Exploring the Effect of Supportive Teacher Evaluation Experiences on U.S. Teachers' Job Satisfaction

Timothy G. Ford
Angela Urick
&
Alison S. P. Wilson
University of Oklahoma
United States

Abstract: Teacher satisfaction is a key affective reaction to working conditions and an important predictor of teacher attrition. Teacher evaluation as a tool for measuring teacher quality has been one source of teacher stress in recent years in the United States. There is a growing body of evidence on how to evaluate teachers in ways which support their growth and development as practitioners. For this study, we inquired: What is the relationship between supportive teacher evaluation experiences and U.S. teachers' overall job satisfaction? To answer this question, we employed a multilevel regression analysis to multiply-imputed data on U.S. lower-secondary teachers' experiences from the 2013 Teaching and Learning International Survey (TALIS). We found a small, positive relationship between the perceptions of supportive teacher evaluation experiences and U.S. secondary teachers' satisfaction after controlling for other important teacher and school characteristics and working conditions. Further, teachers who felt their evaluation

led to positive changes in their practice had higher satisfaction. Teachers whose primary evaluator was a fellow teacher as opposed to the principal also had higher satisfaction on average. We discuss the implications of these findings for school leaders as well as future teacher evaluation policy.
Keywords: teacher evaluation; job satisfaction; teacher feedback; teacher motivation; multilevel modeling; TALIS 2013

Explorando el efecto de las experiencias de evaluación docente en la satisfacción laboral de los docentes de EE.UU.
Resumen: La satisfacción del maestro es una reacción afectiva clave a las condiciones de trabajo y un importante predictor del desgaste del docente. La evaluación de los maestros como una herramienta para medir la calidad de los docentes ha sido una de las causas del estrés docente en los últimos años en los Estados Unidos. Existe un cuerpo creciente de evidencia sobre cómo evaluar a los maestros de maneras que apoyen su crecimiento y desarrollo como profesionales. Para este estudio, preguntamos: ¿Cuál es la relación entre las experiencias de evaluación docente y la satisfacción laboral general de los docentes de EE. UU.? Para responder a esta pregunta, empleamos un análisis de regresión multinivel para datos de imputación múltiple en las experiencias de docentes de secundaria de EE. UU. De la Encuesta Internacional de Enseñanza y Aprendizaje 2013 (TALIS). Encontramos una relación pequeña y positiva entre las percepciones de las experiencias de evaluación docente de apoyo y la satisfacción de los docentes de secundaria de los EE. UU. Después de controlar otras características importantes del docente y la escuela y las condiciones de trabajo. Además, los maestros que sintieron que su evaluación llevó a cambios positivos en su práctica tuvieron una mayor satisfacción. Los docentes cuyo evaluador principal fue un compañero profesor en comparación con el director también tuvieron una mayor satisfacción en promedio. Discutimos las implicaciones de estos hallazgos para los líderes escolares, así como la política futura de evaluación docente.
Palabras-clave: evaluación docente; Satisfacción laboral; motivación del maestro; modelado multinivel; TALIS 2013

Explorando o efeito das experiências de avaliação de professores na satisfação profissional de professores americanos
Resumo: A satisfação do professor é uma reação afetiva fundamental às condições de trabalho e um importante preditor do desgaste do professor. A avaliação dos professores como uma ferramenta para medir a qualidade dos professores tem sido uma das causas do estresse do professor nos últimos anos nos Estados Unidos. Há um crescente corpo de evidências sobre como avaliar os professores de forma a apoiar seu crescimento e desenvolvimento como profissionais. Para este estudo, perguntamos: Qual é a relação entre as experiências de avaliação do apoio ao professor e a satisfação geral do trabalho dos professores de EE. UU? Para responder a essa questão, utilizamos uma análise de regressão multinível para dados de múltiplas imputações nas experiências de professores do ensino médio em EE. UU Do Inquérito Internacional de Ensino e Aprendizagem 2013 (TALIS). Encontramos uma relação pequena e positiva entre as percepções das experiências de avaliação do apoio ao professor e a satisfação dos professores do ensino médio nos EUA. UU Depois de controlar outras características importantes do professor e da escola e condições de trabalho. Além disso, os professores que sentiram que sua avaliação levou a mudanças positivas em sua prática tiveram maior satisfação. Os

professores cujo principal avaliador era um colega professor em comparação com o diretor também tiveram uma satisfação maior em média. Discutimos as implicações dessas descobertas para os líderes das escolas, bem como a futura política de avaliação de professores.
Palavras-chave: avaliação de professores; Satisfação no trabalho; motivação do professor; modelagem multinível; TALIS 2013

Introduction

Teacher satisfaction is an important affective reaction to school working conditions, and has been found to mediate the relationship of working conditions to teacher attrition (Cha & Cohen-Vogel, 2011; Skaalvik & Skaalvik, 2009, 2010, 2011). The most recent *MetLife Survey of the American Teacher* (2013) reports that the percentage of U.S. teachers who report being "very satisfied" in their jobs has dropped 23 percentage points since 2008 to a low of 39%. At the same time, stress among teachers has increased, with over half of teachers reporting that they experience great stress every day or several times a week. This represents an increase of 15 percentage points since 1985—the last time the question was asked. While the causes for these sharp changes in teacher well-being are not localized to one source, the increased demands placed on U.S. teachers due to accountability have likely exacerbated the problem (Hargreaves, 2010). Increased pressure to conform to outside expectations for their work has made it difficult for teachers to pursue the psychic rewards that attracted them to teaching in the first place (Ford, Van Sickle, Clark, Fazio-Brunson, & Schween, 2017; Ford, Van Sickle, & Fazio-Brunson, 2016; Hargreaves, 2003, 2010; Ingersoll, 2003). Furthermore, lack of support to match increased pressure is damaging to teachers' perceptions of self-efficacy and autonomy (Lavigne, 2014)—key sources of teacher satisfaction and intrinsic motivation for improvement (Firestone, 2014; Ingersoll, 2003; Lortie, 1975; Niemiec & Ryan, 2009; Skaalvik & Skaalvik, 2011, 2014; Sylvia & Hutchinson, 1985).

In addition to the effects on teachers themselves, there are several reciprocal consequences for the broader school community whose teachers experience disproportionate job dissatisfaction. Teachers with psychosocial issues stemming from dissatisfaction can negatively affect school climate through strained interactions with coworkers and students (Grayson & Alvarez, 2008; Kokkinos, Panayiotou, & Davazoglou, 2005). Additionally, schools pay the price through increased teacher absences, mental and medical healthcare costs, and compromised teacher performance (Grayson & Alvarez, 2008; Leithwood, Menzies, Jantzi, & Leithwood, 1999).

In addition to salary, working conditions such as stress and lack of collegiality, professional discretion, and administrative support remain key predictors of teachers' decisions to leave the profession (Borman & Dowling, 2008; Pearson & Moomaw, 2005; Shen, Leslie, Spybrook, & Ma, 2012; Skaalvik & Skaalvik, 2009, 2011; Urick, 2016). Approximately 16% of U.S. teachers decided to change schools or leave the profession in 2012 according to the recent Teacher Follow-up Survey collected by the National Center for Education Statistics (Goldring, Taie, & Riddles, 2014). More than two-thirds of these teachers left voluntarily (Goldring et al., 2014), and may not have been replaced by a qualified teacher, which has contributed to growing teacher shortages (see Carver-Thomas & Darling-Hammond, 2017). Whether teachers ultimately decide to leave or not, negative teacher affective outcomes, such as job dissatisfaction, influence overall school working conditions, organizational capacity and, in turn, student outcomes (Cha & Cohen-Vogel, 2011; Evans, 2001).

The Landscape of "New" Teacher Evaluation in the United States

One potential source of teacher support or stress in recent years—teacher evaluation—has received increasing attention in the "new era of accountability" (Murphy, Hallinger, & Heck, 2013, p. 349). This new era, ushered in with the authorization of the No Child Left Behind Act (NCLB) in 2001, marked a shift away from social democratic education policy making towards neoliberal policy solutions to educational problems (Burch, 2009; Hursh, 2007). Neoliberalism emphasizes the efficiencies brought about by market competition, deregulation, and a more explicit focus on the measurement and tracking of performance outcomes for the purposes of incentivizing improvement (i.e., performance management). With Race to the Top (RttT) in 2009, the focus of NCLB largely shifted to the problem of "teacher quality," and, in doing so, teacher evaluation again gained prominence as a proposed lever for school improvement (Lavigne & Good, 2015). Through funding attached to the American Recovery and Reinvestment Act, states were incentivized to implement annual performance-based evaluation systems for teachers which included common standards and assessments, feedback systems, and some measure of effectiveness based on student test scores (including value-added measures [VAMs]; Race to the Top Fund, 2016).

The thrust of current educational accountability policy, including the newest iteration of teacher evaluation, falls within Schneider and Ingram's (1990) classification of authority/incentive policy tools. The underlying rationale of authority/incentive tools is that using rewards or punishment to induce desired behaviors is the most effective way to motivate individuals—an argument with a rich history in the social sciences. For example, in psychology, such an orientation to human motivation is aligned with classic operant theory (Skinner, 1953). However, due to the emphasis on performance over process (Hursch, 2007), current education accountability reinforces *outcomes* not behavior, and this results in both desired behaviors (increased teaching effort/focus, instructional improvement) and undesired behaviors (teaching to the test, narrowing of the curriculum, cheating, etc.) being reinforced equally (Ryan & Brown, 2005; Ryan & Deci, 2017; Ryan & Weinstein, 2009).

In organizational science, this approach to motivation is at the heart of the classic "Theory X/Theory Y" debate of organizational management. Theory X is undergirded by assumption that human beings are inherently averse to work and responsibility and are therefore best motivated by external means (via rewards or punishment). Theory Y, on the other hand, recognizes the limitations of authority as a form of control and instead operates on the assumption that humans have an innate desire to actively seek out and pursue identified goals (McGregor, 1960). While scholars now recognize that this dichotomy has limitations, evident is the degree to which a Theory X mindset has driven much of current education policy making (Harvey, 2014)—this despite the corporate world having largely rejected this approach in favor of more development-oriented approaches (Cappelli & Tavis, 2016).

Since the early days of Race to the Top, and now with the advent of the Every Student Succeeds Act (ESSA), some states have augmented their teacher evaluation policies based on early feedback, but the overall trend points toward the increased use of extrinsic motivational tools in state teacher evaluation policy. As of 2015, 43 out of 50 states new teacher evaluation systems include student achievement as a measure of teacher effectiveness; for 35 of these, student achievement growth is a preponderant or significant criterion in the evaluation. Around one half of states allow the results of teacher evaluation to inform dismissal decisions, 19 states allow evaluation evidence to inform tenure decisions, and in 14 states teacher effectiveness is tied to compensation (i.e., pay-for-performance; Doherty & Jacobs, 2015).

A growing number of education researchers and practitioners are skeptical of the ability of next-generation teacher evaluation systems, as currently designed, to improve teaching and learning

(Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Hallinger et al., 2014; Hewitt & Amrein-Beardsley, 2016; Murphy et al., 2013). Though the evidence on performance or "merit" based pay is mixed, several recent experimental evaluation studies revealed limited, if any, change in student achievement, teacher motivation, attitudes, or instructional practice over time (Marsh et al., 2011; Springer et al., 2012; Yuan et al., 2013). Furthermore, recent scholarship has revealed some unintended consequences of current U.S. teacher evaluation policy, such as: lack of support and/or guidance in the use of teacher evaluation results (Amrein-Beardsley & Collins, 2012; Ford et al., 2017); lack of validity and/or reliability (either real or perceived) of evaluation results (Darling-Hammond et al., 2012; Ford et al., 2016; Jiang, Sporte, & Luppescu, 2015; Longo-Schmid, 2016; Reddy et al., 2017); and evidence of increase in adverse affective states for teachers subjected to high-stakes evaluation, such as high stress and anxiety as well as decreased job satisfaction, professional commitment, and turnover (Ford et al., 2017; Hewitt, 2015; Holloway & Brass, 2017; Ingersoll et al., 2016; Jiang et al., 2015).

Current Study

When considered together, these trends in current education policy mark an important shift in the nature and scope of teacher professionalism, and this has implications for how teachers view their work and the satisfaction they derive from it (Ford et al., 2017; Hargreaves, 2010; Holloway & Brass, 2017; Torres & Weiner, 2018). In her work in the sociology of professions, Evetts (2009, 2011) distinguishes between occupational and organizational professionalism, with the former denoting a professionalism characterized by partnership, collegiality, autonomy, and trust, and the latter a professionalism of bureaucratic and hierarchical control, standardization, and hyper-rationality. She argues that while administrators may espouse the values and approaches of occupational professionalism, the reality of day-to-day work reflects to a greater degree the values and approaches of organizational professionalism. We would argue, as others have, that this holds true for the teaching profession (Hargreaves, 2010; Zeichner, 2010) and by extension, teacher evaluation policy and practice. While district, state, and national teacher evaluation policies may espouse the importance of the benefits to teachers of evaluation (see, e.g., U.S. Department of Education, 2009, Race to the Top, Great Teachers and Leaders subsection D, Part 5), in practice high-stakes, top-down, teacher evaluation reflects an increasing prioritization of the needs of the educational organization for control and certainty over the needs of teachers to feel supported in their learning and development as practitioner-professionals (Holloway & Brass, 2017; Holloway, Sørensen, & Verger, 2017). These shifts have implications for teacher satisfaction, as we will discuss a bit later, because the characteristics of occupational professionalism are those historically that have been the primary sources of attraction and retention within teaching (Cohen, 2011; Lortie, 1975; Taylor & Tashakkori, 1995; Scott, Stone, & Dinham, 2001; Shen et al., 2012). Given these considerations, an important question that remains to be addressed in the literature is: If current evaluation systems were designed to be more supportive of teachers' growth as professionals, would this result in greater teacher satisfaction? While the answer to this question may seem obvious to most, it somewhat surprising to note that, to date, it has received very little empirical attention.

As a first step in this over-arching line of inquiry, we took up the following question for this study: What is the relationship between supportive teacher evaluation experiences and U.S. teachers' overall job satisfaction? To answer this question, we fit a multilevel regression model to data from U.S. lower-secondary teachers' in the 2013 Teaching and Learning International Survey (TALIS). The benefit of using the 2013 TALIS data for this study were three-fold. First, the TALIS data was collected by the National Center for Education Statistics (NCES) at the height of implementation of Race to the Top, so that, by spring 2013 when the TALIS data were collected, those U.S. states first

awarded the grants were already well into implementation of their new evaluation systems (Race to the Top Fund, 2016). Second, as an international survey, TALIS 2013 utilized a multitude of measures designed to capture a wide-range of teacher evaluation practices in schools, not necessarily just those reliant on formal incentives/punitive structures (see Strizek, Tourkin, & Erberber, 2014). Third, it was designed to capture teacher evaluation practice as situated within the context of a broad set of school climate and teacher working conditions.

The results of this investigation could be of benefit to both policy makers and practitioners in the US, as well as other countries employing more "high stakes" approaches to teacher evaluation. Furthermore, because states have more flexibility in augmenting teacher evaluation systems to encourage more routine, useful, and supportive feedback under the Every Student Succeeds Act (ESSA), information derived from this study might potentially be used to make changes to current systems with the enhancement of teacher support in mind. Additionally, because administrators are often responsible for overseeing the evaluation process at their respective school sites, a clearer understanding of the relationship between teacher evaluation policy and practice and teacher affective outcomes like job satisfaction could help school leaders adapt their roles within the evaluation system to foster a more supportive environment that better promotes authentic collaboration, professional growth, and instructional improvement (Davis & Wilson, 2000; Hulpia, Devos, & Rosseel, 2009).

Teacher Job Satisfaction: Measurement and Sources of Influence

Satisfaction is defined as a "...positive (or negative) evaluative judgment one makes about one's job," (Moe, Pazzaglia, & Ronconi, 2010, p. 1145) and informs teacher's feelings of professional commitment and/or motivation (Skaalvik & Skaalvik, 2011). These evaluations involve an interplay between how individuals experience teaching as well as the school environment and what they expect from these experiences (Papaioannou & Christodoulidis, 2007). Sylvia and Hutchinson (1985) defined teacher satisfaction as gratification derived from "higher order" needs— in other words, work elements which are intrinsically rewarding. As a cognitive process, satisfaction occurs for individuals when their capabilities are well-matched with the challenge of a task (Csikszentmihalyi, 1990). Satisfaction for teachers, like workers in other human improvement occupations, is intricately tied to the degree to which they are able to pursue and reap the "psychic rewards" that attracted them to the field (Hargreaves, 2010). For teachers, having autonomy, meaningful relationships with colleagues and students, and seeing their hard work pay off in student success are all key psychic rewards (Cohen, 2011; Scott, Stone, & Dinham, 2001; Shen et al., 2012; Taylor & Tashakkori, 1995).

Job satisfaction is a somewhat ambiguous term, and this has led to some variation in how it has been measured in educational research. In measuring job satisfaction, researchers have tended toward a more global perspective, as preferences about the specific circumstances which satisfy individual teachers can vary (Skaalvik & Skaalvik, 2010). In general, two facets of teacher job satisfaction have comprised the measure in past studies: the extent that teachers have had a positive judgment of their work and/or the profession as well as school working conditions (Judge, Thoresen, Bono, & Patton, 2001; Klassen & Chiu, 2010; Weiss, 2002). For this study, we followed this general operationalization, conceptualizing teacher job satisfaction as containing both positive evaluative judgments about working conditions in schools as well as that of their work and profession more generally.

Beyond the intrinsic factors that primarily drive teachers, there are other sources of influence on teachers' job satisfaction, and these can be classified into two categories: a) aspects of school

context and process, and b) external factors (Dinham & Scott, 1998; Shen et al., 2012; Skaalvik & Skaalvik, 2011). Aspects of school context and process consist of both fixed characteristics of schools such as school size and school composition, and school process variables which encapsulate school working conditions. Both school size and composition (percent poverty/students of color) have been found in prior studies to be negatively related to satisfaction (Perie & Baker, 1997). Yet others have found no association between school size and satisfaction (Skaalvik & Skaalvik, 2009). Being assigned a mentor or participation in teacher induction have also been found to increase satisfaction, intent to stay, and perceptions of working conditions, particularly for new teachers (Pogodzinski, Youngs, & Frank, 2013).

Findings regarding the relationship of other teacher and principal background characteristics such as years of experience, gender, and schooling level have been mixed, but a general assessment of studies on teacher satisfaction is that the relative effects of school process variables like school climate/culture, administrative support, collegiality, empowerment/decision making, and relationships with students are larger than those of teachers and school background characteristics (Shen et al., 2012; Van Maele & Van Houtte, 2012). The one exception to this generalization, however, is in the domain of teacher affective responses to school conditions, and these tend to relatively strong predictors of satisfaction. Already discussed previously, as a corollary affective reaction to satisfaction, teacher burnout in response to adverse working conditions, is moderately-to-strongly related to job satisfaction (Skaalvik & Skaalvik, 2009, 2011). Furthermore, self-efficacy, or a teacher's beliefs about the perceived internal and external resources (various forms of support such as autonomy, professional development, relationships with colleagues) they can draw upon in completing a teaching task, is also positively related to job satisfaction (Skaalvik & Skaalvik, 2009, 2014), suggesting that it mediates the relationship between school climate and satisfaction (Malinen & Savolainen, 2016). These significant relationships with burnout and efficacy suggest that a factor, like teacher evaluation, which also shapes a teacher's perception of their ability, would be directly related to job satisfaction. Moreover, it is likely that the supports or climate surrounding the evaluation would further enhance or exacerbate this relationship.

In the accountability era, factors external to the school related to satisfaction are those related to external pressure and intervention, as well as public perception of schools (Skaalvik & Skaalvik, 2011), and these likely have indirect effects on teacher satisfaction through the principal's enactment of leadership practice that either is (or is not) conducive to positive school climate and working conditions. Few studies in the literature focus on the direct relationships of these factors on teacher satisfaction, perhaps because of the indirect nature of the relationship or because of the lack of direct measures of these effects available for use. Our study, however, considers all the above sources of satisfaction in our models of the effects of supportive teacher evaluation practice on teacher satisfaction.

Theoretical Framework

Self-determination theory (SDT) is a multi-faceted psychological theory of human behavior and personality development (Ryan & Deci, 2017). SDT researchers have amassed, over decades of research, a set of empirically-tested theories and propositions about the various sets of conditions under which individuals are optimally motivated. A key maxim of Self-determination theory (SDT) is that of *dialectical integration*: an intrinsic desire to engage in and interact with the world, exercise capacities, and pursue connectedness toward a more complex sense of self (Deci & Ryan, 2000). Basic Psychological Needs Theory (BPNT), a sub-theory of SDT, predicts that intrinsic drive will remain intact so long as certain key psychological conditions are met, namely the needs for

competence, autonomy, and relatedness. *Competence* refers to the need to experience performances as effectively enacted; thus individuals are driven to build upon existing skills and capacities in anticipation of future performance (Niemiec & Ryan, 2009; Ryan & Deci, 2002). *Autonomy* concerns a perceived internal locus of causality; that is, the taking of action for which impetus derives not from the need to conform to external forces/expectations but rather from self-endorsed or determined values and beliefs (Ryan & Deci, 2000). Finally, *relatedness* refers to the psychological need to feel connected to significant others; i.e., to care and be cared for as well as share a sense of belongingness to others in your community (Ryan & Deci, 2002). When these needs are met, the individual experiences a wide-range of positive outcomes including autonomous motivation, a personal growth orientation, engagement, enjoyment, self-efficacy, satisfaction, and decreased burnout (Kunter & Holtzberger, 2014; Ryan & Deci, 2017).

However, a significant part of daily life is carrying out tasks that are not intrinsically rewarding, and necessitate motivation by external means. In these cases, other SDT sub-theories, like Organismic Integration Theory (OIT), help explain how to move individuals from controlled towards more autonomous (intrinsic) motivation—an orientation to work, as mentioned above, that yields substantially greater positive behavioral and psychosocial outcomes (Ryan & Deci, 2017). The critical take-away in discussing these two distinct sub-theories is that understanding the context underlying the motivation of individuals towards a task is critical to selecting an appropriate motivational approach. At best, the misalignment between context and strategy will have no effect on behavior or performance; at worst, it can have deleterious effects of individual motivation, well-being, and performance (Ryan & Deci, 2017).

BPNT as a sub-theory is particularly relevant to understanding teacher motivation due to the unique characteristics of the teaching profession (Roth, 2014). Because the majority of teachers enter the profession for altruistic reasons (Lortie, 1975; Rosenholtz, 1991; Watt & Richardson, 2014), it is one of the few occupations where individuals exhibit an intrinsic orientation to their work (Kunter & Holtzberger, 2014). In this case, the motivational strategy is simple: create conditions that reinforce and activate existing intrinsic motivation. As was mentioned previously, this is not the approach accountability systems in the U.S. generally take; thus the prevalence of extrinsic rewards/punishment tied to performance creates a fundamental misalignment between the motivational context and the approach (Niemiec & Ryan, 2009). This misalignment manifests itself, among other ways, as increased stress, decreased satisfaction, and increased attrition (Ford et al., 2017; Ryan et al., 2017; Saeki, Segool, Pendergast, & von der Embse, 2017; von der Embse et al., 2016; von der Embse, Schoemann, Kilgus, Wicoff, & Bowler, 2017). Refocusing the design and implementation of teacher evaluation to better reflect the characteristics of occupational professionalism—those things that provide teachers with a deep satisfaction in their work, namely partnership, collegiality, and autonomy—would begin the process of restoring working conditions to schools more conducive to meeting the psychological needs of teachers (Eyal & Roth, 2011; Ford & Ware, 2018; Roth, 2014; Ryan & Weinstein, 2009)

Teacher Evaluation Supportive of Teacher Growth and Development

Despite a continued struggle to remake U.S. teacher evaluation policy anew, there is a growing body of evidence about how to evaluate teachers in ways which support their growth and development as practitioners. There are many purposes for evaluating teachers, but a basic distinction we can draw is between teacher evaluation for accountability (summative) versus professional development (formative) purposes. On one hand, the goal of summative teacher evaluation is to assess the teacher's performance or quality, typically for accountability purposes (i.e., in reaching a decision about whether to apply reward or sanction; Organization for Economic Co-

operation and Development [OECD], 2009). Formative evaluation, on the other hand, involves evaluation for the purposes of teacher support and professional development (Delvaux et al., 2013).

Of course the formative/summative dichotomy can be a misleading one; these evaluation systems are not necessarily incompatible, but over-emphasis of the summative components of evaluation can undermine efforts to provide valid and reliable feedback to practitioners (Campbell, 1979; Ryan & Brown, 2005) as well as negatively affect school climate and culture (Ford et al., 2017; Saeki et al., 2017). Reinhorn, Johnson, and Simon (2017) found that the most successful schools in their sample prioritized formative evaluation, embedded within a supportive, improvement-oriented professional culture, and this in turn influenced teachers' attitudes towards evaluation as a developmental process and helped legitimize summative evaluation processes.

Recommendations for teacher evaluation practice which is supportive of teachers' growth and development can be traced back to the first teacher evaluation movement in the U.S. in the 1980s, but have resurfaced as a result of recent policy developments (The New Teacher Project [TNTP], 2010; Weisberg, Sexton, Mulhern, & Keeling, 2009). First, teacher evaluation should be systematic and frequent and should yield useful, meaningful information about a teacher's practice as well as critical feedback on what to improve (Delvaux et al., 2013; Ford et al., 2016; TNTP, 2010; Tuytens & Devos, 2011). The SDT concept of *functional significance* states that the effects of external events on human motivation hinge on the psychological meaning they have for the recipient (Ryan & Weinstein, 2009). Events have a positive effect on an individual's self-motivation when they have *informational* significance—that is, when they provide feedback that helps learners become more effective but without eclipsing autonomous action (Deci & Ryan, 2000; Ryan & Deci, 2017). On the other hand, events have *controlling* significance if they are experienced as pressure toward specific outcomes, and in these cases, individuals often respond by exerting the least amount of effort needed to gain reward or avoid punishment (Ryan & Weinstein, 2009). Finally, events have *amotivating* significance when the arousal they engender is debilitating or when they contain no inherent rationale for action. For example, events that are too challenging or feedback which is highly negative foster feelings of helplessness (loss of control) or incompetence (i.e., lack of self-efficacy), leading individuals to withdraw effort (Ryan & Brown, 2005). Time and energy are increasingly in short supply in U.S. public schools, so faculty who perceive the evaluation process as a waste of these resources are likely going experience frustration and stress from having to engage in it.

Second, assessments of teacher performance should be based upon a set of high-standards which reflect what is currently understood as good teaching practice (Darling-Hammond, 2013; Lavigne & Good, 2014, 2015). An evaluation approach based on the standards of good teaching supports instructional improvement by establishing the clear expectations necessary to motivate change (Kelly, Ang, Chong, & Hu, 2008) and by providing a common language for evaluators and teachers to discuss instructional feedback (Kraft & Gilmour, 2016). Clear standards can also facilitate perceptions of the evaluation as valid and fair, and this can drive use of information from their evaluation to make changes in practice (Delvaux et al, 2013; Ford et al., 2016; Lavigne, 2014). Perceptions of the validity, fairness, and usefulness of the evaluation process are a strong determinant of teacher satisfaction with the teacher evaluation process (Delvaux et al., 2013).

Third, such information should be based on a thorough assessment of teaching practice—no one measure of performance (whether student test scores or otherwise) is adequate to arrive at a determination of teacher effectiveness and construct a plan for change (Grissom & Youngs, 2016; Lavigne & Good, 2014; Master, 2014; TNTP, 2010). "Objective" measures of success like test scores seem particularly inadequate as measures of high-quality teaching, where objective success is elusive, not easily measured, nor feedback necessarily immediate (Cohen, 2011; Lortie, 1975).

Furthermore, utilizing predominately summative measures of performance such as standardized test scores or other student growth measures to make judgments of teacher performance increase the likelihood that teachers perceive the information generated as controlling as opposed to informational and will be less likely (or able) to use it in making changes to their practice (Adams, Forsyth, Ware, & Mwavita, 2016; Ryan & Weinstein, 2009). Furthermore, the prioritization (whether intentional or unintentional) of one measure of teacher performance may undermine credibility in the evaluation system if teachers do not feel that this measure truly reflects the quality of their instruction (Lavigne, 2014). Such a narrow measurement focus also ignores other valuable contributions of teachers to the development of whole child, such as the cultivation of non-cognitive dispositions or deeper, higher-order thinking (Grissom, Loeb, & Doss, 2016).

Fourth, during the evaluation process, teachers must be properly supported (in terms of time, autonomy, but also collegial support and professional development) for improvement in practice to emerge (Darling-Hammond, 2013; Ford et al., 2017). These three areas of support align with self-determination scholars' identification of three psychological needs which must be met for intrinsic motivation for tasks to be activated: the need for competence, autonomy, and relatedness (Ryan & Deci, 2002). Teacher involvement through meaningful dialogue, goal-setting, and peer support might also help promote investment in the process (Kraft & Gilmour, 2016) while supporting teacher autonomy. It stands to reason that a teacher who has been provided useful, critical feedback without the support needed to utilize the feedback in constructive way to improve his/her practice will likely struggle to change and be frustrated as a result. Innovative approaches to teacher evaluation which involve peer/mentor assistance and support are emphasized in the recent literature (Darling-Hammond, 2013; Darling-Hammond et al., 2012; Hinchey, 2010; Lavigne & Good, 2014).

The nature of the relationship between the evaluator and the teacher is also important to teacher's satisfaction with the evaluation system (Delvaux et al., 2013). If the teacher perceives the evaluator to be incompetent, or the feedback process is compromised by poor communication, this can have a detrimental effect on the perceived usefulness of the process and the likelihood that a teacher will act on the evaluation results (Chow, Wong, Yeung, & Mo, 2002; Kelly et al., 2008). Perceptions of competence of the evaluator are related to the degree that the teacher feels the evaluator is qualified to rate their performance, and this is driven by knowledge of the evaluator's teaching experience, subject matter content/pedagogical knowledge, and training in the evaluation process (Delvaux et al., 2013; Milanowski & Heneman, 2001).

Finally, supportive teacher evaluation processes as a whole should be viewed by teachers and leaders alike as a significant, worthwhile activity (TNTP, 2010). This perception is a more global determination of its validity and importance, and is based on many of the design and implementation features listed above, but also the degree to which these features lead to improved results.

The Role of School Leaders in Supporting Teacher Development through Evaluation

Many of the features of supportive teacher evaluation mentioned in the previous section will not work without concomitant attention to the organizational hierarchies and power dynamics of a school. School leaders can play an important role in this process not as implementers of evaluation systems, but as facilitators of evaluation systems embedded within a school climate which values and emphasizes the norms of authentic occupational professionalism (Murphy et al., 2013). This calls attention to the multifaceted role of the school leader in promoting a school climate conducive to teacher autonomy, satisfaction, and commitment (Dou, Devos, & Valcke, 2017), as well as teacher growth and learning (Drago-Severson, 2012).

While supportive teacher evaluation is one mechanism for teacher growth, teacher development is also enhanced within a professional context (Kraft & Papay, 2014), suggesting the need for a more comprehensive set of supports such as opportunities for collaboration, shared decision making, and professional development (Drago-Severson, 2012; Kraft & Papay, 2014). Additional features of a school climate that supports teacher learning and growth include a school leader's investment in relationships—conveying care, respect, and appreciation for teachers' work—as well as a willingness to model learning (Drago-Severson, 2012). While teacher evaluation is often associated with the instructional dimension of leadership (Blase & Blase, 1999; Grissom, Loeb, & Master, 2013), developing a school climate that supports teacher learning and growth requires school leaders to employ a range of leadership strategies that reflect a balance of instructional, managerial, and visionary leadership approaches (Drago-Severson, 2012). Thus, while school leaders are often charged with formal summative evaluation responsibilities, this role should not be considered in isolation from other important—and demanding—leadership responsibilities as well as the school context in which evaluation practices are enacted.

Another important consideration when evaluating the potential of evaluation systems to support teacher development is the possible discrepancy between espoused or intended implementation and actual implementation practices. Murphy et al. (2013) have suggested that school leaders lack the skills to meaningfully leverage teacher evaluation for improvement purposes. Moreover, Kraft and Gilmour (2016) found that principals' individual goals and attitudes can influence evaluation practices, and other constraints can influence the quality of evaluation feedback, including lack of time amidst competing responsibilities, the nature of support and training for implementation, evaluation systems requirements and design, evaluators' level of general pedagogy versus subject-specific expertise, and the trust required to balance summative and formative interactions with teachers. Grissom, Loeb, and Master's (2013) findings that some practices, such as classroom walkthroughs, might be less effective than more direct support for teacher development through evaluation and coaching have implications for evaluators' time use and priorities and point to the need to consider what instructional leadership practices are more or less effective for supporting improvement. This underscores the complexity of school leaders' efforts to harness evaluation systems to promote collaboration around instructional improvement.

Method

This study is a secondary analysis of the 2013 Teaching and Learning International Survey (TALIS 2013) administered by the Organisation for Economic Co-operation and Development (OECD). TALIS 2013 surveyed a total of 34 countries including the US. The focus of the "core" data collection efforts of TALIS 2013 remained, as they did for TALIS 2008, on teachers and leaders who work in lower secondary schools, level 2 of International Standard Classification of Education (ISCED; Organization for Economic Co-operation and Development [OECD], 2010, 2014). While options to collect data at Levels 1 (Primary education) and 3 (Upper Secondary) were left open for individual countries to pursue, the majority of the OECD countries—including the United States—did not exercise them (OECD, 2014). Thus, our inferences about the relationship of supportive teacher evaluation to teacher job satisfaction in the United States were necessarily constrained to lower secondary teachers as a subgroup.

This administration of TALIS focused on the policy-related matters of both the appraisal of teachers' work in schools and their reported feelings of job satisfaction (OECD, 2014). TALIS 2013 items regarding teacher "appraisal" were replete and elicited information about various aspects of the evaluation/appraisal process. For each country sampled, TALIS 2013 set a target size of 200

schools with 20 teachers per school. Schools were selected according to a national sampling plan which used systematic random sampling with probability proportional to size (PPS) within explicit strata which might include school types, regions or funding (OECD, 2014). In line with the study's stated purpose, only the U.S. sample of schools was utilized (*N*=122).

Measures and Instrumentation

Job satisfaction. TALIS 2013 presents a unique opportunity to study the relationship of aspects of teacher evaluation practice on teacher job satisfaction. OECD created a composite measure of this focal dependent variable based on analyses of reliability and construct validity across countries. Job satisfaction has been measured as the extent that teachers are satisfied with their position and school (Skaalvik & Skaalvik, 2011), and teacher satisfaction has been linked to decisions to move schools or leave the profession (Klassen & Chiu, 2011). The TALIS job satisfaction scale includes satisfaction with current work environment with items, "I enjoy working at this school," "I would recommend my school as a good place to work," "All in all, I am satisfied with my job," and satisfaction with profession, "The advantages of being a teacher clearly outweigh the disadvantages," "If I could decide again, I would still choose to work as a teacher." For both scales, reliability was above .70 (OECD, 2014). The teacher job satisfaction measure was standardized to facilitate effect size interpretation.

Supportive teacher evaluation (SUPPEVAL). This measure was created from TALIS teacher-level items to capture strategies/approaches to teacher evaluation that research suggests support teacher development and build teachers' intrinsic motivation for improvement (Darling-Hammond; 2013; Delvaux et al., 2013; Firestone, 2014; Ford et al., 2017). Using the program WINSTEPS 3.81 (Linacre, 2014), a Rasch rating-scale model was applied to a cluster of items from the teacher questionnaire (TT2G31A-H) that reflect evidence-based approaches to supportive teacher evaluation.

TALIS 2013 utilized confirmatory factor analysis (a classical test theory [CTT] approach) for its measure construction and scaling (OECD, 2014). The Rasch model, in contrast, is an Item Response Theory [IRT] approach, and is distinguished from classical test theory in considering the ability of respondents in tandem with the difficulty (i.e., ease of endorsability) of the items to which they are responding. Advantages and disadvantages of both notwithstanding (see, for example, Singh, 2004), both approaches are useful in the development and scaling of latent measures; in fact, other prominent international education datasets have opted to use an IRT approach for measure construction and scaling (e.g., TIMSS and PISA). Our choice to adopt an IRT approach over CFA was due to the exploratory nature of our development of the SUPPEVAL measure, which required a wider range of information to assess person and item performance than is typically provided using a CTT approach.

In addition to a host of other diagnostic information, the WINSTEPS program produces a scaled-score for each teacher in log-odds units which represents where each teacher's perceptions locates him/her on the continuum of supportive teacher evaluation (low, negative values reflect perceptions of a more punitive/non-supportive system, and high, positive numbers reflect perceptions of a more developmental/supportive teacher evaluation system). We set our threshold at mean-squared values of .5 to 1.5—accepted thresholds for Winsteps analysis (Linacre, 2014). Based on these criteria, one item was discarded as misfitting (TT2G31F), and this item asked teachers about how dismissal was linked to teaching evaluation performance. We hypothesized that a likely possibility for the noisiness of this item (high infit/outfit) in the measurement model is because the applicability of dismissal as a consequence for low performance differs across sampled schools (some schools are subject to these rules and some not due to district/state-level mandate),

similarly rated SUPPEVAL teachers across sampled schools might have had very different responses to this particular item if it does not apply to them.

Information related to the items which comprised the measure and results of the Rasch analysis are all included in Appendix A. Our final Rasch model of SUPPEVAL revealed high person and item reliability, as well as adequate item separation. Item difficulty analysis revealed that the easiest item for teachers to endorse was: "Teacher and appraisal feedback are largely done to fulfill administrative requirements" (TT2G31), and the most difficult to endorse was "Measures to remedy any weakness in teaching are discussed with the teacher" (TT2G31C). In a rating scale model, item difficulty is interpreted as an estimate of the distance of the "balance point" of the data (i.e., where responses on either side of the scale are equal) from the mid-point of the scale (in the TALIS case is midway between 'agree' and 'disagree,' given a 4 point scale). In other words, more teachers agreed than disagreed with the first statement, and more disagreed than agreed with the second statement. Since, item 31C was reverse coded, both of these difficulty statistics reveal that, on average, TALIS teachers perceived their evaluation system as less supportive—hence the negative average SUPPEVAL score in the sample (see Table 1).

Analytical Approach

Handling missing data. Instead of employing list-wise deletion, we employed multiple imputation (MI) techniques to the raw, teacher and school level data utilizing the *mi* statistical package in R (Gelman, Hill, Su, Yajima, & Pittau, 2015). Multiple imputation is substantially more robust than typical list or pair-wise deletion procedures to missing data bias, and results in multiple versions of the same dataset with different plausible values for the missing data based on available variable data and their underlying covariance structure (Enders, 2010). Furthermore list-wise deletion requires that the data are missing completely at random (MCAR) in order to ensure unbiased estimates—MI assumes only that the data are multivariate normal and, at minimum, missing at random (MAR).

The one exception to using MI for handling missing data in our analysis was with respect to the focal outcome, teacher job satisfaction. Teachers in the U.S. sample who were missing a job satisfaction score were removed prior to analysis (i.e., the values were not imputed). In all, this resulted in only 72 U.S. teachers being removed (approximately 3 percent of the total). Teacher job satisfaction, as well as several other similar perception items towards the end of the TALIS teacher survey (such as self-efficacy and climate perceptions) exhibited a unit nonresponse pattern (Enders, 2010). After reviewing the missing data coding procedure in the TALIS technical manual, the missing codes in the dataset indicated that nearly all of the teachers who were missing a job satisfaction score either returned the survey blank or incomplete (OECD, 2014). In this case, we endeavored to determine whether or not there were significant differences between teachers who were missing a job satisfaction score and other measured variables—in other words, could the data be assumed to be missing completely at random (MCAR). To test this assumption, we conducted a series of Bonferroni corrected t-tests and chi-squared tests of independence between teachers who had a job satisfaction score and those for whom it was missing with respect to TALIS teacher and school level variables. We found no significant differences between the groups with respect to these covariates, and thus list-wise deletion was a justifiable missing data handling approach (Enders, 2010).

Once teachers without a job satisfaction score were removed, the remainder of the missing data exhibited a general item non-response pattern (deLeeuw, Hox, & Huisman, 2003). This missing data pattern manifests as gaps in item response that appear to be randomly dispersed throughout the dataset (i.e., MCAR). Because multiple imputation does not require us

to invoke a MCAR assumption for the results to be unbiased, and there was no evidence to suggest they were missing not at random (MNAR), we chose the less stringent assumption of MAR, which allows for missing data on a variable to be related to other measured variables in the analysis (Enders, 2010).

The multiple imputation procedure employed fading noise reduction and resulted in 5 imputed datasets, which achieved convergence at 120 iterations. Since the HLM program can handle multiply-imputed datasets in estimation, separate 2-level HLM analysis files were created for each imputed dataset and then pooled final model estimates were provided via the multiple imputation feature in HLM 7.0. These procedures resulted in a final sample of 1853 teachers nested within 122 schools. Table 1 displays both pre- and post-imputation descriptive statistics for all study variables.

Table 1
Descriptive Statistics for Study Variables

Teacher Characteristics	Original N	Imputed N	Mean	Imputed Mean	SD	TALIS Variable(s)
Female	1852	1853	0.65	0.64	0.48	TT2G01
Years of experience[a]	1840	1853	0.01	0.01	1.03	TT2G05B
Graduate degree	1850	1853	0.02	0.02	0.13	TT2G10
Full time teacher	1853	1853	0.97	0.97	0.17	TT2G03
Permanent contract	1853	1853	0.67	0.67	0.47	TT2G06
Proportion SPED students in class	1853	1853	0.13	0.13	0.33	TT2G09
Participated in an induction program	1847	1853	0.61	0.61	0.49	TT2G19A
Assigned a mentor	1848	1853	0.12	0.12	0.32	TT2G20A
School mgmt. (Hrs/wk)	1751	1853	1.75	1.78	4.87	TT2G18E
Paperwork (Hrs/wk)	1800	1853	3.28	3.52	0.99	TT2G18F
Barriers to PD year	1843	1853	10.82	10.80	3.03	TT2G27C-G ($\alpha = .73$)
Felt prepared to teach	1851	1853	10.59	10.59	1.70	TT2G13A-C ($\alpha = .84$)
Depth of preservice content & methods	1853	1853	8.01	8.02	1.52	TT2G12A-C ($\alpha = .74$)
Feedback led to positive chg in practice	1810	1853	2.24	2.23	0.76	TT2G30A-N ($\alpha = .94$)
Perception of climate and student relations[a]	1850	1853	0.00	0.04	1.00	TT2G45A-D ($\alpha = .81$)
Tch. pcpt. of shared responsibility and decision making[a]	1850	1853	-0.01	0.00	1.00	TT2G44A-E ($\alpha = .89$)
Teacher self-efficacy	1844	1853	12.68	12.59	1.83	TSELEFFS
Supportive teacher evaluation	1853	1853	-2.04	-2.03	1.35	TT2G31A-E,G,H
Job satisfaction[a]	1853	1853	0.00	0.00	1	TJOBSATS

Table 1 (Cont'd.)
Descriptive Statistics for Study Variables

Principal/School Characteristics	Original N	Imputed N	Mean	Imputed Mean	SD	TALIS Variable(s)
Principal female	101	122	0.49	0.44	0.50	TC2G01
Doctoral degree	101	122	0.16	0.15	0.35	TC2G03
Principal experience[a]	101	122	0.04	0.07	0.97	TC2G04B
Village, <3,000	100	122	0.19	0.16	0.37	TC2G09
Town, 3k-100k	100	122	0.45	0.50	0.50	TC2G09
City, 100,001-1,000,000	100	122	0.22	0.20	0.40	TC2G09
Large City, >1,000,000	100	122	0.11	0.09	0.29	TC2G09
Private school	99	122	0.22	0.23	0.43	TC2G10
Stud./teacher ratio	99	122	14.90	14.50	7.26	G12A/G14A
% of non-native speakers[a]	99	122	-0.11	-0.10	0.91	TC2G15C
% disadvantaged homes[a]	99	122	-0.16	-0.14	1.00	TC2G15A
Principal perception of climate[a]	99	122	0.02	0.05	0.84	TC2G30A-F ($\alpha = .80$)
Prin job satisfaction[a]	99	122	0.00	-0.05	1.19	PJOBSATS
Eval. primarily by principal	99	122	0.78	0.82	0.38	TC2G27A
Eval. primarily by fellow teachers (non-mgmt.)	99	122	0.17	0.14	0.34	TC2G27D
Eval. primarily by external individuals	99	122	0.11	0.09	0.29	TC2G27E
Degree of external control over school policies	100	122	17.25	17.31	3.36	TC2G18A5-K5 ($\alpha = .91$)

Note: [a]Variable standardized. Appropriate weights applied for descriptive statistics and reliability analyses.

Two-level HLM analysis. In addition to the focal predictor, SUPPEVAL, we included other important teacher and school characteristics, attitudes, and perceptions about school working conditions and teacher appraisal practices that were presumed to be related to teacher satisfaction based on prior research. School level variables related to teacher satisfaction such as school climate, principal and school characteristics such as school poverty, urbanicity, sector, and principal satisfaction, as well as the primary evaluators in teacher appraisal were used to model between-school variation in teacher job satisfaction. The final 2-level HLM structure is represented in Equations 1 and 2 below:

Level 1 (Teacher):

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{0qj} X_{qj} + r_{ij} \tag{1}$$

Level 2: (School):

$$\beta_{0j} = \gamma_{00} + \sum_{s=1}^{S} \gamma_{0s} W_j + u_{0j} \tag{2}$$

$$\beta_{0qj} = \gamma_{q0} \text{ for } q = 1, \dots, Q. \tag{3}$$

Equation 1 indicates that teacher job satisfaction, Y_{ij}, was modeled with respect to Q number of teacher-level covariates, including SUPPEVAL. At the school level (Equations 2 and 3), variation around the grand mean of teacher job satisfaction, γ_{00}, was modeled with respect to the sum of S number of school level predictors, including various school and principal characteristics, and a term for unexplained school variation in teacher job satisfaction, u_{0j}. The teacher weight, TCHWGT, was incorporated into the final analysis to maintain the intended representativeness of the sample. All other teacher-level effects remained fixed at Level 2, as an analysis revealed that there was little between-school variance in the relationships between each of the teacher-level predictors and job satisfaction (Equation 3).

Results

The central purpose of this paper was to examine the influence of supportive teacher evaluation practice on teacher job satisfaction, and the results of the 2-level HLM analysis are displayed in Table 2. As an important first step in this analysis, we first examined the intra-class correlations (ICCs) with respect to the outcome via an examination of the partitioning of variance in teacher job satisfaction in the unconditional (null) model. Our analysis revealed a statistically significant amount of the variance in teacher job satisfaction (approximately 11%) was at the school level ($p < .001$), supporting our decision to model variance in teacher job satisfaction as containing both between-teacher and between-school components.

The final model of teacher job satisfaction revealed some important findings. First, our final model revealed an effect of SUPPEVAL (albeit small) on teacher job satisfaction (coef. = 0.051, SE = .018, $p < .01$). While controlling for all other teacher and school level factors, for each log-odds

increase in the supportive nature of the evaluation experience, (SUPPEVAL mean= -2.03, range ≈ 10 log odds units), teachers had a corresponding increase in satisfaction of half of a tenth of a standard deviation. Put another way, all other variables held constant, there is an average predicted difference of a half a standard deviation (1 raw job satisfaction point) between teachers on the very low end of the SUPPEVAL scale, and those at the very top.

Table 2
Two-Level HLM of the Effects of Supportive Evaluation on U.S. TALIS Teachers' Job Satisfaction (n = 1853)

Teacher Fixed Effects	coeff	SE[a]	t-ratio
Intercept	0.295	0.149*	1.979
Supportive teacher evaluation (SUPPEVAL)	0.051	0.018**	2.784
Female teacher	0.018	0.041	0.435
Years of experience	0.003	0.024	0.106
Graduate degree	0.120	0.134	0.900
Full time teacher	-0.286	0.097**	-2.942
Permanent contract	0.033	0.051	0.655
Proportion of SPED students in class	0.025	0.069	0.358
Participated in an induction program	-0.054	0.044	-1.238
Has been assigned a mentor	-0.063	0.073	-0.867
School management (hrs./wk.)	-0.005	0.005	-0.940
Hours per week spend on paperwork	-0.011	0.005*	-2.052
High barriers to PD in last year	-0.037	0.008***	-4.818
Teacher felt prepared to teach	0.024	0.013~	1.788
Preservice training in content and methods	0.013	0.017	0.734
Eval. feedback has led to positive changes in practice	0.267	0.039***	6.823
Tch. percept. of climate and student relations	0.157	0.023***	6.720
Tch. shared responsibility and decisionmaking	0.214	0.031***	6.950
Teacher self-efficacy	0.042	0.013**	3.117

School Fixed Effects			
Female principal	-0.127	0.061*	-2.060
Principal doctoral degree	0.105	0.060~	1.753
Principal experience	-0.012	0.032	-0.379
Village, <3000 people	0.121	0.116	1.048
Town, 3001-100,000	0.160	0.090~	1.767
City, 100,000-1,000,000	0.030	0.118	0.258
Private school	0.128	0.095	1.342
Student-teacher ratio	-0.010	0.005*	-2.043
Percent non-native speakers	0.017	0.034	0.503
Percent students from disadvantaged homes	-0.011	0.049	-0.215
Principal perception of climate	0.059	0.032~	1.835
Principal job satisfaction	0.011	0.023	0.479
Eval. primarily by principal	-0.145	0.082~	-1.768
Eval. primarily by fellow teachers (non-mgmt.)	0.204	0.069**	2.945
Eval. primarily by external individuals	-0.126	0.092	-1.365
Degree of external control over school policies	0.018	0.009~	1.952

Table 2 (Cont'd.)
Two-Level HLM of the Effects of Supportive Evaluation on U.S. TALIS Teachers' Job Satisfaction (n = 1853)

		Random Effects	
Level 1	Between Teachers		0.594***
Level 2	Between Schools		0.027***
Level 1	Pseudo-r²		0.31
Level 2	Pseudo-r²		0.75

***$p \leq .001$, **$p \leq .01$, *$p \leq .05$, ~ $p \leq .10$

Note. [a]Robust standard errors reported. Coefficient estimates in this table are the averaged results from the 5 imputed datasets provided by the HLM program, with the teacher weight applied (TCHWGT). The outcome variable, teacher job satisfaction (TJOBSATIS), was standardized for this analysis. Large cities (1,000,000+) were the comparison/holdout group for urbanicity.

A key related finding from the analysis was that teachers who perceived that the feedback from their evaluation prompted positive changes in their practice was associated with higher job satisfaction on average (coef. = 0.267, SE = .039, $p < .001$), and this finding was over 2.5 tenths of a standard deviation. This finding is demonstrated further in Figure 1. While there is a wider range of below average teacher satisfaction responses, teachers who perceived around average or above average satisfaction for their current school and profession had correspondingly higher perceptions of supportive evaluation as well as a higher perceived positive change in practice from their evaluation. In reference to this figure, the positive, three-dimensional relationship is demonstrated by the high clustering of responses at the intersection of the top of the figure, positive end of the scale of teacher satisfaction, and the corner of the cube where the positive sides of both evaluation scales (supportive and positive change in practice) meet.
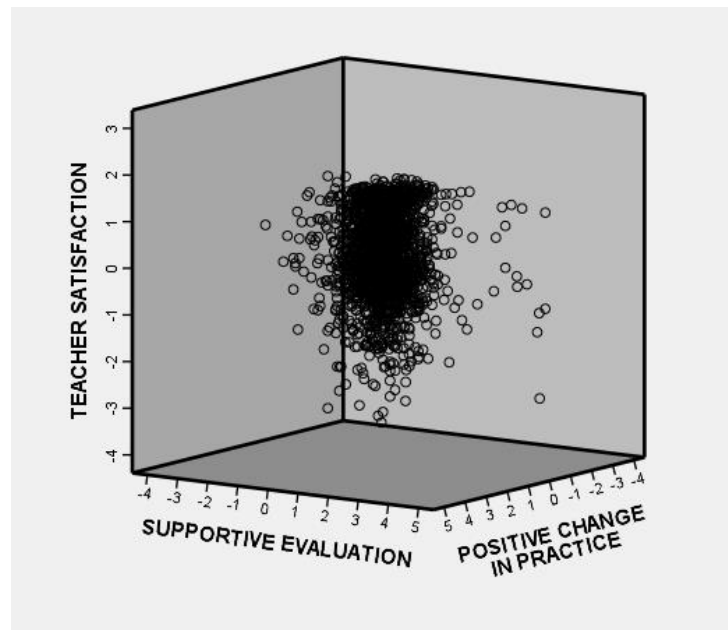
Figure 1. Three-dimensional positive relationship between supportive evaluation, positive change in practice from evaluation, and teacher satisfaction.
Note: Raw data from mean of imputed teacher datasets; standardized variables.

Other notable findings were that teachers' perception of climate and teacher-student relations and shared decision making were strong, independent predictors of teacher job satisfaction (coef. = 0.157, *SE* = .023, *p* < .001 and coef. = 0.214, *SE* = .031, *p* < .001, respectively). Teachers' feelings of self-efficacy were also positively related to job satisfaction (coef. = 0.042, *SE* = .013, *p* < .01). Further, teachers' job satisfaction was negatively related to the perceived barriers to professional development they experienced in the past year (coef. = -0.037, *SE* = .008, *p* < .05). At the school level, most notable was the estimate of the effects on a teacher's satisfaction attributed to their primary evaluator. We found a significant difference in average teacher satisfaction in USA TALIS schools where non-management teaching colleagues were the primary evaluator (coef. = 0.204, *SE* = .069, *p* < .01) as opposed to another group (principal, assigned mentor, other member of school management, or external evaluator). In contrast, in schools where the principal was identified as the primary evaluator, average teacher satisfaction was significantly lower than in schools where the primary evaluator was someone else within the school (coef. = -0.145, *SE* = .082, *p* < .10). Also of interest was the non-significance of principals' job satisfaction and the satisfaction of teachers in their school (coef. = 0.011, *SE* = .023, *p* = n.s), as well as that of principal and teacher experience, assigned mentor, or participation in a mentoring program. Finally, there were also marginal positive associations of the school climate with teacher job satisfaction across schools (coef. = .059, *SE* = .032, *p* < .10), and a small negative influence on satisfaction as student-teacher ratio increased (coef. = -.01, *SE* = .043, *p* < .05).

In summary, this two-level HLM tested the independent influence of teacher evaluation and their surrounding working conditions on satisfaction. The main variable of interest, supportive evaluation, had a significant relationship with satisfaction. While in earlier arguments we suggest the importance of more supportive evaluations, as compared to summative, on teacher outcomes like satisfaction, this study did not seek to compare different forms of evaluation or predict how schools build supportive evaluation within broader working conditions. However, with these results, we do illustrate that a set of simultaneous, significant predictors at the teacher and school level represent working conditions which matter for teacher satisfaction—the nature of evaluation, climate, shared decision making, self-efficacy, professional development, and student-teacher ratio. This study extends the current literature by demonstrating that a supportive approach to teacher evaluation and a teacher's view of the usefulness of the feedback for positive changes in their practice help to explain variance in satisfaction within and across schools even after other working conditions of schools are taken into account.

Discussion

Central to this study was the assumption, rooted in self-determination theory, that the process of teacher evaluation most likely to result in positive benefits for teachers is when it is designed to support their psychological needs as learners and yields meaningful knowledge that can be directed towards meeting challenging goals for practice. The findings of this study provide some evidence for these assertions, revealing that there is a relationship between the perceptions of supportive teacher evaluation experiences and U.S. secondary teachers' satisfaction after controlling for other important teacher and school characteristics and working conditions. Beyond the basic features of the design and implementation of these supportive practices, we also found that the degree to which teachers experienced positive changes in their practice from evaluation was also related to their satisfaction—over two-and-one-half tenths of a standard deviation for every unit increase in perceptions of positive change. As with prior research, these findings suggest that teachers who find utility in the feedback they receive and can use this information to improve their

practice are more likely to be satisfied with the work they are doing and with their place in the profession (Ford et al., 2017; Hewitt, 2015; Ingersoll et al., 2016; Longo-Schmid, 2016).

With respect to prior literature on teacher satisfaction, other findings are important to note as well. Prior research has identified overall school climate as a measure of working conditions to influence satisfaction (Johnson, Kraft, & Papay, 2012; Malinen & Savolainen, 2016). The findings here support prior findings, while noting that the relationship of a principal's perception of the school climate to teacher satisfaction was much smaller as compared to teacher perceptions of the climate, student relations, and shared responsibility and decision making. Furthermore, teacher efficacy, also found in numerous prior studies to be related to teacher satisfaction (Malinen & Savolainen, 2016; Skaalvik & Skaalvik, 2010, 2014; von der Embse, Sandilos, Pendergast, & Mankin, 2016; Wang, Hall, & Rahimi, 2015) was also found to be the case in the TALIS sample as well. To our knowledge, few studies have examined the relationship between a principal's job satisfaction and the satisfaction of the teachers they lead. Interestingly, within the U.S. TALIS sample, principal satisfaction was not related to teachers' job satisfaction, though the link was a plausible one.

Furthermore, our findings demonstrated that average teacher satisfaction was likely higher in schools where someone other than the principal (or individual external to the school) was the primary evaluator—most particularly when the primary evaluator was a fellow teacher, mentor, or other member of school management. This aligns with some findings which suggest that who is evaluating teachers matters for the satisfaction they have with their rating and their perceptions of the fairness of the process (Delvaux et al., 2013). While principals can leverage their general pedagogical expertise to provide quality feedback, fellow teachers are more likely to be able to provide more frequent, subject-specific support (Kraft & Gilmour, 2016; Reinhorn et al., 2017). This does not minimize the role of principals in evaluation systems, but rather reinforces the potential for peer mentoring and instructional coaching to complement principal feedback within a more comprehensive system of teacher support. It also provides principals with opportunities to delegate evaluation responsibilities to others within the school, particularly if more frequent feedback is desired (Lavigne & Good, 2015). It is also possible that teacher evaluation designs that position fellow teachers as evaluators, might increase the likelihood that the evaluation process is seen as a collaborative one and the feedback generated from it safe, valid, and meaningful for improvement (Reddy et al., 2017). However, in the context of the TALIS data, the quality of the "teacher as evaluator" approach is, at best, unclear; thus our conclusions must be tempered accordingly.

A continuing area of concern for school leaders is the development of their instructional leadership and mentorship skills—historically a weaker area of leadership preparation. Self-determination theory suggests that providing feedback requires a delicate interplay between making it substantive and challenging while also staving off arousal that is debilitating to motivation for improvement (Ryan & Brown 2005; Ryan & Deci, 2017). Principals' ability to effectively engage in these instructional leadership practices lies in the nature of the relationship that principals cultivate with teachers (Blase & Blase, 1999). Effective instructional leaders express an authentic interest in teachers' growth and development and leverage trust and mutual respect to engage in supportive interactions (Blase & Blase, 1999; Drago-Severson, 2012). By providing extensive teacher support within a collaborative model of teacher evaluation, principals can promote a professional culture focused on continuous improvement (Reinhorn et al., 2017). If evaluation is seen as an extension of daily practices within a supportive, improvement-oriented context, principals can help minimize the anxiety around evaluation (Reinhorn et al., 2017).

Finding likely gains in job satisfaction among teachers who experience more supportive teacher evaluation processes not only begins to corroborate existing psychological and educational research into what types of experiences build upon teachers' existing intrinsic motivation for

learning, but it also points to a key lever for change in the fight to retain good teachers—helping to keep them satisfied in their school and job. The results suggest we continue to study the ways in which teacher evaluation policies are designed and to what extent they incorporate teacher professional learning and development. We acknowledge that some summative evaluation of performance is necessary, and even desirable. However, ensuring that the teacher evaluation is seen as a fair and valid measure of performance should go a long way to building its credibility as a formative tool for teacher improvement and to ensuring that it is a valued part of the improvement process of a school (Ford et al., 2017; Longo-Schmid, 2016).

One possible limitation to the overall results of the study concerns our inability to tease out the individual effects of the different design features of supportive teacher evaluation that might be most influential to satisfaction. To this, we would argue that it is perhaps best to view these various features as a complementary system for teacher evaluation, one in which the whole is likely greater than the sum of its parts. Furthermore, while we were able to capture some perspectives on the teacher evaluation from teachers experiencing them, the TALIS measures were not equipped to capture the often subtle ways that power is exerted within schools which might affect teachers' experiences and their feelings about evaluation. As some scholars have noted, the years since NCLB have led to a gradual acceptance of the intrusions into their professional lives brought about by top-down accountability policies (Holloway & Brass, 2017). This observation has important implications for future research in this area, as it suggests that the standard by which one reports feeling satsified with one's job might be shifting. Finally, because of our use of observational data and our investigation of supportive evaluation structures on teacher job satisfaction was exploratory in nature, it is important to emphasize that causal links cannot be made. Future research should establish more definitive causal linkages between teacher evaluation approaches and teacher job satisfaction.

Another important caveat to the above findings is the fact that our sample of teachers was limited to the lower-secondary level. Studies have revealed a small correlation between schooling level and satisfaction (Perie & Baker, 1997; Shen et al., 2012; Skaalvik & Skaalvik, 2011), and this coupled with the fact that elementary teachers report higher instances of great stress in a given week (Met Life Inc., 2013) suggest the effects of supportive teacher evaluation practice on satisfaction could be even stronger for elementary teachers. Either way, future research should examine the relationship of supportive teacher evaluation practice at other schooling levels (i.e., elementary and high school levels). One final limitation to our study is based upon recent work which has found that a teacher's satisfaction is causally related to their evaluation rating in Tennessee, on the order of around .08 standard deviations (Koedel, Li, Springer, & Tan, 2017). Because such a variable is unavailable in the TALIS dataset (teachers' prior evaluation ratings), we are necessarily unable to empirically test its relationship to teacher satisfaction. Whether this relationship holds across states due to the variation in the rigor and quality of evaluation systems between them is still an open question worthy of exploration.

Implementing a more supportive teacher evaluation system is easier said than done, however. As Young and Kim (2010) assert, using assessment for formative/supportive purposes is "…is not a beginner activity" (p. 9). While the above findings implicate the actions of states, districts, and also individual school leaders, in crafting teacher evaluation systems that are supportive of teachers' psychological needs, this will necessitate the concomitant growth and development of principals and their knowledge and skills as instructional leaders. While the psychological needs for competence, autonomy and relatedness are universal needs (Ryan & Deci, 2002), providing targeted competence support to each and every lower-secondary level teacher will require at least a modicum of content and pedagogical knowledge for a wide-range of subjects on the part of principals. What

leaders can do, beyond becoming more skilled instructional leaders, is to work to make the school climate one which is more conducive to teacher learning by providing more space and autonomy for teachers wishing to reflect more deeply on their practice, and by providing access to expertise like instructional coaches to assist them in their classroom (Camburn & Han, 2015; Marsh, McCombs, & Martorell, 2010). To meet the demands of a more supportive evaluation system, principals might need assistance in reprioritizing, shifting, or delegating other leadership responsibilities (Kraft & Gilmour, 2016).

Conclusion

Hallinger, Heck, and Murphy (2014) assert that faith in the transformative potential of teacher evaluation as a tool for school improvement has far outstripped the empirical evidence of its effectiveness. These scholars, as well as others, continue to emphasize the importance of making the process (and products) of teacher evaluation supportive of teachers' learning and development as practitioners (Delvaux et al., 2013; Ford et al., 2016; Hallinger et al., 2014; TNTP, 2010; Tuytens & Devos, 2011). Of course, being able to render a summative judgment about a teacher's effectiveness provides information useful for determining a general evaluation of effectiveness, but to affirm and activate teachers' motivation toward improvement, we submit that these systems must also be able to yield actionable knowledge for teachers as well as provide an appropriate support structure as teachers work to make meaningful changes to their practice (Ford et al., 2016). Good teaching is difficult work. As a cognitive process, satisfaction in one's work occurs when capabilities are well-matched with the challenge of a task (Csikszentmihalyi, 1990). An important focus of future policy and practice should attend to how we can apply what we know about teachers' needs for professional learning and development to the creation of an evaluation system which best supports teachers in meeting the challenges of teaching and learning in the 21st century.

References

Adams, C. A., Forsyth, P. B., Ware, J. K., & Mwavita, M. (2016). The informational significance of A-F school accountability grades. *Teachers College Record 118*(7). Retrieved from: http://www.tcrecord.org/Content.asp?ContentId=20925

Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-added Assessment System (SAS-EVAAS) in the Houston Independent School District (HISD): Intended and unintended consequences. *Educational Policy Analysis Archives, 20*(12). Retrieved from: http://dx.doi.org/10.14507/epaa.v20n12.2012

Blase, J., & Blase, J. (1999). Principals' instructional leadership and teacher development: Teachers' perspectives. *Educational Administration Quarterly*, *35*(3), 349-378.

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, *78*, 367-409. https://doi.org/10.3102/0034654308321455

Burch, P. (2009). *Hidden markets: The new education privatization*. New York: Routledge.

Campbell, D.T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*, 67–90.

Camburn, E. M., & Han, S. W. (2015). Infrastructure for teacher reflection and instructional change: An exploratory study. *Journal of Educational Change*, *16*, 511-533. https://doi.org/10.1007/s10833-015-9252-6

Cappelli, P., & Tavis, A. (2016, October). The performance management revolution. *Harvard Business Review,* 58-67.

Carver-Thomas, D., & Darling-Hammond, L. (2017). *Addressing California's growing teacher shortage.* Palo Alto, CA: Learning Policy Institute.

Cohen, D. K. (2011). *Teaching and its predicaments.* Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674062788

Cha, S-H., & Cohen-Vogel, L. (2011). Why they quit: A focused look at teachers who leave for other occupations. *School Effectiveness and Improvement, 22,* 371-392. https://doi.org/10.1080/09243453.2011.587437

Chow, A. P. Y., Wong, E. K. P., Yeung, A. S., & Mo, K. W. (2002). Teachers' perceptions of appraiser–appraisee relationships. *Journal of Personnel Evaluation in Education*, *16*(2), 85-101. https://doi.org/10.1023/A:1020096127713

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York: Harper & Row.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Teachers College Press.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15. https://doi.org/10.1177/003172171209300603

Davis, J., & Wilson, S. M. (2000). Principals' efforts to empower teachers: Effects on teacher motivation and job satisfaction and stress. *The Clearing House*, *73*, 349-353. https://doi.org/10.1080/00098650009599442

de Leeuw, D., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, *19*(2), 153-176.

Deci, E. L., & Ryan, R. M. (2000). The "what" and the "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry ,11*, 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, *36*, 1-11. https://doi.org/10.1016/j.tate.2013.06.011

Dinham, S., & Scott, C. (1998). A three domain model of teacher and school executive career satisfaction. *Journal of Educational Administration*, *36*, 362-378. https://doi.org/10.1108/09578239810211545

Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading, and learning.* Washington, DC: National Council on Teacher Quality.

Dolton, P., & Van der Klaauw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics*, *81*(3), 543-550. https://doi.org/10.1162/003465399558292

Dou, D., Devos, G., & Valcke, M. (2017). The relationships between school autonomy gap, principal leadership, teachers' job satisfaction and organizational commitment. *Educational Management Administration & Leadership*, *45*(6), 959-977. https://doi.org/10.1177/1741143216653975

Drago-Severson, E. (2012). New opportunities for principal leadership: Shaping school climates for enhanced teacher development. *Teachers College Record*, *114*(3), 1-44.

Enders, C. (2010). *Applied missing data analysis.* New York: Guilford Press.

Evans, L. (2001). Delving deeper into morale, job satisfaction and motivation among education professionals: Re-examining the leadership dimension. *Educational Management & Administration*, *29*, 291-306. https://doi.org/10.1177/0263211X010293004

Evetts, J. (2009). The management of professionalism: A contemporary paradox. In S. Gewirtz, P. Mahony, I. Hextall, & A. Cribb (Eds.), *Changing teacher professionalism: International challenges and ways forward.* London: Routledge.

Evetts, J. (2011). A new professionalism? Challenges and opportunities. *Current Sociology*, *59*, 406-422. https://doi.org/10.1177/0011392111402585

Eyal, O., & Roth, G. (2011). Principals' leadership and teachers' motivation: Self-determination theory analysis. *Journal of Educational Administration, 49*, 256-275. https://doi.org/10.1108/09578231111129055

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43*, 100-107. https://doi.org/10.3102/0013189X14521864

Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment and high-stakes teacher evaluation (HSTE) policy in Louisiana. *Educational Policy,* 31(2), 202-248. https://doi.org/10.1177/0895904815586855

Ford, T. G., Van Sickle, M. E., & Fazio-Brunson, M. (2016). The role of "informational significance" in shaping Louisiana elementary teachers' use of high-stakes teacher evaluation data for instructional decision making. In K. K. Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 117-135). New York: Palgrave Macmillan.

Ford, T. G., & Ware, J. K. (2018). Teacher Self-Regulatory Climate (TSRC): Conceptualizing an indicator of leader support for teachers' learning and development. *Leadership and Policy in Schools, 17*(1), 27-51. https://doi.org/10.1080/15700763.2016.1197283

Gelman, A., Hill, J., Su, Y-S, Yajima, M., & Pittau, M. J. (2015). *Package 'mi' for R.* Retrieved from: https://cran.r-project.org/web/packages/mi/mi.pdf

Goldring, R., Taie, S., & Riddles, M. (2014). *Teacher attrition and mobility: Results from the 2012-2013 Teacher Follow-up Survey* (Report NCES 2014-077). Washington, DC: U.S. Department of Education, Institute for Education Statistics, Schools and Staffing Survey. Retrieved from https://nces.ed.gov/pubs2014/2014077.pdf.

Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education*, *24*, 1349-1363. https://doi.org/10.1016/j.tate.2007.06.005

Grissom, J. A., Loeb, S., & Doss, C. (2016). The multiple dimensions of teacher quality: Does value-added capture teachers' nonachievement contributions to their schools? In J. A. Grissom & P. A. Youngs (Eds.). *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 37-50). New York: Teachers College Press.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444. https://doi.org/10.3102/0013189X13510020

Grissom, J. A., & Youngs, P. A. (2016). *Improving teacher evaluation systems: Making the most of multiple measures.* New York: Teachers College Press.

Hakanen, J., Bakker, A., & Schaufeli, W. (2006). Burnout and work engagement among teachers. *Journal of Psychology, 43*, 495-513. https://doi.org/10.1016/j.jsp.2005.11.001

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, *26*, 5-28. https://doi.org/10.1007/s11092-013-9179-5

Hargreaves, A. (2003). *Teaching in the knowledge society.* New York, NY: Teachers College Press.

Hargreaves, A. (2010). Presentism, individualism, and conservatism: The legacy of Dan Lortie's Schoolteacher: A sociological study. *Curriculum Inquiry*, *40*(1), 143-154. https://doi.org/10.1111/j.1467-873X.2009.00472.x

Harvey, J. (2014). Getting beyond the blame game. *Educational Leadership, 71*(5), 24-29.

Hewitt, K. (2015). Educator evaluation policy that incorporates EVAAS Value-Added Measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives, 23*(76). https://doi.org/10.14507/epaa.v23.1968

Hewitt, K., & Amrein-Beardsley, A. (2016). Introduction: The use of growth measures for educator accountability at the intersection of policy and practice. In K. Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 1-25). New York: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_1

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research.* Boulder, CO: National Education Policy Center.

Holloway, J., & Brass, J. (2017). Making accountable teachers: the terrors and pleasures of performativity. Advanced online publication, *Journal of Education Policy*. https://doi.org/10.1080/02680939.2017.1372636

Holloway, J., Sørensen, T. B., & Verger, A. (2017). Global perspectives on high-stakes teacher accountability policies: An introduction. *Education Policy Analysis Archives, 25*(85). http://dx.doi.org/10.14507/epaa.25.3325

Hulpia, H., Devos, G., & Rosseel, Y. (2009). The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment. *School Effectiveness and School Improvement*, *20*, 291-317. https://doi.org/10.1080/09243450902909840

Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal*, *44*(3), 493-518. https://doi.org/10.3102/0002831207306764

Ingersoll, R. (2003). *Who controls teachers? Power and accountability in America's schools.* Cambridge, MA: Harvard University Press.

Ingersoll, R., Merrill, L., & May, H. (2016). Do accountability policies push teachers out? *Educational Leadership*, *73*(8), 44-49.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, *44*, 105-116. https://doi.org/10.3102/0013189X15575517

Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school?. *Educational Researcher*, *44*(2), 117-126. https://doi.org/10.3102/0013189X15573351

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*, 1-39.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*(3), 376. https://doi.org/10.1037/0033-2909.127.3.376

Kelly, K. O., Ang, S. Y. A., Chong, W. L., & Hu, W. S. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration*, *46*, 39-54. https://doi.org/10.1108/09578230810849808

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*, 741-756. https://doi.org/10.1037/a0019237

Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology*, *36*, 114-129. https://doi.org/10.1016/j.cedpsych.2011.01.002

Koedel, C., Li, J., Springer, M. G., & Tan, L. (2017). The impact of performance ratings on job satisfaction for public school teachers. *American Educational Research Journal, 54,* 241-278. https://doi.org/10.3102/0002831216687531

Kokkinos, C. M., Panayiotou, G., & Davazoglou, A. M. (2005). Correlates of teacher appraisals of student behaviors. *Psychology in the Schools*, *42*, 79-89. https://doi.org/10.1002/pits.20031

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, *52*(5), 711-753. https://doi.org/10.1177/0013161X16653445

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476-500. https://doi.org/10.3102/0162373713519496

Kunter, M., & Holtzberger, D. (2014). Loving teaching: Research on teachers' intrinsic orientations. In P. W. Richardson, S. A. Karabenick, & H. M. G. Watt (Eds.), *Teacher motivation: theory and practice* (pp. 83-99). London: Routledge.

Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record, 116(1).* Retrived from: http://www.tcrecord.org/Content.asp?ContentId=17294

Lavigne, A. L., & Good, T. L. (2014). *Teacher and student evaluation: Moving beyond the failure of school reform.* New York: Routledge.

Lavigne, A. L., & Good, T. L. (2015). *Improving teaching through observation and feedback: Beyond state and federal mandates.* New York: Routledge.

Leithwood, K. A., Menzies, T., Jantzi, D., & Leithwood, J. (1999). Teacher burnout: A critical challenge for leaders. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout: A sourcebook of international research and practice* (pp. 85-114). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511527784.006

Leung, D., & Lee, W. (2006). Predicting intention to quit among Chinese teachers: differential predictability of the components of burnout. *Anxiety, Stress & Coping*, *19*(2), 129-141. https://doi.org/10.1080/10615800600565476

Linacre, J.M. (2017). Winsteps® (Version 3.93.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2017. Available from http://www.winsteps.com/

Longo-Schmid, J. (2016). Teachers' voices: Where policy meets practice. In K. Kappler Hewitt & A. Amrein-Beardsley (Eds.), S*tudent growth measures in policy and practice* (pp. 49-71). New York: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_3

Lortie, D. (1975). *Schoolteacher: A sociological analysis.* Chicago: University of Chicago Press.

Malinen, O. P., & Savolainen, H. (2016). The effect of perceived school climate and teacher efficacy in behavior management on job satisfaction and burnout: A longitudinal study. *Teaching and Teacher Education*, *60*, 144-152. https://doi.org/10.1016/j.tate.2016.08.012

Marsh, J. A., McCombs, J., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, *24*, 872-907. https://doi.org/10.1177/0895904809341467

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., & Epstein, S. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Rand Corporation.

Master, B. (2014). Staffing for success: Linking teacher evaluation and school personnel management in practice. *Educational Evaluation and Policy Analysis, 36*, 207-227. https://doi.org/10.3102/0162373713506552

McGregor, D. (1960). *The human side of enterprise.* New York: McGraw-Hill.

MetLife, Inc. (2013). *The MetLife survey of the American Teacher: Challenges for school leadership.* New York: Metropolitan Life Insurance Company.

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*(3), 193-212. https://doi.org/10.1023/A:1012752725765

Moe, A., Pazzaglia, F., & Ronconi, L. (2010). When being able is not enough: The combined value of positive affect and self-efficacy for job satisfaction in teaching. *Teaching and Teacher Education, 26,* 1145-1153. https://doi.org/10.1016/j.tate.2010.02.010

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, *42*, 349-354. https://doi.org/10.3102/0013189X13499625

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7*, 133-144. https://doi.org/10.1177/1477878509104318

Organization for Economic Co-Operation and Development. (2009). *Teacher evaluation. A conceptual framework and examples of country practices.* Retrieved from: http://www.oecd.org/edu/school/44568106.pdf

Organization for Economic Co-Operation and Development. (2010). *TALIS 2008 Technical Report.* Retrieved from: http://www.oecd.org/education/school/44978960.pdf

Organization for Economic Co-Operation and Development. (2014). *TALIS 2013 Technical Report.* Retrieved from: http://www.oecd.org/edu/school/TALIS-technical-report-2013.pdf.

Papaioannou, A., & Christodoulidis, T. (2007). A measure of teachers' achievement goals. *Educational Psychology*, *27*, 349-361. https://doi.org/10.1080/01443410601104148

Pearson, L. C., & Moomaw, W. (2005). The relationship between teacher autonomy and stress, work satisfaction, empowerment, and professionalism. *Educational Research Quarterly*, *29*(1), 38-54.

Perie, M., & Baker, D. P. (1997). Job satisfaction among America's teachers: Effects of workplace conditions, background characteristics, and teacher compensation. Statistical Analysis Report #NCES-97-471. Washington, DC: American Institutes for Research.

Perrachione, B., Rosser, V., & Petersen, G. (2008). Why do they stay? Elementary teachers' perceptions of job satisfaction and retention. *The Professional Educator*, *32*(2). Retrieved from: http://files.eric.ed.gov/fulltext/EJ862759.pdf.

Pillay, H., Goddard, R., & Wilss, L. (2005). Well-being, burnout and competence: Implications for teachers. *Australian Journal of Teacher Education*, *30*, 21-31. https://doi.org/10.14221/ajte.2005v30n2.3

Pogodzinski, B., Youngs, P., & Frank, K. A. (2013). Collegial climate and novice teachers' intent to remain teaching. *American Journal of Education*, *120*, 27-54. https://doi.org/10.1086/673123

Race to the Top Fund. (2016). Program. U.S. Department of Education. Retrieved from https://www2.ed.gov/programs/racetothetop/index.html.

Reddy, L. A., Dudek, C. M., Peters, S., Alperin, A., Kettler, R. J., & Kurz, A. (2017). Teachers' and school administrators' attitudes and beliefs of teacher evaluation: A preliminary investigation

of high poverty school districts. Advance online publication, *Educational Assessment, Evaluation, and Accountability*. https://doi.org/10.1007/s11092-017-9263-3

Reinhorn, S. K., Johnson, S. M., & Simon, N. S. (2017). Investing in development: Six high-performing, high-poverty schools implement the Massachusetts teacher evaluation policy. *Educational Evaluation and Policy Analysis*, *39*(3), 383-406. https://doi.org/10.3102/0162373717690605

Rosenholtz, S. J. (1991). *Teachers' workplace: The social organization of schools*. New York: Teachers College Press.

Roth, G. (2014). Antecedents and outcomes of teachers' autonomous motivation: A self-determination theory analysis. In P. W. Richardson, S. A. Karabenick, & H. M. G. Watt (Eds.), *Teacher motivation: theory and practice* (pp. 36-51). London: Routledge.

Ryan, R. M., & Brown, K. W. (2005). Legislating competence: The motivational impact of high-stakes testing as an educational reform. In. C. Dweck & A. Elliot (Eds.), *Handbook of competence and motivation* (pp. 354-372). New York: Guilford Press.

Ryan, R. M., & Deci, E. L. (2002). An overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester, NY: University of Rochester Press.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York: Guilford Press.

Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*, 224-233. https://doi.org/10.1177/1477878509104327

Ryan, S. V., Nathaniel, P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, *66*, 1-11. https://doi.org/10.1016/j.tate.2017.03.016

Saeki, E., Segool, N., Pendergast, L., & Embse, N. (2017). The influence of test-based accountability policies on early elementary teachers: School climate, environmental stress, and teacher stress. Advanced online publication, *Psychology in the Schools*. https://doi.org/10.1002/pits.22112

Sass, D., Seal, A., & Martin, N. (2010). Predicting teacher retention using stress and support variables. *Journal of Educational Administration*, *49*, 200-215. https://doi.org/10.1108/09578231111116734

Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *Journal of Politics, 52*, 510-529.

Schroeder, R. M., Akotia, C. S., & Apekey, A. K. (2001). Stress and coping among Ghanaian school teachers. *IFE Psychologia*, *9*(1), 89-98. https://doi.org/10.4314/ifep.v9i1.23600

Scott, C., Stone, B., & Dinham, S. (2001). I love teaching but.... international patterns of teacher discontent. *Education Policy Analysis Archives*, *9*(28). http://dx.doi.org/10.14507/epaa.v9n28.2001

Shen, J., Leslie, J. M., Spybrook, J. K., & Ma, X. (2012). Are principal background and school processes related to teacher job satisfaction? A multilevel study using Schools and Staffing Survey 2003-04. *American Educational Research Journal, 49,* 200-230. https://doi.org/10.3102/0002831211419949

Singh, J. (2004). Tackling measurement problems with Item Response Theory: Principles, characteristics, and assessment, with an illustrative example. *Journal of Business Research*, *57*(2), 184-208. https://doi.org/10.1016/S0148-2963(01)00302-2

Skaalvik, E. M., & Skaalvik, S. (2009). Does school context matter? Relations with teacher burnout and job satisfaction. *Teaching and Teacher Education*, *25*, 518-524. https://doi.org/10.1016/j.tate.2008.12.006

Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education, 26,* 1059-1069. https://doi.org/10.1016/j.tate.2009.11.001

Skaalvik, E. M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession: Relations within school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education, 27,* 1029-1038. https://doi.org/10.1016/j.tate.2011.04.001

Skaalvik, E. M., & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological Reports*, *114*, 68-77. https://doi.org/10.2466/14.02.PR0.114k14w0

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Springer, M. G., Pane, J., Vi-Nhuan, L., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. M. (2012). Team pay for performance: Experimental evidence from the round rock pilot project on team incentives. *Educational Evaluation and Policy Analysis, 34,* 367-390. https://doi.org/10.3102/0162373712439094

Strizek, G., Tourkin, S., & Erberber, E. (2014). *Teaching and Learning International Survey (TALIS) 2013: U.S. Technical Report (NCES 2015-010).* U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2015/2015010.pdf.

Sylvia, R. D., & Hutchison, T. (1985). What makes Ms. Johnson teach? A study of teacher motivation. *Human Relations*, *38*, 841-856. https://doi.org/10.1177/001872678503800902

Taylor, D. L., & Tashakkori, A. (1995). Decision participation and school climate as predictors of job satisfaction and teachers' sense of efficacy. *The Journal of Experimental Education*, *63*(3), 217-230. https://doi.org/10.1080/00220973.1995.9943810

The New Teacher Project. (2010). *Teacher Evaluation 2.0.* New York: Author.

Torres, A. C., & Weiner, J. M. (2018). The new professionalism? Charter teachers' experiences and qualities of the teaching profession. *Educational Policy Analysis Archives, 26*(19). http://dx.doi.org/10.14507/epaa.26.3049

Tuytens, M., & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader?. *Teaching and Teacher Education*, *27*, 891-899. https://doi.org/10.1016/j.tate.2011.02.004

Urick, A. (2016). The influence of typologies of school leaders on teacher retention: A multilevel latent class analysis. *Journal of Educational Administration*, *54*(4), 434-468. https://doi.org/10.1108/JEA-08-2014-0090

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

Van Maele, D., & Van Houtte, M. (2012). The role of teacher and faculty trust in forming teachers' job satisfaction: Do years of experience make a difference? *Teaching and Teacher Education*, *28*, 879-889. https://doi.org/10.1016/j.tate.2012.04.001

von der Embse, P., Sandilos, L. E., Pendergast, L., & Mankin, A. (2016). Teacher stress, teaching-efficacy, and job satisfaction in response to test-based educational accountability policies. *Learning and Individual Differences*, *50*, 308-317. https://doi.org/10.1016/j.lindif.2016.08.001

von der Embse, N. P., Schoemann, A. M., Kilgus, S. P., Wicoff, M., & Bowler, M. (2017). The influence of test-based accountability policies on teacher stress and instructional practices: a moderated mediation model. *Educational Psychology*, *37*(3), 312-331. https://doi.org/10.1080/01443410.2016.1183766

Wang, H., Hall, N. C., & Rahimi, S. (2015). Self-efficacy and causal attributions in teachers: Effects on burnout, job satisfaction, illness, and quitting intentions. *Teaching and Teacher Education*, *47*, 120-130. https://doi.org/10.1016/j.tate.2014.12.005

Watt, H. M. G., & Richardson, P. W. (2014). Why people choose teaching as a career: An expectancy-value approach to understanding teacher motivation. In P. W. Richardson, S. A. Karabenick, & H. M. G. Watt (Eds.), *Teacher motivation: theory and practice* (pp. 3-19). London: Routledge.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, NY: The New Teacher Project.

Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resource Management Review*, *12*(2), 173-194. https://doi.org/10.1016/S1053-4822(02)00045-1

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. E*ducation Policy Analysis Archives*, *18*(19). https://doi.org/10.14507/epaa.v18n19.2010

Yuan, K., Vi-Nhuan, L., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2012). Incentive pay programs do not affect teacher motivation or reported practices: Results from three randomized studies. *Educational Evaluation and Policy Analysis, 35,* 3-22. https://doi.org/10.3102/0162373712462625

Zeichner, K. (2010). Competition, economic rationalization, increased surveillance, and attacks on diversity: Neo-liberalism and the transformation of teacher education in the US. *Teaching and Teacher Education*, *26*(8), 1544-1552. https://doi.org/10.1016/j.tate.2010.06.004

Appendix

Table 3
Rasch Measure of Supportive Teacher Evaluation (SUPPEVAL)

Item (4-point scale, *strongly disagree* to *strongly agree*)	D	Infit	Outfit
Teacher appraisal and feedback are largely done to fulfill administrative requirements (R) (TT2G31C)	.56	1.12	1.11
The best performing teachers in this school receive the greatest recognition (TT2G31A)	.33	1.19	1.23
Feedback is provided to teachers based on a thorough assessment of their teaching (TT2G31E)	.05	.80	.79
A development or training plan is established for teachers to improve their work as a teacher (TT2G31D)	-.05	.78	.77
A mentor is appointed to help the teacher improve his/her teaching (TT2G31H)	-.13	1.37	1.16
Teacher appraisal/feedback have little impact upon the way teachers teach in the classroom (R)(TT2G31B)	-.25	1.08	1.10
Measures to remedy any weakness in teaching are discussed with the teacher (TT2G31G)	-.51	.79	.77

Note. Person separation reliability = .83; item reliability = .99. TALIS teacher questionnaire item numbers in parentheses. D = item difficulty. R = item reverse coded.

About the Authors

Timothy G. Ford
University of Oklahoma
tgford@ou.edu
http://orcid.org/0000-0003-2975-4191
Tim Ford is an assistant professor in the Department of Educational Leadership and Policy Studies at the University of Oklahoma. He examines the role of leadership in school improvement and teacher evaluation policy, particularly how leaders can support teacher occupational professionalization.

Angela Urick
University of Oklahoma
urick@ou.edu
http://orcid.org/0000-0002-8868-8297
Angela Urick is an assistant professor in the Jeannine Rainbolt College of Education, Department of Educational Leadership and Policy Studies, at the University of Oklahoma. Her research interests include principal and teacher perceptions of leadership, leadership styles, school climate, teacher retention, and school improvement..

Alison S. P. Wilson
University of Oklahoma
alisonwilson@ou.edu
Alison Wilson is doctoral candidate in the Jeannine Rainbolt College of Education, Department of Educational Leadership and Policy Studies, at the University of Oklahoma. Her research interests include the effects of school teacher and student learning conditions on achievement and school improvement.

education policy analysis archives

Volume 26 Number 59   April 30, 2018   ISSN 1068-2341

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join EPAA's Facebook community at https://www.facebook.com/EPAAAAPE and **Twitter feed** @epaa_aape.