



How Teachers Perceive the Impact of Teacher Feedback: A Latent Class Analysis

Yeonsoo Choi



Alex J. Bowers

Teachers College, Columbia University
United States

Citation: Choi, Y., & Bowers, A. J. (2024). How teachers perceive the impact of teacher feedback: A latent class analysis. *Education Policy Analysis Archives*, 32(38). <https://doi.org/10.14507/epaa.32.8464>

Abstract: How do teachers perceive and experience the impact of evaluation feedback across various domains of their work? This study investigates whether there are qualitatively distinct groups of teachers in the United States based on teacher perceived impact of evaluation feedback using the 2013 Teaching and Learning International Survey (TALIS 2013). Applying latent class analysis to the U.S. sample of lower-secondary school teachers ($n=1,824$), we find that there are four distinctive groups of teachers: Receptive (25%), Compliant (15%), Motivated (27%), and Indifferent (33%). Our analysis shows feedback sources and teacher beliefs about the evaluation's purpose are significant predictors of teacher class assignment. Finally, controlling for various individual and school-level characteristics, these teacher groups vary significantly in their intention to change school.

Keywords: teacher evaluation; evaluation feedback; teacher surveys; latent class analysis

Cómo perciben los docentes el impacto de la evaluación docente: Un análisis de clase latente

Resumen: ¿Cómo perciben y experimentan los docentes el impacto de las respuestas de evaluación en los distintos ámbitos de su trabajo? Este estudio investiga si existen grupos cualitativamente distintos de docentes en los Estados Unidos según el impacto percibido por

los docentes de las respuestas de evaluación utilizando la Encuesta Internacional de Enseñanza y Aprendizaje de 2013 (TALIS 2013). Al aplicar el análisis de clases latentes a la muestra estadounidense de docentes de escuelas secundarias inferiores ($n=1.824$), encontramos que hay cuatro grupos distintivos de docentes: receptivos (25%), obedientes (15%), motivados (27%) y Indiferente (33%). Nuestro análisis muestra que las fuentes de respuestas de la evaluación y las creencias de los docentes sobre el propósito de la evaluación son predictores importantes de la asignación de clases de los docentes. Finalmente, controlando por diversas características individuales y a nivel de escuela, estos grupos de docentes varían significativamente en su intención de cambiar de escuela.

Palabras-clave: evaluación docente; respuestas de evaluación; encuestas a docentes; análisis de clases latentes

Como os professores percebem o impacto da avaliação docente: Uma análise latente da aula

Resumo: Como é que os professores percebem e experienciam o impacto das respostas da avaliação nos vários domínios do seu trabalho? Este estudo investiga se existem grupos qualitativamente distintos de professores nos Estados Unidos com base no impacto percebido pelos professores das respostas da avaliação usando a Pesquisa Internacional de Ensino e Aprendizagem de 2013 (TALIS 2013). Aplicando a análise de classe latente à amostra de professores do ensino secundário inferior dos EUA ($n = 1.824$), descobrimos que existem quatro grupos distintos de professores: Receptivos (25%), Cumpridores (15%), Motivados (27%) e Indiferente (33%). A nossa análise mostra que as fontes de respostas da avaliação e as crenças dos professores sobre o propósito da avaliação são preditores significativos da atribuição das aulas dos professores. Finalmente, controlando diversas características individuais e a nível da escola, estes grupos de professores variam significativamente na sua intenção de mudar de escola.

Palavras-chave: avaliação docente; repostas de avaliação; pesquisas com professores; análise de classe latente

How Teachers Perceive the Impact of Teacher Feedback: A Latent Class Analysis

How do teachers perceive and experience the impact of evaluation feedback across various domains of their work? In the early 2010s, teacher evaluation reforms received much attention as a policy lever to change teacher behavior and improve student learning. At the heart of the teacher evaluation policy reforms were two main theories of change outlining how the new reforms will lead to their intended results (Firestone, 2014). On one hand, the first theory of change was heavily grounded in economics (Amrein-Beardsley & Holloway, 2019; Jabbar & Menashy, 2022), centering the institution of rewards and sanctions based on evaluation performance. In this framework, teacher feedback is directly linked to the distribution of extrinsic rewards and sanctions such as performance pay, and it is through this mechanism that teacher evaluation was to change teacher practices. On the other hand, the second theory of change revolved around the role of intrinsic motivation and direction (Firestone & Donaldson, 2019) and emphasized formative functions of evaluation. Under this account, feedback is the main source of intrinsic motivation and direction, and evaluations are most likely to impact teachers' instructional practices when high-quality, meaningful feedback teachers can use to improve their practices is provided (Ford et al., 2018; Kraft & Christian, 2022; Ridge & Lavigne, 2020).

Despite these well-established theories of change, teacher evaluation reforms have resulted in mixed success. Research examining the effects of teacher evaluation reforms suggests that to date, evidence of the effectiveness of the evaluation remains inconclusive (Bleiberg et al., 2021). Studies have also documented unintended negative consequences of evaluation reforms such as decreases in teacher perceived autonomy, morale, and job satisfaction (e.g., Anagnostopoulos et al., 2021; Ford et al., 2017, 2018).

To account for such results, education researchers have increasingly turned to the policy implementation and sensemaking literature. This body of work has challenged the proposed theories of change by demonstrating that whether teacher feedback actually informs teacher practices is a function of how teachers make sense of and experience the evaluations, with studies hinting at ceremonial policy compliance as the norm (e.g., Lane, 2020; Paufler & Sloat, 2020). The implication here is that teacher sensemaking provides an interpretive lens through which teachers further make sense of and act on evaluation feedback and that they are patterned across different contexts and settings. Despite these important insights, however, existing studies primarily draw from in-depth interviews or participant observation to characterize teacher conception, focusing on a relatively small sample of teachers and schools (Donaldson et al., 2016; Mintrop et al., 2018). It thus remains an open question whether there exist generalizable patterns of teacher perception around the impact of feedback, at a larger scale.

We address this gap by using (a) large-scale survey data tapping into multiple dimensions of teachers' engagement with feedback and (b) analytic methods that partition survey samples based on emerging relations among survey item responses and allow distinct subgroups to emerge inductively. We thus apply latent class analysis (LCA) to the U.S. sample from the 2013 Teaching and Learning International Survey (TALIS 2013). The survey includes a number of questions about teachers' perception of whether and to what extent evaluation feedback positively impacted their work.

In line with the proposed theories of change underlying the evaluation reforms, we first examine how teachers perceive the impact of feedback across various domains of teachers' work (e.g., extrinsic and intrinsic rewards) and use teacher survey responses to identify distinct teacher groups. Having identified distinct teacher groups, we then probe into some of the contextual and individual factors affecting teacher perception of feedback. In particular, we draw from the sensemaking literature which posits teacher perception as reflecting an interplay between individual teacher characteristics and broader school, policy, and social contexts (Spillane et al., 2002). Importantly, we explore individual teacher beliefs about the purpose of teacher evaluation between the identified groups and the relationship between different feedback sources and teacher perceived impact of feedback, informed by past studies suggesting such linkages (Firestone & Donaldson, 2019; Jennings, 2012; Kraft & Christian, 2022; Schneider et al., 2021). In other words, do teachers' evaluation experiences vary depending on the feedback sources and their beliefs about its purpose? Lastly, we anticipate that teachers with different perceptions of teacher feedback and its impacts will vary in their job satisfaction and more concretely, turnover intention. Existing research on the unintended consequences of teacher evaluation (e.g., Ford et al., 2018), in particular, gives us some reasons to suspect that teacher evaluation may be related to job satisfaction. We thus explore how each pattern of teacher understanding is associated with turnover intention, as one manifestation of job satisfaction.

Our research questions are as follows:

- 1) Are there distinct groups of U.S. lower-secondary school teachers based on the perceived impact of feedback across various domains of teachers' work?
- 2) To what extent do individual and school characteristics predict teacher assignment to each group?

- 3) Is there an association between teacher response to evaluation and turnover intention?

In the remainder of the article, we first summarize the literature on different theories of change regarding teacher evaluation reforms, paying attention to how teacher feedback may affect different domains of teachers' work. Then, we synthesize the existing empirical literature on teachers' perception of teacher evaluation and explore possible patterns in how teachers make sense of and respond to evaluations. We then describe our data and method. We then present results from the LCA, which identified four distinct groups of teacher understandings. Subsequently, we analyze the correlates of each group assignment and demonstrate that group assignment is associated with teachers' intention to change school, even after controlling for a range of individual and school characteristics. We end with a discussion of theoretical and practical implications of our findings.

Research Background

Theories of Change Informing Teacher Evaluation Reforms

Since their first adoption in Washington, DC, in 2009, teacher evaluation reforms have been rapidly adopted and implemented across the nation, both at the state and district levels (Bleiberg & Harbatkin, 2020). The new teacher evaluations typically combined multiple components such as value-added models, increased frequency of teacher observations, standard-based observation rubrics, various forms of evaluation methods, and more fine-grained performance labels (Steinberg & Donaldson, 2016). Existing research points to multiple and overlapping influences in shaping the spread of teacher evaluation reforms, including federal initiatives, local education agencies, and philanthropic organizations (Bleiberg & Harbatkin, 2020; Kraft & Gilmour, 2017; Reckhow et al., 2021).

What is clear from these accounts is that evaluation reforms have evolved to represent an amalgam of multiple, and sometimes conflicting, goals and objectives. Broadly speaking, the reforms were intended to serve both summative and formative purposes, with distinct theories of change dictating how they were to achieve each purpose. Notably, each theory of change operates with an implicit model regarding the connection between teacher beliefs, behaviors and evaluation feedback.

First, the summative function of evaluations can be summarized as removal or rewards (Popham, 2013). Hence, in summative evaluations, it is critical that individuals with weak performance are identified and if not improving, removed to enhance the overall quality of the workforce. Summative evaluations can also be used to distribute rewards to high performing teachers, affecting teachers' career prospects. As Firestone (2014) relates, according to this view, extrinsic incentives are a key driver of behavioral changes. Teachers are hypothesized to adjust their classroom instruction when faced with a possibility of penalty (e.g., expedited dismissal) or incentives (e.g., merit pay, career promotion). Strictly speaking, within this paradigm, teacher evaluation feedback need not translate into actionable suggestions on how to improve performance. Rather, it is sufficient that evaluations be used to rank individuals according to shared criteria.

Formative evaluations, on the other hand, are geared towards improvement and development. In contrast to summative evaluations wherein weaknesses in performance can lead to penalty, formative evaluations aim to improve upon those weaknesses and cultivate desirable practices via feedback. Moreover, built into the design of formative evaluations is the premise that people have an inherent desire to develop professionally, even in the absence of external factors (Ford et al., 2018). To fulfill their formative functions, then, evaluations must accompany high-

quality feedback that appropriately identifies weaknesses and strengths in performance (Kraft & Christian, 2022) and guides teachers' professional development.

Conceptually, it is plausible that teacher evaluations achieve both formative and summative purposes (Ford & Hewitt, 2020). For example, a teacher may be motivated to change her classroom practices by the prospect of tangible incentives such as performance pay, while simultaneously perceiving evaluation feedback as validating her professional identity and strengths as a teacher. In practice, however, this is difficult to achieve for a number of reasons. First, when evaluation tools are used simultaneously for development and accountability purposes, teachers are less likely to reveal and acknowledge their weaknesses lest such vulnerability threaten their job security (Popham, 1988). In other words, accountability purposes crowd out the affordances of an evaluation as a professional development tool (Mintrop et al., 2018).

Second, when evaluations include multiple design elements each reflecting a distinct purpose, teachers may misinterpret their respective intents (Spillane et al., 2002). For example, teacher evaluation systems typically pursue both summative and formative functions, with the aim of providing high-quality feedback and simultaneously using evaluation to weed out 'ineffective' teachers. Yet, teachers may not interpret these two design elements as serving separate and distinct purposes, instead viewing evaluation feedback as punitive or interpreting the overall intent behind evaluations as 'out to get them'. Third, even when teacher understandings of policy design elements align with the policymakers' intent, their understandings are likely to shift over time in response to the actual policy implementation and their experiences of it (Bradford & Braaten, 2018). For instance, while teachers may believe that evaluation feedback is designed to help them develop professionally, they may still feel that the actual feedback they received in their school was not accurate or useful, thus having a minimal impact on their work (Paufler & Sloat, 2020). Taken together, the existing literature suggests that whether and how teachers act on evaluation feedback is dependent on the meaning the feedback has for the teacher. In this paper, we thus explore to what extent the theories of change behind teacher evaluations align with teachers' actual perception of the impact of feedback.

Teachers' Understanding of Teacher Evaluation Feedback

Both the existing empirical research and logical extensions of the theories behind the evaluation reforms suggest that the meanings teachers ascribe to evaluation feedback are an important object of study. Several studies examine this topic in depth, using in-depth interviews, participant observation, and survey data. These studies demonstrate that while teachers generally express desire for professional development via the evaluation process, they have mixed perceptions about the actual usefulness and impact of the evaluation feedback they received (Braun & Youngs, 2020; Donaldson et al., 2016; Marsh et al., 2017; Pallas, 2021; Reinhorn et al., 2017).

Particularly relevant to this study, Marsh and colleagues (2017) found three types of teacher responses to a new teacher evaluation system in New Orleans: reflection, distortion, and compliance. In their study, the authors draw from a qualitative case study of eight different schools, with educator interviews as their primary data source. According to the authors, teachers and administrators engaged in reflection used evaluation processes as a means for improving teacher knowledge and practice. Teachers took the reform seriously and viewed evaluations as having a positive impact on their work. On the other hand, distortive responses entailed actively "engaging in strategic behaviors to appear effective according to evaluation criteria" (2017, p. 545) or gaming the system. Teachers who displayed such responses also tended to describe the purpose of the evaluation as summative or geared toward accountability. Reflective and distortive responses can then be thought of as representing opposite ends of a spectrum in terms of whether teacher

feedback effects positive changes in teacher practices and motivation. Additionally, reflective responses most closely resemble the logics of action embedded within arguments for formative evaluations: teachers perceive the evaluation as yielding meaningful feedback and hence incorporate it into their practice. Last in their typology, compliance refers to when educators adhere to the technical requirements of the policy but do not embrace its spirit. Teachers are not committed to using the system as a tool for improvement nor do they engage in strategic behaviors to distort their ratings. Rather, teacher evaluations were experienced as a symbolic ceremony with little stakes attached or as a facsimile with little buy-in (see also Duncan & Everitt, 2022).

How and why do teachers, then, come to hold divergent understandings of the impact of evaluation feedback? The literature points to several factors, oftentimes drawing from the sensemaking theory (Spillane et al., 2002); prior teacher values, beliefs and knowledge, school context including principal leadership (Donaldson & Firestone, 2021; Donaldson & Woulfin, 2018), and policy messages. For example, as Reinhorn and colleagues (2017) argue:

[A] teacher's response to a question about whether she found her evaluator's feedback helpful might include her views about what might be better (individual cognition), a judgment about whether her evaluator had experience teaching her subject (situated cognition), and what she thought the policy was supposed to accomplish (policy stimuli). (p. 392)

Building upon Marsh et al.'s (2017) findings and the insights from the sensemaking literature, Lane (2020) describes teachers' general responses to evaluation reforms as "compliant with punctuated distortion" (p. 39). Drawing from an ethnographic study of three schools, Lane reports that teachers momentarily tweaked technical or superficial elements of their lessons when needed to receive a higher evaluation rating. However, their day-to-day operations remained mostly unaffected. Lane attributes such responses to teachers' collective sensemaking of evaluation as high-stakes, performative, and decoupled from their everyday practices, which can be traced back to their past experiences with the prior evaluation system. That is, teachers' beliefs around the purpose of the evaluation provided context under which teachers' subsequent understandings and experiences of the evaluation unfolded.

Using survey and interview data, Ford and colleagues' (Ford, 2018; Ford et al., 2017, 2018) research also examines conditions under which teachers' experiences with evaluations may lead to more motivation, or inversely, demoralization. In doing so, these studies leverage self-determination theory (SDT) from psychology to provide explanations for patterns of teacher perceptions. According to Ford and colleagues (2018), "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" (p. 7). Their explanation highlights how teachers are driven by intrinsic motivation and psychological needs like competence, autonomy, and/or relatedness or belongingness (Ford et al., 2017). According to these authors, a key condition in explaining teachers' response to the evaluation feedback is informational significance, or whether evaluation feedback carries significant enough information to bear on their psychological needs.

Importantly, following self-determination theory, evaluation feedback with little informational significance may lead to teacher demoralization or may be considered as irrelevant at best. This is because "events have a motivating significance when the arousal they engender is debilitating or when they contain no inherent rationale for action" (Ford et al., 2018, p. 9), consistent with prior findings that show when evaluations are perceived as performative, teachers do not engage with actual feedback (Marsh et al., 2017). Because whether evaluation feedback carries informational significance or not is a subjective process, these studies provide further warrant for

studying teachers' conception of the utility and impact of evaluation feedback. Furthermore, these studies suggest teachers' experiences of evaluation can impact their job satisfaction and motivation. Indeed, existing literature complicates the posited theories of change behind evaluation reforms by demonstrating that teacher responses to evaluation feedback are shaped by multiple factors, including prior values, beliefs, and the informational significance of feedback, and that in general, policy compliance is the dominant perception.

Yet how prevalent is this perception among teachers? To address this question, researchers have used survey data to investigate how teachers perceive the purpose and impact of evaluations (Jiang et al., 2015; Kraft & Christian, 2022; Prado Tuma et al., 2018). For example, Jiang and colleagues (2015) found that during the first two years of reform implementation in Chicago Public Schools, most of the teachers reported their evaluation feedback to be helpful, with beginning and elementary teachers responding significantly more positively than experienced and high school teachers. On the other hand, analyzing survey data from Boston Public Schools teachers, Kraft and Christian (2022) found that even though the teachers generally viewed their evaluators as supportive of their professional growth, only a quarter reported that evaluation feedback led to instructional improvement. Together, the above bodies of work suggest that teachers' perceptions of the purpose and impact of the evaluations are far from monolithic.

Moreover, other important questions remain. First, even though qualitative studies analyzing observation and in-depth interviews provide thick descriptions of how teachers experience the various components of the evaluation processes, they are not equipped to answer to what extent these findings are generalizable. Accordingly, the overall prevalence and generalizability of the qualitatively distinct teacher perceptions of and responses to evaluation feedback have been frequently pointed out as questions for future research.

Second, studies using teacher survey data, while addressing the generalizability issue, typically examine teacher responses to individual survey items in isolation (Jiang et al., 2015) or use principal component analysis (Kraft & Christian, 2022). In doing so, these studies are making an implicit assumption that teacher responses map onto a one-dimensional spectrum. However, as the literature reviewed above suggests, teachers' perceptions do not likely fall neatly on an interval scale of negative to positive; the responses range from full buy-in to symbolic engagement. That the evaluation policies have different design components intended to affect different aspects of teachers' work further complicates how teachers experience evaluations and their impact.

Therefore, we build on and extend the existing literature by using an analytic approach that identifies distinct groups of survey respondents based on relations between survey item responses. Our motivating questions are: Do teachers hold heterogeneous understandings when it comes to the impact of teacher evaluation feedback? How do teacher responses to evaluations compare to the theories of change undergirding the evaluation reforms and different teacher responses identified in prior work such as Marsh et al.'s (2017) study? Further extending insights from existing studies documenting unintended consequences of teacher evaluations, we also ask: does teachers' intent to change school correlate with their perception of feedback?

We begin by using latent class analysis (LCA) to identify subsets of respondents, or subgroups. We then control for teacher and school characteristics to produce more accurate assignments to latent groups (Hagenaars & Halman, 1989); we use the automated Bolck, Croon and Hagenaars (hereafter BCH) three-step process in which latent classes are enumerated first and the relationships between latent classes and covariates are estimated at a later step. Doing so allows us to examine the extent to which individual attributes or contextual factors predict assignment to each class without the shifting of latent class assignment due to inclusion of covariates (Nylund-Gibson & Choi, 2018). As we discuss below, our selection of the covariates is guided by our literature review.

Lastly, we examine to what extent teachers assigned to different groups vary in their intention to change school beyond what one would expect based on teacher and school characteristics.

Data and Method

Data

Data for our study comes from TALIS 2013 administered by the Organisation for Economic Co-operation and Development (OECD). TALIS 2013, which surveyed a total of 34 countries including the U.S., focuses on teachers and school leaders working in lower secondary schools. TALIS asks an extensive set of questions about the learning environment and the working conditions of teachers in schools (Rutkowski et al., 2013). Important for our purpose, TALIS 2013 includes a wide range of questions intended to capture teacher evaluation practices beyond tangible rewards and sanctions, or those most aligned with summative evaluations. While TALIS 2018 included a relevant set of questions probing into teacher evaluation practices, it was less exhaustive than TALIS 2013, thus informing our decision to use TALIS 2013. Indeed, education researchers (e.g., Ford et al., 2018; Ford & Lavigne, 2023; Soncin et al., 2024) have used TALIS 2013 survey data and some of the same survey items as included in our study to examine teachers' perception of evaluation systems. Another benefit of using the 2013 TALIS data is that the timing of data collection (i.e., Spring 2013) coincides with the implementation of Race to the Top, when the U.S. states were implementing new evaluation systems (Steinberg & Donaldson, 2016). This feature allows us to examine how the theories of change map to teachers' perception of feedback during the initial rollout of evaluation reforms.

Data analysis in this study focuses only on the U.S. sample, with a sample size of approximately 1,900 teachers in 122 schools (Strizek et al., 2014). Our final sample for class enumeration included 1,824 teachers. To handle missing data for class enumeration, we relied on missing data imputation using Full Information Maximum Likelihood (FIML) in MPlus (Asparouhov & Muthén, 2013; Enders, 2010; Vermunt & Magidson, 2007). At the second step, where we include covariates and distal outcome to the model, the sample size was reduced to 1,395 due to listwise deletion of observations with missing values on covariates, a common practice in LCA (Nylund-Gibson & Choi, 2018). We also applied teacher weights such that our findings can be generalizable to the U.S. lower secondary school teacher population.

Our inclusion of indicator variables focuses on four domains of teachers' work affected by teacher evaluations and feedback: extrinsic incentives, out-of-classroom responsibilities, teaching and classroom management practices, and intrinsic rewards linked to teacher identity and motivation. We discuss each set of measures below. Table A1 in online Appendix A shows the descriptive statistics for each indicator variable. Of note, for the purpose of running LCA, which was designed for analyzing dichotomous data, each indicator variable (originally a four-point scale of no positive, a small, a moderate, and a large change) was dichotomized so that a value of 1 denotes 'a moderate/large change' and 0 'no positive/small change'.

We chose LCA over latent profile analysis—another mixture modeling method for continuous indicator variables—to better account for the relatively small size and the large number of indicator items in our sample (Dziak et al., 2014). In addition, although the survey items included in the study uses a 4-point scale, it is unclear whether the numbers (1–4) can be construed as reflecting changes on a continuous scale. We believe the indicator variables of no positive, a small, a moderate, and a large change arguably reflect changes in categories, not in a numerical scale, which partly informed our decision to use LCA over LPA. We return to the implications of this methodological choice later in this paper. There were 14 variables in total, and while certainly not

exhaustive, these capture important ways in which evaluation feedback may affect teachers' work, grounded in the theories of change undergirding evaluation reforms.

Extrinsic Incentives

As noted in the literature review above, extrinsic incentives were not only a direct outcome of teacher evaluation ratings but a core mechanism through which teacher evaluations affect classroom behaviors. Specific forms of extrinsic rewards considered in the design of teacher evaluation reforms include career ladder, social recognition, and performance-based merit pay (Firestone, 2014). In TALIS 2013, three relevant survey items included are: "Concerning the feedback you have received at this school, to what extent has it directly led to a positive change in [your public recognition from the principal and/or your colleagues; the likelihood of career advancement; your salary and/or financial bonus]?" Overall, teachers were more likely to report that teacher feedback led to a moderate or a large change in public recognition (42%) than the likelihood of career advancement (27%) or salary and/or financial bonus (12%).

Out-of-Classroom Responsibilities

Teacher feedback can also affect teachers' out-of-classroom work and responsibilities through various processes. One example is using feedback to identify appropriate professional development. Evaluation policies usually recommend evaluation feedback be used to direct teachers to appropriate professional development opportunities (Firestone & Donaldson, 2019). Evaluators can also assign teachers to various out-of-classroom roles after observing teachers' classroom performance, based on what they perceive as individual teachers' strengths. Three TALIS survey items capturing the impact of feedback on out-of-classroom responsibilities are as follows: "Concerning the feedback you have received at this school, to what extent has it directly led to a positive change in [role in school development initiatives; job responsibilities at this school; the amount of professional development]?" Approximately 30% to 40% reported that teacher feedback led to a moderate or large change across these areas.

Classroom Practices

Teacher evaluations are ultimately designed to improve teacher practices and performance. In TALIS 2013, five items examine the extent to which teacher feedback is perceived to positively change instructional practices. In TALIS 2013, survey respondents were asked to rate to what extent teacher appraisal feedback led to a positive change in their "classroom management practices", "knowledge and understanding of your main subject field(s)", "teaching practices," "methods for teaching students with special needs," and "use of student assessments to improve student learning". Half of the respondents believed that feedback led to a moderate or large change in their teaching practices (54%) and use of student assessments to improve student learning (50%). However, only 35% reported that their methods for teaching students with special needs improved as a result of teacher feedback. Further, 42% and 37% expressed that teacher feedback led to a moderate or large change in their classroom management practices and knowledge and understanding of their main subject field(s), respectively.

Intrinsic Rewards

A theory of action underlying formative evaluations posits internal rewards (e.g., motivation and job satisfaction) as the main mechanisms via which evaluations shape people's actions and practices. Teacher feedback, in particular, is 'the main source of intrinsic incentives coming from teacher evaluations' (Firestone, 2014, p. 105). Three relevant survey items in TALIS 2013 are: "Concerning the feedback you have received at this school, to what extent has it directly led to a

positive change in [confidence as a teacher; job satisfaction; motivation]?” A total of 60% reported their confidence as a teacher improving due to teacher feedback, whereas 48% and 52% indicated such changes in their job satisfaction and motivation.

Covariates

In addition to the core indicator items, our model includes a number of teacher and school covariates from both the teacher and principal questionnaires. Table A2 in online Appendix A displays descriptive statistics of our covariates and distal outcome.

First, teacher-level demographic variables included in the model are gender, years of teaching experience, and tenure status. Veteran teachers are less likely to view teacher feedback as having a strong impact as they may perceive evaluations as at odds with their professional identities (Pallas, 2021). Second, school attributes examined in this model are urbanicity (city vs. non-city), percentage of non-English speaker students, special education students, and students from disadvantaged homes. We include these variables, albeit rough proxies and only capturing a small range of related factors, to examine the interplay between individual teacher perceptions and the socioeconomic context of the schools they teach at (see also Ford et al., 2018). While there are multiple conditions and contextual factors that shape teachers' experience of evaluations, there is some evidence that schools and districts serving historically marginalized students offer teachers fewer opportunities to learn about teacher evaluation policies (Donaldson et al., 2016). Due to the limited data collected in TALIS 2013, our analysis does not include the racial/ethnic composition of the student population. It is also important to note that despite the inclusion of school-level attributes, our model is not a multilevel model. We opted to use a single-level model given the relatively small sample size and the large number of covariates and the exploratory nature of our analysis. We return to this point in our limitation section.

Third, and importantly, we also include four survey items that tap into teachers' beliefs about the purpose of evaluation. In particular, these items elicited the extent to which teachers perceive evaluations as supporting teacher development or accountability: “Teachers appraisal and feedback are largely done to fulfil administrative requirements,” “A development or training plan is established for teachers to improve their work as a teacher,” “If a teacher is consistently under-performing, he/she would be dismissed,” and “Measures to remedy any weaknesses in teaching are discussed with the teacher.” Respondents were asked to rate how strongly they agree with the above statements on a four-point scale of strongly disagree, disagree, agree, and strongly agree. Again, these items were dichotomized to account for the small sample size, with 0 indicating strongly disagree or disagree, and 1 as agree or strongly agree. It is noteworthy that even though 61% of the respondents endorsed the view of teacher evaluation as an administrative requirement, even a higher percentage (72%) reported that measures to remedy weaknesses in teaching are discussed with the teacher in their school.

Fourth, we included various sources of feedback as covariates as evaluation systems in the United States are typically designed to use a variety of feedback sources (Firestone & Donaldson, 2019). Feedback sources included in the TALIS 2013 survey are: direct observation of classroom teaching, student surveys about their teaching, assessment of content knowledge, analysis of students' test scores, teachers' self-assessment of their work, and survey or discussions with parents or guardians. The original survey items ask who in the school uses the following methods to provide feedback. For our study, we focus on whether teachers indicate the use of the practice in their school or not. We note that classroom observation was excluded from the covariates for the parsimony of the model, as more than 94% of the teachers in the original survey sample reported the use of classroom observation in their schools.

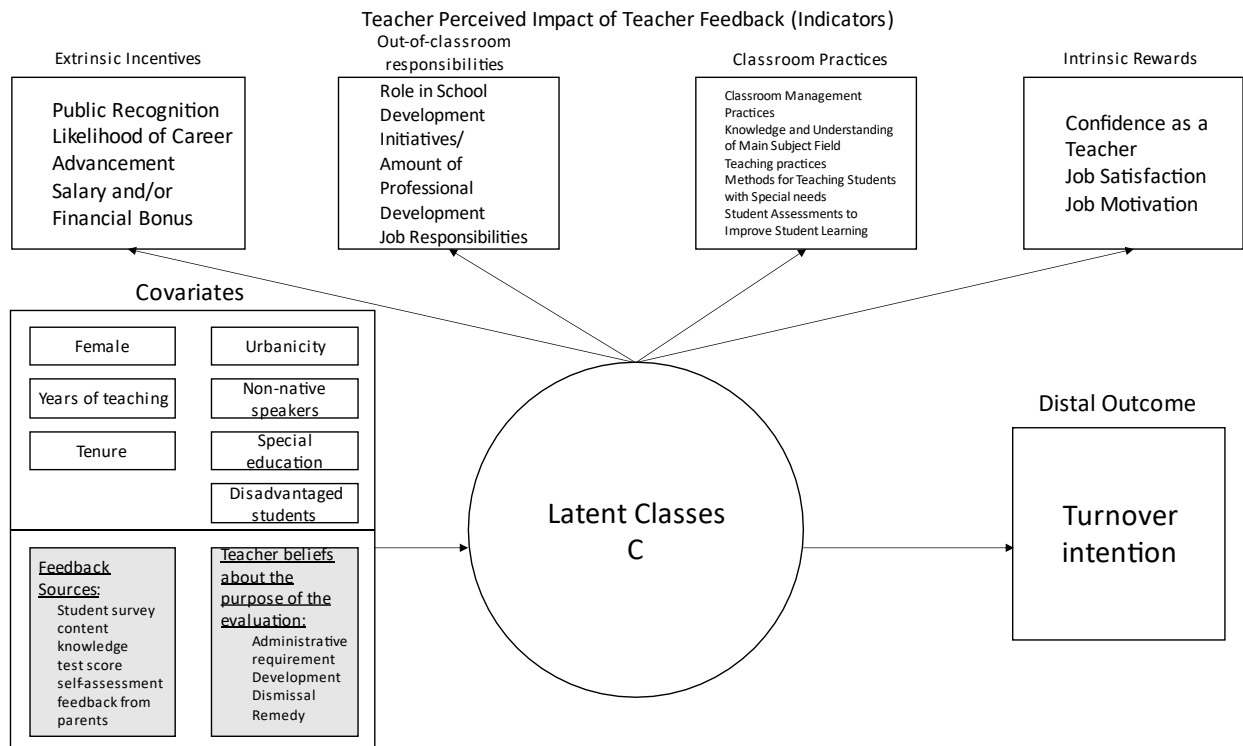
Lastly, based on prior work establishing a relationship between teachers’ evaluation experience and decreased professional commitment, we examined how the teacher group membership identified in our study correlates with teachers’ turnover intention. Though teacher turnover has more formal definitions (Holme et al., 2018), in this study, we use a TALIS survey item that asks about teachers’ intention to change school (see also Qin, 2019). Teachers were asked to rate how strongly they agree or disagree with the following statement: “I would like to change to another school if that were possible.”

Analytic Approach

To address our research questions, we begin by using latent class analysis (LCA). LCA is a person-centric method that is an extension of mixture modeling used to identify latent classes based on a set of observed indicators (Collins & Lanza, 2009; Lanza & Rhoades, 2013). Put simply, LCA is designed for cases in which the observed distribution of survey responses reflects the mixing of two or more statistically different groups. In LCA, there are both meaningful qualitative and quantitative differences among latent classes, or subsets of respondents. As the research questions here center on examining different groups of teachers, LCA is especially well-suited for our research. Figure 1 presents our structural equation model.

Figure 1

Structural Equation Model for Latent Class Analysis of Teachers Reporting Positive Impact of Feedback



Labeled as “Latent Classes C,” the different teacher groups are determined based on the 14 indicator variables described above, which can be regrouped as falling into: Extrinsic incentives, Out-of-classroom responsibilities, Classroom practices, Intrinsic rewards. We then

add 12 individual teacher covariates (identified on the left side of Figure 1 as female, years of teaching, tenure, feedback sources, and teacher beliefs about the purpose of the evaluation) and four school-related covariates (urbanicity, non-native speakers, special education, and disadvantaged students) as control variables. Then, to examine the relationship between teacher groups and turnover intention, we included teachers' intention to change school as a distal outcome (or dependent variable).

The main parameters estimated by the LCA model include the percentage of participants belonging to each class or the relative class proportions, and probabilities of response to each observed indicator variable conditional on latent class membership. The response profiles for each class are then visualized to help characterize each latent class and further drive our interpretation of each latent class, or group (see Figure 2).

Following the latent class analysis literature, we implemented a three-step LCA structural equation modeling framework, specifically, the automated BCH three-step process (Bakk et al., 2013; Nylund-Gibson & Choi, 2018; Nylund-Gibson et al., 2019). First, we performed an LCA using the indicator variables to determine the number of statistically different groups of teachers. In this step, the best-fitting k -class model is determined based on several fit statistics through iterative procedures. To determine model fit, one typically runs multiple models, each with one more class than the previous, and then the model fit statistics of all the resulting models are compared to select the model with the best fit.

While there is no one best method for deciding on the number of latent classes to be estimated, some of the most used statistics are the Bayesian Information Criterion (BIC) and the Lo-Mendell-Rubin (LMR) adjusted likelihood test. When performing the analysis, the BIC of the k class model is compared to the BIC of the $k-1$ class model. The lower the BIC, the better the model fit. Thus, when the specified model has a larger BIC value than the previous, the previously selected model is the best fit. Similarly, the LMR adjusted likelihood test also compares the fit between two neighboring class models (k vs. $k-1$ class model). Again, when the p -value of the test is not significant, this indicates the previously selected model is the best fit. When selecting an LCA model, however, the analyst must consider fit statistics along with interpretability and substantive significance. After selecting the best model fit, each member in the sample is assigned to its most likely latent class, or subgroup, adjusting for classification errors.

In the second step, the 16 covariates are added to the model with the group membership preserved from the first step. This allows us to perform multinomial logistic regression, which estimates the odds of an individual belonging to a group based on the covariates. In the third and final step, we followed a chi-square testing procedure to produce distal outcomes. All statistical procedures were conducted in Mplus, Version 8.4 (Muthén & Muthén, 2017). The Mplus code used for the analysis is included in online Appendix B.

Results

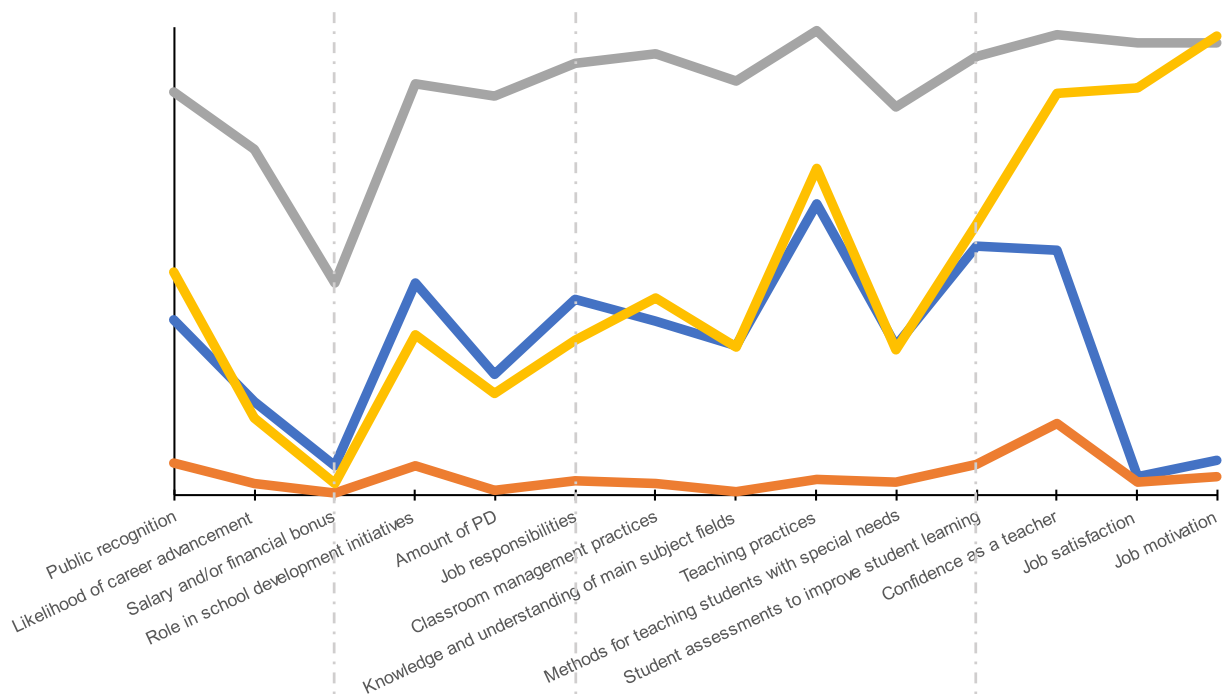
To find the best model fit, we tested a set of iterative models. Table A3 in online Appendix A presents the model statistics for each model. The first non-significant p -value of the LMR test occurred at the five-class model ($p=0.712$), suggesting that the four-class model is the best fit (AIC=22505.682, BIC=22830.700, and entropy= 0.884). The first positive change in BIC, on the other hand, occurred between the eight and nine-class models, suggesting that the eight-class model is the best fit. Ultimately, we chose the more parsimonious four-class model with a significant LMR, as the LMR fit statistic is the more conservative of the measures, especially when the sample size is relatively small (Dziak et al., 2014). Table A4 in online Appendix A displays the classification

probabilities for latent class membership (i.e., the probability of an individual belonging to a particular group when fitting the model). The classification probabilities of the four-class model were high, ranging from 0.854 to 0.965, lending further credibility to the four-class model.

We now characterize four different groups of teachers based on their perceived impact of feedback with reference to class-specific probabilities of responses to the 14 indicators. We named the four identified subgroups of teachers Receptive, Indifferent, Motivated, and Compliant. Of the four groups, two (Receptive and Indifferent) were extreme—notable for high and low values on most of the indicator items. Two other groups—Motivated and Compliant—were less easily placed on a continuum and differed significantly from each other regarding how each group of teachers perceived the impact of feedback on intrinsic rewards. Figure 2 shows the class-specific response probabilities. We also present the relationship of the covariates to class membership by reporting the means and odds ratios for the covariates in Table 1, with the Receptive group as the reference category. We synthesize these results to better describe each group and its attributes.

Figure 2

Indicator Plot of LCA Result by Group



Note: The original survey question for these indicators is “Concerning the feedback you have received at this school, to what extent has it directly led to a positive change in any of the following?” (no positive/small change = 0, a moderate/large change = 1)

Receptive

The Receptive group (25% of all respondents) viewed teacher feedback as having positive impacts across all indicator variables. Not only were they more likely to view feedback as leading to a moderate or large change in extrinsic incentives but also in intrinsic rewards. Notably, teachers belonging to this group were much more likely to report a positive change in their salary and/or financial bonus (45%), a number well above the sample average of 12%. Nearly all the Receptive

group teachers reported improved confidence as a teacher, job satisfaction, and job motivation. The Receptive teachers also outpaced other groups in their perception of the positive impact of teacher feedback on classroom practices; almost every respondent in this group perceived teacher feedback as positively changing their teaching practices (99%) and classroom management practices (95%). They similarly viewed teacher feedback as positively affecting their out-of-classroom work: role in school development initiatives (88%), amount of professional development (85%), and job responsibilities (92%).

Who were these teachers who experienced teacher feedback as positively impacting teachers' work across different domains, and what schools did they work in? As illustrated in Table 1, the typical teacher in this group had less teaching experience compared to the other groups. In relation to evaluation practices in their school, these teachers were much more likely to receive feedback from various sources. Importantly, they were also likely to agree that in their school, measures to remedy any weaknesses in teaching are discussed with the teacher. In other words, they were likely to view teacher evaluations and feedback as serving formative purposes.

Indifferent

The largest group with 33% of the teachers, Indifferent teachers, was least likely to view teacher feedback as leading to positive change in teachers' work. It is especially noteworthy that almost no teachers in this group reported feedback leading to moderate or large positive changes in salary or financial bonus (0.4%), classroom management practices (2.3%), and knowledge and understanding of the main subject fields (0.6%). Not surprisingly, the majority of Indifferent teachers expressed no or small changes happening as a result of feedback when it came to intrinsic rewards (85 to 96%). Together, these patterns suggest that for this group of teachers, teacher feedback neither leads to meaningful changes in their practices nor bears relevance to their self-understandings as a teacher.

The typical teacher in this class is more likely to be a veteran teacher and view teacher evaluation and feedback as being conducted for administrative requirements. In particular, controlling for teacher- and school-level characteristics, teachers who agree that teacher evaluations and feedback are largely conducted for administrative requirements were almost three times more likely to be in the Indifferent than the Receptive group. Indifferent teachers are also less likely to receive feedback based on content knowledge, self-assessments, or parent feedback. Compared to the Receptive group, Indifferent teachers were also less likely to work in urban schools and schools with a higher percentage of students receiving special education services.

Motivated

Unlike Receptive and Indifferent teachers, the two remaining groups did not fall into a monotonic continuum from low to high. One group, which we labeled Motivated teachers, consisted of respondents who expressed moderate levels of positive changes in extrinsic incentives, out-of-classroom responsibilities, and classroom practices. Yet they reported high levels of change in intrinsic rewards. Comprising 27% of the teachers, the Motivated group was the second largest in our sample, closely followed by the Receptive group. Only 2% of Motivated teachers answered that teacher feedback led to a positive change in salary and/or financial bonus, similar to the Indifferent group. However, unlike the Indifferent group, Motivated teachers were more likely to report a change in teaching practices as a result of teacher feedback (70%). Further setting apart the Motivated group from the Indifferent, almost all teachers in this group reported that teacher feedback resulted in moderate or large changes in their confidence as a teacher, job satisfaction, and motivation.

Table 1*Means and Odds Ratios for Covariates with Receptive as the Reference Group*

Variable	Receptive (25%)		Compliant (15%)			Indifferent (33%)			Motivated (27%)		
	Mean	OR	Mean	OR	<i>P</i>	Mean	OR	<i>P</i>	Mean	OR	<i>p</i>
Teacher-level											
Demographic											
Female	0.63	–	0.63	0.58	0.004**	0.70			0.70		
Years of Teaching	12.75	–	13.94	1.04	0.005**	14.70	1.05	<.001***	14.22	1.03	0.010*
Tenure	0.70	–	0.66			0.70			0.69		
Feedback Sources											
Student Survey	0.41	–	0.27			0.18			0.22	0.46	<.001***
Content Knowledge	0.69	–	0.49	0.56	0.008**	0.31	0.28	<.001***	0.50	0.51	0.001**
Test Score	0.80	–	0.69			0.56			0.67		
Self-Assessment	0.72	–	0.50	0.49	<.001***	0.36	0.46	<.001***	0.52	0.52	<.001***
Feedback from Parents	0.61	–	0.38			0.28	0.59	0.013*	0.41		
Purpose of the evaluation											
...Administrative requirement	0.44	–	0.65			0.77	2.95	0.007**	0.55		
...Development or training plan	0.74	–	0.63			0.38	0.52	<.001***	0.66		

...Dismissal	0.57	–	0.45			0.36			0.47	0.68	0.034*
...Remedy any weaknesses in teaching	0.89	–	0.72	0.36	<.001***	0.54	0.25	<.001***	0.76	0.53	0.017*
School-level											
Urbanicity	0.45	–	0.37			0.31	0.59	0.032*	0.35		
% Non-primary Language											
1% to 10%	0.61	–	0.61			0.62			0.63		
10% to 30%	0.14	–	0.15			0.10			0.11		
30% to 60%	0.08	–	0.05			0.06			0.04		
>60%	0.09	–	0.10			0.09			0.07		
% Special Education											
10% to 30%	0.59	–	0.61			0.60			0.61		
30% to 60%	0.05	–	0.03			0.02	0.321	0.001**	0.03		
% Disadvantaged Students											
10% to 30%	0.21	–	0.20			0.22			0.22		
30% to 60%	0.34	–	0.37			0.38			0.39		
>60%	0.35	–	0.32			0.29			0.26		

Note: OR=odds ratio; * $p < .05$; ** $p < .01$; *** $p < .001$

Who were these Motivated teachers? Motivated teachers were more experienced than Receptive teachers and 76% of this group indicated that following teacher evaluations, plans to remedy any weaknesses in teaching are discussed in their school. Interestingly, and somewhat puzzlingly, after controlling for teacher and school characteristics, teachers who agree that a consistently underperforming teacher is likely to be dismissed in their school were 1.5 times (1/0.68) less likely to be Motivated than Receptive. While we cannot explore the reason behind this association in the current study, it is at least possible that when teachers' own assessment of who the underperforming teachers are align with results from the schoolwide teacher evaluations, teachers' trust in the validity of evaluation and feedback may improve. Trust in the validity of evaluations then may lead to positive changes across multiple domains of teachers' work, which characterizes the Receptive group's experience. Compared to the Receptive group, Motivated teachers were also less likely to report that student surveys, assessment of content knowledge, and teachers' self-assessment are used to provide feedback in their school.

Compliant

We labeled the other group Compliant because these teachers reported changes in classroom practices but not in intrinsic rewards. The Compliant group, constituting 15% of the sample, resembles the Motivated group in many ways. Both groups reported similar levels of changes across extrinsic incentives, out-of-classroom responsibilities, and classroom practices. These commonalities notwithstanding, the differences between Compliant and Motivated teachers are most evident in their responses to the questions about whether teacher feedback led to a positive change in their confidence as a teacher (52% versus 86%), job satisfaction (4% versus 87%) and motivation (7% versus 98%). In other words, although Compliant teachers reported changes in their classroom practices in response to teacher feedback just as much as the Motivated teachers did, they did not view such feedback as leading to changes in intrinsic rewards. In fact, they were similar to the Indifferent group, although less extreme, in assessing the impact of teacher feedback on intrinsic motivation and satisfaction.

What factors predict the odds of teachers belonging to the Compliant group? Years of teaching experience and teacher beliefs about the purpose of teacher evaluations again emerged as significant predictors. In particular, teachers who reported that measures to remedy any weaknesses in teaching are discussed with them in their school were 2.7 times less likely to be a Compliant than a Receptive teacher. Teachers whose schools provided feedback based on assessments of content knowledge and self-assessments of their work were 1.8 times and 2 times less likely to be in the Compliant than the Receptive group, respectively.

Turnover Intention

Lastly, we present the association between teacher group assignment and turnover intention in Tables 2 and 3. Previous research gives us much reason to anticipate that different understandings of teacher feedback and evaluations are linked to teachers' turnover intention as teachers may become demoralized by what they perceive as heightened teacher accountability (Brass & Holloway, 2021; Ford et al., 2017, 2018). Unsurprisingly, both Indifferent and Compliant teachers were significantly more likely than Receptive teachers to report that they would like to change to another school. This pattern held true even when compared to the Motivated group. These patterns provide evidence for a significant relationship between teacher perception of feedback and teacher turnover intention. Moreover, these findings further bolster the external validity of the four different groups identified by the LCA analysis, serving as a robustness check, as the above effects are observed after controlling for teacher and school characteristics.

Table 2*Teacher Intent to Change School by Class Controlling for Model Covariates*

Teacher Group	Mean	S.E.
Receptive	0.142	0.020
Compliant	0.227	0.034
Indifferent	0.309	0.024
Motivated	0.131	0.021

Table 3*Pairwise Comparison of Class-Specific Means for Teacher Intent to Change School Adjusting for Model Covariates*

Teacher Group	Motivated	Indifferent	Compliant
Receptive	0.011	-0.167***	-0.085*
Compliant	0.096*	-0.082	
Indifferent	0.178***		

Note: Significance tests are Pearson chi-square; * $p < .05$; ** $p < .01$; *** $p < .001$

Discussion and Conclusion

We aimed to explore whether there is a distinct pattern or a typology of teachers' perceptions of the impact of teacher feedback. To address this question, we used LCA—a relatively underused method in education research—to analyze large-scale survey data with extensive information on teachers' perceptions of teacher evaluations and feedback. While we drew from theory and existing literature to guide our variable selection, LCA has a significant inductive component. Rather than assuming a fixed number of groups or types of teachers a priori, LCA allows distinct groups to emerge from the data itself, much like qualitative research examining distinct patterns in how teachers perceive and respond to evaluations. At the same time, as LCA can be applied to larger datasets, our analysis moves toward answering to what extent teacher response patterns are generalizable across the U.S., complementing insights drawn from qualitative research on teacher evaluation implementation. We further sought to understand how numerous teacher and school characteristics predict assignment to latent groups. We also examined the relationship between teachers' perceptions and their intention to change to another school.

Several findings emerged from our analyses. First, using LCA, we identified four qualitatively different groups of teachers who hold diverging views about the impact of teacher feedback: Receptive, Indifferent, Motivated, and Compliant. The Receptive group reported consistently high levels of changes in response to teacher feedback across various domains of teachers' work. Going back to our literature review (e.g., Firestone, 2014), these teachers seemed to have experienced teacher evaluation feedback as both formative and summative. Almost half of the Receptive teachers

reported that feedback led to a positive change in salary and/or financial bonus. Nearly all the Receptive teachers also indicated that their intrinsic motivation improved. Standing at the opposite end of the spectrum, Indifferent teachers were the most skeptical of the teachers. Their experiences were characterized by consistently low assessments of whether teacher feedback led to a positive impact in multiple domains of teachers' work. For these teachers, teacher evaluation feedback failed to serve either formative or summative purposes.

At the same time, our LCA analyses lend support to the notion that teachers' understandings of the impact of teacher feedback cannot be defined around a simple opposition between positive and negative. In this regard, our analyses build on findings from previous qualitative research on the topic (e.g., Lane, 2020; Marsh et al., 2017). Indeed, only about half of the respondents (60%) belonged to either the Receptive or Indifferent group. The rest of the teachers, whom we describe as Motivated or Compliant, differ markedly not only from the Receptive and Indifferent groups but also from each other. These groups reported little to no positive impact of teacher feedback on extrinsic incentives; yet, approximately 60% to 70% of both groups still viewed teacher feedback as positively impacting their teaching practices. However, whereas the Motivated group reported positive changes in intrinsic rewards due to teacher feedback, Compliant teachers did not appear to have become motivated.

These findings suggest that there are multiple pathways or mechanisms through which teacher evaluations, feedback specifically, are linked to changes in teacher practices. For example, our results tentatively suggest that changes in teaching practices need not accompany changes in teachers' intrinsic motivation. For example, Compliant teachers experience positive changes in their teaching practices, but not always changes in their job satisfaction and motivation as a teacher. Echoing prior research, this group of teachers were less inclined to believe that teacher evaluations provided opportunities to address weaknesses in teaching and less likely to report that their feedback was based on the assessment of content knowledge and self-assessment of their work. While the TALIS 2013 data and measures used in our analysis cannot provide definitive answers, it is possible that for this group of teachers, evaluation feedback led to minor improvements in their teaching practices but did not fundamentally shape or alter their attitudes and beliefs toward teaching practices. This interpretation aligns with findings from qualitative research on teacher evaluation policy implementation emphasizing policy compliance. At the same time, changes in teacher practices may precede changes in teacher motivation and job satisfaction (Spillane et al., 2002). Although we cannot adjudicate this question due the cross-sectional nature of TALIS 2013, it remains a possibility.

Second, our study also sheds light on the individual teacher and school factors that predict teacher understanding of evaluation feedback. Among teacher demographic variables examined in our study, years of teaching experience emerged as a significant predictor, with more experienced teachers less likely to be Receptive teachers. As the sensemaking literature would predict, individual teacher beliefs and feedback sources were also significant predictors of latent class assignment. Teachers who reported that teachers' self-assessment is used in their school for teacher evaluations were significantly more likely to be a Receptive teacher. The same was also found for the assessment of content knowledge. When it comes to teachers' beliefs about the purpose of evaluations, controlling for numerous teacher and school characteristics, teachers who indicated that in their school, measures to remedy any weaknesses in teaching are discussed with the teachers were more likely to be Receptive teachers. We also highlight that Indifferent teachers were much more likely to view teacher evaluations as a mere administrative requirement and less likely to believe that teacher evaluations are intended to support teachers' professional development, compared to the Receptive group.

Finally, we found that the groups identified by the LCA analyses are meaningfully associated with teachers' turnover intention, even after controlling for numerous teacher and school-based characteristics. Our analysis thus contributes to the small but growing literature on the impact of teacher evaluations on 'teachers' psychological needs' (Ford et al., 2018). This has implications for both those involved in designing and implementing teacher evaluations. Evaluation reforms especially during the Obama administration foregrounded ostensible rewards and sanctions as key drivers of changes in teacher practices. Yet decades of research on the teaching profession has pointed out that teachers derive feelings of satisfaction from intrinsic rewards of teaching and professional autonomy (Hargreaves & Goodson, 1996). For those designing evaluation policies, the key to changing teacher practices may not lie in monetary rewards but in creating conditions that encourage teachers to enact their professionalism. Similarly, for school leaders who oversee teacher evaluation processes, it is important that evaluations are conducted in a manner that allows teachers to perceive and experience the positive impacts of evaluations firsthand (Amrein-Beardsley, 2020).

Our findings have other practical and policy implications for school administrators and policymakers committed to creating meaningful evaluation experiences for teachers. For example, this study suggests that sources used to provide evaluation feedback influence how teachers view their impact and that evaluations incorporating different feedback sources may increase their appeal to the teachers. In other words, the types of sources used to provide feedback matter for teacher perception. Our findings also show that teachers who found their feedback most impactful believe that feedback is intended to remedy their weaknesses. While we cannot answer whether teacher beliefs precede teacher perception or vice versa, our findings corroborate the interplay between the two. Thus, for those interested in redesigning teacher evaluations, creating opportunities for collective sensemaking around the purpose of teacher evaluations is important. Likewise, what happens after the evaluation also matters. Teachers are more likely to perceive feedback as impactful when they believe discussions on how to remedy their weaknesses in teaching will ensue. Given this finding, the integration of professional development activities and programs into teacher evaluations is an area of continued discussion and research.

Limitations and Directions for Future Research

Despite these theoretical and practical implications, this research suffers from some limitations. First, to account for the relatively small sample size and a large number of indicator variables, we chose to dichotomize our indicator variables. In general, it is not desirable to collapse information and while in this study this was done to account for the sample size, we propose that future research examining similar topics with bigger datasets or in other national contexts employ methods such as latent profile analysis, which is designed for analyzing continuous indicator variables.

Second, although the TALIS 2013 dataset was one of the most comprehensive datasets we found, it was still imperfect, thereby preventing us from carrying out more complete analyses. For example, teachers' perceptions may have been affected by their evaluation ratings. However, due to the unavailability of such information in TALIS 2013, we were unable to pursue this question in our study. However, given that an overwhelming majority of teachers received high evaluation ratings and continue to do so even after the evaluation reforms (Kraft & Gilmour, 2017), it might be the case that there is not a wide variation in our sample. Nonetheless, without actual data on teacher evaluation ratings, this remains a speculation and we encourage future work examining how teacher evaluation ratings color teachers' interpretations of the reform (see also Pallas, 2021).

Third, we were unable to control for various design features of the evaluation system, which likely varied across districts and states. This is a critical limitation of our research, especially given

that whether a teacher receives a financial bonus as a result of teacher evaluation feedback is almost entirely dependent on the evaluation policy of the district she works in. Likewise, teachers' assessments of the impact of evaluation feedback are likely to be heavily influenced by policy design features. Even though we included feedback sources as our covariates, we acknowledge that this covers but one small component of evaluation designs and is insufficient. For example, a recent study suggests that feedback from fellow teachers versus principals may be received more positively (Ford & Lavigne, 2023). We thus invite researchers to examine additional design features of teacher evaluations. Fourth, our study is based on single-level latent class analysis as an exploratory first step and we invite further multilevel latent class analysis studies (e.g., Soncin et al., 2024) accounting for the nested nature of teachers' evaluation experience. Fifth, and lastly, due to our use of observational data and the inductive nature of our analysis, we cannot and do not make any causal claims about the relationships between teacher conception and their turnover intention. This, too, is an area of future research.

Acknowledgements

The first author would like to thank Na Lor and Aaron M. Pallas for their thoughtful comments on earlier drafts.

References

- Amrein-Beardsley, A. (2020). Policies and practices of promise in teacher evaluation: The introduction to the special issue. *Education Policy Analysis Archives*, 28(55). <https://doi.org/10.14507/epaa.28.5443>
- Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy*, 33(3), 516–542. <https://doi.org/10.1177/0895904817719519>
- Anagnostopoulos, D., Wilson, S., & Charles-Harris, S. (2021). Contesting quality teaching: Teachers' pragmatic agency and the debate about teacher evaluation. *Teaching and Teacher Education*, 98, 103246. <https://doi.org/10.1016/j.tate.2020.103246>
- Asparouhov, T., & Muthén, B. (2013). *Auxiliary variables in mixture modeling: 3-step approaches using Mplus*. [Working paper]. Mplus. <https://www.statmodel.com/download/3stepOct28.pdf>
- Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43(1), 272–311. <https://doi.org/10.1177/0081175012470644>
- Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M. A., & Springer, M. G. (2021). *The effect of teacher evaluation on achievement and attainment: Evidence from statewide reforms* [EdWorkingPaper: 21-496]. Annenberg Institute at Brown University. <https://www.edworkingpapers.com/ai21-496>
- Bleiberg, J., & Harbatkin, E. (2020). Teacher evaluation reform: A convergence of federal and local forces. *Educational Policy*, 34(6), 918–952. <https://doi.org/10.1177/0895904818802105>
- Bradford, C., & Braaten, M. (2018). Teacher evaluation and the demoralization of teachers. *Teaching and Teacher Education*, 75, 49–59. <https://doi.org/10.1016/j.tate.2018.05.017>
- Brass, J., & Holloway, J. (2021). Re-professionalizing teaching: The new professionalism in the United States. *Critical Studies in Education*, 62(4), 519–536. <https://doi.org/10.1080/17508487.2019.1579743>

- Braun, A. M. B., & Youngs, P. (2020). How middle school special and general educators make sense of and respond to changes in teacher evaluation policy. *Education Policy Analysis Archives*, 28(59). <https://doi.org/10.14507/epaa.28.5013>
- Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons. <https://doi.org/10.1002/9780470567333>
- Donaldson, M. L., & Firestone, W. (2021). Rethinking teacher evaluation using human, social, and material capital. *Journal of Educational Change*, 22(4), 501–534. <https://doi.org/10.1007/s10833-020-09405-z>
- Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going “rogue”: How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis*, 40(4), 531–556. <https://doi.org/10.3102/0162373718784205>
- Donaldson, M. L., Woulfin, S., LeChasseur, K., & Cobb, C. D. (2016). The structure and substance of teachers’ opportunities to learn about teacher evaluation reform: Promise or pitfall for equity? *Equity & Excellence in Education*, 49(2), 183–201. <https://doi.org/10.1080/10665684.2016.1144831>
- Duncan, C. P., & Everitt, J. G. (2022). Instantiated recoupling in principals’ enactment of teacher evaluations: Emotion work and new forms of ceremonial conformity in educational institutions. *Symbolic Interaction*, 45(4), 495–516. <https://doi.org/10.1002/symb.612>
- Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for the Bootstrap Likelihood Ratio Test in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 534–552. <https://doi.org/10.1080/10705511.2014.919819>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43(2), 100–107. <https://doi.org/10.3102/0013189X14521864>
- Firestone, W. A., & Donaldson, M. L. (2019). Teacher evaluation as data use: What recent research suggests. *Educational Assessment, Evaluation and Accountability*, 31(3), 289–314. <https://doi.org/10.1007/s11092-019-09300-z>
- Ford, T. G. (2018). Pointing teachers in the wrong direction: Understanding Louisiana elementary teachers’ use of Compass high-stakes teacher evaluation data. *Educational Assessment, Evaluation and Accountability*, 30(3), 251–283. <https://doi.org/10.1007/s11092-018-9280-x>
- Ford, T. G., & Hewitt, K. (2020). Better integrating summative and formative goals in the design of next generation teacher evaluation systems. *Education Policy Analysis Archives*, 28(63). <https://doi.org/10.14507/epaa.28.5024>
- Ford, T. G., & Lavigne, A. L. (2023). Does it matter who evaluates teachers? Principal versus teacher-led evaluation and teacher motivation. *Educational Policy*. <https://doi.org/10.1177/08959048221103796>
- Ford, T. G., Urick, A., & Wilson, A. S. P. (2018). Exploring the effect of supportive teacher evaluation experiences on U.S. teachers’ job satisfaction. *Education Policy Analysis Archives*, 26(59). <https://doi.org/10.14507/epaa.26.3559>
- 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 policy in Louisiana. *Educational Policy*, 31(2), 202–248. <https://doi.org/10.1177/0895904815586855>
- Hagenaars, J. A., & Halman, L. C. (1989). Searching for ideal types: The potentialities of latent class analysis. *European Sociological Review*, 5(1), 81–96. <https://doi.org/10.1093/oxfordjournals.esr.a036509>

- Hargreaves, A., & Goodson, I. (1996). Teachers' professional lives: Aspirations and actualities. In I. Goodson, *Teachers' professional lives* (pp. 9–35). Routledge.
<https://doi.org/10.4324/9780203453988-2>
- Holme, J. J., Jabbar, H., Germain, E., & Dinning, J. (2018). Rethinking teacher turnover: Longitudinal measures of instability in schools. *Educational Researcher*, 47(1), 62–75.
<https://doi.org/10.3102/0013189X17735813>
- Jabbar, H., & Menashy, F. (2022). Economic imperialism in education research: A conceptual review. *Educational Researcher*, 51(4), 279–288. <https://doi.org/10.3102/0013189X211066114>
- Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record*, 114(11), 1–23. <https://doi.org/10.1177/016146811211401108>
- Jiang, J. Y., Spote, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105–116.
<https://doi.org/10.3102/0013189X15575517>
- Kraft, M. A., & Christian, A. (2022). Can teacher evaluation systems produce high-quality feedback? An administrator training field experiment. *American Educational Research Journal*, 59(3), 500–537. <https://doi.org/10.3102/00028312211024603>
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting “the widget effect”: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234–249.
<https://doi.org/10.3102/0013189X17718797>
- Lane, J. L. (2020). Maintaining the frame: Using frame analysis to explain teacher evaluation policy implementation. *American Educational Research Journal*, 57(1), 5–42.
<https://doi.org/10.3102/0002831219848689>
- Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science*, 14(2), 157–168.
<https://doi.org/10.1007/s11121-011-0201-1>
- Marsh, J. A., Bush-Mecenas, S., Strunk, K. O., Lincove, J. A., & Huguet, A. (2017). Evaluating teachers in the big easy: How organizational context shapes policy responses in New Orleans. *Educational Evaluation and Policy Analysis*, 39(4), 539–570.
<https://doi.org/10.3102/0162373717698221>
- Mintrop, R., Ordenes, M., Coghlan, E., Pryor, L., & Madero, C. (2018). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly*, 54(1), 3–46.
<https://doi.org/10.1177/0013161X17696558>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440–461. <https://doi.org/10.1037/tps0000176>
- Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6), 967–985.
<https://doi.org/10.1080/10705511.2019.1590146>
- Pallas, A. M. (2021). The rhetoric of teacher evaluation: New York City teachers' responses to performance labels. *Educational Policy*. <https://doi.org/10.1177/08959048211049432>
- Paufler, N. A., & Sloat, E. F. (2020). Using standards to evaluate accountability policy in context: School administrator and teacher perceptions of a teacher evaluation system. *Studies in Educational Evaluation*, 64, 100806. <https://doi.org/10.1016/j.stueduc.2019.07.007>
- Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(3), 269–273.
<https://doi.org/10.1007/BF00123822>

- Popham, W. J. (2013). On serving two masters: Formative and summative teacher evaluation. *Principal Leadership (Middle Level Ed.)*, 13(7), 18.
- Prado Tuma, A., Hamilton, L., & Tsai, T. (2018). *How do teachers perceive feedback and evaluation systems? Findings from the American Teacher Panel*. RAND Corporation. <https://doi.org/10.7249/RB10023>
- Qin, L. (2019). Factors relating to teachers' intention to change school: A multilevel perspective. *Policy Futures in Education*, 17(3), 318–338. <https://doi.org/10.1177/1478210318822184>
- Reckhow, S., Tompkins-Stange, M., & Galey-Horn, S. (2021). How the political economy of Knowledge production shapes education policy: The case of teacher evaluation in federal policy discourse. *Educational Evaluation and Policy Analysis*, 43(3), 472–494. <https://doi.org/10.3102/01623737211003906>
- 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>
- Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback: A review of the literature. *Education Policy Analysis Archives*, 28(61). <https://doi.org/10.14507/epaa.28.5023>
- Rutkowski, D., Rutkowski, L., Bélanger, J., Knoll, S., Weatherby, K., & Prusinski, E. (2013). *The TALIS 2013 conceptual framework*. Organisation for Economic Co-operation and Development (OECD).
- Schneider, J., Noonan, J., White, R. S., Gagnon, D., & Carey, A. (2021). Adding “student voice” to the mix: Perception surveys and state accountability systems. *AERA Open*, 7, 233285842199072. <https://doi.org/10.1177/2332858421990729>
- Soncin, M., Bowers, A. J., & Agasisti, T. (2024). A global perspective on school leadership: Evidence from a latent class analysis on OECD TALIS data. *Educational Management Administration & Leadership*, 17411432241231902. <https://doi.org/10.1177/17411432241231902>
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431.
- Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, 11(3), 340–359. https://doi.org/10.1162/EDFP_a_00186
- Strizek, G. A., Tourkin, S., & Erberber, E. (2014). *Teaching and Learning International Survey (TALIS) 2013: U.S. technical report* (NCES 2015-010). U.S. Department of Education, National Center for Education Statistics. <http://nces.ed.gov/pubsearch>
- Vermunt, J. K., & Magidson, J. (2007). Latent class analysis with sampling weights: A maximum-likelihood approach. *Sociological Methods & Research*, 36(1), 87–111. <https://doi.org/10.1177/0049124107301965>

About the Authors

Yeonsoo Choi

Teachers College, Columbia University

yc3661@tc.columbia.edu

<https://orcid.org/0009-0009-8668-0468>

Yeonsoo Choi is a doctoral student in the Department of Education Policy and Social Analysis at Teachers College, Columbia University. Her interests include cultural sociology, the sociology of teaching, and critical policy analysis.

Alex J. Bowers

Teachers College, Columbia University

Bowers@tc.edu

<http://orcid.org/0000-0002-5140-6428>

Alex J. Bowers is a professor of education leadership at Teachers College, Columbia University, where he works to help school leaders use the data in more effective ways to help direct the limited resources of schools and districts to specific student needs. His research focuses on the intersection of effective school and district leadership, organization and HR, data-driven decision making, student grades and test scores, student persistence and dropouts.

education policy analysis archives

Volume 32 Number 38

July 23, 2024

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the

derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. **EPAA** is published by the Mary Lou Fulton Teachers College at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

About the Editorial Team: <https://epaa.asu.edu/ojs/index.php/epaa/about/editorialTeam>

Please send errata notes to Jeanne M. Powers at jeanne.powers@asu.edu
