2013

Assessment of Teachers' Reactions to a Knowledge- and Skills-Based Pay Structure at an International School

Joel Courtney Lowe
Lehigh University

Assessment of Teachers' Reactions to a Knowledge- and Skills-Based Pay Structure at an International School

by

Joel Courtney Lowe

A Dissertation
Presented to the Graduate and Research Committee
of Lehigh University
In Candidacy for the Degree of

Doctor of Education
in
Educational Leadership

Lehigh University
May 2013

CERTIFICATE OF APPROVAL

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Education.

_____
Date

_____
Jill Sperandio, PhD
Dissertation Director

_____
Accepted Date

Committee Members

_____
Jill Sperandio, PhD
Associate Professor of Education

_____
Floyd D. Beachum, EdD
Associate Professor of Education

_____
Louise Donohue, EdD
Professor of Practice

_____
Charles Kite, PhD
Deputy Headmaster/Business Manager,
Canadian Academy (Retired)

ACKNOWLEDGEMENTS

There are, of course, many people involved in a large work of research and writing. I could not have completed this work all on my own, and there are many to acknowledge.

The largest effort and sacrifice in service of this final goal has come from my wife and best friend Kristin. We have travelled a very difficult road together throughout the time of my studies, and she has faced it all with her characteristic faith, strength, and perseverance. In the midst of it, she was entirely encouraging and supportive. She is the strongest, smartest person I know. I thank God we embarked on our journey together more than 20 years ago. Not only I have gained strength from her, but so have our daughters Emeline, Quinlan, and Imogen. They have sacrificed much in supporting me in my study and research, and I thank them for their patience and understanding. I look forward to much more time together.

In my academic life, I must gratefully acknowledge the efforts of my committee: Dr. Jill Sperandio, Dr. Floyd Beachum, Dr. Louise Donohue, and Dr. Chuck Kite. Their advice, critique, enthusiasm, and encouragement have been insightful and inspiring, even in the face of adversity. Dr. Qiong Fu provided invaluable support with statistics and analysis, and Dr. Tara Levinson gave selflessly of her time to aide me in completing this study. Dr. Anthony Milanowski, Dr. Herbert Heneman, and Dr. Allan Odden also supplied valuable insight and shared their previous research. My colleagues in the International Program at Lehigh have become life-long friends and a powerful professional network, and I look forward to the continuing dialogue about what we all enjoy most. I also acknowledge the time and input of the many dedicated international educators who took part in the study, when they already had much on their plates.

Proverbs 3:6

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study explores teachers' reactions to a knowledge- and skills-based pay (KSBP) system implemented in a large international school. Such systems are designed to set teacher compensation based on demonstrated professional knowledge and skills as opposed to the traditional scale based on years of experience and degrees attained. This study fills a gap in the research into the reactions to KSBP of highly qualified teachers in a school with high-achieving students and in a competitive market in which the best teachers can choose which schools and countries to work and live in next. Based in expectancy theory (Vroom, 1964), a survey and focus group interviews were used to solicit teacher perceptions and reactions to the system. Exploratory factor analysis confirmed alignment of survey items with theory, and hierarchical multiple regression, bolstered by qualitative analytic induction (Patton, 2002) showed a significant relationship between teachers' perceived impact on their work and implementation issues (fairness, accuracy, and qualified evaluators) on their reactions to the system. Though teachers generally understood the system and agreed with the evaluation standards, the tie to pay was criticized, and teachers did not feel they received enough direct feedback to grow professionally except through their own intrinsic motivation. Teachers' perceptions showed that the system was not motivating overall, though almost everyone received a pay raise. Age was the only significant predictor of motivation, with the oldest teachers reporting least motivation. Finally, no significant difference was found between teachers who had completed the system and those who had not in their intention to return to the school. Overall, implementation issues were more connected to teacher reactions than design issues. The findings indicate that the leverage point for such systems is an increase in perception of professional growth rather than increased pay, followed by assurances that the system is fair and accurate.

CHAPTER 1

Introduction to the Study

The purpose of this study was to assess teacher perceptions of the impact of a knowledge- and skills-based pay system intended to motivate highly qualified teachers in high-achieving schools to improve pedagogy and ultimately improve student learning. For performance pay systems to be effective in increasing student learning, they must motivate teachers to improve their teaching practice. Furthermore, teachers must feel positively inclined toward the system and its ability to help them reach the goal of improved practice and improved student learning. The current study investigated factors under school control that influence perceptions and reactions to pay for performance in one such system in an independent international school. The study was conducted in a single international school because of the unique context it provided. International schools, established in many countries to educate the children of expatriate workers in the host country, are extreme examples of private schools: free from many of the constraints of national public schools such as teacher unions and centralized educational policies and school reform programs. They have more flexibility in the policies they establish and the curricula they follow.

International schools compete vigorously with each other for teachers, students and tuition dollars, making attracting and retaining high quality teachers a top priority. Teachers in this context are free agents, and the best international school teachers are able to pick and choose the schools and countries in which they wish to work. Motivating these teachers through alternative pay to engage in improving their practice must be done very deliberately. An assessment of one system's efforts at motivating and retaining

teachers in such an extreme case context is a litmus test for performance pay schemes more generally. The system must be motivating enough and generate enough positive reaction to retain and satisfy these high achievers. This study investigated teacher perceptions of the motivational effect of such a system, the elements of the system that contributed to their reactions, whether favorable or unfavorable, and the ability of a pay for performance system to retain highly mobile teachers.

Pay for Performance in Schools

In the light of research that increasingly indicates the importance of effective teachers in student learning (Sanders & Rivers, 1996) schools and national education systems are investigating ways to improve student learning through incentives offered to teachers (Sclafani & Tucker, 2006; Springer, 2009; U.S. Department of Education, 2009; Woessmann, 2010). Consequently, schools and researchers are devoting attention to studying the use of performance pay and its effects on students and teachers. Although as many as 96 percent of public schools in the U.S. continue to pay their teachers according to years of service and level of education (Podgursky & Springer, 2007), many schools and districts have experimented with alternative pay systems. Such pay systems may rely on student test scores as measures of performance, but may also tie pay to teacher acquisition of new knowledge and skills, teaching particular strategic subjects (such as mathematics and science) or working in schools in which teacher retention is problematic (Podgursky & Springer, 2007; Springer, 2009).

Proponents of performance pay argue that the traditional single salary scale does not offer teachers substantial motivation to improve, and that it does not reward any of the instructional behavior found to be related to enhancing student learning (Odden &

Kelley, 2002). Proponents further argue that paying according to some measure of performance is a powerful way to motivate teachers to work at improvement of classroom instruction, to attract high-quality teachers, and to encourage less effective teachers to leave the school or the profession (Hanushek & Rivkin, 2004; Hanushek, 2007; Solmon & Podgursky, 2001; Solmon, 2006).

Opponents of performance pay posit several arguments against its use in education. The most common arguments point out that education is highly complex and context-dependent. The factors that contribute to student learning are difficult to measure. They are often the result of collaborative effort, and attempting to isolate these factors may damage that collaboration and undermine the traditional intrinsic motivation of teachers (Johnson, 1984, 1986; Murnane & Cohen, 1986; Pechthalt, 2007).

US Public Schools

The early public school merit pay programs of the 1980s failed primarily because of lack of adequate funding and the use of idiosyncratic measures of effectiveness (Murnane & Cohen, 1986). The current political climate in the US is partially responsible for the resurgence of alternate pay systems. In 1999-2000, fewer than 6 percent of public school districts used some aspect of performance tied to pay, whereas in 2003-04, that number had risen to approximately 13 percent (Podgursky & Springer, 2007). Many states and districts in the US have begun developing their own performance pay systems or adopting systems used in other states. The Obama administration's Race to the Top initiative has also given such programs a boost with its mandate that states provide "opportunities for highly effective teachers and principals…to obtain additional

compensation and be given additional responsibilities" (U.S. Department of Education, 2009, p. 9).

Public Schools Internationally

While the debate in the US continues, several other countries also report at least some use of performance pay for teachers. For the first time in 2003, the Organisation for Economic Cooperation and Development (OECD) surveyed member countries on the use of performance incentives. In that survey, 13 of the 28 member countries indicated some use of performance incentive (Woessmann, 2010). Analysis of a survey administered six years later found that 15 OECD countries and three partner countries "offer an additional payment for outstanding performance in teaching" (Organisation for Economic Co-operation and Development, 2009, p. 397). While the larger number may be an indication the trend is growing, caution must be applied to the analysis. The United States, for example, was listed as one of these countries because some states did use forms of performance incentive. Clearly, however, the U.S. context was far from homogeneous, and this variability is likely found in other countries.

Private Schools

The growth in pay for performance in public schools has been outstripped by its more prevalent use in private schools; however, it is significantly more wide-spread in secular private schools than in religious ones (Ballou, 2001). Because they are not public institutions, private schools have more latitude in confidential decision making, which may account for their more wide-spread use of alternate systems for pay, fringe benefits, and other personnel policies that may not be accepted in a public, unionized context.

Research indicates that teachers had more positive attitudes toward performance pay in private schools than in public ones (Ballou, 2001). Ballou posited several reasons for this difference, including the idea that private schools may attract more entrepreneurially minded teachers from the beginning. Such teachers may be more positively inclined toward systems that compensate them based on their performance. A further possibility is the fact that private schools operate in a more competitive environment than do public schools. A high quality faculty is one distinguishing characteristic schools may use as a competitive advantage to differentiate themselves from one another. Teachers in such an environment may realize that high performance from teachers in general has a positive spillover effect on the school and other faculty in the attraction of students to the school, enhancing quality and reputation.

Certain private school findings thus support the notion that private school teachers may find performance pay motivating. Similarly, some research suggests that younger teachers have more positive attitudes to performance pay than older ones (Coggshall, Ott, Behrstock, & Lasagna, 2010; Milanowski, 2007) Yet it must be acknowledged that pay for performance in practice may also be de-motivating. Indeed, in one large study, private school teachers overall expressed more positive attitudes toward performance pay in general. However, those who taught in schools that actually used performance pay were less positively inclined (Ballou, 2001). The implication is that private school teachers have a more positive attitude toward performance pay until they actually experience it.

International Schools

Though performance pay is clearly more common in private schools than in public schools in the U.S. (Ballou, 2001), it is not yet apparent in international schools.

Many international schools seek accreditation from U.S.-based agencies and consequently follow trends in U.S. education and educational reform; however, pay for performance is a notable exception. Though they exhibit many of the qualities of national private schools with the added dimension of a highly mobile teacher population, international schools have heretofore not served as a research site for performance pay.

Teacher Reactions to Pay for Performance

If the goal of pay for performance programs is to motivate teachers and, by extension, improve student achievement, it is essential that teachers see them as positively motivating and view them positively overall. Schools and districts have attempted variations on these systems in order to eliminate the competition many teachers believe inherent in merit pay (Odden & Kelley, 2002). Knowledge-and skills-based pay (KSBP) systems are one example. KSBP systems are designed to recognize and reward teachers for exhibiting qualities of effective teachers without limits on how many outstanding teachers can be identified. Though studies have used large-scale statistical analysis to detect the relationship between teacher motivation and attitude and the existence of pay-for-performance programs (Ballou & Podgursky, 1993; Ballou, 2001; Heneman & Milanowski, 1999; Heneman, 1998), few qualitative site-based studies have been conducted to explain how and why teachers are opposed to or in favor of them.

Several researchers have investigated the particular conditions under which teachers feel positively inclined toward and accepting of school-based pay-for-performance programs. Based in expectancy theory, which holds that individuals will pursue rewards when they desire the reward and they have a belief they can attain the reward, researchers have determined key factors related to teacher reactions to pay for

performance programs. The presence of enabling conditions, prior success, and perceived fairness of the program all have a significant impact on teacher expectancy (Ha, 2003; Heneman & Milanowski, 1999, 2003; Heneman, 1998; Kelley, Heneman, & Milanowski, 2002; Milanowski & Heneman, 2001). Such studies are helpful for program designers and have contributed to the field through instruments that enable an investigation of the specific factors influencing teacher reactions to performance pay.

Because advancement of student knowledge is the ultimate goal of any educational institution, research into the effect of such pay programs is usually tied to student achievement. Studies in both developing and developed countries have shown weak short-term effects on student achievement when teachers or schools are taking part in a pay for performance scheme, whether rewards are earned by schools, groups of teachers, or individuals (Atkinson et al., 2004; Figlio & Kenny, 2007; Glewwe, Ilias, & Kremer, 2003; Lavy, 2002, 2007; Muralidharan & Sundararaman, 2009). Other studies have shown mixed or no significant effect (Hudson, 2010; Marsh et al., 2011; Springer et al., 2010), while still others have shown a negative impact on student achievement in schools engaging in some form of pay for performance (Fryer, 2011). In summary, the findings of these studies show that pay for performance systems may have some weak effects in schools with relatively low-income students with low academic performance and low parent involvement. It is unclear from the research if such results are transferable to schools that are characterized by affluent, high-performing students and with relatively highly skilled teachers, as is the case with many private and international schools.

Pay for performance systems may have additional goals beyond improving student learning. Motivating teachers to improve their practice and to remain in a school

with an engaging professional evaluation and development program are examples. Accordingly, researchers and policy makers are interested in teachers' reactions to these programs, and this has encouraged studies specifically on teacher reactions and motivations (Ballou & Podgursky, 1993; Goldhaber, DeArmond, & DeBurgomaster, 2010; Ha, 2003; Hart, 1987; Heneman & Milanowski, 1999, 2003; Heneman, 1998; Kelley & Finnigan, 2003; Kelley et al., 2002; Kelley, 1999; Milanowski & Heneman, 2001). Can pay for performance systems motivate teachers to improve their pedagogy with a target of improving student learning?

Purpose and Research Questions

This study examined the question of whether a knowledge- and skills-based pay system can have a favorable motivational impact on teachers by analyzing teacher perceptions of one such system operating in a single independent school with a highly-qualified population of teachers with a wide range of choices regarding where to live and work. The research questions were:

RQ1 To what extent do teachers perceive the use of a performance pay system at their school motivates them to improve their practice?
RQ2 What factors influence teacher reactions toward the performance pay system?
RQ3 Is there a correlation between participation in the pay for performance system and teachers' decisions to renew their teaching contract with the school?

The Research Site

The International School (TIS) is a pre-kindergarten through grade twelve international school in an overseas country. (The name of the school has been replaced with a pseudonym in this study). The school was founded in the 1960s and has been accredited by a US-based school association since the early 1970s. The school currently

serves 2,640 students across all the grade levels and employs 246 teaching faculty and 25 administrators. It is governed by a self-perpetuating Board of Managers. When a member of the board resigns or retires, the board itself chooses a new member to replace him or her; there are no elected members. TIS enjoys a reputation as one of the most sought-after of the roughly 40 international schools in its host country. The school has long waiting lists for enrollment at all grade levels. More than 75% of students are ranked as "above average" (Stanines 7-9) on standardized tests of achievement used at the school, relative to other private school test-takers. Graduates attend the most prestigious universities in the world. Similarly, the faculty is highly qualified by traditional measures. The average length of teaching experience for teachers is 13 years. Many faculty hold advanced degrees in their fields, with 45% holding master's degrees, 12% more than one master's, and three percent holding doctorates. TIS is a unique and rich context in which to study the motivational impact of a performance pay system on high-quality, mobile teachers teaching high achieving students.

The school introduced a knowledge- and skills-based pay system called TIS Knowledge and Skills (TISKS) in 2009. (The name of the program has been replaced with a pseudonym in this study). TISKS determines teachers' level of pay according to their demonstration of a set of knowledge and skills deemed characteristic of effective teachers. (See Appendix A for a detailed explanation of the requirements, processes, and policies pertaining to the system). The nature of TISKS makes for an excellent research site in the commitment to the system as the complete determiner of pay. While many systems use job enlargement or additional duties to determine the amount of a bonus, TISKS scores are the sole determinant of the base salary of teachers at TIS. Teacher

perception of the system is therefore all the more important, and highly compelling from a research perspective. At the time of this study, the school had had three cohorts of teachers complete the system. The three cohorts constituted roughly half the teacher population, which made it an ideal site to investigate teacher perceptions and reactions with the ability to compare results between those who had taken part in the system and those who had not.

Knowledge- and skills-based systems are portrayed as being well-suited to enhance professional development for teachers and align teacher efforts with institutional goals (Odden & Kelley, 2002). TISKS was designed to pay teachers according to their demonstrated level of knowledge and skills that are compatible with the strategic goals and direction of the school. Ultimately, TIS sees TISKS as a system for (a) identifying the strengths and needs of teachers, (b) paying them accordingly, and (c) developing them in the areas of need so they can become the kind of teachers that will continuously move the school forward. Accordingly, the standards against which teachers are assessed were developed from the widely used work of Danielson (1996) on an evaluation framework designed to enhance teachers' professional practice. The school also offers an extensive menu of professional development activities in an effort to support teachers to be successful in the classroom and in the TISKS process. One intention of the system is to create a circumstance in which highly mobile teachers feel motivated and supported enough to remain at TIS as their school of choice.

Summary of Methodology

The present study followed a mixed-method concurrent triangulation design. Teacher perceptions of the effectiveness of the system in achieving its goals of

motivation to improve pedagogy and the factors that contribute to overall perceptions were investigated using two approaches: a quantitative and open-ended survey open to all teachers in The International School and focus group interviews conducted with a group of teachers who had completed the profile year of the system. The survey gathered demographic data on all responding teachers, their self-reported perception of the effectiveness of the system to motivate teachers to improve pedagogy, and their intentions to remain at TIS. Further items gathered perception data on the experience of the system from teachers who had completed the process. These perceptions were gathered across several dimensions: their understanding of the system, their acceptance of the system, the impact of the system on their work, their perception of the role of their evaluators, the fairness of the process, the fairness of the results, the accuracy of the results, the stress associated with the process, the amount of effort required of them by the process, and their overall favorableness toward the process. Data were analyzed for statistical differences between the groups of pay system participants and non-participants and intercorrelations among the dimensions and overall favorableness toward the system for pay system participants.

A group of TISKS teachers from each division in the school were invited to take part in focus group interviews after the survey was administered. The interviews prompted participants with questions about their experience of the system. Data from the interviews was intended to confirm and corroborate findings from the survey. Such triangulation was intended to strengthen findings and contribute to robust analysis.

Rationale and Significance

This research is useful for schools considering implementation of a pay for performance system. More and more school boards are calling for accountability and investigation of alternative methods of paying teachers. Administrators who must develop such systems will need guidance and a research base to determine the extent to which they are viable and effective at motivating teachers to improve their practice, leading to improved student learning. New compensation models may also be sought for strategic cost management in schools competing for teachers. For such purposes, it is essential to determine if the system change has a positive effect on the teachers schools wish to reward and retain. This research study adds insight into whether this is the case in schools with high achieving students and highly qualified teachers.

Definition of Terms

The following definitions provide an explanation of the key terms used throughout the study.

International school – A typically Western-influenced school established in another country to provide a link for expatriate students to their home culture and educational values. Though international schools typically educate students from multiple nationality groups, most follow a North American or International Baccalaureate curriculum model.

Knowledge- and skills-based pay (KSBP) – A pay system rewarding teachers for attaining and demonstrating use of knowledge and skills valued by the school. In such a system, teachers are evaluated according to pre-determined, clear standards.

Pay for performance or **performance pay** – In the context of this study, these synonymous terms refer to pay systems that determine some portion or all of a teacher's compensation according to some performance on the part of teachers rather than solely on the traditional model of paying for years of experience and degrees attained. Examples of performance vary. Examples include, but are not limited to job enlargement, extra duties, an increase in student test scores, or the acquisition and exhibition of knowledge and skills valued by the school.

Teacher – Though the pay system in this study had different rubrics for different faculty roles such as teacher, counselor, teacher-librarian, tech facilitator, and athletic director, for clarity the term teacher is used collectively throughout the discussion to refer to all of these roles.

Teacher reactions – Refers to the extent to which a teacher views an element of a performance pay system, the effect of a performance pay system, or the system itself—either positively or negatively.

Teacher turnover – For the purposes of this study, teacher turnover refers to teachers leaving a school at the end of a current teaching contract.

TISKS teachers – In the context of the research site of this study, refers to the group of teachers who have completed some version of the knowledge- and skills-based pay system in place at The International School. Those who have not yet completed or begun the process are referred to as **Non-TISKS teachers**.

CHAPTER 2

Literature Review

International schools are uniquely free from the constraints of national systems; however, many tend to follow trends in U.S. education, since they originally set out to educate students for matriculation into U.S. universities. It is therefore natural to look predominantly to U.S. trends in teacher compensation to establish context for the present study. This chapter begins with an explanation of expectancy theory, a theory of workplace motivation that serves as the study's theoretical framework. It then presents an overview of changes in teacher compensation in the United States. Research into the effects of alternate pay programs is then summarized in U.S. and international contexts. Finally, a summary of research into the reactions of teachers to such systems, and a summary of teacher motivation research using expectancy theory is presented. The final section connects the literature reviewed to the research questions of the study.

Theoretical Framework: Expectancy Theory

The motivational effect of pay for performance has been researched in light of several theories of motivation. As research into motivation should be grounded in an elaborated theory of motivation, the present study will utilize expectancy theory (Vroom, 1964) in keeping with past research.

Expectancy theory serves not only as a basis for research, but can also inform the design of a pay system to maximize teacher motivation, as it did in the design of the TISKS program at TIS. Figure 1 offers an illustration of the way in which expectancy theory can inform the effect of a pay system on teacher performance (Kelley et al., 2002, p. 3). Though expectancy theory is used in a variety of fields to explain the motivation of

workers, the present discussion will use the term *teachers* exclusively. The theory holds that teachers will be motivated to respond with increased effort and improved performance when three conditions are in place: expectancy, instrumentality, and valence (Heneman, 1998; Heneman & Milanowski, 1999, 2003; Odden & Kelley, 2002).

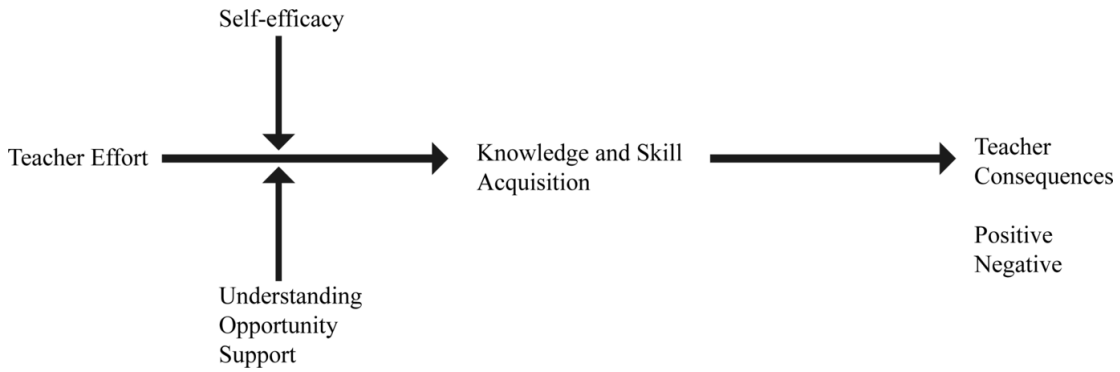

Figure 1. Motivational model for knowledge- and skill-based pay based on expectancy theory.

Expectancy Perception

Teachers must have a "line-of-sight" perception that increased effort will lead to the desired result, specifically that they will acquire the knowledge and skills seen as important by the school and the teachers themselves (Heneman, 1998; Milanowski, 2002; Odden & Kelley, 2002). Several factors influence the teacher's belief that they can acquire the targeted knowledge and skills, including the teacher's sense of self-efficacy. Institutional factors include the extent to which the teacher understands the definitions and ways to demonstrate the knowledge and skills, the institutional support opportunities made available to teachers to develop the knowledge and skills, and the amount of support they receive from their colleagues and administrators. Institutional support includes high-quality, targeted professional development and resources. Peer and administrator support includes the leadership of the principal, feedback (usually from

observation) on progress toward meeting expectations, and a school culture fostering the acquisition of targeted knowledge and skills. If teachers believe they understand what is required, have the right support, and are provided the right opportunities to gain the knowledge and skills, their expectancy perception that they can succeed will be high (Milanowski, 2002; Odden & Kelley, 2002). In Figure 1, the arrow running from teacher effort to knowledge and skill acquisition represents the expectancy perception, with the various factors affecting it also illustrated.

In school contexts, research indicates that teacher expectancy is connected to professional development support offered by the school to improve, the fairness of the performance system, a lack of conflict among goals, and the history of the school in granting awards in the past (Kelley & Finnigan, 2003).

Instrumentality Perception

The arrow in Figure 1 leading from knowledge and skill acquisition to teacher consequences represents the instrumentality perception. The perception reflects the notion that acquiring the knowledge and skills valued by the school and positive consequences such as receiving a pay increase or seeing an increase in student learning are strongly connected. If teachers do not believe the reward is linked to acquiring the proper knowledge and skills, then they will not be motivated to acquire them (Milanowski, 2002; Odden & Kelley, 2002).

The positive strength of the instrumentality perception is affected by several factors. For example, if teachers know the pay increases are well-funded, and if they see their colleagues getting pay increases based on evaluations that indicate they have acquired the requisite knowledge and skills, they will be more likely to be motivated to

acquire the knowledge and skills themselves. An important factor in the equation is the way in which teachers are assessed in their exhibition of the valued knowledge and skills. If teachers do not see the evaluation system and their evaluators as fair, valid, and reliable, they will not be motivated to seek new knowledge and skills. Teachers need to believe their efforts will lead to the promised benefits in a fair, systematic way in order to be motivated to work toward those goals (Milanowski, 2002; Odden & Kelley, 2002).

In the case of TISKS, perceptions of instrumentality may be affected by the outcomes of the process. This ranges from a pay level based on exhibited knowledge and skills (usually hoped to be higher than the current level of pay), recognition that they are valued members of the school community, and higher student achievement, though the school does not directly measure the latter. Furthermore, teachers must believe the ratings they receive are based on a fair, systematic appraisal, consistently applied by qualified evaluators. In other words, it is important that they see themselves as able to succeed and receive the promised reward. As to resources, teachers must believe the system is adequately funded in order to allow all teachers to receive higher pay should their scores warrant; otherwise, the system will appear to be a form of merit pay in which teachers compete for a finite set of rewards.

Valence

The final factor in expectancy theory relates to how valuable the reward is to teachers. A small monetary reward, for example, may not have enough motivational impact to push teachers toward the goals (Johnson, 1986). In this case, valence is low. If the positive consequences are highly valued, expectancy theory says teachers will be more likely to be motivated to gain knowledge and skills leading to those consequences.

Avoidance of negative consequences is also motivating. If teachers' ratings are shared among peers, and teachers do not want to be seen as less than expert in their field, they will be motivated to seek the higher rating for social reasons if nothing else, which also results in high valence (Milanowski, 2002).

Outcomes other than a higher salary may also lead to an increase in valence. Extrinsic motivators such as pay can be bolstered or hindered by intrinsic motivators such as a sense of pride in one's work, collaborative relationships with one's colleagues, and a sense of satisfaction when one sees students improve and succeed (Kelley, 1999; Odden & Kelley, 2002). Valence is highest when intrinsic and extrinsic motivators are aligned. If the definitions and levels of knowledge and skills on which teachers are evaluated are consistent with their own conceptions of high quality and essential abilities of skilled teachers, the system will likely be all the more motivating due to high valence.

Stress, fatigue, and frustration are commonly cited by-products of knowledge- and skills-based pay systems, especially if they include a portfolio component, as does TISKS. These by-products may cause teachers to withdraw their support from a program if they are deemed more detrimental than the reward is worth (Heneman & Milanowski, 1999, 2003; Kelley, 1999; Odden & Kelley, 2002). Valence is the extent to which such factors influence teacher perception that the effort is worth the reward, whether that reward is financial or more intangible. Even in the absence of stress, fatigue, and frustration, creation of the portfolio and a felt need to defend one's actions certainly require mental energy and time. Valence indicates teachers' acknowledgement that such efforts are worth the promised rewards financially and in professional growth.

For the present study, this theory serves as theoretical framework in several ways. Survey items and interview questions designed for data gathering were designed in the frame of expectancy theory, including factors that enable teachers to have high expectancy in the outcome of going through the system (such as administrative support). The design of TISKS was based on expectancy theory, and so is the design of the instruments for the study.

A History of Teacher Compensation Reform

Almost every school in the United States currently pays its teachers according to a single salary scale, typically determined by a teacher's years of experience and degrees and qualifications earned. The single salary scale system has been widely adopted by international schools so this method has become the norm. The single scale structure has been in place since at least 1921 (Odden & Kelley, 2002). The system came into being in order to make pay structures intentionally more rigid. Prior to the single scale, teachers could be paid in wildly variable ways. For example, women were often paid on average much lower than their male colleagues. The same was true for lower-paid primary school teachers when compared to their secondary school counterparts. Likewise, a teacher who had housing provided could receive much less cash compensation. The single scale, inasmuch as it focused on a fixed set of inputs, made strides in putting an end to this discrimination, which led to its wide adoption. Hassel (2002) indicates that approximately 97 percent of schools in the U.S. had adopted such pay scales at the time of his survey in 2002.

Adoption of the single salary scale does not mean it has not been used without problems, or without detractors. Ironically, a perennial criticism is that the structure pays

all teachers equally. While it seems to have leveled inequalities in pay based on demographic factors (age, race, gender, area of teaching) it provides no way to distinguish between teachers who are more and less effective (Ballou & Podgursky, 2001). Furthermore, some argue paying everyone the same provides no incentive to teachers other than the intrinsic rewards of teaching. However, while calls for reform of teacher pay have accelerated in recent years, it is not the case that a call for change is universal. Resistance to change is clear, owing to the fact that for teachers at least, the single pay scale is stable and predictable. Teachers know they can raise their level of pay through longevity and a program of continuing education credits (Mohrman, Mohrman, & Odden, 1996).

Since the call for education reform with *A Nation At Risk* (1983), policy makers have attempted various alternate pay systems with sustained and growing interest since the 1980s. Ballou (2001) reported that in 1993, some 12% of public school districts and 35% of secular private schools in the U.S. were using some form of alternative merit pay systems. While the percentage of schools in the U.S. may seem small, Woessmann (2010) reports that, of the 28 member countries in the Organisation for Economic Co-operation and Development (OECD), 13 countries reported using some form of incentive pay in 2003. Global interest has grown with the proliferation of economic models to judge the value of teachers based on student outcomes.

A growing research base indicates that teacher quality is the strongest contributor to student achievement but that years of experience and degrees attained are not strong indicators of teacher quality (Goldhaber & Brewer, 1999; Goldhaber, 2002; Hanushek, 2003). Further research indicates that a series of less effective teachers can have long-

term and cumulative negative effects on students' learning (Sanders & Rivers, 1996). In question with such studies is a clear definition of teacher effectiveness; however, teachers clearly do make a difference. Since teacher salaries make up anywhere from 50-70% of school budget expenditures (Firestone, 1994; Odden & Kelley, 2002), accountability is high for leaders allocating those dollars. A common rule of thumb states that international schools spend approximately 70% of their operating budgets on teacher compensation. Schools and policy makers therefore seek to recruit, motivate, and retain the most effective teachers to improve student achievement and maximize strategic investment. Reforming the pay system may be one way to achieve these goals (Reichardt & Van Buhler, 2003).

Common Alternative Pay Systems

Pay reforms take many shapes. One common call has been to raise teacher salaries across the board. Hanushek (2003) suggests such raises can actually be counter-productive, since "Current teachers, both good and bad, would be encouraged not to leave teaching" (p. F93). Indeed, Hanushek goes on to point out international test data correlated with national expenditure on education. Using test data from the Program for International Student Assessment (PISA) and data on national expenditure from the OECD, Hanushek finds no relationship between expenditure and higher test results, indicating on one level that more money for teachers in isolation is not the answer for raising student achievement. Such increases would not do anything to rectify the disconnect between teacher and administrator pay and student achievement.

To address this gap, policy makers and researchers have created several different models of performance-based pay systems. Though many are commonly termed "merit

pay," schools use several distinct types including career ladder programs, merit pay, school-based performance awards (SBPA) and knowledge- and skills-based pay. The two most common types today are merit pay and knowledge- and skills-based pay (Springer, 2009), described below.

Career Ladders

Career ladder plans typically emphasize job enlargement, targeted extra compensation for valued duties, and professional development opportunities to further enhance skill and opportunity. Such systems envision teachers in various grades—or on different rungs of a ladder, such as "novice", "professional" or "master" teachers according to observations and evaluations usually conducted by supervisors (Protsik, 1996). As teachers move up the rungs, pay increases, and usually responsibility does as well. Teachers who are "master," for example, may be called upon to be team leaders, department heads, or mentors to new or struggling teachers. Though many of these systems were abandoned in the 1990s, some researchers reported positive effects in teacher motivation. Firestone (1991) reported through an intensive case study approach of two districts with more than 1200 teachers, that they felt more motivated by the career ladder systems of their schools than by any proposed traditional merit pay system. However, most of the career ladder systems popular in the U.S. in the 1980s were eventually dropped (Ballou & Podgursky, 1997; Murnane & Cohen, 1986).

Merit Pay

Whereas the pay reforms and career ladder systems of the 1980s and 1990s can be characterized as systems that focused on inputs (Ballou & Podgursky, 2001; Hanushek, 2003; Springer, 2009), the term "merit pay" most closely matches those systems that

focus on outputs. In the strictest sense, merit pay refers to programs that reward teachers based on observations, student test scores, teacher portfolios, or various other indicators of effectiveness. Rewards may be granted to an individual teacher, groups of teachers, or entire schools.

Programs offering incentives to teachers may tie students to a particular teacher and attempt to discern the value the teacher adds to their learning, paying them accordingly. Value added is often determined through analyzing students' standardized test scores. The approach is not without detractors, and is far from conclusively successful, as discussed below. Other factors used to determine differential pay include signing contracts in hard-to-staff schools and in subject areas in which it is more difficult to retain teachers, typically math and science.

School-Based Performance Awards (SBPA)

Hassel (2002) indicates that school-based performance awards (SBPA) are the most commonly implemented form of merit pay in the U.S. In such a system, schools are rewarded for their collective results, often through forced ranking and shared payments. For example, the top five schools in a district may each be given a bonus amount that is then divided and used according to the vote of the faculty and staff of the school.

In contrast to traditional merit pay programs, school-based performance award programs seek to reward groups of teachers, or the whole school based on meeting particular goals or improving student scores. Such programs may be more suitable for education than other fields, given that teachers are not entirely in control of all the factors that contribute to measured job performance. Merit pay programs are probably best suited to those situations in which individuals are in control of their work and can be more

easily evaluated on their output. Some argue that more strategic design principles must be taken into effect when designing pay systems, including the nature of the work force and its goals (Lawler, 1995; Mohrman et al., 1996; Odden & Kelley, 2002). For schools, awarding bonuses or pay increases based on the work of a group of teachers seems to make sense, since the outcomes for students are most often based on the work of many in concert rather than a single teacher (Kelley, 1997).

Typical models of SBPA programs award bonuses to teachers and non-teaching staff for reaching specific goals with students in a given timeframe (usually an academic year). In some cases, staff vote to decide how to allocate and/or distribute the bonuses (Odden & Kelley, 2002). The underlying principle is that group rewards will motivate teachers to work together, hence supporting the collaborative nature of teaching (Kelley et al., 2002). Several districts and states in the U.S. have implemented SBPA programs with varying degrees of success including Charlotte-Mecklenburg, North Carolina; Dallas, Texas; Maryland, Kentucky, South Carolina, and Douglas County, Colorado.

Knowledge- & Skills-Based Pay Systems

Knowledge-and-skills based pay programs emphasize inputs. The existing single salary scale is a focus on inputs in which years of experience and levels of education stand as proxies for the knowledge and skills of effective teachers (Odden & Kelley, 2002). It is assumed that more experience and more advanced qualifications are equal to greater knowledge, skill, and effectiveness in the classroom. Newer pay systems hold that such traits are subject to empirical observation and need not be assessed solely by proxy. Teachers are therefore rewarded for developing and exhibiting new skills and knowledge considered to enhance instruction and further student learning. Notable examples are the

programs created by the Consortium for Policy Research in Education at the University of Wisconsin-Madison (Odden & Kelley, 2002). Such programs usually involve repeated classroom observations and a portfolio of work to be judged against adopted criteria for teacher effectiveness. They may look to external validation of a teachers' skill such as scores on teacher proficiency exams or certification through the National Board for Professional Teaching Standards. Empirical research on these systems is underway, but still inconclusive (Springer, 2009).

Potential Benefits of Performance Pay

Calls for performance pay in the U.S. have recently been renewed, partially because of support from the U.S. federal government. Such calls are usually associated with thinking that making teachers more effective will in turn raise student achievement results. The concept comes from research in industry. Lazear (2000) shows that a pay for performance program initiated in a windshield installation company increased productivity by some 44%. In a follow-up, Lazear (2001) acknowledges that installing windshields is very different from educating students, but the general finding was that incentives matter. He argues that rewarding effective teachers may cause those more likely to be effective to enter and remain in the profession. Similarly, performance pay may also serve as an incentive for the less productive on staff to leave; as their pay drops because of low performance, they are likely to seek employment in other fields. Others claim that alternate compensation systems may motivate early-career teachers who are likely to be successful to remain in teaching (Hart, 1994; Milanowski, 2007).

Paying teachers for various outputs beyond student achievement may align teacher and institutional interests, benefitting school reform. Hassel (2002) states,

"Research in the private sector suggests that for pay systems to support organizational goals, they need to be aligned with the organization's culture and how people are organized to work toward those goals" (p. 3).

Researchers have found evidence that incentive pay for teachers can have effects on such extended goals as student retention rates. In a study that focused on a performance pay program designed to retain students, Eberts, Hollenbeck, and Stone (2002) compare two similar alternative high schools in Michigan. One implemented a 12-12.5% bonus for teachers who retained 80% of their students each quarter. Furthermore, teachers became eligible for further bonuses if their students rated them highly in reviews. The other school maintained a standard salary scale. Using difference-in-difference analysis, the researchers found that the performance pay program did indeed increase student retention by 40%. However, they also found that students in the merit pay school experienced on average a 0.53-point drop in average GPA as opposed to a 0.37 drop at the control school. Furthermore, the merit pay school evidenced a drop in daily attendance. The researchers posit that the drop in GPA could be related to the retention of more low-achieving students. The drop in daily attendance could be attributed to the lack of incentive to manage daily attendance in favor of retention. The study therefore contributes to the notion that a performance pay program must include all elements deemed most important to a school's community if it is to support teachers in meeting them. In keeping with this premise, a knowledge-and-skills based pay system may be a way to encourage best practices through focused compensation (Mohrman, Mohrman, & Odden, 1996; Odden & Kelley, 2002), but the school or district must clearly articulate those goals they most value.

Potential Problems with Performance Pay

Murnane and Cohen (1986) cite the fundamentally multi-dimensional nature of teachers' work as a reason for the demise of most pay-for-performance programs in education. Goldhaber, DeArmond, Player, and Choi (2008) subsequently titled this argument the "nature of teaching" hypothesis. Among its tenets are (a) the difficulty of measuring a teacher's output, (b) the collaborative work necessary in schools, and (c) the complex constellation of tasks that make up the work of teaching.

Measuring Output

The output of teachers is hard to observe and measure reliably and validly. Recent development in complex statistical models and longitudinal student databases for tracing teacher added value may be diminishing this argument as regards student achievement measured by test scores, but the models' suitability for making decisions about teachers' careers is subject to debate (Armour-Garb, 2009; Ballou, 2002; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). Systems based on such models also consistently show instability in the ranking of schools (Milanowski, 1999), which is troubling for both teachers and administrators, since it may signal random measurement error.

Collaboration

Much of the work of teachers in schools involves collaborating in teams. Critics argue that the competition inherent in many individual teacher-based merit pay programs negatively affects this cooperation and therefore reduces performance (Darling-Hammond & Berry, 1988; Kohn, 1993; Ramirez, 2001). Many modern systems are designed to overcome this problem through offering rewards to groups working together for school goals rather than to individual teachers.

Divided Attention

Teachers' jobs are complex; raising test scores is not their only work. When pay for performance systems use test results to reward teachers, other school goals may be ignored in favor of those tasks that are directly rewarded. Firestone (1994) points out that many teachers consider the practices necessary to succeed in such systems (such as "teaching to the test") to be antithetical to good teaching. Stecher and Barron (2001) present evidence that teachers in an accountability program in Kentucky spent considerably more time on test preparation and narrowly focused on tested techniques and goals at the milepost grade levels at which they were held accountable. The implication was that broad-based institutional goals suffered at those grade levels in favor of the tests.

Aside from the complexity issue, many performance pay systems have failed under their own weight or for more general philosophical reasons. Directly relevant to the TIS context, as noted above, Ballou (2001) finds that private schools in the U.S. use performance pay much more widely than public schools. He posits that this is likely due to differences in schools and the type of teachers they attract, not in the nature of a teacher's job. Private schools enjoy more autonomy than do public schools. They are not as beholden to teacher unions and those groups' general opposition to performance pay. They are also more familiar with strategic resource use, since they operate in a more business-like environment, competing for teachers and students, than public schools. He argues, therefore, that performance pay in public schools is abandoned possibly because of the lack of proper conditions for it to thrive, not its inherent impossibility. It can work in principle if it does so in private schools.

Another important observation from the public vs. private school literature is a finding about the use of performance pay in Christian schools. Though it is true private schools use performance pay more frequently than public schools, Christian schools are notable exceptions (Ballou, 2001). These schools in particular have been resistant to performance pay because the practice contradicts egalitarian notions held as fundamental to the nature of the schools. Furthermore, many Christian schools have faced legal challenges arising from individualized pay practices. Some Christian schools, for example, might pay a teacher who is the head of household higher than a single teacher in order to allow a spouse to stay home and take care of household and child-rearing duties. When these practices were ruled discriminatory by the courts, the reaction was for schools to model themselves increasingly on the practices of public schools in order to avoid problems (Ballou & Podgursky, 1997).

Empirical Research Internationally

Given such active interest in pay reform, many researchers highlight the surprising paucity of rigorous research on the effectiveness of performance pay, particularly for individual teachers (Clotfelter, Ladd, Vigdor, & Diaz, 2004; Eberts et al., 2002; Figlio & Kenny, 2007; Ingvarson, Kleinhenz, & Wilkinson, 2007; Koppich & Rigby, 2009; Podgursky & Springer, 2007). More work has been done on group incentives and their effect on school reform, with mixed results. Figlio & Kenny (2007), in the introduction to their study discussed below, indicated, “In summary, there is no U.S. evidence of a positive correlation between individual incentive systems for teachers and student achievement” (p.902). Accordingly, literature review below includes rigorous studies from India, Israel, Africa, and the United Kingdom in addition to U.S. examples.

India

A group of 500 schools in the rural Indian state of Andhra Pradesh served as the population for an experimental study into the effects of teacher incentives on student achievement. Muralidharan and Sundararaman (2009) tracked the effect of individual and group bonuses (mean bonus 3% of pay) on student outcomes in math and language. To keep teachers from focusing only on those students near the testing threshold, the average gain of all students had to be greater than 5% to qualify for the bonus. The 500 schools were randomly divided into groups that received the group or individual bonuses (n=100 each), those that received other input interventions (an extra contract teacher or a block grant to be spent on student materials, n=100 each), and a control group (n=100).

Year two of the study revealed "students in incentive schools performed significantly better than those in comparison schools by 0.28 and 0.16 standard deviations in math and language tests respectively" (Muralidharan & Sundararaman, 2009, p. 2). Students in the incentive schools scored significantly higher on both mechanical and conceptual questions on the tests, and the schools also performed better on subjects that were not part of the incentive. The results suggested positive direct and spillover effects of the incentive program. Furthermore, the researchers reported the individual incentive schools performed better than group incentive schools in the second year.

Some problems with generalizability are clear, especially given the relative poverty and low baseline performance of the schools. Interestingly, the moderate effects visible even in a rigorous experimental study may reinforce the findings of other studies that students with relatively low levels of performance respond well to teachers motivated by incentive programs.

United States

Ladd (1999) studied the Dallas Independent School District SBPA from its inception in 1991 until 1995. Since only 20% of schools in the district could qualify for bonuses based on value added, the program used multiple regression analyses to limit extant differences in student demographics before ranking schools on a raft of outcomes, including test scores, drop-out rates, and attendance rates. The bonuses equaled $1,000 for each teacher and administrator and $500 for each janitor and secretary.

Results showed that student performance rose in participating schools compared to cities without similar programs. Reading pass rates rose 10-15% over the four years, while math pass rates rose 14-17% in the same period. Importantly, though pass rates were low prior to the initial year of the program, the tests used to measure reading and math proficiency were changed more than once during the time under study, which may have clouded results, since the scores had to be equated from old tests to new. Furthermore, pass rates say more about low-performing students than high-performing ones.

Tennessee's Career Ladder Evaluation System was a model public school merit pay program begun in 1984. The system depended on an intricate system of evaluation to place teachers on career ladder rungs, each of which was associated with an increased level of pay. Dee and Keys (2004) combined data from 79 schools participating in this program and a contemporaneous research project dubbed Project STAR (Student Teacher Achievement Ratio), a 4-year experiment instituted to investigate the role of class size in student achievement. Because students were randomly assigned to classes in the latter experiment, analyzing correlated student achievement data with participation in the

career ladder system controlled for possible connections between teachers and types of students, forming an opportunity for a natural experiment.

The researchers found that students of teachers at the lowest rungs in the career ladder system had math scores three percentile points higher than those in non-career ladder classes. Reading scores did not show a significant difference at higher rungs in the career ladder, though the difference was two percentile points. Given the random assignment of students to classes, the findings are persuasive; however, the correlations could have many causes. It is possible, for example, that students of teachers at the beginning rungs of the ladder simply benefited from younger, better-trained teachers who were required to participate in the program. Since the study only compared teachers to each other, no conclusions can be drawn about whether the incentive program affected teaching at the school level.

Using a rich data set from the U.S. National Education Longitudinal Survey (NELS) paired with contemporaneous data from the Schools Staffing Survey, Figlio and Kenny (2007) analyzed the presence of merit pay programs in 502 U.S. schools. Because the data available to them were sufficiently detailed, they were able to control for school and student characteristics through extensive regression analyses and checking findings against data sets from which areas of potential selectivity bias and endogenous variables had been removed. The results indicated that schools with more selective incentive programs, as opposed to those in which a majority of teachers received stipends, showed a greater, though moderate, impact on student achievement. Significantly, the effects were strongest in schools in which parental involvement was lower than others. The researchers indicate that these are usually lower-income schools.

Schools with teacher incentive programs showed an increase of 1.3 to 2.1 points in test scores. The researchers themselves pointed out that this was a small increase given the 33-point standard deviation in test scores; however, the relationship was statistically significant. The nature of that relationship is questionable. Given the cross-sectional nature of the data set, it is impossible to say whether the incentive program led to the increase or if better schools were using incentive programs. Without a genuine experimental investigation, demonstrating causality between incentive programs and student achievement is difficult.

Marsh (2011) conducted a three-year mixed-method study of the effects on student achievement and teacher behavior of the experimentally-designed group performance bonus system used by New York City public schools. Across 198 participating schools in phase one and 196 in phase 2, the results indicated no effect on student achievement and no effect on the school's overall progress reports for state accountability (environment, progress, performance) when compared to equivalent New York City schools that did not participate in the bonus program. The effects on teacher behaviors are detailed below in the discussion of the motivational power of incentives. It is significant to note that all of the schools in the experimental and control groups were classified as high needs schools. Ultimately, the city abandoned the program.

Fryer (2011) conducted statistical analysis of student achievement in the same set of New York schools taking part in the bonus program. Using student-level administrative data on approximately 1.1 million students in the New York system, Fryer determined the effect of involvement in the bonus program on student outcomes as measured by a raft of subject test scores and attendance and graduation rates. Regressions

controlled for previous levels of student achievement, estimating the effect of the school's participation in the bonus program on learning as measured by scores. Results indicated negative effects of program participation on all student scores in elementary and high school. In middle schools, the negative effect was statistically significant and non-trivial. In addition, students in treatment high schools were 4.4 percent less likely to graduate and 7.4 percent less likely to receive a Regent's diploma than their counterparts in non-treatment schools. Since treatment and non-treatment schools were determined through an experimental design, these results were surprising.

The United Kingdom

Atkinson et al. (2004) studied the effects of a 1999 teacher career ladder program designed to improve student scores in the United Kingdom. Using a difference-in-difference design and attempting to control for student characteristics and teacher experience through regression analysis, the team found that teachers in the scheme showed an increase in value-added of almost one half a GCSE grade per student. The gain was equal to 73% of a standard deviation. The pay scheme actually gave bonuses to 97% of the eligible teachers who applied, so the result contrasts with those of other studies (Figlio & Kenny, 2007) that indicate that selective programs are more successful than more inclusive ones. Results did, however, support the common finding that low-performing students benefit most from pay-for-performance schemes. Because the scheme was open only to teachers with eight years or more of experience, the treatment and control groups differed by nature, which may account for some of the variability.

Israel

Another group of studies to focus on incentives to raise pass rates of students in

low-performing schools comes from Israel. Lavy (2002, 2004, 2007) conducted and evaluated two studies of the effects of tournament-style programs on student achievement and teacher performance. In the first program, 62 religious and secular schools competed for public recognition and U.S.$1.4 million to be awarded to the top third of schools in rank order. They competed on a raft of measures, including student scores, higher pass rates and reduced dropout rates. The measures were carefully designed to reduce possible gaming of the system. For example, to control for schools eliminating low-performing students, schools were measured on their ability both to retain students and to raise their test scores.

Though the studies did not follow experimental designs, the design of the programs themselves controlled for many factors in order to level the tournament playing field, which in turn made it more open to statistical analysis, particularly the pre-test post-test strength of regression-discontinuity design. Lavy (2002) compared the results of the program over two years to those of another 22 Israeli schools that did not take part in the program, but did receive additional resources. Lavy found significantly higher student achievement in incentive schools in the two years between the initiation of the program and the study. Again, Lavy found students from poorer families showed the largest improvement in results. Lavy also noted spillover effects into subjects not included in the competition similar to results from the India study noted earlier.

Lavy (2004) also conducted a much-referenced study of individual teacher incentives. Similar to the above program, the Israeli Teacher Incentive Experiment established a tournament-style competition to award bonuses to 629 high school teachers in 48 schools in Israel. The awards were to be paid out to teachers whose students

showed gains in mathematics and language scores in four levels of forced ranking. Teachers' performance showed a significant increase over teachers in non-participating schools. Though only 302 teachers were awarded bonuses, Lavy reported that performance of all teachers in the treatment schools showed improvement. Overall, the treatment schools showed an 18% increase in mathematics exam-takers and a 10% increase in English exams. Furthermore, pass rates for these exams increased by 5.4% and 4.2% respectively. Lavy (2004) explained several other tests of the data conducted to eliminate possible intervening factors such as teacher selection of students, teacher experience, levels of certification, and student ability. Because of the regression-discontinuity design and the conditions placed on entry in the tournament, these factors are impressively controlled. In the end, Lavy (2004, 2007) concluded the incentive schemes were more cost effective than resource incentives, and that individual incentives were most efficient of all.

Because the individual incentive program ended after one year due to budget reductions, it is difficult to generalize about the program's long-term effectiveness and affordability. The results should also be treated with caution because of the reported teacher behaviors that seemed to contribute to student gains: namely, more test-preparation practice and additional tutoring sessions outside of normal school hours. It is unclear if the program would have sustained such measures over the long term.

Africa

Glewwe, Ilias, and Kremer (2003) studied the effects of a randomized program in Kenya that provided incentives of up to 43% of mean monthly salary to teachers in 50 of 100 rural primary schools. The schools were low performing relative to others in Kenya,

with low student pass rates and notable teacher absenteeism. Group bonuses were paid to schools in the categories of "top-scoring" and "most improved" based on average test scores on several exams given by the district. Results indicated that teachers spent more time preparing students for exams, students sat more exams, and their scores were higher when their teachers were involved in the program. This was only true in the short term, however. The researchers note that this result could be related to the fact the program was explicitly created to be temporary. Teachers may have invested more time and energy into long-term results if they believed the program would continue. Since teacher absenteeism did not decrease and student test scores retreated to the same level as non-treatment schools after the end of the program, the incentive appears to have compelled teachers to prepare students to take tests better, if nothing else.

Summary

The evidence is clearly mixed on the correlation between incentive pay systems and increased student achievement. One finding that is true to varying degrees across studies and countries is that teacher incentives have the most positive impacts on student achievement in schools with lower relative academic achievement or parental involvement and income levels. This is a significant finding when considering incentives in international schools. In general, international schools serve high-achieving students with well-to-do, involved parents. The present study, though it does not directly address student achievement, fills a gap in the research in private, international schools.

Teacher Reactions To Performance Pay

Teachers' reactions to performance pay systems are crucial for such systems to succeed. For a system to motivate teachers to improve their practice for the improvement

of student learning, teachers must accept the system. According to expectancy theory, they must see the connection between their efforts and the positive end results. Firestone (1994) cites several reasons for teachers' mistrust of performance pay, including distrust of evaluators' ability to capture performance, the danger of favoritism in evaluation, and the amount of time such programs take away from teaching. Several large-scale surveys and analyses of large federal databases (Ballou & Podgursky, 1993; Ballou, 2001; Heneman & Milanowski, 1999; Kelley et al., 2002; Milanowski, 2007) found teachers were generally motivated by the bonuses but worried about fairness, and they were not generally supportive of the programs continuing, except those already unhappy with their level of pay. Notably, private school teachers, younger teachers, and those working with low-achieving or disadvantaged students expressed a more positive view.

Milanowski (2007), in an admittedly small study of teachers-in-training at a U.S. university (n=183), found that 78% of the subjects were favorably disposed toward pay for performance in some form, including a knowledge- and skills-based system. The same percentage of teachers was generally not in favor of linking pay to student test scores. Trevor (2010) analyzed survey results from over 2,500 unionized teachers in the U.S. in 2005. He found that most teachers opposed an attachment of test scores to pay increases and in fact supported years of service and level of education as the primary determinants of pay levels.

Coggshall, Ott, Behrstock, & Lasagna (2010) contribute further evidence that younger teachers are more accepting of the notion that pay can or should be tied to performance. The researchers present results from a study involving surveys returned by a national random sample of 890 teachers from Generation Y—those born between 1977

and 1995. The results show that this generation of teachers are much more accepting of many facets of performance-based pay than their older colleagues, including its ability to improve teacher effectiveness and the overall fairness of paying teachers according to the level of their contribution to the school.

If performance pay systems are to succeed, a positive reaction from teachers is paramount. Lawler (1967), in a seminal article on performance appraisal, identified a factor entitled "Attitudes toward fairness and acceptability of the system" as playing an important role in the ability of any system to evaluate performance with any validity. Subsequent researchers attempted operationalization of this variable. Landy, Barnes, and Murphy (1978), in a study using surveys of employees (n = 950) in a large, multidivision manufacturing organization found that 26% of the variance in perceptions about the company's performance appraisal system was accounted for by five factors: (a) that the appraisal was part of a larger plan, (b) frequency of appraisal, (c) plans for improvement discussed between employee and supervisor, (d) the supervisor's knowledge of the employee's job duties, and (e) the supervisor's knowledge of the employee's job duties.

After the critique of other researchers, a follow up study showed that employee attitude in this same group was not significantly affected by the rating attained by the employee (Landy, Barnes-Farrell, & Cleveland, 1980). In other words, the attitude of the system's users was due in some part to features of the system itself and its acceptability to those being evaluated, not by whether or not employees attained high ratings through the process. As discussed above, many models for performance pay programs exist. This finding indicates the choice of a model is important to its success in implementing organizational goals. Indeed, negative attitudes toward a system may hinder rather than

help in attaining the goals of the organization (Dickinson, 1992).

Teachers must accept any system as viable if it is to motivate them to better performance and a focus on the goals of the school (Greenberg, 1987; Hedge & Teachout, 2000; Heneman & Milanowski, 1999, 2003; Keeping & Levy, 2000; Milanowski & Heneman, 2001). Teachers who have negative attitudes about a particular system may still go through the motions of the system but may not experience the possible benefits in genuinely affecting their professional practice. International school teachers may opt out of the system by seeking a job in a new school that does not use such a system. They are generally "free agents," able to move schools (and countries) with relative ease. Clearly, the rewards must appear attractive enough for teachers to persevere with a KSBP system.

Many other factors are involved in determining the positivity or negativity of a teacher's reaction to an evaluation-based system for assigning pay. Keeping and Levy (2000) proposed a set of six dimensions to an employee's performance appraisal reaction. These were validated through a survey-based study involving 208 employees of a large international organization. Because of their demonstrated construct validity, they form the basis for much subsequent survey-based research into appraisal reactions. The dimensions are:

- Satisfaction with the system
- Satisfaction with performance appraisal sessions
- Perceived utility of the system
- Perceived accuracy of the system
- Procedural justice
- Distributive justice

Researchers investigating teacher evaluation systems designed to be attached to pay have determined a wider set of factors that influence teacher reaction (Greenberg,

1987; Hedge & Teachout, 2000; Heneman & Milanowski, 1999, 2003; Keeping & Levy, 2000; Milanowski & Heneman, 2001):

- Understanding of the system
- Acceptance of evaluation standards
- Impact on teaching and work
- Trust of and role of the evaluator
- Fair evaluation process (procedural justice)
- Fair evaluation results (distributive justice)
- Accurate evaluation results
- Stress, pressure, frustration, and fatigue
- Effort expended versus reward gained

Much overlap is evident in these dimensions. In general, when teacher perceptions around all of the factors except stress have a positive value, the teacher regards the system as a whole positively. High levels of stress can negate the positive perceptions around the other factors (Heneman & Milanowski, 1999). Likewise, when stress is high and the other factors are even slightly negative, an overall negative reaction can result. Table 1 illustrates the way these dimensions can be seen to align with the elements of expectancy theory.

Heneman and Milanowski conducted two studies of a system quite similar to that instituted at TIS in the Cincinnati public schools (Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). Through a combination of teacher response surveys, semi-structured interviews, and exit surveys of those leaving the district, the researchers sought to identify, among other observables, the reactions of teachers to the system. The survey items and interview questions were tied to the dimensions identified above from previous research and based in expectancy theory.

Table 1

Suggested Alignment of Expectancy Theory Perceptions and Appraisal Reaction Dimensions

Expectancy Theory Perception	Appraisal Reaction Dimension
Expectancy	Understanding of the system Acceptance of evaluation standards Trust of and role of the evaluator
Instrumentality	Fair evaluation process (procedural justice) Fair evaluation results (distributive justice) Accurate evaluation results
Valence	Impact on teaching and work Stress, pressure, frustration, and fatigue Effort expended versus reward gained

Results of factor analysis from the teacher response surveys indicated that, in the initial year of the program, teachers were somewhat favorable toward it; they "neither fully embraced the system nor rejected it out of hand" (Milanowski & Heneman, 2001, p. 206). Results from the interviews in the first study showed that teachers were more favorable than they were in the survey items. In the subsequent study, after changes were made to the system to focus more on teacher observation and increased feedback from supervisors, favorability to the system dropped overall, though teachers expressed appreciation for some of the changes. Specifically, teachers indicated that the system "created stressful and burdensome experiences" (Heneman & Milanowski, 2003, p. 184). Teachers believed the system created more stress and took more effort than the results were worth. In two factors (procedural justice and accuracy), survey scores decreased significantly ($p < 0.05$) from one year to the next.

Overall, the two studies showed both positive and negative reactions to the evaluation system. In the positive camp, teachers found the four domains for evaluation

to be relevant to their jobs and consistent with what they saw as good professional practice. Teachers therefore understood and accepted the competency model embodied in the evaluation standards. The adverse reactions to the system arose from dissatisfaction with the implementation of the evaluation and pay systems tied to the standards. In fact, many teachers pointed to examples of how the observations and feedback based on these standards had changed their classroom instructional practice. It seems, therefore, that the standards themselves were accepted and understood by teachers generally, but that teachers objected to having their pay tied to these standards.

Milanowski and Heneman considered the negative reactions to the system in Cincinnati in great detail. They concluded that, in part, the new system was too much of a shock for teachers who were used to classroom autonomy and a long-standing single salary scale. Furthermore, teachers felt a lack of formative (as opposed to summative) feedback in the evaluation system.

Kelley, Heneman, and Milanowski (2002) summarized two research studies into the motivational effects of SBPA programs in Kentucky and Charlotte-Mecklenburg schools. The researchers developed surveys after open-ended interviews revealed trends in teacher response related to their selected motivational model of expectancy theory. Surveys were sent to thousands of teachers in hundreds of schools using SBPA models. 1,150 teachers responded in Charlotte-Mecklenburg (39% return rate) and 1,750 in Kentucky (30.9% return rate). Results showed that teachers had both positive and negative views of these school-based systems. Importantly, they found that teachers' motivation was affected by their expectation that the goals of the program were equal to positive goals for students as a whole. Furthermore, teachers needed to believe that their

increased effort would actually result in receiving the bonus offered. Finally, teachers' perceived fairness of the system for determining whether or not the bonus was awarded was another factor in how favorably they viewed the program.

Ha's (2003) doctoral dissertation study of teacher reactions to a bonus pay system is notable in its exploration of the reaction of teachers in Korea to a government-sponsored teacher bonus program. Using survey items gathered from Milanowski and Heneman (2001) and Heneman and Milanowski (2003) and group interview protocols, Ha found that the teachers had variable levels of favorableness toward the system. Furthermore, hierarchical multiple regression tests indicated that the nine dimensions previously identified in the research of Heneman and Milanowski (discussed above) were strong predictors of teacher reactions. The regression model with only control variables (years of experience, education, age, credential or teacher rank, gender) showed these variables accounted for a "somewhat significant amount of variance in overall favorableness (R^2=.16)" (Ha, 2003, p. 114). When the nine dimensions were added in the second model, a significant proportion of variance was explained (R^2=.74). In particular, fair process, the role of the evaluator, fair results, and accuracy were significantly predictive of overall favorableness.

Teacher Turnover

If pay for performance systems have the potential to motivate teachers, they conversely have the ability to de-motivate them, perhaps to the extent teachers choose to leave the school. Little evidence exists on teacher turnover relative to pay for performance. One such investigation was undertaken in a study of the Governor's Educator Excellence Grants (GEEG) program in Texas (Taylor & Springer, 2009). The

program, created according to an experimental design model, distributed grants to 99 high-performing schools serving low-income students over a three-year period. Each school created its own incentive program under guidelines established by the office of the Texas Education Agency. One guideline was the requirement that teachers be actively engaged in the design of the programs. Thus the program overall gave an excellent opportunity to observe the effects of programs designed with teacher preferences in mind. With regard to teacher turnover, the research results indicated no significant difference in teacher turnover in those schools that used incentive programs when compared to predictions based on Texas Education Agency data and data from the National Center for Education Statistics and the U.S. Bureau of Labor Statistics.

More recently, Marsh, et. al (2011) conducted an extensive study of the school-based bonus system implemented in New York City. Among other research questions, they investigated whether the bonus system led teachers to be more involved in professional development and express a desire to remain in their current schools. Taking advantage of the experimental design of the compensation program itself, the researchers used surveys to determine teachers' reactions to these dimensions. Teachers were surveyed in schools that used the SPBA and control schools that did not. The overall teacher response rate was 57 percent, with 1,532 surveys returned. Survey items asked teachers to comment on the amount of time they spent on work outside of working hours and the amount of time teachers devoted to professional development activities they found themselves—i.e. not required by the school. Similarly, survey items asked teachers about their plans to renew their employment in their current school or to look for another school. Results indicated no evidence that receiving a bonus motivated teachers to seek

more professional development or to remain in the school in which they received that bonus. This three-year study, conducted under the auspices of the RAND Corporation, contributed to the city's abandonment of the program.

Fryer (2011), as part of a study discussed above into those same New York City schools, also explored teacher retention in school and in district. His study used Human Resources data supplied by the Department of Education to track actual teacher movement and to control for normal teacher mobility. Though the results were not statistically significant, Fryer reported "no evidence that teacher incentives affect retention in either district or school" (2011, p. 21).

Summary

Research into teacher reactions to performance pay has shown variable results. Teachers involved in these systems may view them more or less favorably depending on several different factors, including how fairly they are administered, how well they are understood, and how much stress and effort are required to comply with the system's requirements. In at least one private school study, teachers were less favorably disposed toward the idea of pay for performance if they were working in a school that used one rather than one that did not (Ballou, 2001). Younger teachers appear to be more favorably disposed toward performance pay than their older colleagues (Coggshall et al., 2010). In the few studies that have addressed the issue, taking part in a performance pay system has not been shown to have any significant effect on teachers' desire to engage in further professional activity or to remain in the school that paid them the bonus (Marsh et al., 2011; Taylor & Springer, 2009).

Conclusion

The growing interest in performance pay for teachers is not yet matched by research into its effectiveness. Even the most rigorous experimental studies show only moderate positive results with specific populations, the causal nature of which is impossible to determine. Programs with wide eligibility but a limited number of rewards show the most promise. Contrary to conventional wisdom, teachers do appear to respond to incentives through changing their methods, putting more effort into teaching, and seeking creative ways to increase student achievement in targeted areas. Those targeted areas must therefore be carefully identified, defined, and rewarded according to the true overall institutional goals, not simply in terms of test scores. Teachers are not immune to gaming any system created, so the design must control for this possibility to avoid the unintended consequences observed in many of these studies. The research also shows that the incentives are most effective in raising achievement of students at lower socio-economic levels and lower levels of academic performance, which may be a limit to broader applicability. Finally, since many programs last only a short time, sustainability must be planned for and tested in further research. While paying teachers according to performance shows some promise in improving certain types of student achievement, planners must carefully consider design and the time and energy necessary to reach desired results.

Teachers' reactions to alternate pay systems are a critical element in their implementation. If new systems do not have the desired effect, the extra expense and person hours required may not be worth the effort. Understanding the nature of teachers'

reactions to alternate pay systems and the factors contributing to these reactions is therefore of great interest to schools and administrators.

Researchers have identified several key factors that explain variation in teachers' reactions to a pay system, all of which are subject to investigation through perception surveys. One avenue of investigation is through instruments created in the context of expectancy theory, a theory of motivation with a large research base in various workplaces. The present study used items and methods informed by these studies and grounded in expectancy theory to probe the relevant research questions.

CHAPTER 3

Method

The present study investigated teacher perceptions of the motivational effect on highly qualified teachers of a knowledge- and skills-based performance pay system, the elements of the system that contribute to their reactions, whether favorable or unfavorable, and the ability of a pay for performance system to retain highly mobile teachers. For such systems to achieve their aims of improved student learning, teachers must feel motivated by and positively disposed toward the system. Previous research into the application of employee appraisal models to the education context was taken into account in the design of the present study. The program at TIS began with great controversy and is still in its infancy, so how well it will serve its purpose and how well it will be received remain open questions. By extension, research into these questions can give valuable insight to school leaders about the efficacy of such systems.

The administrators who designed the TISKS model considered expectancy theory and tried to incorporate its motivational tenets into the system. Teacher perception of expectancy, instrumentality, and valence must be high enough to warrant the effort necessary in an evaluation system if it is to succeed. This study assessed teacher reactions to the system according to its affective goals of providing motivation for continued professional development and motivation of teachers to remain at the school. Furthermore, the overall reaction to the program was evaluated and described using a set of previously identified dimensions shown statistically to correlate with the favorableness of teachers' reactions.

The following research questions guided the study:

RQ1 To what extent do teachers perceive the use of a performance pay system at their school motivates them to improve their practice?
RQ2 What factors influence teacher reactions toward the performance pay system?
RQ3 Is there a correlation between participation in the pay for performance system and teachers' decisions to renew their teaching contract with the school?

This mixed-method study utilized a concurrent triangulation design to gain the fullest picture possible of teacher reaction to the new appraisal and pay system. Concurrent triangulation design is so named because it facilitates the triangulation of data from qualitative and quantitative methods gathered in roughly the same window of time (Creswell, 2003). Though data collection may not be simultaneous, the quantitative and qualitative processes address the research questions separately and results are treated with equal weight and integrated at the interpretation stage. Concurrent triangulation design allows the researcher to bring together the differing strengths of qualitative and quantitative approaches: specifically the ability to garner information efficiently through quantitative analysis of a larger sample and the ability to get in-depth detail from a smaller sub-sample through qualitative analysis (Patton, 2002). The design is "useful for attempting to confirm, cross-validate, and corroborate study findings" (Hanson, Creswell, Clark, Petska, & Creswell, 2005, p. 229). The quantitative and qualitative portions of the study will be addressed separately in this discussion.

Quantitative

Participants

The quantitative portion of the study employed an anonymous survey sent to all teachers at The International School (*N*=246). All members of the population were invited to take part in the survey; however, the survey included additional items for those members of the population who had completed some form of TISKS. For the purposes of

the present study, in keeping with the research questions, all teachers who had taken part in some form of TISKS were considered one group termed TISKS Teachers (*n*=119), and those who had not were considered another group termed Non-TISKS Teachers (*n*=127). Both the Non-TISKS and TISKS Teachers were asked to respond to items related to RQ1 and RQ3. Only TISKS Teachers were asked to respond to items related to RQ2. The timing of data gathering was deliberate and advantageous. The first pilot cohort of teachers to go through the system was handpicked in order to ensure success and a positive experience while troubleshooting the process. This could skew teacher reaction results, as these teachers were presumably more experienced and confident from the outset. Though it can be assumed the more eager teachers would take part earlier, gathering data at a time when roughly half the faculty had completed the profile year and the other half had not was intended to mitigate against this possible impact.

Instrument

The survey solicited demographic information for use as control variables in teachers' reactions to the system: gender, age, years of experience, and education attained. There were also three items for all teachers about their self-reported sense of the extent to which the pay system being implemented in the school motivated them, and whether or not they intended to renew their current contracts at TIS when they expired.

The control variable items: gender, age, level of experience, and level of education attained, have been used in similar studies to understand the involvement of salient teacher characteristics in reactions to appraisal evaluations (Ha, 2003; Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). The demographic information requested was limited to those variables that have been controlled for in previous

research to limit the amount of information collected that may lead a participant to feel he or she could be identified. For example, the survey did not ask teachers in which division of the school they taught, nor whether or not they had a specialist role in the school. Though there are different administrators implementing the TISKS system in each division, they have all been engaged together in many hours of training and calibration experiences across divisions. This training is itself a way to control for variability across divisions, making the solicitation of this information unnecessary.

For all teachers, further items solicited perceptions about the extent to which TISKS exerted an influence on their motivation to improve their practice, whether or not they had taken part in it yet. The item solicited responses on a five-point Likert scale: 1=*Strongly Disagree*, 2= *Disagree*, 3= *Neither Agree nor Disagree*, 4= *Agree*, 5= *Strongly Agree*. The item on intention to remain at TIS after the current contract used a 3-point scale: 1= *No*, 2= *Unsure*, 3= *Yes*.

The survey for TISKS teachers incorporated the Teacher Reaction Scales (TRS), which were developed and described by Heneman and Milanowski (2003). The items for the scales were drawn from a bank of over 200 items previously psychometrically validated to align with the dimensions of appraisal reactions in the work of Keeping and Levy (2000). Specifically in the context of teacher and school appraisal interactions, the scales as used in the survey for the present study have shown construct validity across studies in the U.S. and Korea (Ha, 2003; Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). Heneman and Milanowski (2003) reported satisfactory confirmatory factor loadings (0.40 or greater) on their a priori scale. One cautionary note is that some strong cross-loadings were visible for items on the accuracy and distributive justice scales

due perhaps to similar wording in the items. The items were kept as separate items in the TRS, however, due to their "potential distinctiveness" (Heneman & Milanowski, 2003, p. 183).

The TRS items also showed construct validity in the dissertation work of Ha (2003). In his study of teachers' reactions to a government bonus system in South Korea, Ha used the TRS and conducted an exploratory factor analysis with orthogonal (varimax) rotation to determine if the scales would converge in the same dimensions as in previous studies. The scales did perform as expected. The analysis is relevant to this study not only because the scales proved again to converge on the same dimensions as in previous studies, but also because they did so when used in South Korea, a national education system outside of the United States, where they had previously been validated. It is therefore reasonable to propose they are likely to be valid for an international school environment as well.

For the purposes of this study, the items were adapted to the TIS context. For example, the title "TISKS" was substituted for "the evaluation system" in order to ensure clarity for the survey respondents at TIS; however, the nature of the items was unchanged. Likewise, the term "evaluator" was replaced with the term "panel members" to reflect practice in the TIS system. (See Appendix B for the survey instrument, including the TRS).

The scales were developed and validated by previous researchers (Ha, 2003; Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). Exploratory factor analysis was also conducted in this study to determine its appropriate use with this population of teachers, followed by reliability tests with Cronbach's Alpha in order to

determine the internal consistency of the scales for the population. The reliability coefficients explain the extent to which the items grouped for each dimension represented similar constructs.

The surveys elicited perceptions tied to the dimensions previously identified to have an impact on teachers' reactions to such systems (see the survey in Appendix B for explanation of which items correspond to which dimensions):

- Understanding of the system
- Acceptance of evaluation standards
- Impact on teaching and work
- Trust of and role of the evaluator
- Fair evaluation process (procedural justice)
- Fair evaluation results (distributive justice)
- Accurate evaluation results
- Stress, pressure, frustration, and fatigue
- Effort expended versus reward gained

All items solicited teacher reactions on a 5-point Likert scale: 1=*Strongly Disagree*, 2= *Disagree*, 3= *Neither Agree nor Disagree*, 4= *Agree*, 5= *Strongly Agree*. Teachers were also invited to supply open-ended answers to questions about the identified dimensions. Data collected through these open-ended items were added to the qualitative data for analysis as discussed below.

Because the surveys were designed to gather information from TISKS Teachers and Non-TISKS Teachers, two different forms were used. The form to which a teacher responded depended on the answer to the question: "Have you completed TISKS?" If the answer was no, the survey ended. If the answer was yes, the teacher was then presented with the TRS after indicating in which year they completed the process. Figure 2 illustrates the flow of the survey instrument.

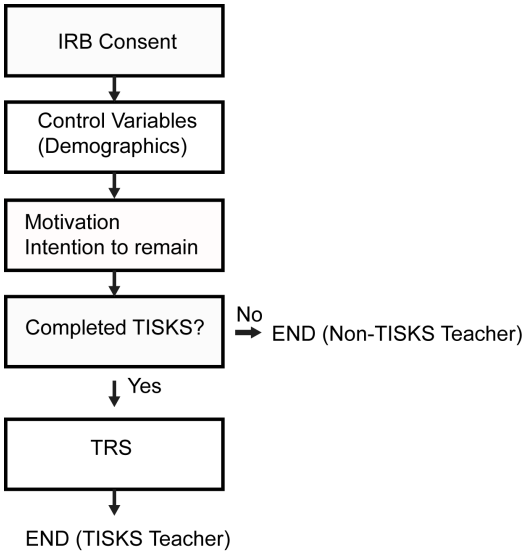

Figure 2. Subject discrimination flow chart for teacher survey.

Data Collection

The survey was constructed and administered through Zoomerang, an online survey tool that allows for anonymous data collection and export of data in a format usable for statistical analysis. The initial online screen of the survey explained the research goals, guaranteed confidentiality, and served as the opt-in/opt-out opportunity for respondents. Only those who opted in were able to complete the survey.

Of the 246 teachers eligible, 182 completed the survey, for a response rate of 74%. Of these respondents, 60% were female, 40% were male. Because teacher age, years of experience, and degrees attained have been correlated with reactions to performance pay in the past (Ballou & Podgursky, 1993; Coggshall et al., 2010; Goldhaber et al., 2010; Ha, 2003; Milanowski & Heneman, 2001; Milanowski, 2007), the survey included items asking teachers to identify themselves in bands relative to these data points. 12 teachers (6.6%) identified themselves as between the ages of 21 and 30, 50 (27.5%) between 31 and 40, 65 (35.7%) between 41 and 50, 40 (22.0%) between 51 and 60, and 15 (8.2%) 61 or over. As to experience, 15 (8.2%) reported one to five years

of teaching or work in schools, 31 (17.0%) from six to ten years, 42 (23.1%) from 11 to 15 years, 39 (21.4%) from 16 to 20, and 55 (30.2%) 21 or more years. Finally, 15 teachers (8.2%) reported a bachelor's degree as their highest degree, 29 (15.9%) some graduate courses but no master's, 91 (50%) a master's, 41 (22.5%) more than a master's but no doctorate, and 6 (3.3%) reported doctorate degrees. Of the 182 total responses, 98 reported having completed TISKS, while 84 had not.

Analysis

Full descriptive statistics are presented for each item and scale on returned surveys, including the mean and the standard deviation. The means indicate average teacher reaction to the scales, and the standard deviation illustrates the general spread of the distribution.

RQ1: Motivation. The first research question related to the motivational effect of the system was explored through a single item on the survey. Item 9 read: "TISKS is a motivator for me to improve my professional practice (whether or not I have completed the process)." The means and standard errors for the Likert scale response on this item were used to test the hypothesis with regard to differences between the two groups in the sample: TISKS Teachers and Non-TISKS Teachers.

Since previous research has indicated age and experience may also have an impact on teachers' self-perception of the motivational impact of pay for performance (Coggshall et al., 2010; Goldhaber et al., 2010; Milanowski, 2007), hierarchical multiple regression was used to test for the effect of these control variables on any relationship that may exist between TISKS experience and motivation. One regression model was run with the self-perception of motivational impact as the dependent variable and age and

experience as a block of predictor variables. The result illustrates to what extent these two demographic variables explain variation in self-reported teacher motivation. A second block added membership in the TISKS Teacher group or Non-TISKS Teacher group to the first model to determine the extent to which having taken part in some form of TISKS may explain variability in self-reported motivation when age and experience are controlled.

RQ2: Reaction and influencing factors. Exploratory factor analysis with a promax rotation was conducted with the TRS to determine if items clustered as in previous research with this population. The scales were also summed into dimensions included in the identified factors.

The first step was to assess the intercorrelations among all elements: the summed scales of the dimensions of the TRS and the control variables (gender, age, years of experience, and education). Intercorrelations are necessary to check for multicollinearity among the dimensions. Limited collinearity supports the notion that each of the dimensions measures a different construct.

The impact of the identified factors and dimensions on variance in overall favorableness was then explored through hierarchical linear multiple regression. The first block included only the control variables to determine the extent to which they accounted for variability in the scale for overall favorableness. The second block entered included each of the identified factors from the TSR dimensions as predictor variables to see if they improved on the prediction of variability in overall fairness.

RQ3: Intention to return. A single item on the survey requested information about a teacher's intention to remain at TIS after the current contract. Item 11 read: "Do

you plan to return to TIS when your current contract ends? 1=*No*, 2=*Unsure*, 3=*Yes*." Since TISKS was designed partly to attract and retain teachers, it is possible to determine whether or not teachers are more likely to remain at the school after completing the program. A Chi-square contingency test was used to discover whether there was any significant difference in intention to return to TIS between those who had completed the process and those who had not. All 182 respondents answered this question.

Qualitative

Participants

The qualitative portion of the study involved two focus group interviews of four and five participants respectively with a stratified sample (n=9) of teachers drawn from the sub-population of only TISKS Teachers (n=119). The sample was designed to contain three teachers who had taken part in TISKS from each of the four school divisions: Lower Primary (grades Pre-K through 2), Upper Primary (Grades 3 through 5), Middle School (Grades 6 through 8) and High School (Grades 9 through 12).

Because the Lower Primary and Upper Primary are on a separate campus fifteen minutes' drive from the Middle School and High School, two focus group interviews were held: one on each campus. One focus group consisted of one teacher from Lower Primary and three from Upper Primary. One Lower Primary teacher who volunteered did not show up for the interview, and another had not completed TISKS and was therefore eliminated. The other focus group consisted of two teachers from Middle School and three from High School. Middle School had a teacher event on the day of the interview, so only two attended.

Stratification through selection of an equal number of teachers from each division was meant to include voices that would help account for any variability in the implementation of the system on the part of the administrators in each division. Though administrators across all divisions had taken part in many hours of calibration and training to ensure consistency, stratification of the respondent pool was used to further control for any variability. The sample was therefore purposeful. Patton (2002) explains that focus groups are an effective method of gathering high-quality data efficiently when the groups are of subjects of similar backgrounds, and optimally of six to ten participants.

The researcher was an administrator in the school at the time of the study and had been involved in the TISKS panels of several teachers; therefore, it was important his involvement in qualitative data collection be managed with complete control to protect participants' identities and to assure them they could share their genuine feelings about the system without having their names attached to data. All TISKS teachers were invited to volunteer to take part in a focus group interview through an email from the third party facilitator not employed by the school and not a parent of students in the school, nor a spouse of anyone employed by the school. The email explained the general nature of the research study and clearly identified the researcher and his role in the school. The third party facilitator then chose teachers to take part on a firs-come first-serve basis from those who responded and coordinated meeting times and locations. Once groups were established, rooms were booked and recording equipment was supplied for the focus group sessions. The researcher was not on campus or was in a completely different part of campus when the interviews took place.

Focus Group Interview Process and Data Collection

A focus group facilitator with a research background and NIH training was engaged to facilitate and record the sessions to protect confidentiality and to minimize participant reactivity (Maxwell, 2005). Because the interviewees knew the researcher, they may have been tempted to anticipate and provide the answers they believed were being sought or to be guarded in their responses because of perceived threat. Such reactivity needed to be controlled through assurances of the confidentiality of responses. The third party intermediary helped to control for the fact that participants knew the identity of the researcher. She digitally recorded the interview discussions (audio only) and kept subjects focused on the topics under discussion.

Each focus group interview was scheduled to last from 45 minutes to one hour. The interview questions were important as "the focus group interview is first and foremost an interview" (Patton, 2002, p. 385). The questions were designed to elicit responses focused on the phenomenon under study, not solve a problem or build consensus. Although the focus group interview draws power from the social interactivity of the participants, Patton (2002) points out that the true power of the method is that it is focused. In this case, the focus was on three questions related to the research questions of the study and the theoretical framework. The questions were:

1. What is your feeling about how much TISKS motivates teachers at TIS to improve in their teaching and work? Is the system worth the effort?
2. To what extent do you feel teachers have the information and support they need to be successful in TISKS?
3. Overall, how fair do you think the TISKS process is?

The questions were designed to be open-ended, encouraging discussion and interaction rather than simple answers. The facilitator allowed and encouraged a maximum of up to

20 minutes of discussion on each question. This was designed to give every member a chance to contribute and to elaborate on or connect to the responses of others. (See Appendix C for the full focus group interview protocol).

The focus group interview is more than simply a discussion, so the facilitator was trained in how to keep the group on topic, how to control those who would dominate, and how to establish a permissive environment that encouraged response. One of the strengths of the focus group method for gathering data is the power of social interaction. Especially in discussing sensitive topics, participants are more likely to share their own thoughts in an interactive environment in which they hear others sharing similar thoughts. Relating to the experiences and feelings of others can lead not only to more data because respondents build on the contributions of others, but also to a more comfortable environment for response than an individual interview.

The facilitator solicited informed consent to take part in the focus group and to have the session recorded before subjects entered the interview room. She began each session with a brief explanation of the goals of the research and assurances of the confidentiality of responses. At the end of the session, the facilitator thanked the participants and closed the session.

Digital audio recordings of the two sessions were sent directly via electronic file transfer to a third-party transcriber in North America who was neither the researcher nor the focus group facilitator and who had no relationship with the school. The transcriber transcribed recordings verbatim, under strict instruction to remove all names and names of schools or divisions that may have been mentioned in the course of the discussion. The anonymous transcriptions were then shared only as text files with the researcher. These

were combined with the open-ended text answers to the survey to provide more detailed insight into teachers' thinking than would have been possible solely through survey responses.

Analysis

Analytic induction (Patton, 2002) was used for content analysis of interview data and open-ended response items from the survey. Analytic induction is a two-stage content analysis process in which the researcher first analyzes data deductively, applying a theoretical framework to identify themes and patterns in the data. In the second stage, the researcher “strives to look at the data afresh for undiscovered patterns and emergent understandings" (Patton, 2002, p. 454). In this study, coding was conducted by hand of all text from the interview transcripts and the open-ended responses.

The first step was to sort and code all of the text responses deductively into the broad categories from the original questions on both the survey and the theoretical framework: expectancy, instrumentality, and valence. All text was printed in landscape-oriented spreadsheets with columns for the text and blank columns for coding. Utterances were identified and classified in the first column in each of the three expectancy perceptions and further coded with a plus or minus symbol to indicate a positive or negative reaction in each perception. Any responses that did not fit these categories were given a unique name in a third column on the page and put aside for subsequent analysis. A second pass through the data was used to sort and code the responses in the same way into the more specific themes of the nine dimensions investigated in the survey data:

- Understanding of the system
- Acceptance of evaluation standards
- Impact on teaching and work
- Trust of and role of the evaluator in sessions

- Fair evaluation process (procedural justice)
- Fair evaluation results (distributive justice)
- Accurate evaluation results
- Stress, pressure, frustration, and fatigue
- Effort expended versus reward gained

Again, each of these was marked in the second column as a positive or negative response based on the context and substance of the text. Any responses that did not fit the theoretical framework were marked, named, and entered into the third column for a final coding pass through the data. In the final inductive pass, themes were marked that may have emerged from the data previously but were not directly accounted for in the theoretical framework. Meaningful names were assigned to the themes and they were coded in the third column in the same way as the other two passes. An example page from the coding notebook is included in Appendix D for clarity. These themes were then checked for connections to one of the nine dimensions, one of the three broad categories in expectancy theory, or if they appeared to be themes independent of the framework. Discussion of newly emergent themes is included in the results section.

A final step in the analysis involved a review of data to increase credibility. The concept of credibility in qualitative methods is roughly analogous to that of validity in quantitative methods. While several means are available to carry out credibility review, analyst triangulation (Patton, 2002) is an effective strategy that can increase credibility. Analyst triangulation involves having "two or more persons analyze the same qualitative data and compare their findings" (Patton, 2002, p. 560). For this purpose, the assistance of a committee member with international school experience was solicited to analyze the collected data and compare coding of the data into established and new categories. Such triangulation was intended to further reduce researcher bias and selective perception and

ensure coding was reflective of objective analysis. Feedback matched initial coding, with a few examples of ambiguous statements that could only be interpreted by someone with first-hand knowledge of the school and the TISKS protocols. Revisions to coding were made to ensure the most representative themes were used in analysis. After analysis, a further triangulation check was conducted to ensure conclusions were founded in the data.

Triangulation

The open-ended qualitative responses to questions about TISKS and the themes that arose deductively through the lens of the theoretical framework and inductively through further content analysis were compared and contrasted with the quantitative findings to gain fuller insight into the research questions. In keeping with the tenets of concurrent triangulation design, the qualitative data were summarized and matched with the TRS dimensions in the discussion of results, bolstering quantitative findings in a necessarily small sample. New themes emerging from the inductive analysis are noted and discussed through the lens of all three research questions as results are discussed.

Underlying Assumptions

Though an expressed goal of the TISKS system is the ultimate improvement of student achievement, this study did not intend to investigate that goal. The study was solely concerned with the motivational power of the system, teachers' reactions to the system, and the power of the system to retain teachers in a fluid market based on one of the first populations of international school teachers to engage in such a system. The research was based on the conceptual framework of expectancy theory, which has a large research base.

An assumption must be made that subjects answered the questions on the surveys and in the interviews honestly. Assurances were in place that all answers were anonymous and would have no impact on job security or salaries, but the possibility of teachers' perceptions of impact must be acknowledged. As in any survey and interview research, the strength of the findings must rest in part on an assumption of accurate and honest responses from the teachers.

Strengths and Limitations of Study

Strengths

The mixed method design of this study provides an opportunity for rich, field-based research in a potentially growing field. Combining a theoretical framework and survey instruments validated through past research with current interview data could lead to previously unidentified dimensions in knowledge- and skills-based pay systems. Since TIS enjoys a strong reputation among international schools, it can be deduced the teachers are representative of highly qualified, mobile teachers who have been purposeful in their decision to work at a prestigious school. Their reactions are therefore likely to be representative of other highly qualified teachers, more so than at a smaller, more isolated school. Generalizability of the results to other schools with highly qualified teachers is therefore enhanced.

Limitations

TIS is a large school, yet the pool of teachers from which to pull the sample was small compared to teachers across many schools. To maximize participation, the opportunity for the entire population to respond to the survey was necessary, though this entailed inherent problems and limited the statistical tools available to identify

correlations. The quality of the analysis was therefore enhanced by the collection of qualitative data to bolster results. However, the unique features of TIS made it a strong research site. A population of highly effective teachers who are more mobile than many teachers casts the system in an important light. It must be all the more effective in motivating and retaining teachers in this market, so it is an ideal case to study, much like an extreme case in qualitative sampling. The careful investigation of details specific to the school's implementation of its knowledge- and skills-based pay system can be instructive for other schools considering such systems.

CHAPTER 4

Results

This chapter presents the results of the investigation through the three research questions, combining findings from the quantitative and qualitative data gathering and analysis.

RQ1: Motivation

The first research question was: " To what extent do teachers perceive the use of a performance pay system at their school motivates them to improve their practice?" All teachers completing the survey were asked to respond on the 5-point Likert scale to the following item: "TISKS is a motivator for me to improve my professional practice (whether or not I have completed the process)." Results are shown in Table 2.

Table 2

Frequency of Teachers' self-reported motivation as a result of TISKS

Item	*N*	*M*	*SD*	1	2	3	4	5
TISKS is a motivator for me to improve my professional practice (whether or not I have completed the process)	182	2.81	1.26	33 (18.1)	51 (28.0)	29 (15.9)	55 (30.2)	14 (7.7)

Note. ( ) indicates the percentage of teachers

The mean on this item was 2.81, which is very close to neutral, though tending toward negative. 46.1% of teachers disagreed with the statement, while 37.9% agreed to some extent. 15.9% reported a neutral response. The mean score in this essentially bi-modal distribution was determined mostly by the fact that more than twice as many teachers indicated they strongly disagreed with this statement than indicated that they strongly agreed. Examination of the open-ended question attached to this portion of the

survey revealed several patterns of note. While not every respondent entered an open-ended response, 60.4% did. Teachers entered varying quantities of text in these open-ended boxes, and some addressed several topics in one box. Responses were ordered in a spreadsheet relative to overall negative response (1 or 2), neutral response (3) or positive response (4 or 5). Each block of text was then analyzed for major themes and observations. Exploration of these responses revealed a few often-repeated observations tied to motivational impact and several that were mentioned only one time. The observations and the number of times they are mentioned are detailed in Table 3.

Table 3

Open-ended Survey Responses Relative to TISKS as Professional Motivator

Response	Negative Observations (# of mentions)	Positive Observations (# of mentions)
1 or 2 (Negative)	Intrinsic motivation preferred (19) Tie to pay is negative (17) Lack of focus on students (6) Lack of useful feedback (6) De-motivating influence (4) Insulting to profession (2) Disagree with the rubric (2) Program too evaluative to motivate (2) Forced participation (1) Too much stress (1) Feeling criticized (1) Subjective/unfair process (1) No new knowledge through program (1) Lack of follow-through (1) Too much gaming of system (1) No lasting motivational effect (1)	
3 (Neutral)	Teachers put on a show for evaluators (2) Tie to pay is negative (1) Feeling defeated (1) Questioning true growth (1)	Forced accountability (1) Increased reflection (1) Increased knowledge/skill focus (1) Tie to pay is positive (1) Presence of others in class (1)
4 or 5 (Positive)	Motivation is extrinsic/false (1) Too much time away from students (1) Lack of useful feedback (1) No lasting effect (1)	Tie to pay is positive (10) Increased reflection (4) Positive professional conversation (3) Structured rubric for reflection (3) Professional recognition (2) Increased observation (1)

Observations most often expressed among those who disagreed with the statement included a preference for intrinsic motivation rather than extrinsic, a disagreement with or negative consequences from a tie to the pay scale, TISKS's lack of focus on student learning or time on task for class preparation, and a lack of feedback through the program. In addition, teachers mentioned the demotivating effect of being evaluated, disagreement with the structure of the rubric used to evaluate teachers, and the lack of choice in participation. Teachers agreeing with the statement cited a positive association with higher pay, the encouragement of the program to reflect on one's practice, positive professional conversations, and the disciplined inquiry enhanced by having areas identified on the evaluation rubric.

In addition to the text entered into the open-ended response boxes on the survey, a question from the focus group interviews concerned motivation. Teachers in the focus groups had all completed the TISKS process and received their final scores and the pay increases that came along with them. Similar themes emerged from the focus group conversation. Participants did not necessarily see TISKS as a motivator, though they did express motivation from a desire to be recognized as a good teacher and to be better compensated. With regard to professional improvement, teachers felt the structure provided by the rubric was a good space within which to reflect on their teaching, but they saw themselves as primarily self-motivated as opposed to being motivated by the program.

TISKS and Non-TISKS Teachers

The overall neutral value of the response from teachers leads naturally to the question of the potential difference in response to the question of motivation from those

who have gone through the TISKS process and those who have not. In order to make such a comparison, intercorrelations were first computed among the variables shown to affect teacher motivation and performance appraisal leading to pay changes in previous research, namely age (Coggshall et al., 2010; Milanowski, 2007) and experience (Ballou & Podgursky, 1993; Goldhaber et al., 2010; Ha, 2003). A binary item on the TIS survey asked if teachers had completed TISKS or not. 84 teachers had not taken part in TISKS, whereas 98 had and had received their final scores from the process. The intercorrelations of these variables are shown in Table 4. Age and years of experience both showed significant negative correlation with one's indication of TISKS as a motivator to improve professional practice. The two were also understandably strongly correlated with one another: the older a teacher is, the more likely it is that they would have more years of experience. Participation in TISKS did not have a significant correlation with the motivation score. The older and more experienced teachers were, the less likely they were to see TISKS as a motivator for improved professional practice. Participation in TISKS appeared to make no difference.

Table 4

Means, Standard Deviations, and Intercorrelations Among Variables

Variable	*Mean*	*SD*	*1*	*2*	*3*
1 Age	2.98	1.05	—		
2 Years Experience	3.48	1.30	.71**	—	
3 TISKS (0=No)	.54	.50	-.01	-.04	—
4 Motivator	2.81	1.26	-0.35**	-.26**	.03

**$p < .01$. $N = 182$.

A hierarchical multiple regression was then conducted. To control for the variables, which may covary with teacher responses, two control variables were used: age (in bands 1 to 5) and years of teaching experience (also in 1 to 5 bands). The regression model included two sets of variables added as blocks: a controls only block (Block 1) and a controls plus independent variable block (Block 2). The results indicate whether participation in TISKS can account for any of the variance explained (R^2) in the reported ability of TISKS to motivate teachers to greater professional development. Results are shown in Table 5. VIF statistics were used to determine that multicollinearity was not an issue. All VIF values were below 10, which is the value Myers (1990) cites as a concern for multicollinearity.

Table 5

Summary of Hierarchical Regression Analysis for Variables Predicting Teachers' perception of TISKS as Motivator for Professional Development (N=182)

	t	*p*	*B*	β	*F*	*df*	*p*	adj. R^2
Block 1					12.30	2, 179	.000	.111
Age	-3.28	.001	-.39	-.32				
Experience	-.33	.744	-.03	-.03				
Block 2					8.20	1, 178	.000	.107
Age	-3.28	.001	-.39	-.33				
Experience	-.31	.756	-.03	-.03				
TISKS (0=No)	.36	.723	.063	.03				

The results showed a small but significant variance explained by the model. Only the age predictor was significantly predictive of the difference in reported motivation. The older the respondent, the less likely they were to feel motivated by TISKS. There was no significant difference between those who had completed TISKS and those who had not in their self-reported sense of motivation to improve as a professional as a result of the school using TISKS.

RQ2: Reaction and Influencing Factors

The second question concerned teachers' overall favorableness toward TISKS, and the sample was drawn only from those teachers who had completed TISKS and received their scores and their pay raises. It is important to remember that almost all teachers received some form of pay raise through participation in TISKS, which is a salient feature of this sample of teachers. The relevant portion of the survey was made up of the Teacher Reaction Scales, a set of items created and validated in previous research (Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). The survey presented several items related to the teachers' perceptions of and experience with TISKS in order to determine the elements of that experience that may have had an effect on their overall impression of the program. Teachers responded to each item on a traditional Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Therefore, scores of 1 or 2 can be interpreted as disagreement while scores of 4 and 5 can be interpreted as agreement. In addition, the survey gave opportunities for teachers to write open-ended comments on the questions in blocks. Finally, the focus group discussions allowed a smaller sample of teachers who had completed TISKS to share their thoughts about the program through a set of focused questions. Transcripts from the interviews were analyzed for alignment

with expectancy theory, the nine a priori dimensions from previous research (Ha, 2003; Heneman & Milanowski, 2003; Milanowski & Heneman, 2001), and then for any new themes emerging. Further analysis included examining frequencies and computing the means of survey items and scales. Exploratory factor analysis was conducted to conclude whether the items clustered as expected, and multiple hierarchical regression was conducted to explore the relationship between the identified factors and dimensions and overall favorableness.

Qualitative Results: Open-ended Questions and Interview Results

In the open-ended survey questions and focus group interviews, several themes were identified in line with expectancy theory and the a priori dimensions. These results are summarized in Table 6.

Table 6

Summary of Open-ended Survey and Interview Results

Dimension	Summary of Teacher Responses
Understanding of System	Most teachers understood the system's processes and the requirements it had of them.
	Teachers generally felt the guidance offered in the form of templates and portfolio guidelines were clear and made compliance easier.
	There was some lingering confusion among teachers about the exact purpose of the system: professional growth, evaluation, baseline data, or placement on the pay scale?
Acceptance of Standards	Most teachers expressed satisfaction with the rubric, with some commenting that none was perfect, but this was the best they had seen.
	A few teachers mentioned that the rubric did not apply well to their specialist assignment or grade level.
	Some teachers expressed dissatisfaction with the vague wording of various performance levels in the rubric, leaving them open to subjectivity.
	Several teachers mentioned a lack of attention in the rubric to the extras or the "X factor" that make teachers excellent outside the classroom.
	A few teachers mentioned that attaining the highest levels in all

	categories was virtually impossible, or would require far too much time and attention away from other areas of their work or lives.
Trust and Role of the Evaluators	Many, but not all, teachers reported having positive interactions and professional conversations with members of their panels. Some teachers expressed concern that their panel members were not familiar with either their grade level or their specialty and therefore not qualified to evaluate them. Several teachers pointed out the lack of specific feedback from their supervisors and panel members. Several teachers expressed a belief that administrators played favorites and results were consequently less than objective. Some teachers expressed the belief that administrators were trapped by TISKS and could not carry out their evaluations unfettered.
Fair Results (Distributive Justice)	Many teachers complained that they had heard results varied greatly across the school divisions, and so were not fair. Within and across divisions, some teachers mentioned rumors of low inter-rater reliability among administrators. Several teachers mentioned the belief that, for financial reasons, there were a fixed number of high scores that could be assigned.
Fair Process (Procedural Justice)	Several teachers expressed a sense of fairness from the process because they received pay raises and were being recognized. Many teachers agreed the process was fair as it brought administrators into the classrooms more. Many teachers expressed dissatisfaction with the way scores were rounded and the algorithm for determining final scores. Several teachers complained that the process was not fair for them as their administrators did not complete the required number of observations.
Accurate Results	Several teachers commented on the accuracy of the results they received—matching their self-assessment and reflecting their practice. Many teachers expressed a lack of understanding of how their scores were determined. Many teachers commented on the disconnect between the conversations they had with their panel members and the surprisingly low scores they ultimately received. Many teachers complained of inaccurate results due to their perception of poor training and practice among the administrators on panels.

Impact on Teaching and Work	Almost all teachers who commented expressed dissatisfaction with the feedback from TISKS, saying it was not direct enough, too summative, and lacked follow-up steps. Some teachers reported having made changes in their teaching and practice. Some teachers recognized changes in themselves and colleagues, but worried the changes were short-term and only directed at doing well in TISKS. Many cited the tie to one's pay as one reason this may be the case.
Effort vs. Reward	Most teachers agreed that TISKS was more effort than the results were worth. Teachers who felt it was worth the effort felt so mostly because they received a raise. Other reasons mentioned were professional growth, validation, and recognition. Several teachers said TISKS was too cumbersome and took them away from students and family too much to be worth the effort.
Stress	Most teachers commented on TISKS causing them a great deal of stress. Teachers cited the direct tie to pay as the main reason the system was so stressful. Many teachers reported seeing their colleagues struggling with the effects of stress while going through TISKS and noted the adverse effect on morale. Among the few teachers who did not feel TISKS was stressful, they believed this was so because they were sure of themselves as teachers and organized in their approach.
Overall Favorableness	The majority of teachers expressed negative feelings toward TISKS and did not recommend the school continue to use it. Reasons not to continue with TISKS included the direct link to pay, the stress involved, and the undermining of morale and collegiality. Many teachers expressed concern with the future shape and direction of TISKS and did not recommend continuing with it until this was settled.

Across all of these a priori dimensions, two crosscutting themes emerged: 1) tie to pay and 2) need for feedback and follow-up.

Tie to pay. Teachers expressed mixed feelings on the tie to pay. Most teachers were critical of the direct connection between results and the amount they were to be paid, even if they received a pay raise through the process. Teachers commented on the

artificial motivation created by the tie to pay, the supposed quota on high scores demanded by financial constraints, which affected beliefs about fairness and trust of the evaluators, and the stress caused by the tie to pay. Many teachers stated overtly that they would be more in favor of TISKS if it were linked to professional development rather than pay.

Need for feedback and follow-up. The need for focused feedback was another theme that crossed comments on several dimensions. Whether teachers felt positive or negative about TISKS, most expressed a need for focused feedback that gave them follow-up directions to pursue for future growth or simply to improve their performance in the TISKS profile year. The lack of feedback was an often-repeated criticism of the system, and it crossed over into many of the dimensions, including understanding of the system, acceptance of the standards, trust and role of the evaluators, fair process, and impact on work.

Quantitative Results

Exploratory factor analysis. Because the sample size was necessarily small, exploratory factor analysis was conducted both to determine the extent to which the survey items clustered in similar ways to previous research in the US (Heneman & Milanowski, 1999, 2003; Milanowski & Heneman, 2001) and in Korea (Ha, 2003) and to reduce the number of dimensions for subsequent analysis. Exploratory factor analysis based on principal axis factoring with promax rotation was used to request seven factors in SPSS. Promax rotation was used because the scales and underlying factors have shown intercorrelations in previous research (Ha, 2003; Milanowski & Heneman, 2001). Expectancy theory suggests three underlying factors have an impact on reactions to

appraisal situations: expectancy, instrumentality, and valence; therefore, a three-factor solution was requested. Three factors were identified with eigenvalues above 1.0 and explaining 70% of variance in the Teacher Reaction Scales. Each identified factor accounted for 7.9% to 50.3% of variance. The Kaiser-Meyer-Olkin measure of sampling adequacy was .901. Bartlett's test of sphericity was significant at $p < .0001$. All communalities were well above .30, confirming that each item shared some common variance with other items. Results from the factor analysis are shown in Table 7.

The factors primarily aligned with the proposed alignment toward expectancy theory, with a notable but explainable exception. The trust and role of evaluators dimension, which had been identified with expectancy, clustered with fair results, fair process, and accuracy, though these three dimensions could be considered elements of instrumentality. This is not unexpected as the cross-loading has been observed before, notably in Ha (2003). Ha suggested this could be due to the identification of an effect (such as a high or low score or a pay increase) with the person conducting the evaluation. Indeed, Ha found through his regression model, with a much larger sample of teachers from Korea, that fair results, fair process, accuracy, and the role of the evaluator were the most predictive dimensions of teachers' favorability toward their evaluation and bonus system. Another exception to the originally conceived structure was evidenced by the weak loading (.43) of the item "Working toward improving performance on the rubric will really help teachers develop as professionals" into the same factor occupied by the valence dimension of "Impact on Work." The item was eliminated on the basis of its low factor loading. Finally, the weak and cross-loading item "I have a clearer idea of what the school expects because of TISKS" was eliminated from the final factors due to the weak

Table 7

Factor Loadings And Communalities Based On Principal Axis Factoring With Promax Rotation For 20 Items From The Teacher Reaction Scales (N = 93)

Item	Factor			Comm-unality
	1	2	3	
I had a good understanding of the rubric and procedures involved in TISKS.		.77		.58
I knew what I needed to do in order to do well in TISKS.		.54		.53
Overall, I have a good understanding of TISKS.		.76		.60
The rubric does a good job of defining good teaching and work.		.68		.75
The rubric describes the kind of teaching that teachers in our school should strive for.		.72		.69
I feel quite satisfied with the discussions of my performance I had with my panel members.	.75			.70
I am satisfied with the feedback I received from my panel members.	.65			.64
Overall, I am satisfied with my interactions with my panel members.	.76			.75
I agree with my final TISKS scores.	.89			.89
The TISKS scores I received were fair.	.93			.88
I got the TISKS scores I deserved.	.89			.89
Overall, the way the TISKS process was run was fair.	.74			.81
My TISKS evaluation was conducted in a fair way.	.80			.75
The processes and procedures of TISKS were fair.	.69			.77
The TISKS scores I received were an accurate evaluation of my performance in the six domains.	.77			.87
My panel's evaluations reflected my true performance.	.76			.89
My panel's evaluation of my work matched my own evaluation.	.76			.74
TISKS helped me learn how I can improve my teaching.			.85	.77
I learned a lot from TISKS.			.87	.68
The TISKS process helped me develop as a professional.			.73	.76

and split loading of the item across two factors. Though these items were not included in the factor scores generated for later regression analysis, they are included in the descriptive discussion in the survey and interview results section.

The three factors were closely related to the expectancy theory perceptions, so the factors were named accordingly. Factor 1 was instrumentality (including the role of the evaluator). Importantly, these dimensions clustering into one factor is directly in line with Ha's (2003) finding that these four dimensions were the most strongly connected to overall favorableness toward a similar evaluation and bonus system for teachers in Korea. Factor 2 aligned with expectancy, and factor 3 aligned with the impact on work items of valence, and so was called "impact on work."

The goal of reducing the large set of variables into a smaller set for regression analysis was therefore accomplished through factor analysis, which supported and confirmed the theoretical framework of the study. It must be acknowledged that the total number of observations (N=93) is lower than traditionally suggested for exploratory factor analysis. However, several indicators support the factor structure finally decided. The subject to item ratio was 4.7:1, each final item had a strong loading on a single factor, each factor had at least three items, and all communalities were consistently high (Costello & Osborne, 2005). It cannot be claimed that the factor structure was as robust as one with a much larger data set, but given theory and the total sample of teachers available, these indicators supported the use of the factors in subsequent regression analysis with the appropriate amount of caution and bolstered by qualitative data.

With the dimensions reduced into three factors, coefficient alphas were computed for each factor to confirm internal consistency. All three factors had Cronbach's Alpha

estimates above .80: 1) instrumentality (α=.96), 2) expectancy (α=.87), and 3) impact on work (α=.91). The elimination of the item "I have a clearer idea of what the school expects because of TISKS" was further supported in the Chronbach's alpha estimate and strengthened the impact on work factor.

Survey Results With Interviews. The survey items, their frequencies in response, their mean scores and standard deviations, and the mean score and standard deviation of each dimension are presented in this section. Results from the coded qualitative data related to each subscale are presented to supplement understanding.

Understanding of the system. The understanding dimension included three items relating to how well the teachers understood the pay system and what was required of them in the process. A summary of the teacher responses is shown in Table 8.

Table 8

Frequency of Teachers' Responses on Understanding of the System Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
I had a good understanding of the rubric and procedures involved in TISKS.	96	3.89	0.93	2 (2.1)	7 (7.3)	14 (14.6)	50 (52.1)	23 (24.0)
I knew what I needed to do in order to do well in TISKS.	96	3.27	1.02	3 (3.1)	22 (22.9)	26 (27.1)	36 (37.5)	9 (9.4)
Overall, I have a good understanding of TISKS.	96	3.82	0.77	0 (0)	7 (7.3)	17 (17.7)	58 (60.4)	14 (14.6)
Average		3.66	0.78					

Note. ( ) indicates the percentage of teachers

The overall mean score on this dimension of 3.66 is midway between neutral and agreement and indicates moderate agreement that teachers felt they had a good

understanding of TISKS, including the rubric and the processes with which they were going to be evaluated. Scores on the individual items reveal some variation in perception. 76.1% of teachers expressed agreement with having a good understanding of the TISKS rubric and procedures. 75% expressed agreement with having a good overall understanding of TISKS. Somewhat less agreement was expressed about knowing what it was they needed to do in order to do well in the program. 46.9% expressed agreement with this statement, while 26% disagreed.

Open-ended responses and data from the focus group interviews gave limited additional information about the scores concerning understanding of the rubric and process. Generally, teachers expressed an understanding of the rubric and the process and procedures required by TISKS. In the interviews, teachers expressed appreciation for the fact there were clear specifications and guidelines for the elements to be included in their portfolios. In addition, they expressed an understanding that these had been revised based partly on teacher feedback through the on-going iterations of the process, making them easier for teachers to follow.

Some comments in the open-ended questions expressed disagreement with the notion that teachers always knew what was required of them in order to do well in TISKS. The comments concerned the lack of feedback provided by the evaluators and their changeable, subjective or misinformed use of the rubric. One teacher summed this up by saying, "The rubric was clear but knowing what the admin wanted was not." Thus, it appears the application of the system and the rubric were the source of disagreement or misunderstanding of what was required in order to do well in TISKS. Two teachers mentioned a perceived shift in the system from its inception as a framework for

professional development to one with a primary purpose of evaluating and placing teachers on a pay scale. This perceived change in mandate appeared to generate some confusion among teachers about the system itself. In both the open-ended questions and the interviews, some teachers mentioned hearing conflicting information directly from administrators about the underlying purpose of the system.

Acceptance of the Standards. Three items in the acceptance of the standards dimension gave teachers the opportunity to express the extent to which they agreed with the standards upon which they were to be evaluated and the ability of those standards to identify and detail what it meant to be a good teacher. Results are summarized in Table 9.

Table 9

Frequency of Teachers' Responses on Acceptance of Standards Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
The rubric does a good job of defining good teaching and work.	96	3.05	1.05	5 (5.2)	29 (30.2)	24 (25.0)	32 (33.3)	6 (6.3)
The rubric describes the kind of teaching that teachers in our school should strive for.	96	3.38	0.92	2 (2.1)	16 (16.7)	29 (30.2)	42 (43.8)	7 (7.3)
Working toward improving performance on the rubric will really help teachers develop as professionals.*	96	3.11	1.12	4 (4.2)	32 (33.3)	20 (20.8)	29 (30.2)	11 (11.5)
Average		3.18	0.91					

Note. ( ) indicates the percentage of teachers *Not included after factor analysis

Overall, the mean score for the acceptance dimension was close to neutral. Closer examination of the individual items reveals a full range of agreement. Teachers were fairly evenly distributed in their agreement with the statement about the rubric's ability to

define good teaching or work. 35.4% of teachers disagreed, 25% were neutral, and 39.6% of teachers agreed. Similar results can be seen with the utility of working toward improved performance on the rubric and its tie to developing as a professional. 37.5% of teachers disagreed, 20.8% were neutral, and 41.7% agreed. Teachers were somewhat more positive toward the statement that the rubric described the kind of teaching toward which teachers at TIS should strive. 51.1% of teachers agreed with this statement, while 18.8% disagreed.

In the open-ended questions and the interviews, teachers had much more to say about the rubric than they did their understanding of the system overall. While many teachers believed the rubric did a good job of defining aspects of work and levels of achievement toward which all teachers should strive, others found the rubric vague, over-crowded with details, open to subjectivity, and in a few cases missing important aspects of their jobs. Some teachers saw the rubric as too focused on TIS and a particular definition of good teaching to be applicable beyond the school.

Some teachers reported an organizing effect from simply reading through the rubric. Seeing what categories were included and the various descriptors gave these teachers a framework within which to analyze and reflect on themselves and their work. In addition, the rubric challenged some teachers to improve and made them aware of areas to improve in their teaching. On the other hand, other teachers pointed out that even though the rubric did a good job of describing a high level of performance, they doubted that teachers could sustain performance at those high levels for all areas and over time. Several teachers mentioned a proclivity for short bursts of high performance in order to score well, followed by regression to a normal level of performance. Similarly, several

teachers commented on the subjective use of the rubric. Whether the rubric was well-developed or not, these teachers felt administrators did not know how to apply the rubric to the teachers they were observing, or, in some cases, that they preferred to rely on relationship or on "favorites." The perception was that this caused some administrators to bypass the rubric altogether.

Another category of comments on the rubric involved its perceived inadequacies. In both the open-ended responses and in interviews, several teachers believed the rubric did not apply well to their specialties or grade levels. Specifically, these teachers mentioned portions of their job that were not reflected on the rubric at all, though they did not elaborate.

More generally, several teachers made mention of the fact the rubric did not include for any teacher activities they pursued outside of the classroom that contributed to the school such as extra-curricular and cross-curricular activities. Furthermore, a few teachers mentioned the unquantifiable nature of some important aspects of being a good teacher such as maintaining a good relationship with students. In both the open-ended questions and the interviews, the sections dealing with communication and collaboration were singled out for their subjectivity and their attempt to measure qualities that are too difficult or impossible to measure.

Trust and Role of Evaluators. The dimension about the role of the panel members in the process contained three items seeking perception about the discussions, feedback, and other interactions teachers had with their panel members during the profile year process. Results are summarized in Table 10.

The overall mean score of 2.99 on the dimension is almost exactly neutral. An examination of the individual items reveals a somewhat more nuanced picture. While the mean scores on the items concerning discussions and interactions with panel members were neutral tending toward agreement (3.20 and 3.13 respectively), the item concerning satisfaction with the

feedback received from panel members was below neutral, tending toward disagreement (2.62). The result is indicative of a lower level of satisfaction with feedback than with discussion and other interaction. Indeed, 51.6% of teachers disagreed with a feeling of satisfaction with the feedback they received from their panel members.

Table 10

Frequency of Teachers' Responses on Trust and Role of Evaluators Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
I feel quite satisfied with the discussions of my performance I had with my panel members.	93	3.20	1.13	8 (8.6)	18 (19.4)	23 (24.7)	35 (37.6)	9 (9.7)
I am satisfied with the feedback I received from my panel members.	93	2.62	1.32	24 (25.8)	24 (25.8)	16 (17.2)	21 (22.6)	8 (8.6)
Overall, I am satisfied with my interactions with my panel members.	93	3.13	1.26	14 (15.1)	15 (16.1)	20 (21.5)	33 (35.5)	11 (11.8)
Average		2.99	1.11					

Note. ( ) indicates the percentage of teachers

Input from the open-ended questions and interviews echoed the survey responses. Many teachers commented on how positive, hard working, and committed their panel members were. Several teachers commented on having great conversation on important

topics with their panel through the profile year, and more than one indicated having felt valued, appreciated, and supported. One teacher acknowledged and thanked panel members for giving up valuable time for observations and input and for building good trust. A few comments teachers made about having been "lucky" with the make-up of their panels were indicative of the belief that panel members' roles in the process were distinctly subjective. More than one teacher indicated they needed consciously to keep a positive relationship with their supervisor and not "argue" or "disagree" as the supervisor's score carried the most weight in determining the teachers' salaries, and there was no requirement for the panel members to refer directly to the rubric. Several teachers advanced the idea of hiring outside evaluators to conduct the observations in order to guarantee objectivity of the panel members.

Consistent with the survey responses, most less-than-favorable comments centered on the need for focused feedback. They attributed some panel members' inability to give useful targeted feedback either to the administrators' lack of subject knowledge, too few observations, or direct instruction from "the system" not to give such feedback or discuss scores on the rubric. The lack of feedback was the single most repeated criticism of panel members in both the open-ended responses and the interviews. Several teachers who made very positive comments about their panel members personally and professionally still highlighted the need for better feedback. A few teachers indicated the overwhelmingly positive interactions with their panel members became awkward when scores inconsistent with the tenor of those conversations subsequently shocked them.

Additional criticisms focused on the inconsistent effort and helpfulness from the various panel members, though a few of these comments were offered with the caveat

that administrators are generally busy and seemed well meaning. While some teachers indicated they had helpful collaboration only with their peers on the panel, others indicated their peers were not helpful at all, or that only one of their panel members was helpful and the others absent or disinterested. Several cited having been observed inconsistently by some panel members, and two indicated at least one panel member never observed them at all.

Fair Results (Distributive Justice). The fair results dimension contained items soliciting teacher agreement with the extent to which the scores they ultimately received were indicative of a process that gives results fairly to participants. In other words, the scores are distributed fairly to those who deserve them. The results on this dimension are summarized in Table 11.

Table 11

Frequency of Teachers' Responses on Fair Results (Distributive Justice) Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
I agree with my final TISKS scores.	95	2.85	1.21	13 (13.7)	30 (31.6)	18 (18.9)	26 (27.4)	8 (8.4)
The TISKS scores I received were fair.	95	2.95	1.21	10 (10.5)	30 (31.6)	21 (22.1)	23 (24.2)	11 (11.6)
I got the TISKS scores I deserved.	95	2.82	1.13	10 (10.5)	33 (34.7)	23 (24.2)	22 (23.2)	7 (7.4)
Average		2.87	1.14					

Note. ( ) indicates the percentage of teachers

With mean scores just below neutral, it is clear there is a full range of opinion on whether or not the TISKS score distribution was a fair one. 35.8% of teachers responding indicated they agreed with their final scores, while 45.3% did not agree. Similarly, 35.8% of responding teachers indicated their scores were fair, while 42.1% of teachers indicated

they did not agree that their final scores were fair. On the final item in the dimension, 30.6% of responding teachers agreed they got the TISKS scores they deserved, whereas 45.2% of teachers believed they did not get the scores they deserved.

Comments from the open-ended survey questions and the focus group interviews relative to fair results focused mostly on the inaccuracy of results based on the subjectivity of the administrators scoring the teachers and the variability of scores across administrators from the various divisions of the school. Several teachers mentioned that administrators in each division had their favorite teachers or those with whom they had particularly good relationships and that those teachers faired better with TISKS scores. There were also several mentions of a shared belief across the school that some divisions were easier than others when it came to achieving high scores. Similarly, more than one teacher cited what they believed to be low inter-rater reliability among administrators when using the rubric. Teachers also expressed uncertainty about whether or not the rubric and scores could be evenly applied across teachers in different stages of life. For example, some were unsure teachers with small children or other family commitments could consistently perform at the highest levels of the rubric across all six domains, and they did not find this fair. Several teachers referenced the lack of transparency in how scores were determined, so they were unsure if the scores were fair.

At least six teachers mentioned the belief that there were a fixed number of high scores available to be awarded to teachers through TISKS. These teachers felt the pre-determined scores were the result of financial decisions and limited financial resources at the school, which undermined the fairness of the results.

Fair Process (Procedural Justice). The fair process dimension solicited teacher opinion about the process and procedures they experienced that led to the scores they ultimately received. This dimension is contrasted with the fair results dimension in that it asked teachers to comment specifically on the process and the way in which the evaluation was conducted as opposed to the specific scores they received. Results from the three items in the dimension are summarized in Table 12.

Table 12

Frequency of Teachers' Responses on Fair Process (Procedural Justice) Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
Overall, the way the TISKS process was run was fair.	95	2.68	1.16	16 (16.8)	31 (32.6)	19 (20.0)	25 (26.3)	4 (4.2)
My TISKS evaluation was conducted in a fair way.	95	2.96	1.18	12 (12.6)	25 (26.3)	20 (21.1)	31 (32.6)	7 (7.4)
The processes and procedures of TISKS were fair.	95	2.74	1.10	14 (14.7)	27 (28.4)	28 (29.5)	22 (23.2)	4 (4.2)
Average		2.79	1.06					

Note. ( ) indicates the percentage of teachers

The overall mean score on this dimension is only slightly lower than that on the fair results dimension. It also tends toward disagreement, slightly below neutral. The lowest mean score was evident on the first item in the dimension concerning the overall fairness of the process for TISKS. 30.5% of the teachers answering the question agreed it was overall run in a fair way; however, 49.4% disagreed. The second item, which asked teachers whether their TISKS evaluation was conducted in a fair way, received slightly higher agreement. 40% of teachers agreed that their evaluation was conducted in a fair way, while 38.9% of the teachers responding disagreed. On the last item in the

dimension, 27.4% of responding teachers believed the processes and procedures of TISKS were fair. 43.1% disagreed with the fairness of the processes and procedures.

There were repeated mentions in the survey and the interviews of a positive sense of the TISKS process specifically because the teachers received a financial reward. Since the process ended with a pay increase teachers felt they were due, many teachers attributed this result to the fair process overall. Several teachers mentioned the positive attention they received from their administrators during the process and indicated that this meant the process was fair for them. Many teachers offered caveats during the interview and on the surveys, stating they had heard the process was less than fair for some of their colleagues, but they could not comment directly on what that meant.

The most repeated negative comment among teachers in the surveys and interviews was the lack of focused feedback resulting from the process coupled with an absence of follow-up between profile years. Teachers had the sense they were left unsure of the process by which they received the scores they did, and they had no idea other than their own to guide them in what they could do to improve. This tainted the process for many teachers and, as they often stated directly, led them to select the lower agreement scores on the survey. Other procedure-specific issues that were mentioned included dissatisfaction with the way scores from the three panel members were rounded mathematically to arrive at a final score. Teachers were not made aware of the scores of the individual panel members, but they did know the algorithm used to compute the final result, and they did not agree with it.

Other criticisms offered by multiple teachers included the proportion of the final score allotted to panel members' scores, the uneven experience or expertise of individual

panel members, and the fact that some administrators did not complete the required number of observations, thus subverting the fairness of the process. A few teachers mentioned uncertainty with the idea a panel could gain a comprehensive picture of a teachers' performance through a few observations—especially if they did not complete them as required or if they only saw lessons teachers had prepared to show themselves off to others. Finally, more than one teacher mentioned dissatisfaction with the appeals process when they disagreed with scores. They believed there was no recourse and the appeals process was moot.

Accurate Results. The accurate results dimension sought to determine how teachers perceived the accuracy with which the scores they received depicted the quality of their teaching and work in the school. Three items asked teachers to react to the scores they received at the end of the TISKS profile year relative to their performance and the match with their own evaluation of their performance. Results are summarized in Table 13.

The overall mean score on the dimension was 2.80, which is just below neutral and can be characterized as moderately disagreeing to neutral. The means of the individual items are also all moderately disagreeing to neutral, with a slight difference between the final item and the first two. 34.4% of teachers surveyed agreed that their final score was an accurate reflection of their performance in the six domains of the rubric. 45.2% of teachers did not. Similarly, 36.6% of teachers agreed that the evaluation from their panel reflected their true performance while 45.2% of teachers did not agree. The final item asked teachers to make a judgment about the match between their own evaluation of their work and that of the panel that evaluated them. Here, 26.9% of

teachers agreed that these evaluations matched, while 55.9% of teachers did not agree that these evaluations matched.

Table 13

Frequency of Teachers' Responses on Accurate Results Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
The TISKS scores I received were an accurate evaluation of my performance in the six domains.	93	2.86	1.10	8 (8.6)	34 (36.6)	19 (20.4)	27 (29.0)	5 (5.4)
My panel's evaluations reflected my true performance.	93	2.87	1.15	10 (10.8)	32 (34.4)	17 (18.3)	28 (30.1)	6 (6.5)
My panel's evaluation of my work matched my own evaluation.	93	2.66	1.13	11 (11.8)	41 (44.1)	16 (17.2)	19 (20.4)	6 (6.5)
Average		2.80	1.07					

Note. ( ) indicates the percentage of teachers

Accuracy of the results was the subject of several comments on the survey and in the interviews. Several teachers mentioned how accurate they found their scores to be. The panel members' final results matched their own, and they felt as if they had been accurately evaluated. For others, though they were happy with the results they received and did not contest them, the profile across the six domains did not match their own assessment of their performance, so they questioned the accuracy in this regard.

The most common criticism was the inaccuracy implied by the fact their panel discussions seemed disconnected from the final results received. Teachers felt they were led during the process and panel discussions to believe that their scores would be higher than those they received in their score report. The fact that this report did not include any useful feedback further fed this perception. Many teachers again commented on the

inaccuracy of the scores being the result of a "rigged game." In this perceived scenario, the school decided it could only afford to give a fixed umber of high scores, as higher scores meant higher salaries. Scores were therefore arbitrarily held low in service to this financial goal. They were not intended to be accurate reflections of true performance.

Finally, some sense of inaccuracy was attributed to the role or duties of the panel members themselves. A number of teachers related that their panel members did not observe them enough to give them accurate scores. In one case, a teacher commented that a panel member was not aware of who they actually were, mentioning several aspects of practice that were not reflective of observations or the teacher's portfolio. The teacher concluded the administrator was thinking of someone else and was not scoring them accurately.

Impact on Work. The dimension related to impact on work comprised four items giving teachers the opportunity to reflect on what they gained from TISKS. The dimension included questions about teachers gaining a clearer vision of what is required of them in their jobs, an improvement in teaching, further learning, and development as a professional as a result of participation in TISKS. Results are summarized in Table 14.

The overall mean score of 2.98 in this dimension shows a neutral response to impact on work. Across all the items in the dimension, teachers expressed neither overall agreement nor disagreement with the assertion that participation in TISKS had a positive impact on their work. The highest mean score was recorded on the first item, asking teachers whether they agreed they had a clearer idea of what TIS expects from teachers through participation in TISKS. The mean score was slightly above neutral, tending toward agreement. 50.0% of teachers agreed they had a clearer picture of TIS

expectations because of their involvement in TISKS. 30.8% of teachers disagreed. This item was not included in the mean calculation, however, as it was eliminated during exploratory factor analysis due to weak and cross-loading between instrumentality and valence. Somewhat less agreement was found on the next two items, with mean scores of 2.96 and 2.97—almost exactly neutral. 38.3% of teachers surveyed agreed that TISKS helped them learn how they could improve their teaching. 39.4% disagreed. Similarly, 33% of teachers agreed they had learned a lot from TISKS, whereas 41.4% of teachers disagreed with this statement. On the final item, 39.4% of teachers agreed the TISKS process had helped them develop as professionals. 40.4% of teachers disagreed that TISKS played a role in developing them professionally.

Table 14

Frequency of Teachers' Responses on Impact on Work Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
I have a clearer idea of what the school expects because of TISKS.*	94	3.19	1.14	8 (8.5)	21 (22.3)	18 (19.1)	39 (41.5)	8 (8.5)
TISKS helped me learn how I can improve my teaching.	94	2.96	1.18	11 (11.7)	26 (27.7)	21 (22.3)	28 (29.8)	8 (8.5)
I learned a lot from TISKS.	94	2.97	1.17	7 (7.4)	32 (34.0)	24 (25.5)	19 (20.2)	12 (12.8)
The TISKS process helped me develop as a professional.	93	3.00	1.16	8 (8.5)	30 (31.9)	19 (20.2)	28 (29.8)	9 (9.6)
Average		2.98	1.07					

Note. ( ) indicates the percentage of teachers. *Not included after factor analysis

In the survey responses and interviews, impact on work received a wide variety of comments. A large segment of those commenting cited a forced, organized time of reflection and conversation about professional practice that they found helpful and for which they were grateful. Some commented on their belief this led to better teaching and a better experience for the students in their classes. A few teachers cited specific areas among the six domains that they felt they had improved during the TISKS process. Another positive attribution to impact on work was the increased presence of administrators in classrooms and the fact they now knew what teachers were doing with students. In the focus group interviews, this observation was tempered by the shared belief that administrators were spending time only in the rooms of those who were currently going through the TISKS profile year.

The theme of divided administrator attention carried through in other interview and survey comments. Several teachers commented on the limited value of the impact on their work as there was no mechanism for follow-up between profile years. Teachers felt valued and attended to during the profile year, but ignored immediately thereafter. Several reported not having been observed before or since their profile year. Similarly, several teachers mentioned being left to their own devices to develop. They did not see this as much different from their profile year in that they had their own professional development plans and developed themselves in spite of having to take part in TISKS.

Negative reactions relative to impact on work related to what several teachers identified as a change in morale and collegiality due to competition and the stress many felt as they went through the profile year since it determined their salary. A few teachers mentioned that they saw more effort from their colleagues as they went through the

profile year, but they felt this effort was directed predominantly at "doing well" or "earning more" than it was directed toward becoming a better teacher.

Effort Vs. Reward. A single item on the survey asked teachers whether they agreed that TISKS took more effort than the results were worth. The results are summarized in Table 15.

The mean score on this item of 3.59 indicates moderate agreement that TISKS required more effort than the results were worth for teachers. 57.5% of teachers surveyed agreed this was the case, while 24.4% disagreed, believing the effort was not more than the worth of the results.

Table 15

Frequency of Teachers' Responses on Effort Vs. Reward Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
TISKS took more effort than the results were worth.	94	3.59	1.26	5 (5.3)	18 (19.1)	17 (18.1)	25 (26.6)	29 (30.9)

Note. ( ) indicates the percentage of teachers

Teachers on the survey responses and in the interviews generally agreed that TISKS took a great deal of effort and was often tiring because of additional responsibilities beyond teaching and preparation. Whether the effort was worth the reward caused several teachers to comment on what rewards they gained. Teachers who believed the effort was worth the reward believed so because they grew as professionals or because they received a pay increase. Among those who commented on this worth, the increase in pay was the most commonly referenced reason the effort was worth it. Professional development, validation, a lasting product in the portfolio, and greater

recognition for one's craft were less often mentioned, but were other reasons given for the reward being worth the effort.

Several teachers commented on the reasons they did not find the reward worth all the effort required by TISKS. Among these, time spent on TISKS that could have been spent on work for or with students was the most commonly cited, followed by time away from family at home due to all the reflection and writing necessary. One teacher mentioned having chosen time with family over TISKS and the disappointment that came when seeing others who had not made that choice doing better and receiving higher scores. Other teachers commented on the requirements for the portfolio and observations being too cumbersome and focusing simultaneously on too many aspects of work and teaching to be useful in the long term. One teacher commented on the disappointment experienced when during the panel review, it appeared the members of the panel had not read the material upon which he or she had worked so hard.

Stress. A single item on the survey gave teachers the opportunity to agree with a statement about TISKS causing them a lot of stress. The results from this item are summarized in

Table 16.

Teachers clearly agreed with this statement, as evidenced by a mean score of 4.04. A full 73.4% of teachers agreed that TISKS caused them a lot of stress. Of those, 44.7% strongly agreed. 11.7% of teachers disagreed, indicating TISKS did not cause them a lot of stress.

Table 16

Frequency of Teachers' Responses on Stress Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
TISKS caused me a lot of stress.	94	4.04	1.09	2 (2.1)	9 (9.6)	14 (14.9)	27 (28.7)	42 (44.7)

Note. ( ) indicates the percentage of teachers

Comments from the survey and interviews clearly pointed to the most common reason for stress being the direct link of TISKS scores to pay. This created a "burden of anxiety" that many saw as counterproductive to genuine growth and honesty with their panels. Several commented on the fact that teachers should not have to worry about providing for themselves and their families on the basis of several observations and a portfolio of evidence about their practice.

In the interviews, several mentions were made of the cumulative effect of watching colleagues break down, develop nervous ticks, and compromise their health through the profile year. Others reported ill health effects for themselves, anxiety around the writing component of the system, a lack of time with and understanding from family members, and a sense they were engaging in a game of survival. One teacher commented that the most stressful aspect was the sense he or she had already been judged from the beginning.

On the other end of the spectrum, teachers who reported limited stress from TISKS attributed this to their comfort with themselves as teachers and the ability to keep the process in perspective.

Overall Favorableness. The final dimension on the survey asked teachers to respond with their overall favorableness toward TISKS after having gone through the process of the profile year. The dimension included three items about teachers'

satisfaction with the evaluation system and whether or not they believed the school should continue to use it. Results are summarized in Table 17.

Table 17

Frequency of Teachers' Responses on Overall Favorableness Dimension

Item	*N*	*M*	*SD*	1	2	3	4	5
In general, I feel that the school has an excellent performance evaluation system in TISKS.	92	2.32	1.06	24 (26.1)	30 (32.6)	25 (27.2)	11 (12.0)	2 (2.2)
Overall, I am satisfied with TISKS.	92	2.52	1.17	21 (22.8)	30 (32.6)	15 (16.3)	24 (26.1)	2 (2.2)
The school should continue to use TISKS.	92	2.32	1.28	35 (38.0)	18 (19.6)	19 (20.7)	15 (16.3)	5 (5.4)
Average		2.38	1.08					

Note. ( ) indicates the percentage of teachers

The overall mean score of 2.38 on this dimension shows a degree of disagreement with the statements about a positive feeling toward TISKS. All three items in the dimension have scores approaching disagreement as well. With mean score of 2.32, 14.2% of teachers believed TIS had an excellent performance evaluation system in TISKS, and 58.7% of teachers surveyed disagreed. Overall satisfaction with TISKS was reported by 28.3% of teachers, while 55.4% of teachers did not agree that they were satisfied overall with the system. 16.3% of teachers were neutral on this question. When asked whether TIS should continue to use TISKS, 21.7% of teachers surveyed indicated they believed so. 57.6% of teachers indicated they disagreed with the continued use of TISKS.

This dimension garnered many comments from teachers in the survey questions. Those who agreed that the system was satisfying and should be continued focused on the

fact this system was not ideal but was the best they had experienced. This was attributed to higher pay for teachers and an increased presence in the classroom for administrators. Though they believed the system should continue to be used, few did so without also suggesting that the system be revised to make it less subjective, less onerous, and more focused on on-going improvement and growth.

As the majority of teachers expressed lower than neutral feelings toward TISKS overall, there were many more comments about why the system should be discontinued. The most commonly used was the unpopular direct link to pay. Some teachers suggested this one aspect of the program could be changed and all of TISKS would improve as a system for reflection and professional growth. Other comments were more varied, from the introduction of stress and competition into faculty culture to the emphasis on evaluation and justification that took time away from students in class, extra-curricular activities, and family. Stress was a repeated reason not to continue with TISKS, including stress among administrators that teachers could perceive. Some teachers commented that everyone was a victim in TISKS, including administrators. Finally, several teachers mentioned that they did not believe TISKS should be continued because they had not received any word about how the system would be changed for them when they went through it a second time. This was coupled with a continued confusion about the purpose of TISKS overall: was it a system for professional growth or primarily for determining compensation?

Taken as a whole, the survey, the open-ended survey responses, and the interviews reveal faculty who are generally dissatisfied with TISKS, as evidenced by the mean score in overall favorableness of 2.38—almost two-thirds of a point below neutral.

The scores on the other dimensions cluster around the neutral point, with the exception of stress, which is the highest of the dimensions and has a mean score of 4.04—the only one to rise above the level of agreement. Teachers generally reported understanding the system and agreeing somewhat with the standards upon which they were to be evaluated, but there was an expression of concern about the evaluators, the fairness of the process and the results, and the accuracy of the results. The role of the evaluators is especially salient. Evaluators were mentioned in narrative responses relative to every dimension related to implementation, which helps explain why the evaluator dimension clustered strongly with the other instrumentality dimensions. Teachers reported a neutral effect on their work, but also higher agreement with the notion that TISKS was more effort than it was worth and caused a lot of stress.

Correlation Analysis. In preparation for regression analysis, due to the small sample size, age and level of education were re-coded into dummy variables to aide in calculations. Effort and stress were also standardized as z scores so that all variables except the dummy-coded controls had a mean score of zero and a standard deviation of 1. Inter-correlations among all the factors and variables are shown in Table 18.

As expected, many of the variables and factors are significantly intercorrelated ($p<.05$), suggesting the underlying variables are not entirely independent. Significant intercorrelations among the Teacher Reaction Scales were also found by Heneman and Milanowski (2003; 2001) and by Ha (2003). The intercorrelations were generally moderate, with the highest being the correlation between the impact on work and instrumentality factors (.593). It is not surprising, then, that impact on work and instrumentality were also similarly strongly correlated with the dimension of overall

favorableness. All of the factors and variables were significantly correlated with overall favorableness, with the exception of teacher age. Level of education, effort, and stress showed significant negative correlations to overall favorableness.

Table 18

Correlations Among Extracted Factors and Variables (N=92)

	1	2	3	4	5	6	7
1 Age (0=<41)	--						
2 Education (0=some graduate but no master's)	.04	--					
3 Instrumentality	-.18*	-.27**	--				
4 Expectancy	.06	-.26**	.52**	--			
5 Impact on Work	-.06	-.34**	.59**	.43	--		
6 Effort	.07	.17	-.58**	-.43**	-.47**	--	
7 Stress	.15	.17	-.45**	-.11	-.37**	.49	--
8 Favorableness	-.02	-.38**	.68**	.50**	.68**	-.53**	-.45**

*p<.05, **p<.01

Regression Analysis. Using the three identified factors and the single-item scales of effort and stress, a hierarchical multiple regression was conducted to explore the amount of variation in overall favorableness attributable to the control variables and the extracted factors and remaining variables. Results are summarized in Table 19.

Results indicate a significant and moderate (R^2 = .123) amount of variance in overall favorableness explained by the control variables block, which included age and level of education. Level of education was significantly negatively predictive of overall favorableness. The higher the level of education among educators who had gone through TISKS, the lower was their overall favorableness toward the system. When the factors and effort and stress were entered as a block, a significant and large proportion of

variance was explained (R^2 = .604). The increase in variance explained beyond the controls only block was also significant. VIF statistics were consulted to ensure no problems of multicollinearity. All VIF values were below 10, which is the value Myers (1990) cites as a concern for multicollinearity.

Table 19

Summary of Hierarchical Regression Analysis for Factors Predicting Teachers' Overall Favorableness toward TISKS (N=92)

Variables	*t*	*p*	*B*	β	*F*	*df*	*p*	adj. R^2
Block 1					7.41	2, 89	.001	.123
Age (0=<41)	-.06	.952	-.01	-.01				
Education (0= some graduate but no master's)	-3.84	.000	-.93	-.38				
Block 2					20.83	5, 84	.000	.604
Age (0=<41)	1.11	.271	.18	.08				
Education (0= some graduate but no master's)	-1.67	.098	-.29	-.12				
Expectancy	1.51	.135	.15	.13				
Instrumentality	3.00	.004	.33	.30				
Impact on work	3.82	.000	.38	.33				
Effort	-.75	.453	-.07	-.07				
Stress	-1.73	.087	-.16	-.14				

An examination of the standardized coefficients revealed that the two factors of impact on work and instrumentality were significantly predictive of overall favorableness, even to a greater extent than level of education. Teachers who saw

themselves learning and growing professionally from TISKS and who believed they received accurate results through a fair process and were satisfied with the interactions with their panel members were more favorably disposed toward the system.

RQ3: Intention to return

The final research question concerned the ability of TISKS to encourage teachers to renew their contracts and remain at TIS. A single item on the survey addressed this question. Results from the item are shown in Table 20.

Table 20

Teachers' Self-reported Intention to Return to TIS After Their Current Contract

Item	*N*	*M*	*SD*	1	2	3
Do you plan to return to TIS when your current contract ends?	182	2.41	.750	29 (15.9)	50 (27.5)	103 (56.6)

Note. ( ) indicates the percentage of teachers

To investigate the relationship between participation in TISKS and intention to sign another contract, the number of respondents falling into each category was recorded. These data are summarized in Table 21.

Table 21

Summary of Contingency Data: Intention to Renew Contract

Completed TISKS	No	Unsure	Yes
Yes (*n*=98)	18 (18.4%)	21 (21.4%)	59 (60.2%)
No (*n*=84)	11 (13.1%)	29 (34.5%)	44 (47.5%)

A chi-square contingency test was performed. The results were insignificant, χ^2 (2, *N*=182) = 4.10, *p*=.129. There was no significant association between teachers who

had taken part in TISKS and an intention to return to TIS after the end of one's current contract.

CHAPTER 5

Analysis

The surveys and interviews conducted revealed a wealth of detail about the ways international school teachers reacted to this performance pay innovation. This was an opportune time to sample the perceptions of teachers at TIS. It is one of the largest international schools in its region, and approximately half of the eligible teachers had completed the profile year of the program at the time data were gathered. The roughly equal populations of teachers who had and had not completed the program provided an opportunity to ascertain reactions to the program from both sides of the final stage of receiving scores and, ultimately, the impact on one's earnings.

Overall, teachers before and after the final panel discussion and pay consequences reported a less than positive sense of motivation from the program. Similarly, teachers who had gone through the program and received their scores and pay raises had predominantly negative views of the program and its continuation at the school, despite overall neutral average reactions to many specific aspects of the program and its implementation. Finally, participation in the program did not result in any significant difference in teachers' choices about remaining at TIS. The present chapter considers these findings and their practical considerations in depth.

Teacher Motivation

Though almost half of the teachers surveyed reported that TISKS was not a motivator for them to improve their professional practice, almost 40% reported that it was. At the highest level, it can be said that teachers had varied responses relative to motivation. As in other studies (Heneman & Milanowski, 2003; Milanowski & Heneman, 2001), teachers did not entirely support or reject the ability of such a program to motivate

them. That the strongest predictor of this reaction among teachers was their age is not a surprise. A similar pattern has been seen in previous research in which younger teachers report higher motivation from performance-based pay schemes than older teachers (Coggshall et al., 2010; Milanowski, 2007). It stands to reason that younger teachers, anticipating a long path ahead in the traditional model of moving up the pay scale by adding years of experience, would feel motivated by a system that promises to move them higher on the pay scale faster. Furthermore, some younger teachers expressed an appreciation of the observation-heavy system's focus on mentor relationships and a structured approach to considering one's strengths and areas for growth as a teacher. Though teachers with more experience also acknowledged this systematic approach to reflection, other considerations led to a less than favorable response from teachers on average.

The overall neutral-to-negative response relative to motivation is also borne out in the work of Ballou (2001). He found that teachers in private schools were more likely to express interest in pay for performance than their public school counterparts, but that this interest was not evident among private school teachers who worked in schools that actually used such a system. There can be many reasons for this, and the present study investigated this in more depth in the second research question. For the present question of motivation, however, the finding that there was no significant difference in the motivational effects reported by those who had completed the profile year and those who had not is an important one. It would be reasonable to assume that those who had benefitted financially from such a program would report being motivated by such an incentive. Clearly, they did not do so to any significant degree.

It is important to consider that there were many teachers who did report being motivated by TISKS. Aside from those younger teachers who reported a sense of motivation from working with mentors through the program, there were clearly those who felt motivated by the attention and recognition from administrators and the pay raise they received. The single most reported motivator (among those who wrote narrative responses) was the difference in pay afforded by the program. It is important to recognize, however, that the tie to pay was not positively viewed by all and was in fact high on the list of reasons teachers reported not being motivated by TISKS.

Those who were not motivated by TISKS and wrote narrative responses most often cited a preference for intrinsic motivation for professional development as opposed to external motivators such as pay. Many specifically highlighted that higher pay was not a motivator for them, but that it was assumed to be the primary motivator by the TISKS model of evaluation. The "artificial" or "forced" external motivator flew in the face of many teachers' motives for entering the profession and ran counter to their professional pride in their work and their calling. Other de-motivators cited were the time the process took away from students and learning and the lack of useful feedback provided by evaluators. Many teachers questioned the learning value for students and teachers of spending so much time on a system that did not provide useful information for moving forward professionally. Even teachers who had not yet gone through the profile year commented on having heard this either directly or second-hand from those who had. These de-motivators had taken on a life of their own in the faculty, whether based in fact or not.

One can conclude from these results that motivation through pay is not axiomatic. TISKS was designed to determine teachers' placement on the pay scale, but also to serve as a vehicle for professional reflection and development. Results from this portion of the survey, open-ended questions, and interviews indicate that these two goals are necessarily in tension for many teachers. The tie to pay induced in many an underlying sense of competition. While some felt the process helped them to reflect systematically on their practice, even many of those who espoused this belief explained how the tie to pay somehow clouded the entire process for them.

The finding that teacher age was most strongly and significantly predictive of positive motivation can be seen as a useful criterion for school administrators. One way to motivate young, well-qualified teachers is to establish a system by which they can attain the highest levels of the pay scale without having to wait for the traditional model based on years of experience to increase their pay. The same system does not, however, appear to motivate well-qualified teachers who are older. In fact, results from the present study indicate that such a system can de-motivate older teachers. As several factors beyond age are likely to affect teacher reaction to such systems, a deeper investigation is necessary.

Teacher Reactions and Influencing Factors

The second research question made use of expectancy theory to seek further insight into teachers' varied reactions to TISKS. Their reactions were operationalized as overall favorableness: whether they were satisfied with the system, believed the school should continue to use it, and found it to be of high quality. Seven factors, plus effort exceeding reward and stress, had in previous research been shown to relate to teachers'

overall favorableness toward similar systems (Ha, 2003; Heneman & Milanowski, 2003; Milanowski & Heneman, 2001). It is therefore useful in an analysis of results to indicate similarities and differences in the findings of this study when compared with past studies. The following discussion is organized around the three perceptions of expectancy theory: expectancy, instrumentality, and valence.

Expectancy: Understanding and Acceptance

Teachers at TIS generally understood the TISKS system and were accepting of the standards upon which teachers would be evaluated, with some notable exceptions. These results are consistent with the findings of Heneman and Milanowski (2003) in Cincinnati, Ohio in the US, but contrast with those of Ha (2003) in Korea. Though teachers understood the system, there were questions for some about the goal and identity of the system. These teachers wondered if it was truly a professional development system or if it was designed primarily to determine teachers' pay. This could be due to what several teachers revealed in the survey and interviews as a perceived change in the messages from administration about the goal. The comment from several that earlier assurances TISKS was a professional development program contrasted with later messages that TISKS was about baselines for compensation. This shift was confusing. Nonetheless, teachers generally reported an understanding of what was required of them and good guidance in the creation of their portfolio and other requirements.

Teachers expressed similar agreement with the standards on which they were to be evaluated. They felt in general that the rubrics embodied good practice for professional teachers and other roles in the school. The few exceptions did not believe the rubric captured their specialty well or that the rubric did not give enough attention to the

qualities and activities that made teachers at international schools outstanding outside of the classroom. This is in sharp contrast with the teachers in Ha's (2003) study, in which teachers expressed the opposite concern: that their competency model put too much emphasis on duties outside of the classroom. As international school teachers are regularly hired not only for classroom performance, but also for areas in which they can add value outside the classroom, this finding makes sense and is an important distinction in this study concerning international school teachers.

If teachers generally did not favor TISKS, it did not appear to be a result of their understanding of the system or their acceptance of the standards. Indeed, during the course of the program, changes to the requirements and rubrics were made in response to teacher and administrator feedback. These changes were mentioned positively in interviews and open-ended questions, which is consistent with a generally favorable response to these dimensions. Statistical analysis indicated that these two dimensions did not account for a significant amount of the reaction teachers had to TISKS. Though the results were overall neutral-to-positive in these dimensions, other factors were stronger predictors of teacher reactions. This finding is consistent with those of other studies (Ha, 2003; Milanowski & Heneman, 2001). Whether viewed positively or negatively, the design of a system appears less important than issues of implementation.

Instrumentality: Evaluators, Fair Process, Fair Results, Accurate Results

The role of the evaluators, fair process, fair results, and accurate results clustered together into a single factor that did account for a significant amount of the variance in teachers' reactions to TISKS. This finding is consistent with previous studies and speaks to the implementation of the program as opposed to its design. Insights from each of

these dimensions are considered in the following discussion. Teachers were more likely to see TISKS favorably to the extent that they trusted their evaluators, believed the process was fair, and believed results were fair and accurate.

Teachers generally viewed their trust of and the role of their administrators and panel members with neutrality. It is important to point out, however, the clustering of the evaluator dimension with other dimensions on the subject of fairness in this study. Teachers appeared to see their panel members as playing a role in carrying out a fair process rather than as a support to attain high levels of performance, moving their role from one of establishing expectancy to one of supporting instrumentality. To that end, there were several suggestions to replace administrators with impartial, subject- or grade-level-trained impartial evaluators brought in specifically to administer the scoring of teachers. The administrator's dual role of evaluator and professional development support seemed problematic for many teachers.

In interviews and open-ended questions, teachers made much of the positive interactions they had with their panel members, but they commented frequently on the lack of useful feedback they received and the disconnect between their final scores and the positive tenor of the conversations they had with panel members. Numerous teachers also mentioned the importance of a positive relationship with administrators, but primarily in the service of receiving high scores. Conversely, not maintaining a positive relationship or being one of the "favorites" of an administrator was cited as a reason for low scores. Such comments put the panel members in the role of arbiter rather than support, and the imbalance between useful feedback and positive relationship puts administrators in an ambiguous situation. Several teachers mentioned the belief their

administrator panel members were not allowed to give them direct feedback, implying a disempowerment of the administration, which could also account for the clustering of the evaluator dimension with other features of the fairness of the system. In summary, teachers' views of the role of their evaluators were seemingly affected by their perceptions of favoritism and the empowerment of the administrators by the school to freely give feedback. Whether or not teachers believed administrators were free to give high quality feedback, they did not generally believe they received it. This is consistent with findings in other studies (Ha, 2003; Heneman & Milanowski, 2003).

The lack of feedback also figured prominently in the interviews and open-ended questions relative to the dimensions on fair process and fair results. Teachers were generally neutral-to-negative on these dimensions, not least because they felt like they went to a lot of effort to receive very little useful information about their professional practice. Though several attributed their high scores and pay raises to a fair process and fair results, many of those same teachers were puzzled about exactly how they received those results, as they were given no feedback tied to specific criteria—positive or negative. A change made to the protocol in the second year by which teachers received a few points from the rubric as feedback was seen as a token gesture and not helpful in follow-up reflection. Overall, teachers commented that the limited feedback was almost exclusively summative, with no formative value. Such lack of feedback was exacerbated among teachers who reported that their administrators had not conducted the requisite number of observations.

Other, more mechanical reasons were cited for a negative response. Many teachers disagreed with the mechanics of averaging and rounding panel members' scores

to arrive at the final score. Transparency was thus criticized in the way panel members arrived at scores, but the transparent but unpopular algorithm for combining these scores into final scores was equally criticized. Over all of these concerns, the direct tie to pay emerged again and again as an important reason why the evaluators must be professional and impartial and the process and results must be fair. Because the stakes were high, many teachers felt the system should be refined beyond partiality.

The sense of partiality extended beyond evaluators to the different divisions in the school. Teachers had heard stories about the relative ease or difficulty of attaining high scores in different divisions. Though administrators did not openly discuss the scores of individual teachers, the teachers themselves clearly did, and stories had evolved about which divisions gave better scores and consequently higher pay. This undermined the notion of fair results for many teachers.

Similar patterns around feedback and the role of evaluators emerged in the consideration of whether or not results attained were accurate. Several teachers reported being happy with their results and finding them to be in concert with their own assessment of their strengths and areas for growth. Others did not. One would expect this with any evaluation, but many teachers commented on the fact that the mismatch in scores across the domains, combined with a lack of feedback undermined trust in accuracy and confused them. Again, these results agree with finding in other studies (Heneman & Milanowski, 2003; Kelley et al., 2002; Milanowski & Heneman, 2001).

Valence: Impact on Work, Effort, and Stress

Impact on work proved to be the most predictive factor of teachers' overall favorableness toward TISKS. Teachers were most likely to see TISKS favorably to the

extent they thought the process helped them develop professionally, taught them a lot, and showed them ways to improve their teaching and work. This finding fits well with expectancy theory and matches results from Milanowski and Heneman (2001) in Cincinnati, but is a significant difference from the results of Ha (2003) in Korea. His study found no significant impact of this dimension on overall favorableness. Effort being greater than reward and the amount of stress caused by TISKS did not have a significant effect on teachers' favorableness toward TISKS.

In keeping with the findings on teachers' sense of motivation, this group of international school teachers sees itself as fundamentally intrinsically motivated. Professional development and improving their craft are their self-reported primary motivators. The finding of impact on work bears this out empirically. Teachers gave many reasons why they did not rank TISKS's impact on their work more highly, including the transience of its effects, the lack of follow-up, and the direct tie to pay. The quantitative and qualitative results combined demonstrate teachers' favorability toward TISKS would increase if they felt it had a more lasting impact on their development as professionals, gave them more targeted feedback on improving their work, and did not result in competition and gaming associated with a direct tie to pay. Many teachers believed they made their own professional development plans. They looked upon TISKS favorably to the extent it helped them to organize their thinking for those plans. Overall favorableness was negative for the majority of teachers, however, so it can be said that these teachers, as several expressed in their own words, learned and grew in spite of the issues with TISKS that caused them to dislike it and its influence on the school.

Effort and stress did not have a significant impact on teachers' overall favorableness, though there were many comments in open-ended questions and interviews about apparent stress and difficulties with teachers and their colleagues attributed to TISKS. Ha (2003) also found no significant contribution of these dimensions to overall favorableness in his final regression model. One reason this could be true is the general stressful nature of international and private school work, or of teaching in general. Teachers expect a degree of stress in their work. However, successful international school teachers such as those hired by TIS are likely to be resilient and used to a degree of stress. Participation in TISKS likely added stress and effort that was not seen as resulting in a reward, but teachers probably expected this, and so their reactions to these items did not influence their overall impressions of TISKS per se, though they did show strong and significant negative correlation in the variable-level correlations.

Overall, the results from the investigation into teachers' reactions aligned with expectancy theory. Teachers' expectancy perception was higher than the other perceptions. They understood the system's processes and procedures, and they accepted the standards upon which they would be evaluated. Reaction to the role of evaluators, a fair process, and fair and accurate results was overall neutral to negative, indicating lower than neutral instrumentality. In other words, teachers for various reasons did not strongly believe they would reach the high levels of performance set by the system. They felt their evaluators were subjective, that the results were variable across the school, that the process was not adequately followed, and that the results were limited by poor evaluator training or institutional limits beyond teachers' and administrators' control. Finally, their valence perceptions were neutral to negative. Some believed TISKS was valuable in

giving them a framework in which to consider their practice and opportunity with administrators and peers to do so. When this was the case, favorableness was higher. However, many felt this impact was transient or not maintained. Most felt the effort was greater than the reward and that stress was too high. All of these factors combined led to an overall low degree of favorableness toward TISKS, despite tangible rewards such as pay raises.

Intention to Return to the School

The final research question addressed the stated goal of TISKS to attract and retain the best teachers in a highly mobile and competitive pool of well-qualified international school educators. The results reported indicate that there was no significant difference between those who had completed TISKS and those who had not in teachers' reported intention to return to TIS. Previous research has found that the teacher turnover rate at international schools is anywhere between 14% (Odland & Ruzicka, 2009) and 17% per year (Mancuso, Roberts, & White, 2010). The sample from TIS revealed 15.9% of teachers were intending not to return to the school after their current contract ended. The overall results from responses on this item were therefore not a cause for surprise and fitted well with known rates. Participation in TISKS was not related to teachers' intention to return to TIS at the end of their current contract. Decisions about renewal or non-renewal for the teachers surveyed were likely influenced by other factors than TISKS. The desired purpose of retaining teachers is not being met. The findings of this study offer some support to other findings (Coggshall et al., 2010; Milanowski, 2007) that young teachers can be attracted to pay reform, but more research is necessary to identify the ability of reforms to attract and retain a variety of teachers.

Overall Summary

This study investigated the ability of a knowledge- and skills-based pay system to motivate highly qualified teachers in an international school. Furthermore, it investigated the specific factors that contributed to teacher reactions to such a system within the framework of expectancy theory. A final goal was to determine if such a system had the ability to retain teachers in a highly mobile and competitive hiring environment.

Results indicated that overall teachers were not motivated by the KSBP system. Analysis showed that age was a significant factor in this result. The older teachers were, the less motivated they were by the KSBP system. With reference to overall favorable feelings toward the system, teachers who felt the system taught them a lot, helped them learn how they could improve their teaching, and helped them develop as professionals were more likely to have a positive reaction to the system. To a lesser degree, teachers who thought the processes involved were fair, that the results they received were fair and accurate, and that their evaluators interacted with them well and gave them good feedback were more likely to view the system favorably. Understanding the system, accepting the standards, and the amount of stress and effort involved in going through the evaluation process did not have statistically significant impact on teachers' reactions to the system. Teacher reaction was variable, but overall average reaction was negative, despite the fact almost all teachers benefitted financially. Finally, participation in the KSBP system and the pay raises that accompanied it had no significant impact on teachers' expressed intention to return to the school after their current contract.

Concurrent qualitative investigation through open-ended text responses to the survey items and two focus group interviews revealed further detail about teachers'

reactions. Teachers reported a lack of motivation from the program because they were predominantly intrinsically motivated to improve their practice. Many teachers saw the evaluation's direct tie to a position on the pay scale as counter to intrinsic motivation. In addition, the direct tie to pay increased stress, led to gaming of the system by colleagues, and undermined teacher morale, according to many teachers. Those who did report motivation cited an increase in pay and professional recognition as factors.

The direct tie to pay also had an impact on teacher favorability toward the system. Relative to expectancy, teachers reported an understanding of the system and a general acceptance of the evaluation standards, with a few exceptions. Some were confused about the real goal of the system, and others took issue with specific parts of the rubrics applied to their positions. Overall, expectancy was positive. Instrumentality was overall negative, and this seemed to be impacted most profoundly by a lack of focused, useful feedback toward professional improvement. The lack of feedback caused questions for teachers about their evaluators, about the fairness of the process, and about accuracy. In addition, teachers worried about the transparency of the processes involved even though they expressed understanding of them. They worried that administrators were under-qualified or too busy to conduct the evaluations fairly and accurately. Many also believed the administrators could not score the way they would like because of financial constraints on overall packages. While several doubted anyone could attain the highest levels of the rubrics on all skill and knowledge areas, many others doubted the highest scores would be awarded in any case, because the school could not afford to pay teachers accordingly. Finally, teachers who saw a positive impact on their work did so mostly because of the structured way the system caused them to think about their practice. The majority of

teachers interviewed did, however, give examples of negative impacts on themselves and their colleagues due to stress induced by the system and the amount of effort that did not justify the reward received, though almost all teachers did received financial rewards.

Probably because of the sensitivity of the issue, only one teacher commented in the qualitative data on the intention to return to the school. Quantitative data showed no significant impact on teacher intentions to return.

Practical Implications and Future Research

The results from this study and its place in the research literature imply several steps that can be taken to make a KSBP system more motivating and attractive to highly qualified teachers. While it is clear some design implications can be derived from this study, implementation issues are most important.

Design

In keeping with expectancy theory, good design can go a long way toward raising teachers' positive perceptions of expectancy.

Evaluation instruments and competency model. When designing an evaluation and pay system, it is important that the evaluation be based on a teacher competency model upon which teachers and administrators can agree (Heneman & Milanowski, 2003). TIS went through such a process, involving teachers and specialists in the creation of the various rubrics, and also adjusting them according to feedback from teachers as they were used for evaluation. When teachers agree on the model for good teaching, the use of the evaluation instrument is more likely to be accepted, not least because professional conversations and understanding are enhanced by a shared vocabulary. The instrument and the descriptors it entails need to be clear and robust, as it is the primary

key to which all conversations, feedback, and professional development should return in order to remain consistent with the established vision for good practice espoused by the school. The tool should also be easy to use consistently, as it will be the determining factor for a teacher's placement on the pay scale.

Purpose. Beyond the agreement on a competency model and a tool, the purpose of the system must be clear and transparent. Teachers at TIS held to many different stories about the "true purpose" behind TISKS. Some believed it was a professional development system without any follow-up professional development. Some believed it was a feedback system with only summative and no formative feedback. Others believed it was a cost-saving measure designed to distribute a fixed pool of salary budget to favorite teachers or teachers with good reputations. Given the specific purpose communicated several times in many forums, confusion remained. Though teachers expressed an understanding of the processes involved in the evaluation system, there was less understanding about its true purpose. In the absence of this clear message and genuine follow-up data to support it, anxiety and suspicion grow, undermining acceptance and motivation. Heneman and Milanowski (2003) recommend that, before any summative use of such a system, "all implementation and formative elements of the system be in place and functioning smoothly" (p.191). The results from this study corroborate this recommendation.

The tie to pay of TISKS, though part of the mandate behind the program, is its weakest element. Teachers believed the system was originally a professional development model that was later co-opted to become a placement tool for the pay scale. Matters of pay are necessarily summative. It is unclear from the results of the present

study if enough direct formative feedback would make a difference in the summative nature of the system reported by many teachers, but it is clear this is something they felt was lacking. This leads to recommendations on matters of implementation.

Implementation

Once a model is agreed upon and an instrument designed, careful implementation is crucial. A solid instrument couched in a clear purpose can still be misunderstood if implementation problems subvert trust and cause teachers to doubt fairness and validity. In this case, teachers' perception of instrumentality will be low. Results from this study show a clear path to several recommendations in the creation of positive instrumentality.

Evaluator preparation. Administrators must be well prepared to take on the extra duties and added responsibilities a system like TISKS entails. Like the implementation of the system itself, such preparation should not be limited to the use of the tool. Administrators at TIS spent many hours in training around the use of the rubrics and having conversations with teachers, but such training clearly did not prepare them to give direct, targeted feedback that teachers found useful. This lack of useful feedback has been noted in other systems (Ha, 2003; Heneman & Milanowski, 2003; Kelley et al., 2002; Milanowski & Heneman, 2001), so it cannot be discounted out of hand as a specific failing of TISKS. Such lack of feedback is detrimental to system success and positive teacher reaction. Targeted feedback is also a useful mechanism in reducing teacher perception of administrators having favorites or being harder on some teachers than others. Just as with students, targeted feedback grounded in a clear set of performance standards is more difficult to put down to subjectivity than vague or non-existent feedback and only final scores. An examination of administrators' perceptions of

their preparation and role in TISKS would be a useful addition to the practical research literature.

Create a complete cycle. Not only did teachers criticize the lack of focused feedback, they also criticized the lack of follow-up. Being intensely observed and evaluated in the profile year, receiving limited or no feedback, and then being largely ignored in the intervening years created a vacuum in teachers' instrumentality perceptions. Many were left with a feeling of something having been done to them with vague or tenuous connections to the rubric and their practice. A complete cycle would include focused, actionable feedback in the profile year, professional development targeted at improvement in identified areas in intervening years, and aligned induction and mentoring structures when new faculty join the school. Teachers pointed out the need for these in this study, and past researchers have made similar recommendations (Heneman & Milanowski, 2003). Such interventions may raise teachers' instrumentality perceptions and improve their motivation and reaction to KSBP systems.

Plan for the extra time and effort. For both teachers and evaluators, standards-based evaluation systems take large amounts of time. Not taking these time commitments fully into account and budgeting the time available will most likely add to stress and feelings of unfulfilled effort on the part of teachers. Furthermore, as teachers at TIS reported, the additional time may be counter-productive to time on task for the improvement of student learning. A lack of planning for this increased workload and burden can lead to burnout and dissatisfaction by all involved, as reported by the majority of teachers who had gone through TISKS. A follow-up study exploring the reactions of administrators to the workload in TISKS would add value here as well.

Rewards

The recommendations above can be applied equally to a system that ends in a placement on the pay scale and one that does not. Teachers voiced strong opinions about not wanting to be motivated by pay or being "forced" to be motivated by money. They were far more willing to be motivated by their own intrinsic desire to improve, and this is not an unreasonable assertion for highly qualified, competitive teachers. However, TISKS is by definition a pay system. The results of this study show that it is therefore crucial to create a system that is not only couched in agreed-upon terms and perceived as fair and fairly administered by professional and competent evaluators, but also one that offers the kinds of rewards self-motivated teachers seek. The rewards, or valence, are most important. Results showed the positive impact on work to be the most predictive factor of teacher reaction. Specifically, teachers want to feel the program shows them how to improve and gives them clear direction for professional development. A fair system, fair and accurate results, and professional evaluators they trust make up the factor of instrumentality that is most predictive of a positive reaction. Though a well-designed system and low stress are also likely to have a positive effect, these were not shown to have a significant impact in this study. If these are well-balanced and fine-tuned for the most positive impact, it may be possible to mitigate against the de-motivating influence of the tie to pay for all but the youngest teachers.

References

Armour-Garb, A. (2009). Should "value-added" models be used to evaluate teachers? *Journal of Policy Analysis and Management*, *28*, 692–693.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2004). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, *16*, 251–261.

Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review*, *20*(1), 51–61. doi:10.1016/S0272-7757(99)00060-6

Ballou, D. (2002). Sizing up test scores. *Education Next*, *2*(2), 10–15.

Ballou, D., & Podgursky, M. J. (1993). Teachers' attitudes toward merit pay: Examining conventional wisdom. *Industrial and Labor Relations Review*, *47*(1), 50–61.

Ballou, D., & Podgursky, M. J. (1997). *Teacher pay and teacher quality*. Kalamazoo, MI: W.E. Upjohn Institute.

Ballou, D., & Podgursky, M. J. (2001). Let the market decide. *Education Next*, *1*(1), 1–7.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, *23*, 251–271.

Coggshall, J. G., Ott, A., Behrstock, E., & Lasagna, M. (2010). *Retaining teacher talent: The view from generation Y*. New York: Public Agenda. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED508142

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Reasearch & Evaluation*, *10*(7). Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.

Darling-Hammond, L., & Berry, B. (1988). *The evolution of teacher policy.* Santa Monica, CA: The RAND Corporation.

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

Dickinson, T. L. (1992). Attitudes about performance appraisal. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 141–161). Hillsdale, NJ: Lawrence Erlbaum.

Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources*, *37*, 913–927.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*, 901–914. doi:10.1016/j.jpubeco.2006.10.001

Firestone, W. A. (1991). Merit pay and job enlargement as reforms: Incentives, implementation, and teacher response. *Educational Evaluation and Policy Analysis*, *13*, 269–288.

Firestone, W. A. (1994). Redesigning teacher salary systems for educational reform. *American Educational Research Journal*, *31*(3), 549–574.

Fryer, R. G. (2011). *Teacher incentives and student achievement: Evidence from New York City public schools* (Working Paper No. 16850). Cambridge, MA: National Bureau of Economic Research.

Glewwe, P., Ilias, N., & Kremer, M. (2003). *Teacher incentives* (Working Paper No. 9671). Cambridge, MA: National Bureau of Economic Research.

Goldhaber, D. D. (2002). The mystery of good teaching. *Education Next*, *2*(1), 50–55.

Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom & C. E. Finn, Jr. (Eds.), *Better Teachers, Better Schools* (pp. 83–102). Washington, D.C.: The Thomas B. Fordham Foundation.

Goldhaber, D. D., DeArmond, M., & DeBurgomaster, S. (2010). *Teacher attitudes about compensation reform: Implications for reform implementation* (Working Paper No. 50). Washington, D.C.: Center for Analysis of Longitudinal Data in Education Research, Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/1001434-teacher-attitudes-reform.pdf

Goldhaber, D. D., DeArmond, M., Player, D., & Choi, H. (2008). Why do so few public school districts use merit pay? *Journal of Education Finance*, *33*, 262–289.

Greenberg, J. (1987). Reactions to procedural injustice in payment distributions: Do the means justify the ends? *Journal of Applied Psychology*, *72*, 55–61.

Ha, B.-W. (2003). *Assessment of teacher reactions to the performance evaluation system and the motivational effects of teachers to a performance-based bonus pay program in Korea* (Doctoral dissertation). The University of Wisconsin - Madison, United States -- Wisconsin. Retrieved from ProQuest Dissertations and Theses. (3101321)

Hanson, W. E., Creswell, J. W., Clark, V. L. P., Petska, K. S., & Creswell, J. D. (2005). Mixed Methods Research Designs in Counseling Psychology. *Journal of Counseling Psychology*, *52*, 224–235. doi:10.1037/0022-0167.52.2.224

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, *113*(485), 64–98.

Hanushek, E. A. (2007). The single salary schedule and other issues of teacher pay. *Peabody Journal of Education*, *82*, 574–586.

Hanushek, E. A., & Rivkin, S. G. (2004). *How to improve the supply of high-quality teachers* (pp. 7–25). Washington, D.C.: Brookings Institution Press.

Hart, A. W. (1987). A career ladder's effect on teacher career and work attitudes. *American Educational Research Journal*, *24*, 479–503.

Hart, A. W. (1994). Work feature values of today's and tomorrow's teachers: Work redesign as an incentive and school improvement policy. *Educational Evaluation and Policy Analysis*, *16*, 458–473.

Hassel, B. C. (2002). *Better pay for better teaching: Making teaching compensation pay off in the age of accountability*. Washington, D.C.: Progressive Policy Institute. Retrieved from

http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED467050

Hedge, J. W., & Teachout, M. S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group & Organization Management*, *25*(1), 22–44.

Heneman, H. G. (1998). Assessment of the motivational reactions of teachers to a school-based performance award program. *Journal of Personnel Evaluation in Education*, *12*, 43–59.

Heneman, H. G., & Milanowski, A. T. (1999). Teachers' attitudes about teacher bonuses under school-based performance award programs. *Journal of Personnel Evaluation in Education*, *12*, 327–341. doi:10.1023/A:1008063827691

Heneman, H. G., & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, *17*, 173–195.

Hudson, S. (2010). *The effects of performance-based teacher pay on student achievement* (SIEPR Discussion Paper No. 09-023). Stanford, CA: Stanford University. Retrieved from http://www.stanford.edu/group/siepr/cgi-bin/siepr/?q=system/files/shared/pubs/papers/09-023_Paper_Hudson.pdf

Ingvarson, L., Kleinhenz, E., & Wilkinson, J. (2007). *Research on performance pay for teachers*. Canberra, Australia: Department of Education, Science and Training.

Johnson, S. M. (1984). Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review*, *54*, 175–85.

Johnson, S. M. (1986). Incentives for teachers: What motivates, what matters. *Educational Administration Quarterly*, *22*(3), 54–79.

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, *85*, 708–723.

Kelley, C. J. (1997). Teacher compensation and organization. *Educational Evaluation and Policy Analysis*, *19*, 15–28.

Kelley, C. J. (1999). The motivational impact of school-based performance awards. *Journal of Personnel Evaluation in Education*, *12*, 309–326.

Kelley, C. J., & Finnigan, K. (2003). The effects of organizational context on teacher expectancy. *Educational Administration Quarterly*, *39*, 603–634.

Kelley, C. J., Heneman, H. G., & Milanowski, A. T. (2002). Teacher motivation and school-based performance awards. *Educational Administration Quarterly*, *38*, 372–401.

Kohn, A. (1993). Why incentive plans cannot work. *Harvard Business Review*, *71*(5), 54–63.

Koppich, J. E., & Rigby, J. (2009). *Alternative teacher compensation: A primer*. Berkeley, CA: Policy Analysis for California Education.

Ladd, H. F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, *18*(1), 1–16. doi:10.1016/S0272-7757(97)00044-7

Landy, F. J., Barnes, J. L., & Murphy, K. R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. *Journal of Applied Psychology*, *63*, 751–754.

Landy, F. J., Barnes-Farrell, J. L., & Cleveland, J. N. (1980). Perceived fairness and accuracy of performance evaluation: A follow-up. *Journal of Applied Psychology*, *65*, 355–356.

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *The Journal of Political Economy*, *110*, 1286–1317.

Lavy, V. (2004). *Performance pay and teachers' effort, productivity and grading ethics* (Working Paper No. 10622). Cambridge, MA: National Bureau of Economic Research.

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

Lawler, E. E. (1967). The multi-trait multi-rater approach to measuring managerial job performance. *Journal of Applied Psychology*, *51*, 369–381.

Lawler, E. E. (1995). The new pay: A strategic approach. *Compensation & Benefits Review*, *27*(4), 14–22.

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

Lazear, E. P. (2001). *Paying teachers for performance: Incentives and selection*. Unpublished Paper, Hoover Institution and Graduate School of Business, Stanford University, Palo Alto, CA.

Mancuso, S. V., Roberts, L., & White, G. P. (2010). Teacher retention in international schools: The key role of school leadership. *Journal of Research in International Education*, *9*(3), 306–323.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J. E., … Peng, A. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG1114

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publications, Inc.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67–101.

Milanowski, A. T. (1999). Measurement error or meaningful change? The consistency of school achievement in two school-based performance award programs. *Journal of Personnel Evaluation in Education*, *12*, 343–363. doi:10.1023/A:1008015911761

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

Milanowski, A. T., & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, *15*, 193–212.

Mohrman, A. M., Mohrman, S. A., & Odden, A. (1996). Aligning teacher compensation with systemic school reform: Skill-based pay and group-based performance rewards. *Educational Evaluation and Policy Analysis*, *18*, 51–71.

Muralidharan, K., & Sundararaman, V. (2009). *Teacher performance pay: Experimental evidence from India* (Working Paper No. 15323). Cambridge, MA: National

Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15323

Murnane, R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review*, *56*(1), 1–17.

Myers, R. (1990). *Classical and modern regression with applications* (2nd ed.). Boston: Duxbury.

Odden, A., & Kelley, C. J. (2002). *Paying teachers for what they know and can do* (2nd Edition.). Thousand Oaks, CA: Corwin Press.

Odland, G., & Ruzicka, M. (2009). An investigation into teacher turnover in international schools. *Journal of Research in International Education*, *8*(1), 5–29.

Organisation for Economic Co-operation and Development. (2009). *Education at a glance 2009: OECD indicators*. Paris: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/dataoecd/41/25/43636332.pdf

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Pechthalt, J. (2007). No merit to merit pay. *United Teacher*, *37*(3), 11.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, *26*, 909–950.

Protsik, J. (1996). History of teacher pay and incentive reforms. *Journal of School Leadership*, *6*(2), 265–289.

Ramirez, A. (2001). How merit pay undermines education. *Educational Leadership*, *58*(5), 16–20.

Reichardt, R., & Van Buhler, R. (2003). *Recruiting and retaining teachers with alternative pay*. Aurora, CO: Mid-Continent Research for Education and Learning.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sclafani, S., & Tucker, M. S. (2006). *Teacher and principal compensation: An international review*. Washington, DC: Center for American Progress.

Solmon, L. C. (2006). The case for merit pay. In P. Peterson (Ed.), *Choice and competition in American education* (pp. 102–110). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Solmon, L. C., & Podgursky, M. J. (2001). *The pros and cons of performance-based compensation*. Santa Monica, CA: Milken Family Foundation.

Springer, M. G. (2009). *Performance incentives: Their growing impact on American K-12 education*. Washington, DC: Brookings Institution Press.

Springer, M. G., Hamilton, L., McCaffrey, D. F., Ballou, D., Le, V. N., Pepper, M., Lockwood, J. R., Stecher, B. M. (2010). Teacher pay for performance: Experimental evidence from the project on incentives in teaching. *National Center on Performance Incentives*, 86.

Stecher, B. M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment*, *7*, 259–281.

Taylor, L. L., & Springer, M. G. (2009). *Optimal incentives for public sector workers: The case of teacher-designed incentive pay in Texas* (Working Paper). Nashville, TN: National Center on Performance Incentives.

Trevor, C. O. (2010). *What teachers want: Teacher preferences regarding nontraditional pay approaches* (Briefing Paper No. 266). Washington, D.C.: Economic Policy Institute. Retrieved from http://www.epi.org/page/-/pdf/bp266.pdf?nocdn=1

U.S. Department of Education. (1983). *A Nation At Risk: The Imperative for Educational Reform*. Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html

U.S. Department of Education. (2009). *Race to the Top program executive summary*. Washington, D.C. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Woessmann, L. (2010). *Cross-country evidence on teacher performance pay* (Working Paper No. 3151). Munich: Ifo Institute for Economic Research at the University of Munich.

APPENDIX A – The TISKS System

In 2005, fuelled by a mandate from the Board of Managers, the school administration of The International School (TIS) began researching and developing a pay for performance program in order to pay teachers in a different way to traditional methods. The rationale behind creating a new system included several goals: (a) to pay teachers according to their level of demonstrated knowledge and skill, (b) to attract and retain the best teachers possible from the international school teaching pool, and (c) to motivate teachers to continuous improvement in professional practice.

Knowledge and Skills

The initial rationale behind the system was that it should provide pay for teachers according to their demonstrated knowledge and skill, and not solely based on their years of experience or degrees earned. At the time, the school used a traditional step-based pay scale. Teachers were placed on the pay scale according to the number of years they had taught and any additional degrees earned at the time of their hire. They then moved up the scale one step per year. A step was roughly a two percent increase on the base salary from the previous step. Prior to 2004, administrators determined teachers' original placement using all of a teachers' previous years of experience. After 2004, placement was based on up to nine years of experience only. Consequently, some teachers felt they were relegated to waiting for the traditional system to raise their pay to the level of colleagues doing the same job who had arrived at the school prior to the "nine-year cap." Members of the Board of Managers and the senior administration believed a system that paid teachers based on their performance could level the pay structure so teachers who were equally proficient would be paid equally, regardless of experience, number of degrees, or the time at which they were hired.

Attract and Retain the Best

Another stated motivator for investigating new ways to pay teachers was to ensure pay levels that would attract and retain high quality teachers. The world of international schools is a fluid one; teachers move regularly from country to country and school to school. They may move because they are searching for adventure in a new country, or because they are looking for a more satisfying experience in a different school—personally, professionally, or financially. The new pay structure was intended to ensure teachers would be paid well enough that they would be less likely to leave the school solely for financial reasons.

Professional Development

To account for professional reasons for leaving, the design team also planned the system to be grounded in professional development and to aid in collaborative effort toward school improvement. Designers reasoned that, if teachers felt they were being rated according to accurate, understandable criteria that were recognized as best practices in the profession, they would have professional satisfaction from going through the rating process and a clear way to discover areas in which they could improve.

The system is undergirded by the framework of teacher standards created by Charlotte Danielson (1996), which had been in use as part of the school's performance appraisal program since 2004. Ms. Danielson consulted on site at the school when TIS adopted and adapted her framework for that earlier program. Through collaborative administrator and teacher effort, the system evolved into a standards-based knowledge- and skills-based system called TISKS, designed to contribute to teacher professional development and to give teachers an avenue toward higher job satisfaction and reward

without leaving the classroom. Based heavily on the work of the Consortium for Policy Research in Education, the program became the first of its kind in the school's regional association.

Administrator Training

Because the program relies on the professional evaluation of teachers by administrators, the school prioritized training for the administrative panel members. Division principals, associate principals, and school-wide administrators clocked many hours of training since program inception. Such training included how to document evidence on the common rubric, how to conduct learning-focused conversations with teachers, and how to write final reflections and feedback to share with teachers. Furthermore, all administrators were engaged in many hours of cross-divisional calibration of observation technique and collecting evidence according to the rubric. Administrators from the primary schools observing and commenting on a video of a teachers' lesson together with administrators from the middle school is one example of how the consistency of system execution was addressed.

Implementing the System

A pilot of the program was conducted in the 2009-10 school year, with 28 teachers taking part. These 28 pilot cohort teachers were asked to share their experiences to inform modifications to the system. Based on feedback from teachers and administrators, changes were made to the number of requirements between the pilot year and the first full year of implementation. The system saw its first full year of implementation in the 2010-11 school year with 47 teachers. During the 2011-12 school

year, 55 teachers were taking part. The remaining eligible teachers were going through the profile year in the 2012-13 school year while this study was being conducted.

Components and Process for the Teacher Profile

TISKS is a knowledge- and skills-based pay system. The system is designed to place teachers on a pay scale according to how they demonstrate the knowledge and skills of outstanding teachers identified on a rubric, which is a distillation of knowledge and skills from the work of Danielson (1996), with modifications specific to TIS's international, Christian context. Teachers demonstrate and are rated on these knowledge and skills areas in two ways: through (a) regular classroom observations and (b) a teacher-created portfolio featuring reflective writing and samples of student and teacher-created work. The observers and evaluators of the portfolio are a panel of one teacher-nominated faculty peer and two administrators, one of whom is the teacher's supervisor and the other an administrator with Pre-K through grade 12 responsibility.

The system was being phased in during the time of this study, with the plan for all teachers having gone through some version by the end of the 2012-13 school year. An exception was any teacher who held certification through the National Board for Professional Teaching Standards. These teachers were exempt for five years. The intended cycle for the system included a profile year during which teachers take part in the evaluation process explained below. Once teachers go through the process, the following two years are spent in self-directed professional development, ideally addressing any areas indicated for growth in the evaluation process. After these two years, the teacher again enters a profile year, and the cycle repeats, though the exact shape of the second phase was not determined at the time of the study.

The knowledge and skills rubric. The system's assessment tool is a rubric of knowledge and skills designed to describe an effective teacher. Separate rubrics were developed for other specialists, including counselors, technology facilitators, and teacher librarians. The rubric contains descriptors on six domains:

- Knowledge of Content, Pedagogy & Students
- Instructional Strategies: Delivery Of Content And Engagement Of Students
- Assessment: Assessment Types, Feedback And Use Of Results
- Management
- Communication
- Collaboration And Relationships

The rubric contains descriptors for each of these domains divided into 4 categories. Teachers are assessed and scored 1 through 4 on each category.

Observations. Each member of the panel observes the teacher several times during the course of the academic year. Panel members conduct types of observation: (a) formal, (b) video, (c) informal, and (d) walkthrough. Administrators give teachers feedback on the observations through the use of a common template laying out the characteristiTISKS of a teacher at all four levels on each component of the TISKS rubric. In addition to the written feedback, each observation is followed by a debrief session in which the administrator(s) facilitate a learning-focused conversation conducted according to a cognitive coaching model and designed to help the teacher reflect on the lesson observed. All administrators are regularly trained in the facilitation of such conversations.

***Formal observation*.** Teachers arrange a single formal observation jointly with their supervisor and the school-wide administrator. The teacher knows when the observation will be taking place and must meet to discuss the plan and intended outcomes with the administrators before the observation. The discussion is framed by a standard set of reflective questions provided to the teacher. The observation must include a complete

lesson with a clear beginning, middle, and end. After the observation, the administrators use the established template to provide written feedback and also meet together with the teacher to discuss his or her own reflections on the lesson. All written artifacts are retained for inclusion in the teacher's portfolio.

Video observation. Like the formal observation, teachers and administrators schedule the video observation. The lesson should be complete, with a beginning, middle, and end. One or both of the administrators, or an optional third person takes video of the entire lesson. Subsequently, the teacher watches the entire, unedited video and reflects on the lesson with a set of reflective questions provided to him or her. At a follow-up meeting, the teacher and administrators discuss their reflections and the administrators, after watching the video, provide written feedback using the common template. All written artifacts and the unedited video are retained for inclusion in the teacher's portfolio.

Informal observations. Administrators conduct informal observations unannounced. They may do so jointly or individually, but each administrator must complete at least three informal observations. The teacher peer must complete two and is not required to supply written feedback. Administrators must observe at least 30 minutes of any given lesson, but it is recommended they stay for the entire lesson. The administrators provide written feedback using the common template, and meet with the teacher for a learning-focused conversation around the observation.

Walkthroughs. Each administrator should conduct three five- to fifteen-minute unannounced walkthroughs in the teacher's classroom during the TISKS profile year.

Walkthroughs are intended to provide administrators with a general feel for the classroom atmosphere aside from observation lessons. No written feedback is necessary.

Portfolio. In addition to the observations, teachers must prepare an electronic portfolio of work including, but not limited to the written feedback items arising from the observations. The additional items are written reflections on each of the domains in the teacher rubric and teacher-collected and arranged artifacts giving evidence of their work in each of the areas. The evidence includes examples of:

- student work
- activities created by the teacher
- instructional units written by the teacher
- feedback to students on their work
- communication with parents, students, colleagues
- collaboration with colleagues at TIS or other schools

The final piece is a written statement of the teacher's personal philosophy of education. All items are collected and presented for review in a Web-based digital portfolio format created and stored at the school.

Panel Interview. In April or May of the profile year, the teacher, administrators, and the teacher peer convene a one-hour panel interview. The focus of the discussion is comprehensive, including the portfolio and the observations. It is an opportunity for the panel members to ask any clarifying questions and for the teacher to explain anything he or she thinks helps the panel to understand him or her as a teacher.

Results. After the panel discussion, the panel members and teacher have two weeks to determine their scores (1-4) for the teacher on each aspect of the rubric. They enter these scores anonymously into an electronic database. The final scores for the teacher are computed based on a formula giving weight to the scores of the panel members:

- Teacher her- or himself: 25%
- Supervisor: 40%
- Other Administrator: 20%
- Peer: 15%

Once the scores are entered and computed, the teacher's profile is established. The overall profile, reflecting the teacher's scores on each domain, determines his or her position on the pay scale for the next three years, until the conclusion of the next profile year. It is expected that a skilled teacher will have a mixture of twos and threes in the profile. For many teachers, a mix of twos and threes in the profile can result in a sizable increase in pay; however, it is conceivable that a teacher's salary could drop from the current pay level. Such would be the case should a teacher who has risen to the top of the traditional pay scale receive a profile from her panel that corresponded to a lower salary than her present one. To guard against this during phasing in, the school guaranteed in the first round that a teacher's salary would at the very least remain the same. In subsequent rounds, this restriction will be removed, and salaries will go down should the panel ratings warrant.

An important aspect of the system is the commitment of teachers beyond their profile year. Not only are they expected to be involved in self-directed professional development, but they must also commit to completing their two-year contracts with the new pay level. Teachers may take part in the system mid-contract and be paid the difference in salary retroactively for the entire year in one lump sum should the profile result in a pay raise. If the teacher subsequently decides not to complete his or her contract, the lump sum payment is returned to the school.

APPENDIX B – Survey Including Teacher Reaction Scales

Information and Instructions: Assessment of Teachers' Reactions to a Knowledge- and Skills-Based Pay Structure at an International School

Dear Colleague:

You are invited to take part in a research study of teachers' reactions to TISKS. You were selected as a participant because you are a teacher at The International School, whether or not you have completed TISKS. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Joel Courtney Lowe, under the direction of Dr. Jill Sperandio, College of Education, Lehigh University. Mr. Lowe is the current Director of Curriculum and Assessment at The International School.

Procedures
If you agree to be in this study, we would ask you to complete this survey, which should take approximately 10 to 15 minutes.

Risks and Benefits of being in the study
All responses to this survey are anonymous. There is no way for anyone to connect you with your responses, and there is no risk to you in participation.

The benefit to participation is a contribution to an understanding of the effect TISKS has on teachers and an exploration of the factors underlying teacher reactions. The school will be made aware of anonymous results in aggregate, and this may lead to improvements in the process with regard for teacher input.

Confidentiality
The records of this study will not include any identity information about you. We will not include any information that will make it possible to identify a subject in any published reports. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study
Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the school or the researchers.

Should you have any questions or concerns, please contact either Courtney Lowe at jcl205@lehigh.edu or Dr. Jill Sperandio at jis204@lehigh.edu. If you would like to talk to someone other than the researcher(s), you are encouraged to contact Susan E. Disidore at (610)758-3020 (email: sus5@lehigh.edu) or Troy Boni at (610)758-2985 (email: tdb308@lehigh.edu) of Lehigh University's Office of Research and Sponsored Programs. All reports or correspondence will be kept confidential.

In order to participate, you will need to indicate on the next page your informed consent. Please click on the Start Survey button below and answer the informed consent questions to begin.

Page 1 - Question 1 - Choice - One Answer (Bullets)

I have read the instructions and understand the study.

- ○ Yes, I have read the instructions and I understand the study.
- ○ No, I do not understand and wish to exit the study. [Skip to End]

Page 1 - Question 2 - Choice - One Answer (Bullets)

Informed consent to participate in the study.

- ○ Yes, I give my informed consent to participate in this study.
- ○ No, I do not give my informed consent to participate in this study. [Skip to End]

Page 2 - Heading

Demographic Information

Page 2 - Question 3 - Choice - One Answer (Bullets)

Please select your Gender

- ○ Male
- ○ Female

Page 2 - Question 4 - Choice - One Answer (Bullets)

What is your age in years?

- ○ 21-30
- ○ 31-40
- ○ 41-50
- ○ 51-60
- ○ 61 and over

Page 2 - Question 5 - Choice - One Answer (Bullets)

How many years have you been teaching, including this year?

- ○ 1-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21 or more

Page 2 - Question 6 - Choice - One Answer (Bullets)

What is the highest level of formal education you have completed?

- ○ Bachelor's
- ○ Some graduate but no master's
- ○ Master's
- ○ More than master's but no doctorate
- ○ Doctorate

Page 3 - Question 7 - Choice - One Answer (Bullets)

Have you completed TISKS?

- ○ No [Survey Ends for these respondents after Question 10]
- ○ Yes

Page 3 - Heading

For the next two items, please select the number that best reflects your agreement with the statement.

Page 3 - Question 8 - Rating Scale - Matrix

TISKS is a motivator for me to improve my professional practice (whether or not I have completed the process).

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 3 - Question 9 - Open Ended - Comments Box

Feel free to share any additional information about the effect of TISKS on motivation to improve your professional practice.

Page 3 - Question 10 - Choice - One Answer (Bullets)

Do you plan to return to TIS when your current contract ends?

- ○ No
- ○ Unsure
- ○ Yes

Page 6 - Heading

Please select the number that best reflects your agreement with each statement below about TISKS.

Page 6 - Question 11 - Rating Scale - Matrix

I had a good understanding of the rubric and procedures involved in TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 12 - Rating Scale - Matrix

I knew what I needed to do in order to do well in TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 13 - Rating Scale - Matrix

Overall, I have a good understanding of TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 14 - Rating Scale - Matrix

The rubric does a good job of defining good teaching and work.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 15 - Rating Scale - Matrix

The rubric describes the kind of teaching that teachers in our school should strive for.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 16 - Rating Scale - Matrix

Working toward improving performance on the rubric will really help teachers develop as professionals.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 6 - Question 17 - Open Ended - Comments Box

Feel free to include any additional information about any of your answers above.

Page 7 - Heading

Fair Process and Fair Results

Page 7 - Heading

Please select the number that best reflects your agreement with each statement below about TISKS.

Page 7 - Question 18 - Rating Scale - Matrix

I agree with my final TISKS scores.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 19 - Rating Scale - Matrix

The TISKS scores I received were fair.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 20 - Rating Scale - Matrix

I got the TISKS scores I deserved.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 21 - Rating Scale - Matrix

Overall, the way the TISKS process was run was fair.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 22 - Rating Scale - Matrix

My TISKS evaluation was conducted in a fair way.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 23 - Rating Scale - Matrix

The processes and procedures of TISKS were fair.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 7 - Question 24 - Open Ended - Comments Box

Feel free to include any additional information about any of your answers above.

Page 8 - Heading

Impact on Teaching and Work, Effort and Stress

Page 8 - Heading

Please select the number that best reflects your agreement with each statement below about TISKS.

Page 8 - Question 25 - Rating Scale - Matrix

I have a clearer idea of what the school expects because of TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 26 - Rating Scale - Matrix

TISKS helped me learn how I can improve my teaching.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 27 - Rating Scale - Matrix

I learned a lot from TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 28 - Rating Scale - Matrix

The TISKS process helped me develop as a professional.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 29 - Rating Scale - Matrix

TISKS took more effort than the results were worth.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 30 - Rating Scale - Matrix

TISKS caused me a lot of stress.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 8 - Question 31 - Open Ended - Comments Box

Feel free to include any additional information about any of your answers above.

Page 9 - Heading

Accuracy, Role of Panel

Page 9 - Heading

Please select the number that best reflects your agreement with each statement below about TISKS.

Page 9 - Question 32 - Rating Scale - Matrix

The TISKS scores I received were an accurate evaluation of my performance in the six domains.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 33 - Rating Scale - Matrix

My panel’s evaluations reflected my true performance.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 34 - Rating Scale - Matrix

My panel’s evaluation of my work matched my own evaluation.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 35 - Rating Scale - Matrix

I feel quite satisfied with the discussions of my performance I had with my panel members.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 36 - Rating Scale - Matrix

I am satisfied with the feedback I received from my panel members.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 37 - Rating Scale - Matrix

Overall, I am satisfied with my interactions with my panel members.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 9 - Question 38 - Open Ended - Comments Box

Feel free to include any additional information about any of your answers above.

Page 10 - Heading

Overall Favorableness

Page 10 - Heading

Please select the number that best reflects your agreement with each statement below about TISKS.

Page 10 - Question 39 - Rating Scale - Matrix

In general, I feel that the school has an excellent performance evaluation system in TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 10 - Question 40 - Rating Scale - Matrix

Overall, I am satisfied with TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 10 - Question 41 - Rating Scale - Matrix

The school should continue to use TISKS.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
m 1	m 2	m 3	m 4	m 5

Page 10 - Question 42 - Open Ended - Comments Box

Feel free to include any additional information about any of your answers above.

Thank You Page

Thank you for your participation in this study.

APPENDIX C – Focus Group Interview

Questions for TISKS Teachers

[Distribute and collect signature on IRB consent form before subjects enter room]

Introduction

We are conducting an in-depth analysis of teachers' reactions to TISKS. As part of our evaluation, we are talking to teachers involved in the first three cohorts. This study is being conducted by Joel Courtney Lowe under the direction of Dr. Jill Sperandio in the College of Education, Lehigh University. Mr. Lowe is the current Director of Curriculum and Assessment at The International School.

As focus group interview convener, I have no connection to any student, teacher, or employee of the school. I have been engaged by Mr. Lowe, not the school, for the sole purpose of facilitating this focus group interview.

I would first like to test the recording equipment to make sure I can record everyone clearly. When I give the direction, can each participant please say, clearly and comfortably "Mary had a little lamb?" [Record participants and play back to make sure everyone can be heard. Adjust microphone or seating as necessary.]

Questions

Now let us begin.

1. What is your feeling about how much TISKS motivates teachers at TIS to improve in their teaching and work? Is the system worth the effort? (Allow up to 20 minutes)

2. To what extent do you feel teachers have the information and support they need to be successful in TISKS? (Allow up to 20 minutes)

3. Overall, how fair do you think the TISKS process is? (Allow up to 20 minutes)

LEHIGH UNIVERSITY

CONSENT FORM

Assessment of Teachers' Reactions to a Knowledge- and Skills-Based Pay Structure at an International School

You are invited to take part in a research study of teachers' reactions to TISKS. You were selected as a participant because you are a teacher at The International School who has completed the TISKS process. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Joel Courtney Lowe, under the direction of Dr. Jill Sperandio, College of Education, Lehigh University. Mr. Lowe is the current Director of Curriculum and Assessment at The International School.

Procedures
If you agree to be in this study, we would ask you to take part in a focus group interview with 5 other teachers which should last no more than an hour. During the discussion, several questions will be posed about TISKS, and we will ask you to speak freely based on your experience.

The focus group interview will be conducted by an outside facilitator with no affiliation to any student, parent, or employee of the school, past or present. The facilitator has undertaken US National Institutes of Health training in protecting participants in research. Furthermore, the facilitator fully understands the purpose of the research, including its sensitive topic, and has been trained in gathering the most relevant and complete information while safeguarding the confidentiality of participants.

Risks and Benefits of being in the study
If you choose to participate, all of your responses will be kept confidential. No one outside of the interview room will have access to your responses in any form that can in any way be connected to you personally. Interview participation poses no risk to you or your relationship with the school. The session will be recorded in audio only. The recording from this focus group will be sent to a third-party transcriber in another country not related to the school or the researchers in any way. The transcriber will transcribe all conversation into text only, removing any accidental mention of names, grade levels, or any other identifying details that may occur. Only then will the answers be shared completely anonymously and in text format with the researcher. The digital recording will be deleted immediately upon transcription.

The benefit to participation is a contribution to an understanding of the effect TISKS has on teachers and an exploration of the factors underlying teacher reactions. The school will be made aware of anonymous results in aggregate, and this may lead to improvements in the process with regard for teacher input.

Confidentiality
The records of this study will not include any identity information about you. We will not include any information that will make it possible to identify a subject in any published reports. Research records will be stored securely and only researchers will have access to the records. This consent form will be kept securely by the principal investigator at Lehigh University, and the researcher will not have access to the identity information included here.

Voluntary Nature of the Study
Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the school or the researchers.

Should you have any questions or concerns, please contact either Courtney Lowe at jcl205@lehigh.edu or Dr. Jill Sperandio at jis204@lehigh.edu. If you would like to talk to someone other than the researcher(s), you are encouraged to contact Susan E. Disidore at (610)758-3020 (email: sus5@lehigh.edu) or Troy Boni at (610)758-2985 (email: tdb308@lehigh.edu) of Lehigh University's Office of Research and Sponsored Programs. All reports or correspondence will be kept confidential.

You will be given a copy of this information to keep for your records.

Statement of Consent

I have read the above information. I have had the opportunity to ask questions and have my questions answered. I consent to participate in the study.

Signature: _____ Date: _____

December 5 Transcript

INTERVIEW Q1

INTERVIEW FACILITATOR

Umm...let us begin. The first question that "Name DELETED" is interested in finding more information out is basically what's you're feeling about how much the career structure motivates teachers at HKS to improve in their teaching and work? Is this system really worth the effort?			
Umm...I would go out and say I don't know how much it is as far as improving your work as a teacher, but I think a lot of people...umm...myself included, probably thought well these are the hoops I need to jump through to prove that I'm good or this good or that good or you know whatever but umm...I don't	VALENCE +/-	IMPACT ON WORK +/-	
Not being a part of the school, what are those hoops?			
I mean I was in one of the, I was in the pilot group and we had, *unintelligible*, it's vastly different than it is now. The amount of work, it's still a good amount of work but I mean I'd say probably all of us spend 200 plus hours, wrote 90 pages, no joke, I mean but it was very, I mean it ruined that whole...and we didn't even get our, our documentation what to do until December, it was the day after, Monday after Thanksgiving so we only had four months to do it. It was very, very different than it is now, so...that was just like slogging through...	VALENCE - EXPECTANCY -	EFFORT VS. REWARD - UNDERSTANDING OF SYSTEM -	TOO MUCH TIME -
Was this 90 pages of information, was it a self-report format? Was it reflection of yourself things, reflections of your building or colleagues, like...what would make it...how would that contribute to your teaching?			
hmm...			
There were specifications...	EXPECTANCY +	UNDERSTANDING OF SYSTEM +	
Okay...			
For each letter that you had to respond to. And in terms of written document, video, that sort of thing. Umm...and anything that you were constructing, there's a template to follow.			
uh hmm...			
So the amount that we had to do turned into that amount of work. Cause it really was like writing up a unit, lesson plans, the lesson plans for each day, had to be in that format and it's very different now, so...going through it again next year...like...I will say like the one area I have to redo one area but I have to choose two...and the one area I'd be most interested doing is my second one, is the one I won't do cause it's the one with the most work. It'd be assessment which I'm most interested in but I don't think that's going to move me to become a four in LD SCORE	① VALENCE - EXPECTANCY - ② VALENCE -	EFFORT VS. REWARD - ACCEPTANCE OF STANDARDS - EFFORT VS. REWARD -	

APPENDIX E - VITA

Joel Courtney Lowe
207 S. Chickadee Trail
Michiana Shores, IN 46360 USA

Education

Doctor of Education
Educational Leadership — May, 2013
Lehigh University, Pennsylvania

Master of Science
Library and Information Science — May 2000
University of Illinois at Urbana-Champaign

Master of Arts
Teaching English as a Second Language — May 1993
Northern Arizona University

Bachelor of Arts
English — May 1991
Northern Arizona University

Professional Experience

Director of Curriculum and Assessment
School Name Omitted[1] — 2009-Present

Associate Principal, Academics, High School
School Name Omitted[1] — 2005-2009

Head Librarian, High School
School Name Omitted[1] — 2002-2005

English Department Head, Head Librarian
Canadian Academy, Kobe, Japan — 1996-2002

Instructor of English as a Foreign Language
Kwansei Gakuin University, Nishinomiya, Japan — 1994-1996

USIA English Teaching Fellow
Kyrgyz-American Faculty, Kyrgyz State National University
Bishkek, Kyrgyzstan — 1993-1994

[1]*School name omitted to protect confidentiality of research site*