform.org/working-papers/

Bowen, D.H., & Mills, J.N. (2013). Does merit pay attract teachers with different risk preferences? (Working Paper No. 2013-05). Retrieved from University of Arkansas Department of Education Reform Working Papers site: http://www.uaedreform.org/working-papers/

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2007, March). Teacher attrition, teacher effectiveness and student achievement. Paper presented at the Annual Conference of the American Education Finance Association, Baltimore, MD.

Calik, M., Ayas, A., Coll, R.K., Unal, S., & Costu, B. (2007). Investigating the effectiveness of a constructivist-based teaching model on student understanding of the dissolution of gases in liquids. *Journal of Science Education and Technology, 16*(3), 257-270.

Chen, C.C., Greene, P.G., & Crick, A. (1998). Does entrepreneurial self-efficacy distinguish entrepreneurs from managers? *Journal of Business Venturing, 13*, 295-316.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (NBER Working Paper No. 17699). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w17699

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Davis, J.B. (1994). A closer look at those who have decided to teach. *High School Journal, 77*(4), 274–79.

Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105-115.

Deci, E.L., Koestner, R., & Ryan, R.M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627-668.

Di Carlo, M. (2011, March 29). Teacher quality is not a policy. Retrieved from http://shankerblog.org/?p=2230

Dohmen, T., Falk, A., Huffman, D., &Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260.

Eckel, C. C. & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization, 55*, 447-465.

Fairhurst, A. M., & Fairhurst, L. L. (1995). *Effective teaching, effective learning: Making the personality connection in your classroom*. Boston, MA: Nicholas Brealey Publishing.

Faff, R., Mulino, D., & Chai, D. (2008). On the linkage between financial risk tolerance and risk aversion. *Journal of financial research, 31*(1), 1-23.

Feng, L., & Sass, T. (2008, March). Teacher quality and teacher mobility: Paper presented the Annual Conference of the American Education Research Association, New York, NY.

Finn, Jr. C.E. (2001). Getting better teachers—and treating them right. In T. M. Moe (Ed.) *A primer on America's schools* (pp. 127-150). Stanford, CA: Hoover Institute.

Fox, J.N. (1984). Restructuring the teacher work force to attract the best and the brightest. *Journal of Education Finance, 10* (2), 214-237.

Goldhaber, D., DeArmond, M., Player, D., & Choi, H. J. (2008). Why Do So Few Public School Districts Use Merit Pay?. *Journal of Education Finance*, 262-289.

Goldhaber, D., Gross, B., & Player, D. (2010). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best. *Journal of Policy Analysis and Management, 30*(1), 57-87.

Green, E. (2010, March 2). Building a better teacher. *The New York Times*. Retrieved from http://www.nytimes.com/2010/03/07/magazine/07Teachers-t.html?pagewanted=all

Hannaway, J. (Interviewee). (2009). Five questions for Jane Hannaway. Retrieved from the Urban Institute site: http://www.urban.org/toolkit/fivequestions/JHannaway_teachingreform.cfm?renderforprint=1

Hanushek, E.A., Kain, J.F., O'Brien, D.M., & Rivkin, S.G. (2005). The market for teacher quality. National Bureau of Economic Research Working Paper No. 11154. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E.A., & Rivkin, S.G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267-271.

Harrison, G. W., Johnson, E., McInnes, M. M., and Rutstrom, E. E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1(1), 31-35.

Hills, T., Stroup, W., & Wilensky, U. (2005). Patterns of Risk Seeking and Aversion Among Pre-Service Teachers: Mathematical Decisions, Preference, Efficacy, and Participation.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655.

Jensen, N.C. (2012). *Merit pay in Arkansas: An evaluation of the Cobra Pride Incentive Program in the Fountain Lake School District*. (Order No. 3505369, University of Arkansas). *ProQuest Dissertations and Theses*, 232. (1013441890).

Kane, T.J., Rockoff, J.E., & Staiger, D.O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615-631.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of research in science teaching*, *38*(6), 631-645.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75-86.

Krieg, J.W. (2006). Teacher quality and attrition. *Economics of Education Review, 25*, 13-27.

Kristof, N.D. (2012, January 11). The value of teachers. The New York Times. http://www.nytimes.com/2012/01/12/opinion/kristof-the-value-of-teachers.html?_r=0

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The future of children, 17*(1), 87-109.

Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal, 47*(4), 800-832.

Liang, L. L., & Gabel, D. L. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers. *International Journal of Science Education, 27*(10), 1143-1162.

Oyer, P., & Schaefer, S. (2005). Why do some firms give stock options to all employees?: An empirical examination of alternative theories. *Journal of Financial Economics, 76*, 99-133.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Economic Policy Institute, 1660 L Street, NW, Suite 1200, Washington, DC 20035.

Rinehart, J.S., & Young, I.P. (1996). Effects of teacher gender and principal gender on ratings of teacher performance. *Journal of Personnel Evaluation in Education, 10*(4), 313-323.

Rockoff, J.E., Jacob, B.A., Kane, T.J., & Staiger, D.O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy, 6*(1), 43-74.

Rubin, D.B., Stuart, E.A., & Zanutto, E.L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103-116.

Rushton, S., Morgan, J., & Richard, M. (2007). Teacher's Myers-Briggs personality profiles: Identifying effective teacher personality traits. *Teaching and Teacher Education, 23*(4), 432-441.

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal, 7*(7), 935-947.

Wagner, T. (2001). Leadership for learning: An action theory of school change. *Phi Delta Kappan, 82*, 378–83.

Weiber, A., Rost, K., & Osterloh, M. (2010). Pay for performance in the public sector—Benefits and (hidden) costs. *Journal of Public Administration Research and Theory, 20*(2), 387-412.

West, M.R., & Chingos, M.M. (2009). Teacher effectiveness, mobility, and attrition in Florida. In M.G. Springer (Ed.), *Performance incentives: Their growing impact on American K-12 education* (pp. 251-271): Washington DC: Brookings Institution Press.

Table 1: Sample Descriptive Statistics

Variable	Overall Sample Mean	Range	
		Minimum	Maximum
Teacher Quality Measures			
Student Achievement	1.17 (0.57)	-0.92	2.17
Performance Rating	0.85 (0.09)	0.55	1.00
Performance Pay			
Prefer Pay Based on Student Performance vs. Experience	3.43 (1.33)	1	7
Would Seek Merit Pay at Different School	3.05 (1.06)	1	5
Prefer Pay Based on Individual Performance vs. Team	4.46 (1.85)	1	7
Risk Preference			
# of Safe Choices	4.44 (2.45)	0	10
Teaching Experience	12.4 (10.3)	0	39
School			
Traditional Public	0.65	0	1
School Level			
Elementary	0.43	---	---
Middle	0.29	---	---
High	0.28	---	---
Female	0.78	0	1
N	93	---	---

Note: Standard deviations presented in parentheses. For the first performance pay measure a participant would get a 7 for indicating a preference for pay based 100% on student performance, a 4 for 50% student performance and 50% experience, and a 1 for 100% based on teacher experience. For the second pay measure a 5 is given for "Strongly Agree" and a 1 for "Strongly Disagree". The third measure is scored such that a 7 represents the desire for performance pay based entirely on individual efforts, a 4 for 50% team and 50% individual, and a 1 for entirely based on team performance.

Table 2: Teacher Pay and Risk Preferences and Student Value-Added Scores

Student Achievement	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Performance Pay Preference	-0.20*	-0.21*	-0.21*		
	(0.12)	(0.11)	(0.11)		
Risk Preference	0.06	0.07		0.06	
	(0.04)	(0.04)		(0.04)	
Individual Pay Preference	0.11	0.09			0.09
	(0.11)	(0.10)			(0.10)
Charter School	0.53*	0.55**	0.48*	0.36	0.38
	(0.28)	(0.25)	(0.24)	(0.23)	(0.24)
School Level					
Middle School	0.13	0.17	0.21	0.14	0.16
	(0.26)	(0.25)	(0.25)	(0.25)	(0.25)
High School	-0.84**	-0.81**	-0.93	-0.96***	-0.94***
	(0.34)	(0.33)	(0.32)	(0.32)	(0.33)
Experience	-0.01				
	(0.01)				
Female	0.11				
	(0.29)				
Consistent	0.10				
	(0.33)				
Constant	-0.49	-0.44	-0.10	-0.28	-0.06
	(0.56)	(0.28)	(0.19)	(0.27)	(0.20)
N	70	70	70	70	70
R-Squared	0.18	0.21	0.20	0.18	0.16

Note: Numbers in parentheses below coefficient estimates are standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% p-value levels, respectively, in a two-sided alternative to the null that the coefficient value is 0.

Table 3: Teacher Pay and Risk Preferences and Year-End Performance Evaluations

Performance Evaluation	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Performance Pay Preference	0.13 (0.10)	0.03 (0.11)	0.03 (0.11)		
Risk Preference	0.04 (0.04)	0.03 (0.05)		0.03 (0.04)	
Individual Pay Preference	-0.01 (0.10)	-0.03 (0.10)			-0.02 (0.10)
Charter School	0.81*** (0.27)	0.78*** (0.26)	0.77*** (0.25)	0.82*** (0.24)	0.79*** (0.24)
School Level					
Middle School	-0.53** (0.26)	-0.54* (0.27)	-0.51* (0.27)	-0.52* (0.25)	-0.51* (0.27)
High School	-0.09 (0.25)	-0.13** (0.26)	-0.12 (0.25)	-0.12 (0.24)	-0.13 (0.25)
Experience	0.01 (0.01)				
Female	0.78*** (0.26)				
Consistent	0.04 (0.31)				
Constant	-1.06 (0.47)	-0.15 (0.28)	-0.03 (0.20)	-0.18 (0.27)	-0.03 (0.20)
N	79	79	79	79	79
R-Squared	0.25	0.16	0.17	0.18	0.18

Note: Numbers in parentheses below coefficient estimates are standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% p-value levels, respectively, in a two-sided alternative to the null that the coefficient value is 0.

Appendices

Note: Actual document was formatted to take up the front and back of only one page.

Please complete this page first, then proceed to the back to complete the study. Upon completion you will be paid in cash for your participation as described below.

For each of the 10 lottery pairs listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option, either A or B, for each lottery.

In each lottery pair you will be selecting between a lottery that will pay either $6.00 or $4.80 (Option A) and a lottery that will pay either $11.55 or $0.30 (Option B). In the first lottery there is a 10% chance of receiving the larger payout for lottery 1 and a 90% chance of receiving the smaller amount. In each subsequent lottery pair the chance of earning the higher payout increases by 10%.

After you complete this study, the experimenter will roll a 10-sided die to randomly select which lottery will be used.

Next, the experimenter will roll the same die a second time to determine your actual payoff based on the option you chose for that particular lottery.

The number on the die for the second roll will determine whether you receive Payout 1 or Payout 2.

Example: If the experimenter's first roll is "3," then your payoff will based on Lottery 3. If you chose Option B for Lottery 3 and the second roll is "1", "2", or "3", you will receive a Payout of $11.55, but if the second roll is "4", "5", "6", "7", "8", "9" or "10", you will receive a Payout of $0.30.

Please answer the following question which will not impact your payoff but is intended to ensure you understand this task. Suppose the experimenter rolls a "2" first and then rolls a "9". If you have selected Option A for Lottery 2, what will your payout be? _____

Please notify the experimenter when you have answered this question before you continue with the experiment.

Lotteries. For each of the 10 lotteries listed below, please indicate if you would prefer Option A or Option B by inserting a check mark in the ☑ column. Please select only 1 option per lottery.

Lottery	Option A		Possible Roll(s) for Payout	Option B	
	☑	Payout		Payout	☑
1.		$6.00	1	$11.55	
		$4.80	2,3,4,5,6,7,8,9,10	$0.30	
2.		$6.00	1,2	$11.55	
		$4.80	3,4,5,6,7,8,9,10	$0.30	
3.		$6.00	1,2,3	$11.55	
		$4.80	4,5,6,7,8,9,10	$0.30	
4.		$6.00	1,2,3,4	$11.55	
		$4.80	5,6,7,8,9,10	$0.30	
5.		$6.00	1,2,3,4,5	$11.55	
		$4.80	6,7,8,9,10	$0.30	
6.		$6.00	1,2,3,4,5,6	$11.55	
		$4.80	7,8,9,10	$0.30	
7.		$6.00	1,2,3,4,5,6,7	$11.55	
		$4.80	8,9,10	$0.30	
8.		$6.00	1,2,3,4,5,6,7,8	$11.55	
		$4.80	9,10	$0.30	
9.		$6.00	1,2,3,4,5,6,7,8,9	$11.55	
		$4.80	10	$0.30	
10.		$6.00	1,2,3,4,5,6,7,8,9,10	$11.55	
		$4.80	---	$0.30	

Background Information

1. Year of Birth: _____

2. Gender (please circle): Female Male

3. Race/Ethnicity (please circle): White Black Asian Hispanic

Other (please specify): _____

4. Undergraduate Degree (please specify field and degree):_____

5. Undergraduate Institution:_____

6. What is your mother's highest level of education? (please circle)

Less than H.S. H. S. Some College Bachelor's Degree Graduate Degree

7. What is your estimated annual household income? (please circle)

Less than $35k $36K-$50k $51k-65k $66k-80k $81k-$95k $96k-$110k

$111k-$125k Greater than $125k

8. How many years total have you taught prior to this school year? _____

9. What is your primary teaching subject? _____

10. Did you participate in this school's performance pay program? (please circle) Yes No

11. If you participated in the performance pay program, were you employed at this school ***before it began***? (please circle) Yes No

For #12-15, please indicate how you feel about the following statements: (please circle)

12. The performance pay program was the deciding factor for me working at this school.

Strongly Disagree Disagree Neutral Agree Strongly Agree N/A

13. If I were looking to work in a different school, I would want to teach in a school that offered teacher performance pay.

Strongly Disagree Disagree Neutral Agree Strongly Agree

14. I believe that teachers should be paid based on student performance.

Strongly Disagree Disagree Neutral Agree Strongly Agree

15. I plan to spend the rest of my career as a teacher.

Strongly Disagree Disagree Neutral Agree Strongly Agree

On the scales below, please indicate how you, as a teacher, would prefer to be compensated.

16. ○ ○ ○ ○ ○ ○ ○

100% Student Performance 50/50 100% Teacher Experience

17. ○ ○ ○ ○ ○ ○ ○

100% Just My Performance 50/50 100% Team Performance

18. Please briefly list/describe any jobs that you had **after** college but before becoming a teacher:

_____

_____

Name Last: _____ Name First: _____

October 15, 2012

MEMORANDUM

TO:	Daniel Bowen Jonathan Mills Jay Greene
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	12-10-142
Protocol Title:	*Does Merit Pay Attract Teachers with Fundamentally Different Characteristics?*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 10/15/2012 Expiration Date: 10/14/2013

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (http://vpred.uark.edu/210.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 150 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 210 Administration Building, 5-2208, or irb@uark.edu.

Conclusion

Summary of Results

Research Question 1: Are teachers more risk averse?

Based on the findings of the first study, there appears to be a significant relationship between the decision to enter the teaching profession and risk aversion. This finding confirms a stereotype that is often attributed to teachers. This finding along with prior research on public sector employees suggests that low-risk, more predictable compensation formats are likely responsible for attracting more risk-averse individuals. Moreover, it does not appear as though socialization is likely the reason for teacher risk aversion. The first study finds that individuals entering the profession are more risk-averse than graduate students entering other professions. The results of the second study include an interesting finding with regard to the significant, positive relationship between years of teaching experience and risk aversion. While this relationship could be used to argue the case for socialization facilitating risk aversion, it is also reasonable to expect that this relationship exists in other professions. In other words, the decision to remain in the same profession for an extensive period of time probably reflects a life decision that is not unique to the teacher workforce.

Research Question 2: Does performance pay attract more risk-loving individuals to the profession?

The evidence from the second study is mixed and fairly inconclusive. The first set of analyses in the second study identifies teachers as having opted in to a performance pay program if they were hired with the program in place. The results that use this definition as the means for identifying the desire for performance pay appear to demonstrate that performance-based pay has

no significant impact on the composition of the workforce. However, it can be argued that these "opt in" teachers may not have seriously considered performance-based pay in their employment decision. In order to get a better sense of whether a strong desire for performance pay significantly predicts risk preferences, at least at the time of employment, a second set of analyses that incorporated participants' survey responses were conducted. These results do not affirm the results from the first set of analyses. Rather, using this alternative criterion for determining whether a teacher was seeking performance pay appears to confirm the notion that individuals who claim to have based their employment decisions on the performance-based pay format are indeed more risk loving. However, with there being so few participants indicating this intention (11% of the sample), these results remain fairly inconclusive.

Research Question 3: Are teachers' pay and risk preferences related to teacher quality?

Probably the most intriguing finding from this research is the main result from the third study that teachers who are attracted to performance-based pay are less effective than teachers who are more opposed to working under this compensation format. This result is counterintuitive because the most effective teachers have the most to gain, at least financially, under a performance-based pay system. This finding probably cannot be dismissed on the grounds that the teachers were simply unaware of their effectiveness. Most of the teachers in this sample have at least a few years of feedback for how well they perform in terms of student achievement. One possible explanation for this finding could be that more effective teachers are more mission driven and/or intrinsically motivated and this unobserved characteristic negatively influences participants' perceptions of performance-based pay.

Teachers who are more risk averse also appear to be more effective in terms of students' learning gains. However, this effect does not achieve traditionally accepted levels of statistical significance. The other measure of teacher quality, principal ratings on year-end performance evaluations, is not found to be significantly related to either teachers' pay or risk preferences.

Limitations

There are some noteworthy limitations to these studies. First and foremost are the concerns with regard to the sample sizes of these studies. While the sample sizes are fairly reasonable when compared to other studies of this nature, it is still important to keep in mind the limitations with regard to the generalizability of these results. Aside from sample size, there is also the general concern with regard to the representativeness of the two samples. The first sample consists entirely of graduate students enrolled at one university. Also, the comparison group of this study is confined to MBA and law school students; a broader set of prospective professionals would provide a more comprehensive comparison group. The sample for the second and third studies is confined to teachers employed at one of two school districts; one of which is a charter district and the other is a relatively rural, traditional public school district that was willing to implement a performance-based pay program. Therefore, these districts and the teachers who work at these schools probably have unique, unobserved characteristics that have not been fully taken into account in these analyses. Replication of these studies in different settings would help to determine whether these results are indeed more generalizable.

Another limitation to these studies is the inability to determine causal relationships between teachers' characteristics and preferences. The first and third study results are entirely based on descriptive statistics. Although there are significant relationships between occupational choice, risk characteristics, and pay preferences, the nature of how these different components

interact is not conclusive, at least based solely on the results from these studies. For example, the results of the third study cannot prove definitively whether a characteristic that influences teachers' perspectives of performance pay makes them more effective in the classroom or if this relationship works in the opposite direction.

The design of the second study provides the opportunity to get closer to making a causal argument by taking advantage of the relatively exogenous implementation of the performance pay program, in the fact that there was a sharp discontinuity in the type of compensation format being implemented at the school districts. However, there are still at least three major limitations to this study. One limitation is the extent to which the desire for performance-based pay is appropriately identified. The study attempts to get around this issue by running analyses with two definitions for teachers having chosen to work under performance-based pay. The fact that the results are different depending on which definition is used make it difficult to know with any certainty which result, if either, is more believable.

The second issue is that there is likely substantial attrition of teachers who entered with performance pay in place and then left later for whatever reason. Some of these teachers may have exited because they did not care for the performance pay program; some may have exited because they decided they no longer wanted to teach for reasons unrelated to the compensation format. It is highly plausible that excluding these attriters significantly biases the results from this study; although the direction of this bias is unclear. Unfortunately, it is not possible to track down and administer the risk task and survey with these teachers.

Finally, it would probably be more useful to have a measure of teachers' interests in performance-based pay at the application or hiring phase of a teacher's employment with the school. In many cases, the participants of this study are being asked to reflect on a decision they

made several years ago. It is very likely that teachers' perspectives of performance pay change with their experiences, and positive or negative experiences may produce a hindsight bias that prevents these participants from accurately recalling whether performance pay truly was or was not a deciding factor for their employment decision as well as whether they initially supported or opposed the program.

Implications

Despite the limitations of these studies, I believe that some of these results could have important policy implications. Probably the biggest implication or takeaway from these studies is that education policymakers should be mindful about the consequences regarding the potential impacts on the teacher workforce that stem from transitioning to a performance-based pay program. Individuals who go into teaching are relatively risk averse and, as a result, enacting a teacher performance pay policy could prove to be costly, both financially and politically. Furthermore, while changing the format of teacher compensation will likely attract a new crop of individuals to the profession, whether or not this reform improves teacher quality appears to be uncertain at best. At worst, performance-based pay could deter effective teachers from entering and remaining in the classroom.

If more mission-driven or intrinsically-motivated teachers are indeed more effective, then this finding could have additional education policy implications. Such a result would provide a case for further marketing and defining the job as a vocation more than a profession. This strategy would likely benefit from deemphasizing the importance of teacher wages and other financial benefits and focusing more on the altruistic, sacrificial nature of the occupation. Theoretically, allowing schools to have greater autonomy over defining their missions in

addition to developing their own hiring criteria would also likely improve teachers' abilities to select school environments where they are likely to exhibit greater efficacy.

I also hope that two additional implications from these studies can be drawn from the general methodological approaches incorporated in this research. I hope that one of the potential implications taken from this dissertation is that more controlled, laboratory-based research is seen as having value in its application to education policy research.. The other potential implication I hope for is that this research might serve as an example for the benefit of incorporating tools and methodological approaches from other disciplines into education policy studies. I do not wish to imply that these studies are the first to incorporate such methodologies. I also do not wish to remotely suggest that these studies are in anyway revolutionary in terms of their impacts on the future of education policy research. However, I hope that these studies can simply serve as examples for the potential benefits of incorporating interdisciplinary research techniques and the use of experimental measures in education policy, especially when field studies are unfeasible or a research question dictates the need to collect outcome measures that are not available in preexistent datasets.