Environment and Social Psychology (2025) Volume 10 Issue 10 ISSN: 2424-8975 (O)
doi: 10.59429/esp.v10i10.4056 2424-7979 (P)

RESEARCH ARTICLE

A comparative study of teacher evaluation's impact on professional identity: Evidence from secondary school teachers in China and Japan using TALIS 2013

Xue Lian Wang[1], I-Hua Chen*

[1] *International College, Krirk University, Bangkok, 10110, Thailand*

*** Corresponding author:** I-Hua Chen, aholechen@gmail.com

ABSTRACT

China and Japan share similar contexts for reforming teacher evaluation in basic education. However, existing studies on teacher evaluation in both countries suffer from limited sample sizes and methodological limitations. This study addresses these gaps by employing large-scale empirical data and multilevel modeling to re-examine the influence of teacher evaluation systems. Drawing on self-determination theory (SDT), we establish a multilevel model examining how teacher evaluation influences teachers' professional identity through both mediation and moderation mechanisms involving basic psychological needs. Using TALIS 2013 data, our findings reveal that: (1) satisfaction of basic psychological needs (autonomy, competence, and relatedness) positively correlates with teachers' professional identity in both countries; (2) the pathways of evaluation effects differ markedly between countries - for Chinese teachers, developmental evaluation indirectly enhances professional identity by first promoting perceived participation in school decision-making (autonomy need) and collegial cooperation (relatedness need), while for Japanese teachers, performance evaluation directly diminishes professional identity without mediation through psychological needs; (3) psychological need satisfaction also moderates evaluation effects - for Chinese teachers, higher teaching self-efficacy (competence need) amplifies the positive effect of developmental evaluation on professional identity, while for Japanese teachers, higher collegial cooperation (relatedness need) ironically intensifies the n

.0egative effect of performance evaluation on professional identity. These findings highlight the complex interplay between teacher evaluation systems and psychological need satisfaction across different cultural contexts.

Keywords: teacher evaluation; Japan; China; TALIS 2013; professional identity; self-determination theory

1. Introduction

The evolution of education policies across nations reveals that performance-based teacher evaluation, promoted by the United Kingdom and United States in the 1980s, has profoundly influenced basic education systems globally[1]. However, evaluation systems linking performance responsibilities to teacher salaries have generated considerable controversy, particularly in East Asian countries with Confucian traditions that traditionally emphasize teachers' spiritual development over material compensation. In addition to policy design and institutional structures, cultural and historical factors provide crucial context for understanding

ARTICLE INFO
Received: 4 August 2025 | Accepted: 17 October 2025 | Available online: 29 October 2025

CITATION
Wang XL,Chen IH.A Comparative Study of Teacher Evaluation's Impact on Professional Identity: Evidence from Secondary School Teachers in China and Japan Using TALIS 2013. *Environment and Social Psychology* 2025; 10(10): 4056. doi:10.59429/esp.v10i10.4056

how teacher evaluation influences professional identity in China and Japan. Rooted in the Confucian tradition, both societies have long regarded teachers not only as transmitters of knowledge but also as moral exemplars and spiritual guides[2,3]. This cultural tradition emphasises "teaching as a mission," whereby teachers' professional identity is more closely tied to moral authority, social responsibility, and spiritual fulfilment than to material rewards. Confucian educational values, through their emphasis on the teaching vocation, moral cultivation, and collective responsibility, shape expectations of the teaching profession in both countries: teachers are required not only to impart knowledge but also to "serve as role models" and uphold order and public morality. As a result, competence often encompasses moral and social responsibilities rather than being limited to pedagogical skills. At the same time, the Confucian emphasis on authority and hierarchy situates teachers within strong institutional and social expectations. When external mechanisms of quantifiable accountability and performance-driven evaluation intrude, teachers' sense of autonomy is easily constrained, undermining their intrinsic motivation and professional identity[4-6]. In the Chinese context, Confucian-related work values have been found to exert a dual influence: on the one hand, perseverance and responsibility enhance teachers' sense of competence, but under conditions of highly quantitative performance pressure, these same values may also increase the risk of burnout [7]. Moreover, Chinese scholars have further argued that the recent performance-oriented evaluation system tends to prioritise quantifiable indicators such as students' examination results, which not only risks undermining teacher dignity and eroding intrinsic motivation, but is also perceived by many teachers as unfair and overly dependent on external student performance metrics [8-10]. In the Japanese context, research shows that the introduction of teacher evaluation policies has reinforced the demand for performance and quantification, leading teachers to feel instrumentalised and to experience destabilisation of their professional identity[6].

Japan, sharing cultural and educational similarities with China, has similarly implemented new teacher evaluation systems emphasizing the correspondence between performance evaluation results and teacher salaries[11]. Japanese scholars echo concerns raised by their Chinese counterparts, arguing that new performance evaluation systems typically focus on student academic performance, often leaving teachers feeling like pawns in an evaluation system with diminished professional identity[12,13,14,15].

Despite these shared concerns, existing empirical evidence remains inconclusive. Research in both countries predominantly relies on small-scale qualitative interviews [16,17,12,13], yielding inconsistent results. Furthermore, quantitative studies rarely account for the nested nature of data or employ appropriate statistical methods, potentially compromising the accuracy of research conclusions [18].

As countries worldwide continue deepening teacher assessment reforms, clarifying the effects of different evaluation types and examining their respective mechanisms through large-scale empirical data and appropriate statistical analysis becomes urgent. Comparative analysis of Chinese and Japanese data offers particularly meaningful insights, given their similar educational reform histories and comparable challenges in implementing new teacher performance evaluation systems.

This study conducts a secondary analysis using data from the 2013 Teaching and Learning International Survey (TALIS 2013) conducted by the Organisation for Economic Co-operation and Development (OECD)[19,20]on secondary schools and teachers in China and Japan. We identify two major gaps in current research and formulate research questions accordingly.

1.1. Inconsistent findings

Empirical studies examining teacher performance evaluation in China and Japan have yielded contradictory results. Zhang and Ng conducted in-depth interviews with principals and teachers in three Shanghai public schools, concluding that teacher performance evaluation systems demonstrated positive

effects[16]. Liu and Zhao similarly found through interviews and surveys in two Beijing basic education schools that linking evaluation to salary did not reduce teacher motivation; conversely, evaluations without material consequences were perceived as superficial[18].

In contrast, Katsuno interviewed Japanese secondary and primary school teachers and principals, finding that while school leaders acknowledged evaluation's role in facilitating teacher communication, teachers generally maintained negative attitudes toward evaluation systems[12,13]. These divergent findings may reflect substantial inter-school variations in basic education systems, highlighting the need for large-scale quantitative research with broader scope and enhanced generalizability.

Large-scale quantitative studies have also produced mixed results. A 2005 survey by the Tokyo Metropolitan Teachers' Union and Board of Education found that 57.4% of senior high school teachers believed newly introduced performance evaluation weakened collegial cooperation[14]. Conversely, Ichinose and Yamazaki surveyed administrators in 78 middle schools in Shizuoka Prefecture, finding that teacher performance evaluation improved job satisfaction and collegial cooperation[22]. Rong'sstudy of 2,950 teachers across three Chinese provinces showed that evaluation helped teachers achieve positive teaching outcomes while causing evaluation-related exhaustion[8]. Liu et al. examined secondary school teachers in China's northernmost province, finding that higher perceived positivity in teacher evaluation correlated with increased job satisfaction[18].

1.2. Methodological limitations in previous quantitative studies

Most quantitative studies fail to account for nested data characteristics and employ inappropriate statistical methods. Teacher evaluation studies typically use cluster sampling, selecting schools first, then sampling teachers within schools (e.g., Liu et al. selected four middle schools in one city[18]). While convenient for questionnaire distribution, this approach overlooks the dependency between samples created by teachers within the same school experiencing identical evaluation systems. Consequently, traditional statistical tests used in these analyses would violate independence assumptions, potentially biasing results[23].

Additionally, studies focusing on teacher-perceived variables accompanying performance evaluation explore relationships with teacher attitudes and well-being [18,24]. However, when data sources are limited to teachers and items appear in the same questionnaire, common method bias may confound variable associations. For instance, Liu et al. examined relationships between teachers' negative perceptions of performance evaluation and job satisfaction[18], while Chang and Chang analyzed whether teachers' trust in evaluators affected satisfaction with evaluation systems[24]. The exclusive reliance on teacher-reported data prevents excluding potential confounding from common source variance.

1.3. Study aims

To overcome limitations in previous quantitative studies' data analysis, we employ hierarchical linear modeling (HLM) to analyze the nested data of school teacher evaluation and teacher-perceived variables. The teacher evaluation data at the school level uses principals as the information source, while teacher-perceived variables come from teachers themselves, which helps avoid bias from common method variance. In our HLM, teacher professional identity serves as the dependent variable, given its significance as a source of teachers' work motivation and a key factor in becoming an effective teacher. For the explanatory variables, we draw on self-determination theory (SDT) to highlight the empirical model structure. Specifically, we set the school's teacher evaluation system as explanatory variables at the school level, and teachers' satisfaction of basic psychological needs as mediators and moderators at the teacher level. SDT is particularly suitable for addressing teacher evaluation issues in this study, especially for explaining relationships between external rewards, satisfaction of individuals' basic psychological needs, and motivational variables.

2. Theoretical framework

2.1. Self-determination theory

SDT distinguishes three motivation types: amotivation, controlled motivation, and autonomous motivation [25]. Controlled motivation derives from environmental stimuli beyond individual control, while autonomous motivation originates from inherent individual drives supported by intrinsic interest and enjoyment. SDT conceptualizes motivation as a continuum, with individuals internalizing external rules and values, potentially transforming controlled motivation into autonomous motivation.

According to SDT, motivation internalization depends on environmental satisfaction of three basic psychological needs: autonomy, competence, and relatedness. Autonomy satisfaction involves individuals having decision-making power or control over their activities. Competence satisfaction means individuals feel capable of completing tasks, similar to self-efficacy. Relatedness satisfaction involves feeling emotional support from the environment, facilitating internalization of group values.

Recent studies confirm SDT's relevance in teacher professional identity formation. For instance, Wong and Liu explored the development of teacher professional identity in pre-service teachers through the SDT perspective, highlighting how psychological needs satisfaction facilitated this development[26]. Liu and Zhang examined the gap in teacher professional identity between coursework learning and teaching practice, emphasizing the importance of SDT in bridging this gap[27]. Ma emphasised the crucial role of motivation and commitment in teacher professional identity, arguing that these factors are essential for fostering teachers' sustained engagement, professional development, and capacity to cope with career challenges[28]. Weiß et al. examined the relationship between vocational education and training (VET) trainee teachers' professional identity and their basic psychological need for autonomy, finding that autonomy support plays an important role in the positive development of teacher professional identity and further enhances their intention to remain in the profession[29].

SDT suggests that performance rewards' impact on autonomous motivation depends on their effect on basic psychological need satisfaction[30], While performance rewards may undermine autonomy needs due to their external control nature, they may simultaneously affirm individual competence. This dual effect creates context-specific outcomes requiring empirical investigation.

2.2. The proposed model and hypotheses

Based on SDT's conceptualization of autonomous motivation and its relationship to professional outcomes[31], teachers' professional identity can be understood as a manifestation of autonomous motivation. Building on this foundation, we propose a comprehensive model examining three distinct pathways (see Figure 1): (1) direct relationships between basic psychological need satisfaction and professional identity; (2) cross-level mediation, where school-level teacher evaluation affects individual psychological need satisfaction, which subsequently influences professional identity; and (3) cross-level moderation, where psychological need satisfaction moderates the relationship between teacher evaluation and professional identity.

Following the typology commonly adopted in teacher evaluation research[21,22], we distinguish between two evaluation types: performance evaluation and developmental evaluation. Performance evaluation emphasizes measurable teaching outcomes, typically using student academic achievement as primary indicators, and links results to rewards and sanctions. Developmental evaluation focuses on identifying teaching challenges and providing support for professional development. This distinction is crucial because existing research has predominantly emphasized evaluation's negative consequences[12,13]while overlooking

potential benefits for professional development. Examining both types enables a comprehensive understanding of evaluation's multifaceted effects.

Basic Psychological Needs and Professional Identity

Within the SDT framework, basic psychological need satisfaction provides the foundation for autonomous motivation. Empirical research demonstrates strong correlations between autonomous motivation and job satisfaction, as well as between professional identity and job satisfaction[31,33,34]. Recent evidence further supports this linkage: SDT-based interventions and motivational support significantly enhance teacher professional identity and career commitment [26,27,28,29].Accordingly, we hypothesize:

Hypothesis 1: Higher satisfaction of the three basic psychological needs (autonomy, competence, and relatedness) will be associated with stronger professional identity among teachers in both China and Japan (path a in Figure 1).

Cross-Level Mediation Pathways

We propose that teacher evaluation influences professional identity indirectly through its effects on psychological need satisfaction. Evidence regarding the effects of developmental evaluation remains limited and inconsistent. Although theory predicts that developmental evaluation should enhance professional development, empirical support is sparse. Wang and Liu reported positive findings among university teachers[32]; however, these results may not generalize to basic education contexts. Ichinose and Yamazaki's findings are limited by their reliance on administrator surveys rather than direct teacher responses[22]. International evidence also raises questions about developmental evaluation's effectiveness. For example, Gratton's study in New Zealand revealed that teachers perceived evaluation processes as bureaucratic exercises that failed to provide meaningful professional benefits[33].

Research on performance evaluation yields particularly mixed results. Katsuno documented that Japanese teachers reported performance-based evaluation systems undermined their professional convictions and strained collegial relationships[13]. Cuevas et al. found that performance evaluation pressure reduced autonomous motivation and increased exhaustion[35]. Conversely, Liu et al. found no significant correlation between negative perceptions of performance evaluation and job satisfaction among Chinese teachers[18].

These inconsistent findings may stem from methodological limitations, as previous studies often failed to account for the nested structure of school data, potentially biasing their results. Despite these empirical inconsistencies, SDT theory suggests that developmental evaluation should enhance psychological need satisfaction and professional identity by supporting professional growth, while performance evaluation may undermine these outcomes through external pressure, feelings of incompetence (since teachers have limited control over student scores), and damaged collegial relationships.

Hypothesis 2: Developmental evaluation will enhance professional identity both directly (path b in Figure 1) and indirectly through increased satisfaction of basic psychological needs (paths c and a in Figure 1).

Hypothesis 3: Performance evaluation will diminish professional identity both directly (path b in Figure 1) and indirectly through decreased satisfaction of basic psychological needs (paths c and a in Figure 1).

Hypothesis 4: These mediation pathways will differ between Chinese and Japanese contexts due to environmental and cultural factors that shape how external evaluation influences individuals.

Cross-Level Moderation Effects

Psychological need satisfaction may also moderate how teachers respond to different evaluation types (path d in Figure 1). Teachers with varying levels of need satisfaction may experience evaluation differently. For example, teachers with high competence satisfaction may benefit more from developmental evaluation's feedback, while teachers with low autonomy satisfaction may be particularly harmed by performance evaluation's controlling nature. Similarly, performance evaluation's emphasis on rankings and comparisons may have stronger negative effects on teachers who value collegial relationships (high relatedness satisfaction).

Hypothesis 5: The satisfaction of basic psychological needs will moderate the relationship between teacher evaluation types and professional identity.

Hypothesis 6: These moderation patterns will differ between Chinese and Japanese teachers, reflecting cultural variations in how psychological resources buffer or amplify evaluation effects.

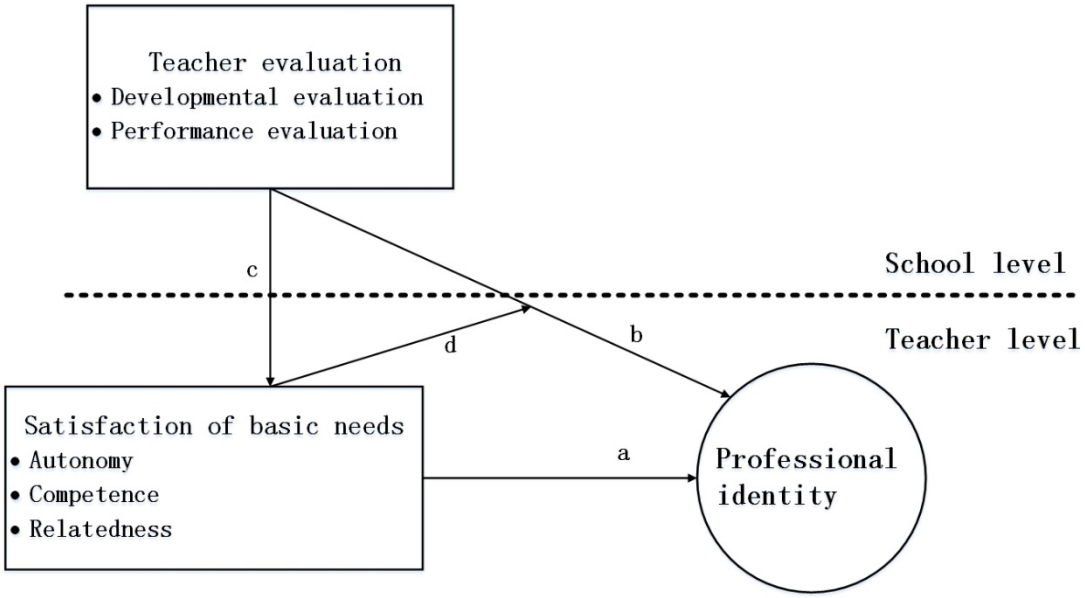

Figure 1. The proposed model

3. Research method

3.1. Demographic background characteristics of subjects in this study

In the TALIS 2013 dataset, 3,925 Chinese and 3,484 Japanese secondary school teachers participated, along with 199 and 192 principals from their respective schools. Principals served solely as information providers for school environmental characteristics and teacher evaluation systems; therefore, their personal background variables are not analyzed in this study.

After excluding incomplete responses, the final valid sample comprised 3,661 Chinese and 3,184 Japanese teachers, with 188 and 183 corresponding principals. Table 1 presents the demographic characteristics of participants from both countries. The gender distribution shows a higher proportion of female teachers in Chinese schools, while Japanese schools have a higher proportion of male teachers. Regarding qualifications and experience, teachers in both countries hold bachelor's degrees or higher, with Japanese teachers averaging approximately seven years more teaching experience than their Chinese counterparts. The participating schools were predominantly public in both countries. Notable differences emerged in school size, with Chinese schools averaging 973 students compared to considerably smaller Japanese schools.

Table 1. Demographic background characteristics of participants in China and Japan

	China (teacher and school=3661/188)	**Japan (teacher and school=3184/183)**
Gender (male); n (%)	1078 (27.9%)	2017 (60.7%)
Educational degree (bachelor degree); n (%)	3580 (97.8%)	3168 (99.5%)
Teaching experience; *m*±*SD*	10.23 ± 7.12	17.51 ± 10.96
School type (public); n (%)	153 (81.4%)	170 (92.9%)
School size; *m*±*SD*	972.53 ± 611.07	453.53 ± 290.38

3.2. Measures

Table 2 summarizes the TALIS 2013 data used in this study, which can be divided into three types. The first is the composite scale indices. These scores are the factor scores calculated by the TALIS team with confirmatory factor analysis (CFA). These data are provided in the released file of TALIS 2013. In this study, "TJSPROS", "SEINS" and "TCEXCHS" are used representing teachers' professional identity, teaching self-efficacy (reflecting the satisfaction of teachers' need for competence) and school colleagues' cooperation (reflecting the satisfaction of teachers' need for relatedness). Second, we created our own variables, which are teachers' perceived participation in decision-making in school affairs (reflecting the satisfaction of teachers' need for autonomy) and two kinds of teacher evaluation. In this part, we chose the corresponding data from several items in the released file that can reflect these constructs. More specifically, we calculated these items with re-weighting according to the factor loading, or directly summed them up to come up with the new measurement indices. The third part includes variables of teachers' backgrounds and the characteristics of schools. In this part, teachers' gender, teaching experience, educational degree, school type and school size are all controlled variables.

In order to confirm whether the TALIS data used in this study can accurately reflect the construct, a CFA with a four-factor (i.e., professional identity, perceived participation in decision-making, teaching self-efficacy, and colleagues' cooperation) model was used to test the reliability and validity of the respective data. We provide the evidence of Cronbach's α, convergent and discriminant validity. Among these, the convergent and discriminant validity are provided by calculating the factor structure and factor loading of this CFA. The following is a detailed description of these parts.

Table 2. Variables used in this study and the respective indicators in TALIS 2013

	Variables	**Corresponding data of TALIS 2013**		**Description**
		Indicator	**Name in the released file**	**Example**
DV	Professional identity	TJSPROS [a]	TT2G46A,46B, 46D and46F	The advantages of this profession clearly outweigh the disadvantages
MV	Autonomy satisfaction (Perceived participation)	Weighted by the factor loading of the right items [b]	TT2G44A and TT2G44D	● This school provides staff with opportunities to actively participate in school decisions ● This school has a culture of shared responsibility for school issues
	Competence satisfaction (Teaching self-efficacy)	SEINSS [a]	TT2G34C、TT2G34J、TT2G34K 及TT2G34L	Craft good questions for my students

	Variables	Corresponding data of TALIS 2013		Description
		Indicator	Name in the released file	Example
IV School level	Relatedness satisfaction (Colleagues' cooperation)	TCEXCHS [a]	TT2G33D—TT2G33G	Exchange teaching materials with colleagues
	Developmental evaluation	Sum the right column scores [b]	TC2G29A1、TC2G29B2、TC2G29D4	● Measures to remedy any weaknesses in teaching are discussed with the teacher ● A development or training plan is developed for each teacher ● A mentor is appointed to help the teacher improve his/her teaching
	Performance evaluation		TC2G29C3、TC2G29F6、TC2G29G7	● If a teacher is found to be a poor performer, material sanctions such as reduced annual increases in pay are imposed on the teacher ● A change in a teacher's salary or a payment of a financial bonus ● A change in the likelihood of a teacher's career advancement
CV Teacher level	Gender	Take the right column value	TT2G01	
	Teaching experience		TT2G05B	
	Educational degree		TT2G10	
CV School level	School type		TC2G10	
	School size		TC2G14	

Table 2. (*Continued*)

a: Complex Scale Indices computed by TALIS team, b: computed by this study; DV means dependent variable; MV means mediator variable; IV means independent variable; CV means control variable

Professional identity

In TALIS 2013, teachers' professional identity is represented by four questions, and Likert's four-point scoring is adopted (from 1 (strongly disagree) to 4 (strongly agree)). The TALIS team uses CFA to obtain the scores of latent factors of teachers' professional identity. When the score in this part is higher, it indicates that the degree of professional identity is higher. In order to confirm the reliability and validity of teachers' professional identity, we used TALIS 2013 data to calculate the reliability and validity of middle school teachers in China and Japan. In terms of reliability, Cronbach's α coefficients of middle school teachers in China and Japan were 0.77 and 0.81, respectively. The evidence of convergent validity shows that the composite reliability (CR) and average variance extracted (AVE) of middle school teachers in China and Japan are 0.77, 0.82 (for CR) and 0.47, 0.54 (for AVE). Fornell and Larcker [26] stated that convergent validity is supported if CR and AVE of each construct are higher than 0.70 and 0.50, respectively. Therefore, we can see that the reliability and validity of teachers' professional identity in China and Japan are acceptable.

Perceived participation of decision-making in school affairs

Perceived participation in decision-making in school affairs reflects the satisfaction of autonomy needs. This is the reason we selected this indicator. Since TALIS did not define perceived participation in decision-making—which means teachers think they have the right to take part in the decision-making of school affairs—this study selected two items from the raw data of TALIS (see Table 1 for the items). The responses

were reported on a Likert 4-point scale from 1 (strongly disagree) to 4 (strongly agree). From the perspective of content validity, the two items measure the extent to which schools provide teachers with the opportunity to participate in decision-making and the culture of shared responsibility. Therefore, these two items can represent teachers' perception that schools give them the right to participate in decision-making in school affairs, thus demonstrating good content validity. According to the empirical evidence of this study, Cronbach's α coefficients for Chinese and Japanese teachers were 0.78 and 0.65. CR for Chinese and Japanese teachers is 0.79 and 0.67, respectively. AVE for Chinese and Japanese teachers is 0.65 and 0.50. We can see that the reliability and validity of the measurement of this variable reached an acceptable level. Finally, we calculated the scores of teachers' perceived participation in decision-making by weighting the factor loadings of these two items to produce one score as a measurement indicator.

Teaching self-efficacy

The TALIS team used CFA with four questions (Likert 4-point scale from "not at all" (1) to "a lot" (4)) to obtain the latent factor scores of teachers' teaching self-efficacy. The higher the score, the higher the degree of teaching self-confidence perceived by the teacher. This score can also be seen as the degree of satisfaction of the need for competence by teachers. Based on our analysis of TALIS 2013 data, the Cronbach's α coefficients of teaching self-efficacy of middle school teachers in China and Japan are 0.86 and 0.81, CR is 0.87 and 0.81, and AVE is 0.62 and 0.52. These values are higher than the generally accepted standards, indicating that teachers in both countries have good reliability and validity in the measurement of teaching self-efficacy.

Colleagues' cooperation

In TALIS 2013, there are 4 items (Likert 6-point scale: never (1), once a year (2), 2-4 times a year (3), 5-10 times a year (4), 1-3 times a month (5), and once a week (6)) used to calculate the latent factor scores of colleagues' cooperation in the school. When this score is higher, it indicates that the degree of cooperation among school colleagues perceived by the teachers is higher, and it also represents that the satisfaction of the need for relatedness is higher. Similarly, we examined the reliability and validity of this variable ourselves. The results show that the Cronbach's α coefficients for middle school teachers in China and Japan are 0.77 and 0.68, CR is 0.78 and 0.69, and AVE is 0.47 and 0.37. It can be found that the reliability and validity evidence of teachers in China is quite satisfactory, but Japanese teachers' measurement in this part has only reached a barely acceptable level.

Developmental and performance evaluation

Developmental and performance evaluations are variables at the school level. In this study, the related data were measured by the principal. For each of these two variables, three items are selected as representatives (Likert 5-point scale: never (1), less than once every two years (2), once every two years (3), once per year (4), and twice or more per year (5)). In terms of connotation, the items we chose corresponding to developmental assessment concern how schools can improve teachers' teaching based on the evaluation, aiming to help teachers' professional growth. The items for performance evaluation concern how schools give teachers corresponding rewards and punishments based on the evaluation. From these items, we can see the content validity for these two variables. However, the reliability of the two factors and the results of the CFA test show that the evidence of reliability and validity of middle school teachers in China and Japan is not satisfactory. In the developmental evaluation, the Cronbach's α coefficients of Chinese and Japanese middle school teachers are 0.64 and 0.56, the CR is 0.66 and 0.58, and the AVE is 0.40 and 0.32. In the part of performance evaluation, the Cronbach's α coefficients of teachers in Chinese and Japanese middle schools are 0.70 and 0.61, CR is 0.73 and 0.64, and AVE is 0.49 and 0.39.

The CFA results revealed unsatisfactory measurement validity for both developmental and performance evaluation constructs. This finding prompted us to investigate potential explanations for these weak psychometric properties. Our analysis identified systematic variation in evaluation practices across schools as the primary issue. Specifically, certain items showed consistently low factor loadings across both countries. For instance, the item "a senior teacher will be arranged to help improve teaching" yielded factor loadings of only 0.50 in China and 0.47 in Japan—substantially lower than other items in the scale. This pattern suggests that mentorship-based evaluation systems are not universally implemented across schools, creating measurement inconsistency.

Despite these psychometric limitations, we maintain that the selected six items possess strong content validity and adequately represent the theoretical construct of teacher evaluation. The items comprehensively cover the key dimensions of how schools assess and support teacher development. The weak statistical properties appear to stem from institutional heterogeneity rather than conceptual inadequacy.

To address this methodological challenge, we modified our analytical approach. Rather than using factor-weighted scores—which would amplify the measurement problems caused by varying institutional practices—we adopted a simple summation method. This approach treats all items equally, preventing schools without certain evaluation practices from being underrepresented in the analysis. The resulting sum scores provide a straightforward interpretation: higher total scores indicate more frequent and comprehensive evaluation practices, regardless of the specific methods employed by individual schools.

3.3. Data analysis

This study employs a two-pronged analytical strategy. First, we present descriptive statistics (means, standard deviations, and Pearson correlation coefficients) to characterize the variables and their bivariate relationships. Second, we test the hypothesized path model (Figure 1) using Hierarchical Linear Modeling (HLM) to examine and compare the relationships between variables for Chinese and Japanese middle school teachers. All analyses were conducted using SPSS 25.0 for descriptive statistics and HLM 6.0 for multilevel modeling. Our multilevel analytical framework follows the systematic approach proposed by Wen and Chiou for testing cross-level mediation and moderation effects, consisting of five sequential steps that build upon each other to establish the presence and nature of hypothesized relationships[37].

The first step establishes whether multilevel modeling is warranted through null model testing. We calculate the Intra-class Correlation Coefficient (ICC = $\tau_{00} / (\tau_{00} + \sigma^2)$) for professional identity and the three mediators (satisfaction of autonomy, competence, and relatedness needs). Here, τ_{00} represents between-school variance while σ^2 represents within-school variance. Following established guidelines, ICCs exceeding 0.059 indicate sufficient between-school variation to justify multilevel analysis. Only variables meeting this threshold proceed to subsequent analytical steps.

The second step employs a random coefficient model to examine whether teacher-level mediators explain variance in professional identity while allowing these relationships to vary across schools. Importantly, school-level evaluation variables are excluded at this stage to isolate teacher-level effects.

Teacher Level (Equation 1-1): Professional Identity_ij = β_{0j} + β_{1j}(Gender_ij) + β_{2j}(Teaching Experience_ij) + β_{3j}(Educational Degree_ij) + β_{4j}(Perceived Participation_ij) + β_{5j}(Teaching Self-efficacy_ij) + β_{6j}(Colleagues' Cooperation_ij) + ε_ij

School Level (Equation 1-2): $\beta_{0j} = \gamma_{00} + \gamma_{01}$(School Type_j) + γ_{02}(School Size_j) + u_{0j} $\beta_{1j} = \gamma_{10} + u_{1j}$; $\beta_{2j} = \gamma_{20} + u_{2j}$; $\beta_{3j} = \gamma_{30} + u_{3j}$; $\beta_{4j} = \gamma_{40} + u_{4j}$; $\beta_{5j} = \gamma_{50} + u_{5j}$; $\beta_{6j} = \gamma_{60} + u_{6j}$

The random slopes (u_1j through u_6j) allow the effects of teacher characteristics and psychological needs satisfaction to vary across schools, capturing contextual differences in these relationships.

The third step implements an intercept-as-outcome model to test whether school-level evaluation practices (developmental and performance evaluation) directly influence professional identity and the three mediators. Each outcome is examined separately while controlling for teacher demographics.

Teacher Level (Equation 2-1): Professional Identity (or Perceived Participation, Teaching Self-efficacy, Colleagues' Cooperation)_ij = β_{0j} + β_{1j}(Gender_ij) + β_{2j}(Teaching Experience_ij) + β_{3j}(Educational Degree_ij) + ε_ij

School Level (Equation 2-2): $\beta_{0j} = \gamma_{00} + \gamma_{01}$(School Type_j) + γ_{02}(School Size_j) + γ_{03}(Developmental Evaluation_j) + γ_{04}(Performance Evaluation_j) + u_{0j} $\beta_{1j} = \gamma_{10} + u_{1j}$; $\beta_{2j} = \gamma_{20} + u_{2j}$; $\beta_{3j} = \gamma_{30} + u_{3j}$

Through steps 2 and 3, we establish whether the regression coefficients constituting the multilevel mediation effect reach statistical significance. These include the effects of teacher evaluation on psychological needs satisfaction and professional identity (Step 3), and the effects of psychological needs satisfaction on professional identity (Step 2).

The fourth step tests for mediation by combining school-level predictors and teacher-level mediators in a full model. This crucial analysis examines how the direct effect of evaluation practices on professional identity changes when psychological needs satisfaction variables are included in the model.

Teacher Level (Equation 3-1): Professional Identity_ij = β_{0j} + β_{1j}(Gender_ij) + β_{2j}(Teaching Experience_ij) + β_{3j}(Educational Degree_ij) + β_{4j}(Perceived Participation_ij) + β_{5j}(Teaching Self-efficacy_ij) + β_{6j}(Colleagues' Cooperation_ij) + ε_ij

School Level (Equation 3-2): $\beta_{0j} = \gamma_{00} + \gamma_{01}$(School Type_j) + γ_{02}(School Size_j) + γ_{03}(Developmental Evaluation_j) + γ_{04}(Performance Evaluation_j) + u_{0j} $\beta_{1j} = \gamma_{10} + u_{1j}$; $\beta_{2j} = \gamma_{20} + u_{2j}$; $\beta_{3j} = \gamma_{30} + u_{3j}$; $\beta_{4j} = \gamma_{40} + u_{4j}$; $\beta_{5j} = \gamma_{50} + u_{5j}$; $\beta_{6j} = \gamma_{60} + u_{6j}$

We focus on comparing the coefficients for evaluation practices between equations where mediators are included (Equations 3-1 and 3-2) versus excluded (Equations 2-1 and 2-2). Complete mediation is supported when the direct effect of evaluation becomes non-significant after including mediators, while the mediator effects remain significant. Partial mediation occurs when the direct effect is reduced but remains significant, with mediators also showing significant effects.

The final step investigates cross-level moderation by testing whether school evaluation practices moderate the relationships between psychological needs satisfaction and professional identity. This analysis reveals whether the strength of teacher-level relationships varies systematically based on school evaluation practices.

Teacher Level (Equation 4-1): Professional Identity_ij = β_{0j} + β_{1j}(Gender_ij) + β_{2j}(Teaching Experience_ij) + β_{3j}(Educational Degree_ij) + β_{4j}(Perceived Participation_ij) + β_{5j}(Teaching Self-efficacy_ij) + β_{6j}(Colleagues' Cooperation_ij) + ε_ij

School Level (Equation 4-2): $\beta_{0j} = \gamma_{00} + \gamma_{01}$(School Type_j) + γ_{02}(School Size_j) + γ_{03}(Developmental Evaluation_j) + γ_{04}(Performance Evaluation_j) + u_{0j} $\beta_{1j} = \gamma_{10} + u_{1j}$; $\beta_{2j} = \gamma_{20} + u_{2j}$; $\beta_{3j} = \gamma_{30} + u_{3j}$ $\beta_{4j} = \gamma_{40} + \gamma_{41}$(Developmental Evaluation_j) + γ_{42}(Performance Evaluation_j) + u_{4j} $\beta_{5j} = \gamma_{50} + \gamma_{51}$(Developmental Evaluation_j) + γ_{52}(Performance Evaluation_j) + u_{5j} $\beta_{6j} = \gamma_{60} + \gamma_{61}$(Developmental Evaluation_j) + γ_{62}(Performance Evaluation_j) + u_{6j}

The cross-level interaction terms (γ_{41}, γ_{42}, γ_{51}, γ_{52}, γ_{61}, γ_{62}) indicate whether psychological needs satisfaction variables (Perceived Participation, Teaching Self-efficacy, and Colleagues' Cooperation) moderate the relationship between school evaluation practices and professional identity.

This systematic approach enables a comprehensive understanding of how school evaluation practices influence teacher professional identity both directly and indirectly through psychological needs satisfaction, while revealing potential differences between Chinese and Japanese educational contexts.

4. Results

Table 3 presents the descriptive statistics for all variables. At the school level, Chinese schools implement both developmental and performance evaluation significantly more frequently than Japanese middle schools (developmental evaluation: $t(369) = 20.74$, $p < 0.01$; performance evaluation: $t(369) = 14.47$, $p < 0.01$). At the teacher level, Japanese middle school teachers reported significantly higher professional identity than their Chinese counterparts ($t(7177) = 19.07$, $p < 0.01$). However, Chinese teachers scored significantly higher on perceived participation in decision-making, teaching self-efficacy, and colleagues' cooperation ($t = 37.36$, 70.99, and 6.12, respectively, all $p < 0.01$).

Table 3. Descriptive statistics and Pearson correlations for variables

Variables	Descriptive statistics	Correlation coefficient matrix			
	M±SD	1	2	3	4
		School level			
Developmental evaluation	China 8.98±1.67 Japan 5.74±1.31				
performance evaluation	China 5.41±1.62 Japan 3.42±0.93				
		Teacher level			
1.Professional identity	China 10.83±1.57 Japan 11.60±1.86	1			
2. Perceived participation	China 4.61±0.97 Japan 3.84±0.76	.450 .204	1		
3. Teaching self-efficacy	China 12.43±2.14 Japan 9.04±1.87	.233 .215	.210 .088	1	
4. School colleagues' cooperation	China 10.97±2.29 Japan 10.67±1.75	.262 .114	.331 .173	.306 .209	1

Prior to conducting the HLM analysis, we tested null models to verify the suitability of multilevel modeling. The ICC values for professional identity were 0.12 for Chinese and 0.05 for Japanese middle school teachers. For perceived participation in decision-making, the values were 0.11 and 0.18; for teaching self-efficacy, 0.04 and 0.03; and for colleagues' cooperation, 0.09 and 0.08, respectively. Since the ICC values for teaching self-efficacy in both countries fell below the 0.05 threshold, this variable was excluded from subsequent HLM analyses examining path c and cross-level mediation effects. The remaining variables showed ICC values between 0.05 and 0.18, indicating substantial between-school variation and within-school similarity in teacher responses, thus justifying the use of HLM to properly distinguish effects at different levels. Table 4 summarizes the HLM analysis results.

4.1. The cross-level mediation effect

We began by examining whether teachers' satisfaction of psychological needs was associated with their professional identity, as this relationship forms the foundation for any mediation effects. Table 4 reveals that all three psychological needs indicators showed significant positive associations with professional identity in both countries, though with notable variations in strength. The strongest predictor was autonomy satisfaction (measured through perceived participation in decision-making), with Chinese teachers showing a particularly robust relationship ($\gamma = 0.63$) compared to Japanese teachers ($\gamma = 0.49$), suggesting that feeling involved in school decisions is especially important for Chinese teachers' professional identity. Competence satisfaction (indicated by teaching self-efficacy) showed a contrasting pattern, with Japanese teachers demonstrating a stronger association ($\gamma = 0.19$) than Chinese teachers ($\gamma = 0.08$), implying that confidence in teaching abilities contributes more substantially to professional identity among Japanese educators. Relatedness satisfaction (reflected in colleagues' cooperation) showed the weakest associations in both countries ($\gamma = 0.05$ for China; $\gamma = 0.06$ for Japan), suggesting that while collegial relationships matter, they are less central to professional identity than autonomy and competence.

Having established these foundational relationships, we tested whether school evaluation practices influence professional identity through these psychological pathways. An important limitation emerged: teaching self-efficacy showed insufficient variation between schools (ICC < 0.05) in both countries, preventing its inclusion in the mediation analysis. Therefore, we could only test mediation through autonomy and relatedness satisfaction.

The effects of developmental evaluation differed dramatically between China and Japan. In China, developmental evaluation demonstrated a complete mediation pattern. Initially, developmental evaluation showed a significant positive direct effect on professional identity ($\gamma= 0.06$, $p < 0.01$) and also significantly enhanced both autonomy satisfaction ($\gamma = 0.05$, $p < 0.01$) and relatedness satisfaction ($\gamma = 0.08$, $p < 0.01$). When we included these mediators in the full model, the direct effect on professional identity disappeared (becoming non-significant), while the mediators retained their strong effects. This complete mediation indicates that developmental evaluation works entirely through psychological pathways—by making teachers feel more involved in decisions and enhancing collegial cooperation, which in turn strengthens their professional identity. In contrast, developmental evaluation showed minimal effects in Japan. No direct effect on professional identity was observed, and among the potential mediators, only relatedness satisfaction (colleagues' cooperation) was positively influenced ($\gamma = 0.12$, $p < 0.01$). However, this single pathway was insufficient to create an indirect effect on professional identity.

Performance evaluation also produced markedly different results across countries. In China, performance evaluation appeared neutral, with no significant effects emerging on professional identity, either directly or indirectly, and teachers' autonomy and relatedness satisfaction remained unaffected. This null finding challenges theoretical predictions that performance-based systems would harm teacher motivation. In Japan, however, performance evaluation proved detrimental, with a significant negative direct effect on professional identity ($\gamma = -0.10$, $p < 0.01$). Surprisingly, this negative impact was not mediated through psychological needs—neither autonomy nor relatedness satisfaction were significantly affected. This direct negative effect suggests that performance evaluation may threaten aspects of Japanese teachers' professional identity beyond the psychological needs we measured.

4.2. The cross-level moderation effect

Following the mediation analysis, we tested for cross-level moderation effects. For Chinese teachers who demonstrated cross-level mediation, we found no evidence of moderated mediation. Neither the

interaction between developmental evaluation and perceived participation (γ_{41}) nor the interaction between developmental evaluation and colleagues' cooperation (γ_{61}) reached statistical significance.

However, we identified two significant cross-level moderation effects independent of mediation. First, among Chinese teachers, teaching self-efficacy positively moderated the relationship between developmental evaluation and professional identity ($\gamma_{51} = 0.02$, $p < 0.01$). In schools with more frequent developmental evaluation, the positive association between teaching self-efficacy and professional identity was strengthened. Second, among Japanese teachers, we found a negative moderation effect of colleagues' cooperation on the relationship between performance evaluation and professional identity ($\gamma_{62} = -0.03$, $p < 0.01$). In schools with more frequent performance evaluation, the positive association between colleagues' cooperation and professional identity was weakened.

Table 4. Result of HLM

	Random coefficient model: For path a	**Intercept- as outcome model: For path b and path c**			**Mediation effect**	**Moderation effect**
Dependent variable	**Professional identity**	**Professional identity**	**Perceived participation**	**School colleagues' cooperation**	**Professional identity**	**Professional identity**
Fixed effects						
γ00	10.35(0.12)* 10.97(0.16) *	10.26(0.15)* 11.21(0.19) *	4.48(0.09) * 4.60(0.10) *	10.90(0.23) * 11.44(0.32) *	10.36(0.12) * 10.96(0.17) *	10.40(0.11) * 10.96(0.17) *
γ01 school type	0.48(0.08) * 0.34(0.10) *	0.58(0.11) * 0.06(0.12)	0.10(0.06) -0.65(0.08) *	0.31(0.17) -0.47(0.27)	0.47(0.08) * 0.35(0.11) *	0.45(0.08) * 0.35(0.11) *
γ02 school size	0.00(0.00) 0.00(0.00)	0.00(0.00) 0.00(0.00)	0.00(0.00) 0.00(0.00)	0.00(0.00) 0.00(0.00)	0.00(0.00) 0.00(0.00)	0.00(0.00) 0.00(0.00)
γ03 developmental evaluation		0.06(0.02) * 0.03(0.04)	0.05(0.02) * 0.01(0.02)	0.08(0.04) * 0.12(0.04) *	0.01(0.02) 0.01(0.03)	0.02(0.02) 0.01(0.03)
γ04 performance evaluation		0.01(0.02) -0.10(0.04) *	0.01(0.01) 0.01(0.03)	-0.01(0.04) -0.006(0.04)	0.01(0.02) -0.11(0.04) *	0.01(0.02) -0.10(0.04) *
γ10gender	-0.08(0.05) 0.15(0.07) *	-0.09(0.06) 0.20(0.07) *	0.002(0.03) -0.03(0.03)	-0.22(0.08) * -0.17(0.07) *	-0.09(0.05) 0.15(0.07) *	-0.09(0.05) 0.15(0.07) *
γ20 teaching experience	-0.01(0.002) * -0.01(0.002) *	-0.01(0.002) * -0.01(0.003) *	-0.001(0.001) 0.004(0.001) *	-0.01(0.003) * -0.01(0.003) *	-0.01(0.002) * -0.02(0.003) *	-0.01(0.003) * -0.01(0.03) *
γ30 educational degree	-0.12(0.14) -0.19(0.16)	0.03(0.15) -0.19(0.18)	0.03(0.11) -0.01(0.06)	0.34(0.29) 0.47(0.18) *	-0.11(0.13) -0.19(0.17)	-0.04(0.14) -0.19(0.17)
γ40 Perceived participation	0.63(0.03) * 0.49(0.04) *				0.63(0.03) * 0.49(0.05) *	0.63(0.03) * 0.49(0.05) *
γ41 Perceived participation×DE						0.02(0.02) -0.01(0.04)
γ42 Perceived participation×PE						-0.02(0.02) -0.02(0.03)
γ50 teaching self-efficacy	0.08(0.01) * 0.19(0.02) *				0.08(0.01) * 0.19(0.02) *	0.08(0.01) * 0.19(0.02)*
γ51 teaching self-efficacy×DE						0.02(0.01) * -0.01(0.01)
γ52 teaching self-efficacy×PE						-0.01(0.01) 0.02(0.03)
γ60 school	0.05(0.01) *				0.05(0.01) *	0.05(0.01) *

	Random coefficient model: For path a	**Intercept- as outcome model: For path b and path c**			**Mediation effect**	**Moderation effect**
Dependent variable	**Professional identity**	**Professional identity**	**Perceived participation**	**School colleagues' cooperation**	**Professional identity**	**Professional identity**
colleagues' cooperation	0.06(0.01) *				0.06(0.02) *	0.06(0.02) *
γ61 school colleagues' cooperation×DE						-0.003(0.01) -0.002(0.01)
γ62 school colleagues' cooperation×PE						0.001(0.01) -0.03(0.01) *
Random effects						
Within school						
σ2	1.60 2.89	2.16 3.22	0.85 0.46	4.77 2.79	1.60 2.89	1.60 2.89
Between school						
τ00	0.18 0.28	0.22 0.32*	0.09 0.15*	0.43 0.29	0.18 0.24	0.20 0.24
τ11	0.06 0.08	0.01 0.08	0.004 0.03	0.01 0.07	0.06 0.08	0.06 0.08
τ22	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
τ33	0.16 0.13	0.03 0.22	0.01 0.01	0.59 0.47	0.16 0.14	0.18 0.14
τ44	0.08* 0.03*				0.08* 0.03*	0.08* 0.03*
τ55	0.01 0.02*				0.01 0.02*	0.01 0.02
τ66	0.00 0.00				0.003 0.003	0.003 0.003

Table 4. (*Continued*)

Note: *The first value presented is for Chinese and the second is for Japanese; DE: Developmental evaluation; PE: Performance evaluation*

5. Discussion

Our HLM analysis of TALIS 2013 data reveals striking differences in how teacher evaluation affects professional identity in China and Japan. Within the Self-Determination Theory framework, we operationalized the three basic psychological needs through specific indicators: perceived participation in decision-making as an indicator of autonomy satisfaction, teaching self-efficacy as an indicator of competence satisfaction, and colleagues' cooperation as an indicator of relatedness satisfaction. Our findings show that contrary to our hypotheses, performance evaluation did not negatively affect Chinese teachers' professional identity or their satisfaction of these psychological needs. In contrast, performance evaluation directly harmed Japanese teachers' professional identity without mediating through psychological needs satisfaction. Developmental evaluation enhanced Chinese teachers' professional identity specifically through

improved autonomy satisfaction (measured by perceived participation) and relatedness satisfaction (measured by colleagues' cooperation), but showed no benefits for Japanese teachers.

The Differential Effects of Developmental Evaluation

According to SDT, evaluation systems should influence professional identity by satisfying individuals' basic psychological needs for autonomy, competence, and relatedness[25, 30]. This mechanism was confirmed only for Chinese teachers, where developmental evaluation indirectly enhanced professional identity through two specific pathways: increased autonomy satisfaction (operationalized as perceived participation in school decision-making) and enhanced relatedness satisfaction (operationalized as colleagues' cooperation). The absence of mediation through competence satisfaction (measured by teaching self-efficacy) was expected, given that self-efficacy showed insufficient between-school variation for multilevel analysis.

This positive finding contrasts sharply with international research from New Zealand[32] and the United States[37], where evaluation systems often fail to enhance teachers' psychological needs satisfaction due to formalism, resource constraints, and implementation challenges. Shanghai's success may reflect its status as a pioneer in China's educational reform, with more thorough implementation strategies that genuinely support teachers' autonomy needs (through participatory decision-making) and relatedness needs (through collaborative practices).

Zhang et al. similarly found that Shanghai's evaluation systems effectively support teacher development by providing opportunities for meaningful participation and collaboration[16]. In contrast, while Japanese administrators report positive effects of developmental evaluation[22,38], our findings align with teacher-centered research showing no benefits for ordinary teachers' psychological needs satisfaction or professional identity. This suggests a significant gap between administrative intentions and teachers' actual experiences of autonomy, competence, and relatedness support in Japanese schools.

The Cultural Context of Performance Evaluation and Psychological Needs

Performance evaluation's effects on psychological needs satisfaction also varied dramatically between countries. Despite theoretical predictions that performance evaluation would undermine all three psychological needs, we found no negative effects on Chinese teachers' autonomy satisfaction (perceived participation), competence satisfaction (teaching self-efficacy), or relatedness satisfaction (colleagues' cooperation) [8, 9,10,39]. This unexpected finding aligns with recent empirical studies by Liu et al. showing that transparent evaluation systems may even help Chinese teachers recognize evaluation's value without threatening their psychological needs[17].

Zhao et al. suggest that Chinese educational culture, with its emphasis on hierarchy and external evaluation, may socialize teachers to accept controlling elements without experiencing them as threats to autonomy[40]. Consequently, performance evaluation may not undermine Chinese teachers' sense of participation in decision-making or their feelings of competence and relatedness.

Conversely, Japanese teachers experienced direct negative effects on professional identity from performance evaluation, though interestingly, not through diminished psychological needs satisfaction as we hypothesized. The direct negative impact suggests that performance evaluation may threaten aspects of professional identity beyond the three basic needs captured by our indicators.

The Moderating Role of Psychological Needs Satisfaction

Our moderation analyses revealed how psychological needs satisfaction shapes teachers' responses to evaluation. Among Chinese teachers, higher competence satisfaction (measured by teaching self-efficacy) strengthened the positive effects of developmental evaluation on professional identity. Teachers who feel

more capable apparently benefit more from developmental feedback, suggesting that competence satisfaction serves as a resource that enhances receptivity to professional development opportunities.

Among Japanese teachers, the negative moderation effect involving relatedness satisfaction (colleagues' cooperation) presents a cautionary tale. In environments with strong collegial bonds (high relatedness satisfaction), performance evaluation appears particularly damaging to professional identity. This finding challenges assumptions that satisfying relatedness needs universally buffers against stress. Instead, when evaluation systems rank or compare teachers, high relatedness satisfaction may intensify the psychological threat by transforming supportive relationships into competitive ones. Additionally, gender may moderate the relationship between teacher evaluation and professional identity. Future research should examine whether male and female teachers respond differentially to developmental versus performance-based evaluation systems, potentially revealing important gender-evaluation interaction effects that could inform more nuanced policy approaches.

Theoretical and Practical Implications

These findings extend STD by demonstrating that the relationship between psychological needs satisfaction and professional outcomes is culturally contingent and contextually complex. While SDT posits that satisfying autonomy, competence, and relatedness needs universally supports motivation and well-being, our results suggest that how these needs interact with external evaluation varies significantly across cultural contexts.

Practically, our findings underscore that evaluation systems must align with how psychological needs are understood and valued within specific cultural contexts. Shanghai's success with developmental evaluation appears linked to its ability to genuinely support teachers' autonomy (through participatory structures) and relatedness (through collaborative practices). The failure of similar systems in Japan suggests that identical structures may not translate into genuine psychological needs satisfaction when they conflict with existing professional cultures.

Our findings yield clear implications for policy and practice: Educational authorities in China would benefit from expanding developmental evaluation practices that enhance teacher autonomy and foster collegial support, while Japanese policymakers should carefully reconsider how performance-based evaluation may inadvertently undermine professional identity, particularly within collaborative school cultures.

The differential effects of performance evaluation are particularly instructive for policy. The absence of negative effects on Chinese teachers' psychological needs satisfaction suggests that teachers socialized in hierarchical systems may maintain their sense of autonomy, competence, and relatedness despite external evaluation. Conversely, the harmful effects observed in Japan, especially when relatedness satisfaction is high, warn that performance evaluation may be most damaging precisely in collaborative environments where it disrupts the social foundations of professional identity.

While our study used validated indicators from TALIS for psychological needs satisfaction, future research should employ measures specifically designed to capture SDT constructs. Additionally, examining other potential mediators beyond the three basic needs may help explain the direct negative effects of performance evaluation on Japanese teachers' professional identity. Understanding these mechanisms is essential for developing culturally responsive evaluation policies that support rather than undermine teachers' psychological needs and professional commitment.

While this study provides important quantitative evidence from TALIS 2013 data regarding the relationship between teacher evaluation and professional identity, several limitations merit consideration. First, our reliance on quantitative data alone precludes deeper exploration of teachers' lived experiences and the nuanced meanings they attach to evaluation processes. Future research would benefit from mixed-methods approaches, including thematic analysis of policy documents and in-depth interviews with teachers from both China and Japan. Such qualitative inquiry could illuminate how teachers navigate the tensions between traditional cultural values and contemporary evaluation systems, providing richer contextualization of the quantitative patterns identified here. Moreover, triangulating multiple data sources would enhance both the credibility and explanatory depth of cross-cultural comparisons.

Additionally, it is important to acknowledge that students play a fundamental role in shaping teachers' professional identity—a dimension not captured in our evaluation-focused analysis. As Shin and Alpern demonstrate, students' expectations, feedback, and classroom interactions significantly influence teachers' self-perception and professional identity formation[41]. Pre-service teachers often enter education programs with pre-formed conceptions of teacher identity shaped by their own educational experiences, which subsequently influence how they understand and commit to their professional roles. In the context of our findings, future research should examine how student-teacher interactions mediate the relationship between evaluation systems and professional identity, particularly exploring whether developmental evaluation approaches that incorporate student voice might strengthen teachers' sense of professional purpose across different cultural contexts.

6. Conclusion

This study analyzed TALIS 2013 data from 3,661 Chinese and 3,184 Japanese secondary school teachers, revealing striking cultural differences in how teacher evaluation affects professional identity. Developmental evaluation enhanced Chinese teachers' professional identity through increased autonomy and relatedness satisfaction, while showing no benefits for Japanese teachers. Performance evaluation had no negative effects on Chinese teachers but directly undermined Japanese teachers' professional identity. Additionally, Chinese teachers with higher self-efficacy benefited more from developmental evaluation, whereas Japanese teachers in collaborative environments experienced greater harm from performance evaluation. These findings challenge SDT's assumption that external evaluation universally undermines psychological needs, suggesting that cultural context fundamentally shapes how teachers experience and respond to evaluation systems.

The differential effects observed between China and Japan underscore that successful evaluation reform requires cultural alignment rather than standardized approaches. Shanghai's success with developmental evaluation demonstrates the benefits when systems genuinely support teacher participation and collaboration, while Japan's negative experience with performance evaluation warns against implementing systems that contradict professional cultures valuing collective harmony. Future research should explore additional mediating mechanisms and longitudinal effects to better understand these complex relationships. As nations continue reforming teacher evaluation, policymakers must recognize that effective systems cannot be transplanted across cultures but must be designed to align with local understandings of autonomy, competence, and relatedness. It is important to acknowledge that certain psychological constructs in this study demonstrated measurement limitations, including suboptimal item reliability and limited between-school variance, which may affect the robustness of our multilevel analyses. Future research should address these limitations through improved scale design, expanded item pools, multi-method measurement

approaches, and validation across diverse cultural contexts and larger samples to enhance both reliability and generalizability.

Future research should pursue several concrete directions: (1) longitudinal designs to establish causal relationships and temporal dynamics between teacher evaluation and professional identity formation; (2) examination of additional mediating mechanisms beyond basic psychological needs to develop more comprehensive theoretical models; (3) cross-national comparative studies investigating how diverse cultural norms and educational traditions shape teachers' responses to evaluation systems; and (4) mixed-methods approaches combining quantitative and qualitative data to capture the nuanced experiences of teachers under different evaluation frameworks.

These findings emphasize that teacher evaluation reforms must be culturally responsive: China would benefit from development-focused systems that support autonomy and relatedness, while Japan should exercise caution with performance-linked evaluations that may erode professional identity.

Author Contributions

Conceptualization, Xue Lian Wang and I-Hua Chen; methodology, Xue Lian Wang; software, Xue Lian Wang; validation, Xue Lian Wang and I-Hua Chen; formal analysis, Xue Lian Wang; investigation, Xue Lian Wang; resources, Xue Lian Wang; data curation, Xue Lian Wang; writing—original draft preparation, Xue Lian Wang; writing—review and editing, Xue Lian Wang and I-Hua Chen; visualization, Xue Lian Wang; supervision, I-Hua Chen; project administration, Xue Lian Wang; funding acquisition, none. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare no conflict of interest

References

1. Murphy J, Hallinger P, Heck R. Leading via teacher evaluation: The case of the missing clothes? Educational Researcher, 2013;42(6): 349–354. https://doi.org/10.3102/0013189X13499625
2. Huang, S. (2020). Confucian Culture and the Institutional Reform of Chinese Schools. In: Peters, M. (eds) Encyclopedia of Educational Philosophy and Theory. Springer, Singapore. https://doi.org/10.1007/978-981-287-532-7_714-1
3. Liang ZM. The theoretical structure of moral education in Japan and the role of Confucian thought. Journal of Educational Research, 2002; 48: 45–72.
4. Tucker JA. Japanese Confucian philosophy. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy (Fall 2008 Edition). Stanford: Stanford University; 2008. https://plato.stanford.edu/entries/japanese-confucian/
5. Tran TT. Is the learning approach of students from the Confucian heritage culture problematic? Asia Pacific Education Review, 2013; 14(4): 507–515. https://doi.org/10.1007/s10671-012-9131-3
6. Katsuno M. Teacher evaluation policies and practices in Japan: How performativity works in schools. Abingdon: Routledge; 2016.

7. Gao X, Liu Y, Wang C. Cultural foundations of teacher occupational health: The impact of Confucian work dynamism on burnout through grit among Chinese teachers. BMC Psychology, 2025; 13: 390. https://doi.org/10.1186/s40359-025-02724-x
8. Rong ZH. The spiritual source and solutions of the identity crisis of primary and secondary school teachers. Curriculum, Teaching Material and Method, 2017; 37(3): 103–109.
9. Yan XL. Problems, causes and countermeasures of teacher evaluation in primary and secondary schools in China. Education Observation, 2018; 7(24): 84–86.
10. Shi GC, Li Y. The loss and pursuit of teachers' sense of professional meaning. Teaching and Management, 2014; (12): 90–92.
11. National Council on Educational Reform. Report of the National Council on Educational Reform: 17 Proposals to Change Education [M]. Tokyo: Office of the National Council on Educational Reform, Cabinet Secretariat; 2000 Dec 22.
12. Katsuno M. Teacher evaluation in Japanese schools: An examination from a micro-political or relational viewpoint. Journal of Education Policy, 2010; 25(3): 293–307. https://doi.org/10.1080/02680931003700431
13. Katsuno M. Teachers' professional identities in an era of testing accountability in Japan: The case of teachers in low-performing schools. Education Research International, 2012. https://doi.org/10.1155/2012/862193
14. Ura T. From the Results of a Questionnaire Survey of Principals and Teachers in Tokyo Public Schools [M]. Tokyo: Dōjidai-sha; 2002.
15. Lai YT. Contradictions and issues inherent in Japan's teacher evaluation system: Focusing on discussions from the National Council on Educational Reform to the Central Council for Education. Annual Report of the Graduate School of Education, Tohoku University, 2013; 61(2): 129–144.
16. Zhang XF, Ng HM. An effective model of teacher appraisal: Evidence from secondary schools in Shanghai, China. Educational Management Administration & Leadership, 2015; 45(2): 1–23. https://doi.org/10.1177/1741143213510508
17. Liu S, Zhao D. Teacher evaluation in China: Latest trends and future directions. Educational Assessment, Evaluation and Accountability, 2013; 25(3): 231–250. https://doi.org/10.1007/s11092-013-9171-4
18. Liu S, Xu X, Stronge J. The influences of teachers' perceptions of using student achievement data in evaluation and their self-efficacy on job satisfaction: Evidence from China. Asia Pacific Education Review, 2018; 19(4): 493–509. https://doi.org/10.1007/s12564-018-9541-6
19. OECD. TALIS 2013 database. 2014a. Available online: https://www.oecd.org/education/school/talis-2013-data.htm (accessed on 13 August 2025).
20. OECD. TALIS 2013 Technical Report. Paris: OECD Publishing; 2014b. https://doi.org/10.1787/9789264196261-en
21. Tokyo Metropolitan Teachers' Union, Tokyo Metropolitan Board of Education. Survey on Teacher Performance Evaluation and Collegiality among Metropolitan High School Teachers [R]. Tokyo: Tokyo Teachers' Union; 2005.
22. Ichinose A, Yamazaki Y. Current situation and issues of the teacher evaluation system and school evaluation system in teacher professional growth: A fundamental survey in Shizuoka Prefectural high schools. Studies in Subject Development, 2013; 1: 81–90. https://doi.org/10.14945/00006329
23. Aarts E, Verhage M, Veenvliet JV, Dolan CV, van der Sluis S. A solution to dependency: Using multilevel analysis to accommodate nested data. Nature Neuroscience, 2014; 17(4): 491–496. https://doi.org/10.1038/nn.3648
24. Chang TM, Chang XS. The influence of evaluator composition on compulsory education teachers' performance-based pay preferences. Journal of Shaoxing University (Education Edition), 2017; 37(2): 1–11.
25. Deci EL, Ryan RM. Self-determination theory. International Encyclopedia of the Social & Behavioral Sciences, 2015; 91(2): 486–491. https://doi.org/10.1016/B978-0-08-097086-8.26036-4
26. Wong CYE, Liu WC. Development of teacher professional identity: Perspectives from self-determination theory. European Journal of Teacher Education, 2024. https://doi.org/10.1080/02619768.2024.2371981
27. Liu B, Zhang Y. Closing the gap of teacher professional identity in student teachers between coursework learning and teaching practice: Perspectives from self-determination theory. Journal of Education for Teaching, 2025. https://doi.org/10.1080/02607476.2025.2532603
28. Ma D. The role of motivation and commitment in teachers' professional identity. Frontiers in Psychology, 2022; 13: 910747. https://doi.org/10.3389/fpsyg.2022.910747
29. Weiß JK, Bottling M, Kärner T. Professional identification in the beginning of a teacher's career: A longitudinal study on identity formation and the basic psychological need for autonomy in VET teacher training. Frontiers in Psychology, 2023; 14: 1196473. https://doi.org/10.3389/fpsyg.2023.1196473
30. Deci EL, Koestner R, Ryan RM. Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. Review of Educational Research, 2001; 71(1): 1–27. https://doi.org/10.3102/00346543071001001
31. Canrinus ET, Helms-Lorenz M, Beijaard D, Buitink J, Hofman A. Self-efficacy, job satisfaction, motivation and commitment: Exploring the relationships between indicators of teachers' professional identity. European Journal of Psychology of Education, 2012; 27(1): 115–132. https://doi.org/10.1007/s10212-011-0069-2

32. Wang ZJ, Liu LD. Can performance appraisal promote breakthrough academic innovation of university teachers? An empirical study based on self-determination theory. Higher Education Research, 2017; 38(4): 52–60.
33. Gratton R. Teacher appraisal: A lesson on confusion over purpose. International Journal of Educational Management, 2004; 18(5): 292–296. https://doi.org/10.1108/09578230410554056
34. Skaalvik EM, Skaalvik S. Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion. Psychological Reports, 2014; 114(1): 68–77. https://doi.org/10.2466/14.02.PR0.114k14w0
35. Cuevas R, Ntoumanis N, Fernandez-Bustos JG, Bartholomew K. Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? Journal of School Psychology, 2018; 68: 154–162. https://doi.org/10.1016/j.jsp.2018.03.002
36. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: Algebra and statistics. Journal of Marketing Research, 1981; 18(1): 39–50. https://doi.org/10.1177/002224378101800104
37. Wen FX, Chiu HC. Multilevel Model Methodology—Key Issues and Tentative Solutions of Hierarchical Linear Modelling [M]. Taipei: Chien Cheng Culture; 2012.
38. Katsuno M. The working of performativity: How is the teacher assessment enacted? [C] In: Proceedings of the Annual Conference of British Educational Research Association; 3–6 September 2008; Edinburgh, UK. Heriot-Watt University; 2008.
39. Yuan MM. The performance appraisal and performance pay system for basic education teachers urgently needs institutional innovation. Shanghai Educational Evaluation Research, 2016; 5(3): 76–79.
40. Zhao YM, Zhang ZT, Liu N, Ding MZ. A review of new developments in self-determination theory. Journal of Management, 2016; 13(7): 1095–1104.
41. Shin M, Alpern LW. Drawing teacher identity: Perspectives of students in a teacher education program. Social Sciences & Humanities Open, 2024; 10: 101036. https://doi.org/10.1016/j.ssaho.2024.101036