Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Student behaviour stress and teacher well-being: classroom management self-efficacy, workload and measurement content in TALIS 2024

Li Ma
Chengdu University
Bosheng Li
202411516202@cdu.edu.cn

Chengdu University

Research Article

Keywords: teacher well-being, occupational stress, classroom management, self-efficacy, professional collaboration, TALIS, cross-sectional study

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11291313/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Student behaviour stress and teacher well-being: classroom management self-efficacy, workload and measurement content in TALIS 2024

Li Ma, Bosheng Li*

College of Teachers, Chengdu University, No. 2025 Chengluo Avenue, Longquanyi District, Chengdu, Sichuan 610106, P. R. China

**Corresponding author*

Bo-sheng Li

College of Teachers, Chengdu University

No. 2025 Chengluo Avenue, Longquanyi District

Chengdu, Sichuan 610106, P. R. China

Email: 202411516202@cdu.edu.cn

ORCID (Bo-sheng Li): https://orcid.org/0009-0005-2191-8800

ORCID (Li Ma): https://orcid.org/0009-0004-7173-413X

Phone: +86-177-7557-4001

Abstract

Background

Student behaviour demands are a recurring source of teacher stress. Classroom management self-efficacy may correspond to these demands, but their shared content also complicates the interpretation of moderation. We examined stress–well-being relationships alongside workload and compared estimates across stress measures.

Methods

We analysed Teaching and Learning International Survey (TALIS) 2024 lower-secondary teacher data from Shanghai (People's Republic of China), Japan, Korea and Singapore. The primary sample comprised 8743 teachers in 731 schools. Weighted regressions included student behaviour stress, classroom management self-efficacy, professional collaboration, their interactions, education-system indicators and teacher covariates. Inference used 100 balanced repeated replication weights with Fay's factor 0.5. Further models added workload stress and compared outcomes, item content and specifications.

Results

After workload adjustment, student behaviour stress remained associated with adverse workplace well-being (standardized coefficient 0.267; 95% confidence interval 0.232 to 0.303), compared with 0.464 before adjustment. The stress-by-efficacy interaction decreased from −0.035 to −0.0241 (normal-based $p = 0.049$; $t(99)$-based $p = 0.051$); the primary interaction block increased weighted R^2 by 0.0013. The behaviour stress index closely reflected discipline stress ($r = 0.959$). Interactions varied across alternative item composites, although the index-versus-three-item contrast was inconclusive ($p = 0.229$). The expanded quadratic model gave an imprecise linear efficacy interaction ($p = 0.065$). Collaboration and job-satisfaction interactions were not statistically significant.

Conclusions

Behaviour-related stress was associated with adverse workplace well-being alongside workload. The modest efficacy interaction varied across stress measures and became less precise with workload and specification changes. Task correspondence and shared measurement content are both relevant to its interpretation. These findings support examining specific classroom demands together with workload and using clearly defined measures when evaluating professional resources.

Keywords: teacher well-being; occupational stress; classroom management; self-efficacy; professional collaboration; TALIS; cross-sectional study

Background

Teachers' well-being includes work-related strain, perceived health effects and satisfaction with professional life (1, 2). These dimensions can diverge: a teacher may remain satisfied with the professison while experiencing distress during daily work. Recent research in Chinese primary and secondary schools links broad work stress with well-being through affective rumination and engagement (3). Distinguishing strain from occupational evaluations, and specific classroom demands from general work pressure, can clarify the role of professional resources.

Student behaviour demands require teachers to maintain instruction while responding to disruption, conflict and intimidation. Evidence links student misbehaviour to burnout, exhaustion and reduced enthusiasm, with teacher–student relationships providing one potential pathway (4-6). The job demands–resources framework and conservation of resources theory explain how effortful demands and available resources relate to these experiences (7-11). Studies of teachers apply these perspectives to burnout, engagement and occupational well-being (12-15). Yet behaviour-related pressure occurs alongside lesson preparation, marking and administrative work. Accounting for workload is therefore necessary to establish whether classroom demands have a distinct conditional association with well-being (16).

Self-efficacy refers to beliefs about the capacity to act in a particular setting (17); teacher efficacy is domain-specific (18, 19). Classroom management efficacy concerns maintaining order and responding to difficult behaviour. Reviews and empirical studies relate teacher efficacy to classroom processes, well-being, burnout and satisfaction (20-23), while longitudinal research indicates that satisfaction can also precede changes in efficacy (24). A recent study of early-career teachers linked social and emotional competencies to burnout through efficacy (25). The triple-match principle suggests that resource–strain relationships depend partly on correspondence between the content of demands and resources (26). Classroom management efficacy and behaviour stress share this correspondence, but their related wording could also reflect common appraisal or measurement content. Comparing stress measures with different item content helps evaluate these competing interpretations.

Professional collaboration provides a second resource encompassing joint teaching, feedback and shared learning. Research on professional learning communities and TALIS identifies relationships between collaboration, efficacy, satisfaction and leadership (27-30). Its broader focus may support well-being through professional learning and collegial support rather than through a behaviour-specific interaction. Recent expert consensus on educator well-being also emphasises relationships, leadership and sustained organisational support (31). Thus, a comparison with collaboration can help distinguish the observed patterns of classroom-specific efficacy from those of collective professional practice.

The remaining question is how the efficacy interaction changes when concurrent workload and the content of stress measures are considered. We address this question using a common survey framework in Shanghai, Japan, Korea and Singapore, recognising that professional beliefs can differ across education systems (32). Theory motivates two expectations: greater behaviour stress is associated with more adverse workplace well-being (H1), and the stress–well-being slope is less positive at higher classroom management efficacy (H2). We treat the comparison with collaboration and the item-content contrasts as exploratory. Unlike studies of broad stress pathways or efficacy

mediation, the analysis directly tests interactions, compares them across stress measures and distinguishes work-related strain from job satisfaction.

Methods

Study design and participants

We conducted a cross-sectional secondary analysis of the OECD TALIS 2024 public-use lower-secondary teacher data (33-35). TALIS uses a two-stage design, selecting schools and then teachers within schools. The four-system dataset included 14089 teachers in 731 schools: 4078 teachers in 205 schools in Shanghai, 3558 in 201 in Japan, 3098 in 169 in Korea, and 3355 in 156 in Singapore (Table 1).

Table 1 Retained lower-secondary teacher samples by education system

Education system	Teachers	Schools	Median teachers/school	P10	P90
Shanghai	4078	205	20.0	19.0	20.0
Japan	3558	201	20.0	12.0	20.0
Korea	3098	169	19.0	12.8	24.0
Singapore	3355	156	22.0	19.0	23.0

Note. Counts describe the retained four-system dataset before model-specific complete-case restrictions. P10 and P90 are the 10th and 90th percentiles of sampled teachers per school. The primary model contains 8743 teachers in all 731 schools.

The OECD classified participation quality as Good in Shanghai, Japan and Singapore and Fair in Korea (36). Overall teacher participation was 98.4%, 96.2%, 76.2% and 95.3%, respectively. We included an analysis excluding Korea to examine its influence on the pooled estimates.

Student behaviour stress questions were administered in two of the three teacher questionnaire versions. Analyses used respondents who received these questions and had complete data for the relevant model. Of 9398 teachers assigned the relevant questions, 8743 had complete primary-model

data; the loss was 655 (6.97%), including outcome and covariate missingness. The primary workplace well-being model covered all 731 schools; the job-satisfaction, workload-adjusted and item-content models included 8740, 8739 and 8650 teachers, respectively. Questionnaire groups had similar observed characteristics, with absolute standardized mean differences no greater than 0.033. Observed balance does not establish that remaining missing values are completely or conditionally random; complete-case estimates therefore depend on the selection process.

Measures

Workplace well-being was assessed using the OECD workplace well-being and stress scale. Its four items address general work stress, time for personal life, and perceived effects of work on mental and physical health. Higher scores indicate more adverse workplace well-being. The secondary outcome, job satisfaction, combines satisfaction with the profession and the work environment; higher scores indicate greater satisfaction (37).

Student behaviour stress was measured using the OECD index covering responsibility for student achievement, maintaining classroom discipline, and intimidation or verbal abuse by students. Higher scores reflect greater perceived stress from these demands. We retain the OECD construct name because the index includes achievement responsibility and intimidation as well as discipline; its strong empirical correspondence with the discipline item is examined below. Primary analyses used the released OECD scale scores.

Classroom management self-efficacy assessed teachers' perceived capacity to manage disruption, clarify behaviour expectations, secure adherence to classroom rules and calm disruptive students. Professional collaboration assessed joint teaching, peer observation and feedback, activities across classes, and collaborative professional learning (37). Models adjusted for the public-use categories of teaching experience (≤5, 6–10, 11–20 and >20 years) and current school count. The primary specification standardizes their published numeric codes and enters them as linear covariates; an additional model uses category indicators. Current school count concerns simultaneous employment,

not career mobility. The workload-adjusted model also included the OECD workload stress scale, which addresses lesson preparation, teaching load, marking and administrative work (37).

Continuous variables were standardized using the unweighted complete-case mean and standard deviation (SD), with the latter calculated using divisor N. These coefficients describe sample-standardized associations, rather than population-standardized effect sizes. Descriptive statistics used survey weights. Regression coefficients are expressed in the standardized units used for estimation. We also examined standardization within education systems. OECD scale construction uses international calibration to support comparison (37); it does not substitute for an item-level invariance test in the selected four-system sample. System indicators control mean differences, while within-system scaling assesses sensitivity to the units used within each system. Neither specification estimates the effect of system-level differences in stress.

Measurement diagnostics

Classroom management efficacy was concentrated at the upper endpoint, which contained approximately 23% of responses. The behaviour stress index correlated 0.959 with the discipline item and 0.596 with the intimidation item; the raw discipline/intimidation pair had alpha = 0.638. These diagnostics informed the nonlinear and item-content analyses.

For the four workplace well-being source items, corrected item–total correlations ranged from 0.275 to 0.673. The unweighted raw composite had alpha = 0.777 and approximate one-factor omega = 0.855. The weighted correlation between workplace well-being and behaviour stress was 0.456, and the partial correlation after covariate and system adjustment was 0.439. To examine overlap in stress-related wording (38), we estimated an additional model using a three-item well-being composite that excluded the general work-stress item.

Statistical analysis

Weighted least-squares regression estimated the primary model:

$$Y_i = \beta_0 + \beta_1 P_i + \beta_2 E_i + \beta_3 C_i + \beta_4 (P_i \times E_i) + \beta_5 (P_i \times C_i) + \gamma X_i + \delta_s + \varepsilon_i$$

Y denotes standardized workplace well-being, P student behaviour stress, E classroom management efficacy, C professional collaboration, X the two teacher covariates, and δ education-system fixed effects. The same specification was fitted for job satisfaction. The stress coefficient represents its association at mean resource levels; the interaction coefficients represent changes in that slope per SD increase in the resource. Shanghai was the reference system, and estimation used the final teacher weight.

Following the TALIS 2024 user guide, standard errors were obtained from 100 balanced repeated replication (BRR) weights with Fay's factor f = 0.5 (34). The variance estimator was $V(\beta) = [R(1-f)^2]^{-1} \sum_r (\beta_r - \beta)^2$, where R = 100. Primary confidence intervals used the normal critical value 1.96, consistent with inference described in the OECD analytical notes (37). Calculations use the underlying estimates and standard errors before rounding; a displayed endpoint of zero can represent a small nonzero value. We also calculated t(99)-based inference for the workload-adjusted interaction and school-clustered standard errors for comparison.

The two focal tests concern H1 and H2 in the workplace well-being model; their nominal p-values are reported without multiplicity adjustment. An exploratory resource contrast compared the two interaction coefficients using their covariance; its p-value is unadjusted and is interpreted as exploratory evidence requiring independent evaluation. Weighted R^2 was compared between additive and interaction specifications. Model-based stress slopes were evaluated at −2, −1, 0 and +1 SD efficacy and at the observed maximum (+1.35 SD), holding collaboration at its mean. The original +2 SD evaluation lies above the observed maximum and is treated as extrapolation.

Sensitivity analyses examined nonlinear specification, school effects, sample composition and measurement. The restricted quadratic model added stress squared, efficacy squared and stress-squared-by-efficacy. The school fixed-effect model used weighted within-school transformation. Weighted school intraclass correlations (ICCs) were 0.156 for workplace well-being and 0.187 for

job satisfaction. Further analyses omitted experience, standardized within systems, excluded Korea, adjusted for workload and used the three-item well-being composite.

Additional analyses in revision

We added stress-by-efficacy-squared to the restricted quadratic specification, retaining all lower-order terms, and fitted a sensitivity model using survey-weighted means and SDs for standardization. We also replaced classroom management efficacy with efficacy in instruction or student engagement while using the intimidation/special-education/diversity stress composite. These replacements are exploratory and change both constructs; they do not isolate a single source of shared content.

Further covariate analyses used categorical experience and school-count controls and school socioeconomic intake reported by principals (a seven-category estimate of the proportion of students from disadvantaged homes), and separate blocks for class-size category and perceived classroom disruption, or relational leadership, followed by workload. Class-size and leadership items have no jointly observed cases among respondents with behaviour stress under the questionnaire design. Each block therefore uses its own fixed common sample and a baseline fitted to that same sample. Gender is unavailable in the Japan and Korea public-use records, so the gender sensitivity analysis covers Shanghai and Singapore. We do not impute these unavailable fields.

Predicted outcome contrasts compared the observed 10th and 90th efficacy percentiles and, separately, efficacy at −2 and +2 SD at fixed stress levels of −1, 0 and +1 SD, with collaboration and continuous controls at their means. For an efficacy span d, the contrast is $d\beta_2 + dp\beta_4$ and its SE uses the full covariance matrix. For the four-SD span, for stress level p, the high-minus-low contrast is $4\beta_2 + 4p\beta_4$. These estimates describe fitted contrasts, not effects of changing efficacy through an intervention.

Item-content and system comparisons

Exploratory item-content models compared the released behaviour stress index with intimidation alone; the mean of achievement responsibility and intimidation; the mean of special-education lesson adaptation and diversity/equity-related work; and the mean of intimidation, special-education adaptation and diversity/equity-related work. The latter two composites represent broader student-related demands.

All five models used the same 8650 teachers, with predictors standardized within this sample. Holm adjustment covered the four alternative-item interaction tests. Paired replicate estimates were used for contrasts with the released index. The achievement/intimidation pair had weighted alpha = 0.404, and the three-item intimidation/special-education/diversity composite had alpha = 0.657. Reliability was considered when interpreting differences between these measures.

System-specific regressions and efficacy-quartile analyses used school-clustered inference. System fits retain pooled primary-sample standardization; their interaction heterogeneity was tested by an independent-system inverse-variance Wald test. We assessed heterogeneity across the four education systems and reported Holm-adjusted quartile contrasts alongside the omnibus test. These comparisons were exploratory.

Use of generative AI during manuscript preparation

AI was used for language polishing and copy-editing to improve readability of the manuscript. All conceptual content, data analysis, interpretation and conclusions were completed by the authors, who take full responsibility for the final manuscript. Reported estimates were calculated by executing that code on the OECD public-use data. The authors are responsible for verifying the code, results, references and final interpretation.

Results

Descriptive characteristics

Table 2 reports weighted means and standard deviations for the four systems. The pooled mean workplace well-being and stress score was 9.74 (SD 1.90); the student behaviour stress mean was 8.92 (SD 1.88). Workplace well-being scores ranged from 9.19 in Shanghai to 9.85 in Singapore, while student behaviour stress ranged from 8.55 in Shanghai to 9.34 in Korea. High discipline stress was reported by 29.2%, 35.5%, 48.8% and 38.0% of teachers in Shanghai, Japan, Korea and Singapore, respectively.

Table 2 Weighted descriptive statistics across the four education systems

Measure	Shanghai	Japan	Korea	Singapore	All four
Workplace well-being and stress (higher = more adverse)	9.19 (1.74)	9.84 (1.95)	9.75 (1.81)	9.85 (1.79)	9.74 (1.90)
Student behaviour stress scale	8.55 (1.88)	8.86 (1.85)	9.34 (1.92)	8.84 (1.85)	8.92 (1.88)
Classroom management self-efficacy	13.55 (2.08)	10.75 (2.00)	12.08 (2.14)	12.38 (2.01)	11.46 (2.27)
Professional collaboration	10.78 (1.54)	9.21 (1.74)	8.95 (1.46)	9.14 (1.59)	9.36 (1.74)
High discipline-stress item (%, >=3)	29.2%	35.5%	48.8%	38.0%	37.7%

Note. Values are survey-weighted means (standard deviations), except the discipline-stress percentage. High discipline stress denotes a response of at least 3 on the four-point source item. Higher workplace well-being and stress values indicate more adverse well-being. These weighted standard deviations are not used to standardize the regressions.

Primary workplace well-being model

In the primary model, higher student behaviour stress was associated with more adverse workplace well-being at mean resource levels (b = 0.464, standard error [SE] = 0.016, 95% confidence interval [CI] 0.433 to 0.495, $p < 0.001$;Table 3). Classroom management efficacy had a near-zero main

coefficient at mean stress (b = −0.004, SE = 0.015, p = 0.793), whereas collaboration had a small favourable conditional association (b = −0.039, SE = 0.018, 95% CI −0.073 to −0.004, p = 0.028).

The stress-by-efficacy interaction was negative (b = −0.035, SE = 0.013, 95% CI −0.061 to −0.009, p = 0.008). The collaboration interaction was not statistically significant (b = 0.013, SE = 0.016, p = 0.433). The secondary efficacy-minus-collaboration contrast was −0.048 (SE = 0.023, 95% CI −0.093 to −0.003, unadjusted p = 0.036). This exploratory contrast favoured a more negative interaction for efficacy; it does not establish superiority of one resource as an intervention.

Table 3 Primary weighted model of adverse workplace well-being

Term	Estimate (SE)	95% CI	p
Intercept	−0.133 (0.024)	[−0.181, −0.085]	<0.001
Student behaviour stress	0.464 (0.016)	[0.433, 0.495]	<0.001
Classroom management efficacy	−0.004 (0.015)	[−0.034, 0.026]	0.793
Professional collaboration	−0.039 (0.018)	[−0.073, −0.004]	0.028
Stress × efficacy	−0.035 (0.013)	[−0.061, −0.009]	0.008
Stress × collaboration	0.013 (0.016)	[−0.019, 0.045]	0.433
Teaching-experience category	0.002 (0.014)	[−0.025, 0.030]	0.869
Current school-count category	0.001 (0.015)	[−0.029, 0.031]	0.936
Japan	0.235 (0.035)	[0.166, 0.303]	<0.001
Korea	0.068 (0.036)	[−0.003, 0.140]	0.060
Singapore	0.209 (0.035)	[0.140, 0.277]	<0.001

Note. N = 8743 teachers in 731 schools. All continuous variables use unweighted complete-case standardization. All fitted coefficients are displayed, including education-system indicators and the standardized numeric codes of the two categorical covariates. Shanghai is the reference system. SEs use BRR with Fay factor 0.5; CIs use 1.96. Reported weighted R^2 = 0.221. Main resource coefficients apply at mean stress, and the stress coefficient applies at mean resource levels.

The fitted stress slope declined from 0.535 at −2 SD efficacy (95% CI 0.475 to 0.594) to 0.417 at the observed maximum (+1.35 SD; 0.369 to 0.464; Figure 1). At −1 SD and +1 SD it was 0.500 and

0.429, respectively. The central-range slope difference, −1 versus +1 SD efficacy, was approximately 0.071 standardized units (95% CI 0.019 to 0.123, calculated from the unrounded interaction estimate and variance); the stress association remained positive throughout the displayed range. The weighted R^2 was 0.221 compared with 0.219 in the additive specification; the interaction block's unrounded increment was 0.0013. The interactions thus accounted for little additional outcome variation.

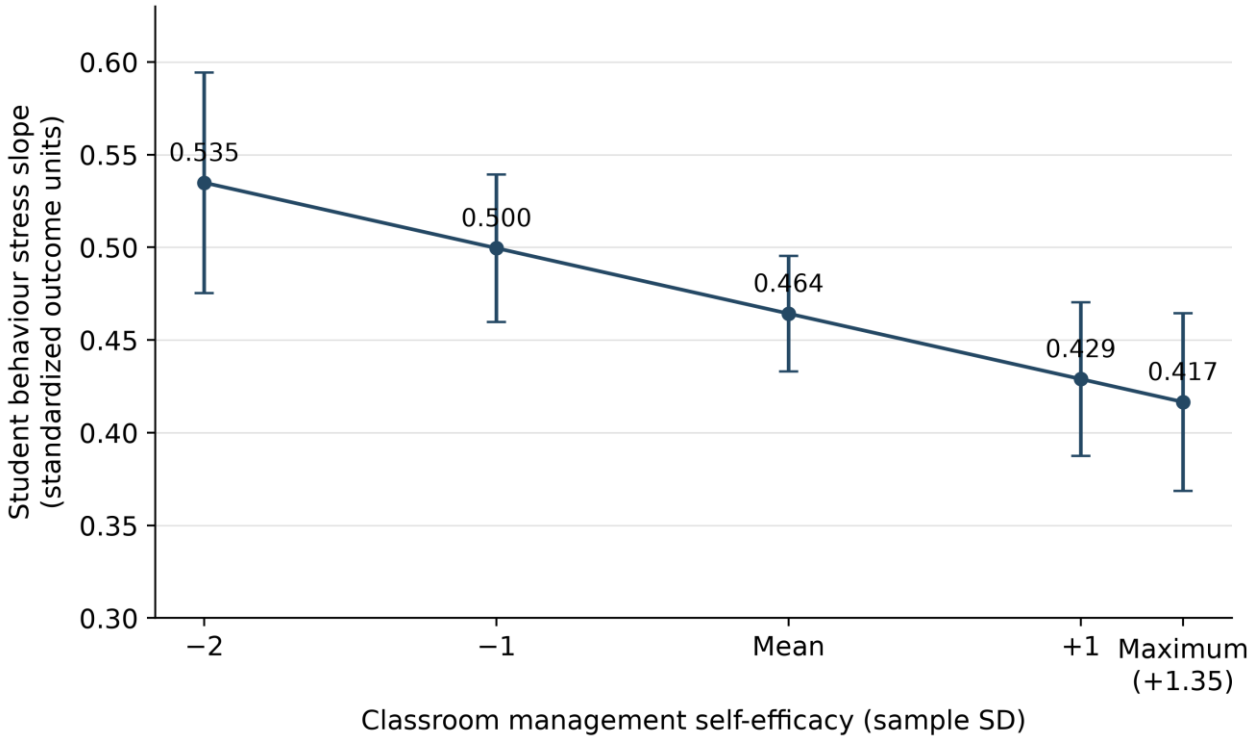


Figure 1 Student behaviour stress slopes at different classroom management efficacy levels.

Model-based slopes from the primary weighted linear workplace well-being model at efficacy values of −2, −1, 0 and +1 SD and the observed maximum (+1.35 SD), with professional collaboration held at its mean. Higher outcome values indicate more adverse workplace well-being. Error bars are 95% confidence intervals (CIs) using BRR-Fay inference and the normal critical value 1.96. Both stress and the outcome use unweighted complete-case standardization.

Workload-adjusted model

After workload stress was included, the student behaviour stress coefficient remained positive (b = 0.267, SE = 0.018, 95% CI 0.232 to 0.303, $p < 0.001$; Table 4). Workload stress also had a positive

conditional association (b = 0.380, SE = 0.018, 95% CI 0.345 to 0.416, p < 0.001). The stress-by-efficacy interaction was reduced to −0.024 (SE 0.012; normal-based 95% CI −0.048 to 0.000 at three-decimal precision). The upper endpoint is −0.00015 before rounding. Its normal-approximation p-value was 0.049, compared with 0.051 under t(99) inference. The collaboration interaction was close to zero (b = 0.003, p = 0.849).

Table 4 Workplace well-being model additionally adjusted for workload stress

Term	Estimate (SE)	95% CI	p
Intercept	−0.112 (0.022)	[−0.155, −0.068]	<0.001
Student behaviour stress	0.267 (0.018)	[0.232, 0.303]	<0.001
Classroom management efficacy	−0.024 (0.015)	[−0.052, 0.005]	0.108
Professional collaboration	−0.027 (0.016)	[−0.058, 0.005]	0.099
Stress × efficacy	−0.024 (0.012)	[−0.048, 0.000]	0.049
Stress × collaboration	0.003 (0.014)	[−0.025, 0.031]	0.849
Teaching-experience category	−0.003 (0.013)	[−0.028, 0.022]	0.820
Current school-count category	0.002 (0.014)	[−0.025, 0.030]	0.861
Japan	0.165 (0.034)	[0.099, 0.231]	<0.001
Korea	0.163 (0.033)	[0.098, 0.227]	<0.001
Singapore	0.115 (0.033)	[0.050, 0.180]	<0.001
Workload stress	0.380 (0.018)	[0.345, 0.416]	<0.001

Note. N = 8739 teachers in 731 schools. Controls, weights and standardization follow Table 3. Estimates, SEs and CIs are displayed to three decimals. All calculations use unrounded values and normal critical value 1.96. For stress × efficacy, the upper CI endpoint is −0.00015 before rounding: displayed 0.000 is a rounded negative value. For this interaction, normal-based p = 0.049 and t(99)-based p = 0.051.

Job satisfaction and outcome-content checks

Student behaviour stress was associated with lower job satisfaction (b = −0.317, SE = 0.014, 95% CI −0.344 to −0.289, p < 0.001). Efficacy and collaboration had positive main associations (b = 0.177

and 0.115; both $p < 0.001$). The efficacy interaction was near zero ($b = 0.005$, SE = 0.015, $p = 0.739$), as was the collaboration interaction ($b = 0.002$, $p = 0.864$). The model R^2 was 0.174.

The stacked common-sample contrast between the efficacy interactions for job satisfaction and adverse workplace well-being was 0.040 (SE = 0.024, 95% CI −0.007 to 0.088, $p = 0.095$). The estimated interaction contrast was positive, with its confidence interval including zero. Removing the general work-stress item from the outcome yielded a student behaviour stress coefficient of 0.424 and an efficacy interaction of −0.033 (95% CI −0.059 to −0.007, $p = 0.014$).

Measurement-content comparisons

In the common-sample item-content analyses (Table 5), the released index retained a negative efficacy interaction (−0.033, SE 0.013, $p = 0.013$). Intimidation alone and the achievement/intimidation pair yielded estimates of 0.005 and 0.009, with neither interaction statistically significant. The special-education/diversity pair yielded −0.031 (unadjusted $p = 0.019$; Holm $p = 0.076$), while the three-item intimidation/special-education/diversity composite yielded −0.018 (unadjusted $p = 0.180$; Holm $p = 0.540$). The direct index-minus-three-item contrast included zero (−0.015, 95% CI −0.039 to 0.009, $p = 0.229$), whereas the index-minus-achievement/intimidation contrast was negative (−0.042, 95% CI −0.060 to −0.025, $p < 0.001$). The achievement/intimidation pair had lower reliability than the three-item composite.

Table 5 Common-sample comparison of stress measurement content

Stress measure	Interaction (SE)	95% CI	p	Holm p	N / schools
OECD student-behaviour stress index	−0.033 (0.013)	[−0.058, −0.007]	0.013	—	8650 / 731
Student intimidation or verbal abuse	0.005 (0.014)	[−0.023, 0.033]	0.713	1.000	8650 / 731
Achievement responsibility + intimidation	0.009 (0.014)	[−0.019, 0.037]	0.511	1.000	8650 / 731
Special-education lesson adaptation + diversity/equity work	−0.031 (0.013)	[−0.058, −0.005]	0.019	0.076	8650 / 731
Intimidation + special-education adaptation + diversity/equity work	−0.018 (0.013)	[−0.044, 0.008]	0.180	0.540	8650 / 731

Note. N = 8650 teachers in 731 schools in every model. The outcome is adverse workplace well-being. Models include the primary controls, weights and system effects. SEs and CIs use BRR-Fay and normal critical value 1.96. Holm adjustment covers the four alternative-item interaction tests. Labels identify the stress-item composites. The direct index-minus-three-item contrast was −0.015 (95% CI −0.039 to 0.009; p = 0.229). The index-minus-achievement/intimidation contrast was −0.042 (95% CI −0.060 to −0.025; p < 0.001).

The calibrated index and the alternative raw composites differ in content and reliability. A significant interaction for one measure and a nonsignificant interaction for another does not, by itself, establish a difference between them. The paired contrasts provide the relevant comparisons.

Specification and measurement sensitivity

In the restricted quadratic model, stress squared was positively associated with the outcome (b = 0.032, SE = 0.011, p = 0.004), and the joint test of stress squared and its efficacy interaction gave $\chi^2(2) = 8.11$, p = 0.017. The linear stress-by-efficacy term remained negative (b = −0.031, SE = 0.014, p = 0.030); the stress-squared-by-efficacy term was not statistically significant (p = 0.782). The restricted model indicates modest stress curvature. Adding stress-by-efficacy-squared gave 0.009 (SE 0.013, 95% CI −0.017 to 0.035, p = 0.496); the linear stress-by-efficacy coefficient in this

expanded model was −0.026 (SE 0.014, 95% CI −0.053 to 0.002, p = 0.065). The added term does not provide evidence of efficacy curvature, while the linear interaction is imprecise under this specification.

The efficacy interaction remained negative when experience was omitted (−0.035, SE 0.013), when variables were standardized within systems (−0.033, SE 0.013), and with school fixed effects (−0.029, SE 0.013). Excluding Korea yielded −0.037 (SE 0.016) under pooled scaling and −0.036 (SE 0.016) under within-system scaling. School-clustered inference for the primary interaction produced an SE of 0.014 and p = 0.015, compared with BRR-Fay SE 0.013 and p = 0.008. The interaction remained negative across these specifications.

Additional covariates, scaling and fitted contrasts

Using category indicators for experience and school count yielded an efficacy interaction of −0.036 (p = 0.007). Survey-weighted standardization gave −0.034 (95% CI −0.059 to −0.009, p = 0.008). The change in coefficient magnitude under weighted scaling reflects different units; it does not supply an independent robustness test.

In the school-intake block (8455 teachers), the efficacy interaction was −0.035 in its matched baseline and after intake adjustment; adding workload yielded −0.023 (95% CI −0.047 to 0.001, p = 0.061). In the classroom-condition block (4331 teachers), the interaction was −0.034 in its matched baseline and −0.032 after adding class size and perceived disruption; adding workload yielded −0.021 (95% CI −0.062 to 0.019). In the leadership block (4375 teachers), it was −0.033 in the matched baseline, −0.034 after leadership adjustment and −0.025 after workload (95% CI −0.054 to 0.003). The Shanghai–Singapore gender block (4553 teachers) gave −0.014 before and after gender adjustment, and −0.019 after workload (95% CI −0.051 to 0.013). These estimates remain negative but are less precise in the restricted samples.

Replacing both stress and efficacy content yielded near-zero exploratory interactions for instructional efficacy (0.001, 95% CI −0.026 to 0.029, p = 0.934; N = 8668) and student-engagement efficacy

(−0.001, 95% CI −0.028 to 0.025, p = 0.915; N = 8665). This pattern is consistent with content dependence but does not distinguish task matching from shared appraisal or differences in measurement reliability.

For the observed 90th-versus-10th efficacy percentile contrast, the fitted difference at +1 SD behaviour stress was −0.108 outcome SD (95% CI −0.203 to −0.014); at mean stress it was −0.011 (−0.095 to 0.072), and at −1 SD stress it was 0.086 (−0.037 to 0.209). At +1 SD behaviour stress, the fitted high-minus-low efficacy contrast was −0.158 outcome SD (95% CI −0.295 to −0.020). At mean stress it was −0.016 (−0.137 to 0.105), and at −1 SD stress it was 0.125 (−0.054 to 0.304). These contrasts use the primary model and represent a four-SD efficacy span; the +2 SD level exceeds the observed efficacy maximum, so this contrast includes extrapolation. They indicate that the relation between efficacy and the outcome depends on stress level, with uncertainty particularly evident near and below mean stress.

Exploratory system and quartile patterns

The system-specific efficacy interactions were −0.019 in Shanghai (95% CI −0.060 to 0.021), −0.050 in Japan (95% CI −0.096 to −0.004), −0.023 in Korea (95% CI −0.074 to 0.029), −0.033 in Singapore (95% CI −0.089 to 0.024), using the pooled primary-sample standardization and school-clustered SEs. The independent-system Wald heterogeneity test gave $\chi^2(3) = 1.06$, p = 0.787. This test did not detect heterogeneity; it does not demonstrate equivalent interactions across systems.

Compared with the lowest efficacy quartile, the stress-slope contrasts were −0.019 in quartile 2 (95% CI −0.089 to 0.051), −0.093 in quartile 3 (−0.168 to −0.018) and −0.069 in quartile 4 (−0.152 to 0.013), with school-clustered SEs. The corresponding nominal p-values were 0.594, 0.015 and 0.100; Holm-adjusted values were 0.594, 0.045 and 0.200. The omnibus interaction test gave p = 0.070. The smaller highest-quartile contrast indicates a non-monotonic pattern, but does not establish the role of endpoint concentration.

Discussion

Behaviour-related demands alongside workload

Behaviour stress remained positively associated with adverse workplace well-being after workload adjustment. Its coefficient decreased when workload entered the model, indicating that the unadjusted estimate partly reflects concurrent occupational demands. This pattern accords with demand–resource accounts and evidence linking misbehaviour to teacher burnout (5, 8, 13). It also complements recent Chinese research on broader work stress and affective rumination (3) by distinguishing a specific classroom demand from work volume. Teachers' perceived difficulty managing classroom behaviour warrants attention alongside preparation, marking and administrative pressure.

Task correspondence and measurement content

The efficacy interaction was modest, added little explained variation and became less precise after workload adjustment. The item-content comparisons are central to its interpretation. The released stress index closely reflected discipline stress, which corresponds to the behaviours addressed by classroom management efficacy. This correspondence is compatible with the triple-match principle (26), but the same pattern can arise from shared appraisals or related questionnaire content. The current self-report design cannot distinguish these explanations.

The alternative measures do not form a controlled experiment in which only the weight of the discipline item changes. They assess different demands and have different reliability. Although intimidation and achievement/intimidation produced near-zero interactions, the index-versus-three-item contrast did not establish a difference. The special-education/diversity estimate was negative but did not survive Holm adjustment. Responding to diverse classroom needs may draw on management skills, yet this exploratory estimate does not establish that mechanism or refute task

correspondence. Future studies can contrast management efficacy with efficacy in instruction and engagement while measuring classroom conditions independently.

Removing the general work-stress item from the outcome retained a negative interaction. The pattern therefore extends beyond the most directly overlapping stress wording to perceived personal-time and health effects. It does not eliminate common-method variance (38). Classroom observations, administrative indicators and repeated measures would help distinguish exposure to disruption, appraisal of demands and changes in well-being.

Efficacy, collaboration and different outcomes

The near-zero efficacy coefficient for adverse well-being and positive coefficient for job satisfaction address different outcomes at the mean level of stress. They are not contradictory. Confidence in classroom management may accompany favourable evaluations of professional life while relating to strain through an interaction with demands. Satisfaction concerns occupational evaluation, whereas the workplace well-being scale emphasises strain (1, 39). The satisfaction result accords with research linking efficacy to occupational evaluations (21, 23); recent early-career teacher research likewise concerns efficacy and burnout through a mediation model, rather than the moderation tested here (25).

A TALIS 2018 study in Türkiye reported efficacy moderation of stress and satisfaction (40). Differences in the demands and efficacy domains measured, the outcome and the education context could contribute to the contrasting estimates. These possibilities are not separately tested here. Our between-outcome interaction contrast included zero, so the observed pattern does not establish that efficacy moderation is specific to adverse workplace well-being. Professional collaboration showed favourable main associations but little evidence of moderation, consistent with a broader role in shared learning and support (27). Its frequency measure also leaves the quality and responsiveness of collaboration unmeasured.

Practical interpretation

The primary stress slope was lower at higher efficacy, but remained positive across the evaluated range. The confidence interval for the central-range slope contrast indicates uncertainty around a small difference; it is not an estimate of an intervention effect. The limited increment in explained variation and attenuation with workload give little basis for prioritising efficacy training over improvements in working conditions from this study alone.

Classroom management coaching can be evaluated together with disciplinary processes, leadership responsiveness and workload arrangements. Expert consensus on educator well-being highlights sustained support, relationships and leadership (31). Evaluations should track classroom practice, perceived demands and well-being over time, allowing improvements in skills to be distinguished from reductions in exposure to demanding conditions.

Strengths and limitations

The large four-system sample, teacher weights and replicate-weight inference support a detailed assessment of associations within the TALIS framework. Paired interaction contrasts and common-sample item comparisons provide more informative evidence than comparisons of significance labels. Workload adjustment, school fixed effects and within-system scaling examine different aspects of model sensitivity.

The cross-sectional design leaves temporal ordering unresolved. Additional models incorporate class size, classroom disruption and relational leadership, but questionnaire availability prevents joint adjustment for all these conditions. Gender adjustment is restricted to two systems, and school intake is measured in broad categories reported by principals. School fixed effects address shared school conditions but do not resolve classroom or individual confounding. Complete-case selection and Korea's lower participation may affect population interpretation, and observed balance cannot rule out selection on unobserved factors.

Endpoint concentration in efficacy and differences in item reliability also matter. The expanded quadratic model tests nonlinear moderation by efficacy but provides no clear evidence for that term. The non-monotonic quartile pattern alone cannot be attributed to a ceiling effect. OECD calibration supplies a common score framework; four-system item-level invariance was not re-estimated (41-43). Within-system scaling retained a negative interaction, but does not establish measurement equivalence or absence of between-system variation.

Conclusions

Student behaviour stress was associated with adverse workplace well-being alongside workload in four TALIS 2024 education systems. Classroom management efficacy showed a modest negative interaction that attenuated after workload and additional covariate adjustment and was imprecise in the expanded quadratic model. Estimates varied across stress measures, with task correspondence, shared content and reliability all relevant to interpretation; the index-versus-three-item contrast did not establish a difference. Behaviour-related demands warrant attention within broader working conditions. The results support precise measurement and joint evaluation of professional resources and organisational support, rather than a general claim that efficacy training buffers teacher stress.

List of abbreviations

BRR: balanced repeated replication; CI: confidence interval; ICC: intraclass correlation coefficient; OECD: Organisation for Economic Co-operation and Development; SD: standard deviation; SE: standard error; TALIS: Teaching and Learning International Survey.

Declarations

Ethics approval and consent to participate

This study was a secondary analysis of publicly available, de-identified OECD TALIS 2024 data. The authors did not recruit or contact participants and did not access directly identifying information. Original data collection was conducted through the TALIS programme; its technical standards require voluntary participation, protection of respondents' privacy and compliance with applicable research ethics procedures(44).

Consent for publication

Not applicable. No identifiable individual information is reported.

Availability of data and materials

The TALIS 2024 public-use data and supporting documentation are available from the OECD at https://www.oecd.org/en/data/datasets/talis-2024-database.html. Code for reproduction of the primary models and the additional revision analyses accompanies the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

No specific funding was received for this study.

Authors' contributions

LM contributed to supervision, interpretation and review and editing.BL contributed to conceptualization, methodology, formal analysis, data curation and writing of the original draft,and is the corresponding author.

Acknowledgements

We thank the OECD for making the TALIS 2024 public-use data and documentation available.

Supplementary information

Additional file 1 (.docx): Supplementary tables and diagnostic summaries. Contains 18 tables covering sample flow, participation, variable definitions, slopes, outcome and model comparisons, full quadratic and scaling analyses, covariate blocks, content replacements, predicted contrasts and complete system and quartile estimates. The analysis code and numerical outputs are supplied separately.

References

1. Hascher T, Waber J. Teacher well-being: A systematic review of the research literature from the year 2000–2019. Educational Research Review. 2021;34:100411.
2. Viac C, Fraser P. Teachers' well-being: A framework for data collection and analysis. Paris: OECD Publishing; 2020. Contract No.: No. 213.
3. Shao Y, Jiang W, Zhu H, Zhang C, Xu W. The relationship between work stress and well-being among Chinese primary and secondary school teachers: The chain mediation of affective rumination and work engagement. BMC Psychology. 2025;13(1):337.
4. Aldrup K, Klusmann U, Lüdtke O, Göllner R, Trautwein U. Student misbehavior and teacher well-being: Testing the mediating role of the teacher-student relationship. Learning and Instruction. 2018;58:126–36.
5. Aloe AM, Shisler SM, Norris BD, Nickerson AB, Rinker TW. A multivariate meta-analysis of student misbehavior and teacher burnout. Educational Research Review. 2014;12:30–44.
6. Spilt JL, Koomen HMY, Thijs JT. Teacher Wellbeing: The Importance of Teacher–Student Relationships. Educational Psychology Review. 2011;23(4):457–77.
7. Bakker AB, Demerouti E. The Job Demands-Resources model: state of the art. Journal of Managerial Psychology. 2007;22(3):309–28.
8. Bakker AB, Demerouti E. Job demands–resources theory: Taking stock and looking forward. Journal of Occupational Health Psychology. 2017;22(3):273–85.
9. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. Journal of Applied Psychology. 2001;86(3):499–512.
10. Hobfoll SE. Conservation of resources: A new attempt at conceptualizing stress. American Psychologist. 1989;44(3):513–24.
11. Hobfoll SE, Halbesleben J, Neveu J-P, Westman M. Conservation of Resources in the Organizational Context: The Reality of Resources and Their Consequences. Annual Review of Organizational Psychology and Organizational Behavior. 2018;5(1):103–28.
12. Collie RJ. Teacher well-being and turnover intentions: Investigating the roles of job resources and job demands. British Journal of Educational Psychology. 2023;93(3):712–26.

13. Dicke T, Stebner F, Linninger C, Kunter M, Leutner D. A longitudinal study of teachers' occupational well-being: Applying the job demands-resources model. Journal of Occupational Health Psychology. 2018;23(2):262–77.
14. Hakanen JJ, Bakker AB, Schaufeli WB. Burnout and work engagement among teachers. Journal of School Psychology. 2006;43(6):495–513.
15. Skaalvik EM, Skaalvik S. Job demands and job resources as predictors of teacher motivation and well-being. Social Psychology of Education. 2018;21(5):1251–75.
16. OECD. Results from TALIS 2024: The State of Teaching. Paris: OECD Publishing; 2025.
17. Bandura A. Self-efficacy: The exercise of control. New York, NY: W. H. Freeman; 1997.
18. Emmer ET, Hickman J. Teacher Efficacy in Classroom Management and Discipline. Educational and Psychological Measurement. 1991;51(3):755–65.
19. Tschannen-Moran M, Hoy AW. Teacher efficacy: capturing an elusive construct. Teaching and Teacher Education. 2001;17(7):783–805.
20. Duan S, Bissaker K, Xu Z. Correlates of teachers' classroom management self-efficacy: A systematic review and meta-analysis. Educational Psychology Review. 2024;36(2):43.
21. Klassen RM, Chiu MM. Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. Journal of Educational Psychology. 2010;102(3):741–56.
22. Skaalvik EM, Skaalvik S. Teacher self-efficacy and teacher burnout: A study of relations. Teaching and Teacher Education. 2010;26(4):1059–69.
23. Zee M, Koomen HMY. Teacher Self-Efficacy and Its Effects on Classroom Processes, Student Academic Adjustment, and Teacher Well-Being. Review of Educational Research. 2016;86(4):981–1015.
24. Burić I, Kim LE. Job satisfaction predicts teacher self-efficacy and the association is invariant: Examinations using TALIS 2018 data and longitudinal Croatian data. Teaching and Teacher Education. 2021;105:103406.
25. Pikić Jugović I, Marušić I, Matić Bojić J. Early career teachers' social and emotional competencies, self-efficacy and burnout: a mediation model. BMC Psychology. 2025;13(1):9.
26. de Jonge J, Dormann C. Stressors, resources, and strain at work: A longitudinal test of the triple-match principle. Journal of Applied Psychology. 2006;91(6):1359–74.
27. Christensen AA, Jerrim J. Professional learning communities and teacher outcomes. A cross-national analysis. Teaching and Teacher Education. 2025;156:104920.
28. Fackler S, Malmberg L-E. Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects. Teaching and Teacher Education. 2016;56:185–95.
29. Fackler S, Malmberg L-E, Sammons P. An international perspective on teacher self-efficacy: Personal, structural and environmental factors. Teaching and Teacher Education. 2021;99:103255.
30. Sun A, Xia J. Teacher-perceived distributed leadership, teacher self-efficacy and job satisfaction: A multilevel SEM approach using the 2013 TALIS data. International Journal of Educational Research. 2018;92:86–97.
31. Patrick P, Reupert A, Berger E, Morris Z, Diamond Z, Hammer M, et al. Initiatives for promoting educator wellbeing: a Delphi study. BMC Psychology. 2024;12(1):220.
32. Vieluf S, Kunter M, van de Vijver FJR. Teacher self-efficacy in cross-national perspective. Teaching and Teacher Education. 2013;35:92–103.
33. OECD. TALIS 2024 database and supporting materials Paris: Organisation for Economic Co-operation and Development (OECD); 2025 [Available from: https://www.oecd.org/en/data/datasets/talis-2024-database.html.
34. OECD. Teaching and Learning International Survey (TALIS) 2024 user guide Paris: Organisation for Economic Co-operation and Development (OECD); 2025 [Available from: https://one.oecd.org/document/EDU/TALIS/GB%282025%299/en/pdf.
35. OECD. Teaching and Learning International Survey (TALIS) 2024 conceptual framework. Paris: OECD Publishing; 2025.
36. OECD. Technical notes on sampling, participation rates and adjudication Paris: OECD Publishing; 2025 [Available from: https://www.oecd.org/en/publications/results-from-talis-

2024_90df6235-en/full-report/technical-notes-on-sampling-participation-rates-and-adjudication_1067298b.html.
37. OECD. Technical notes on analyses in this volume Paris: OECD Publishing; 2025 [Available from: https://www.oecd.org/en/publications/results-from-talis-2024_90df6235-en/full-report/technical-notes-on-analyses-in-this-volume_286d707e.html.
38. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology. 2003;88(5):879–903.
39. Skaalvik EM, Skaalvik S. Teacher job satisfaction and motivation to leave the teaching profession: Relations with school context, feeling of belonging, and emotional exhaustion. Teaching and Teacher Education. 2011;27(6):1029–38.
40. Uslukaya A. Öz-yeterliğin Öğretmen Stresi ve İş Tatmini İlişkisindeki Düzenleyici Rolü: TALIS 2018 Bulguları. İnönü Üniversitesi Eğitim Fakültesi Dergisi. 2025;26(3):1824–53.
41. Desa D. Evaluating measurement invariance of TALIS 2013 complex scales: Comparison between continuous and categorical multiple-group confirmatory factor analyses. Paris: OECD Publishing; 2014. Contract No.: No. 103.
42. Rutkowski L, Svetina D. Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys. Educational and Psychological Measurement. 2014;74(1):31–57.
43. Vandenberg RJ, Lance CE. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. Organizational Research Methods. 2000;3(1):4–70.
44. OECD. TALIS 2024 technical standards. Paris: OECD Publishing; 2025. Contract No.: EDU/TALIS/GB(2023)11/REV1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile.docx