

Associations of Teacher Workload and Work Motivation with Self-Reported Teacher Performance in an Indonesian Madrasah

Hendrafit Hongki ¹, Hima Barima ^{2*}, Dyah Supriatin ³

^{1,2*,3} Department of Management, Faculty of Economics and Business, Universitas Perwira Purbalingga, Purbalingga Regency, Central Java Province, Indonesia.

Email: hendrafit08@gmail.com ¹, himabarima@gmail.com ^{2*}, dyah.raisha@unperba.ac.id ³

Article History

Received: 22 Agustus 2026.

Revised: 20 Oktober 2026.

Accepted: 1 November 2026.

Available online: 20 November 2026.

Published: 1 Januari 2027.

Keyword

Teacher Workload; Work Motivation; Teacher Performance; Madrasah.

Kata Kunci

Beban Kerja Guru; Motivasi Kerja; Kinerja Guru; Madrasah.

ABSTRACT

This single-site study examined workload and work motivation in relation to self-reported teacher performance in one Indonesian madrasah. All 34 eligible teachers joined a cross-sectional census. The primary analysis used 32 items. Regression accounted for 44.6% of performance-score variation, $F(2, 31) = 12.494$, $p < 0.001$. Workload was nonsignificant ($B = -0.068$, $p = 0.557$; H1 not supported), whereas motivation was positive and significant ($B = 0.693$, $p < 0.001$; H2 supported). Decisions persisted with HC3, leave-one-out, and the exploratory 29-item model. These single-site self-report associations are not causal effects.

ABSTRAK

Studi situs tunggal ini menguji hubungan beban kerja dan motivasi kerja dengan kinerja guru berdasarkan laporan diri pada satu madrasah di Indonesia. Seluruh 34 guru yang memenuhi syarat mengikuti sensus potong lintang. Analisis primer memakai 32 butir. Model regresi menjelaskan 44,6% variasi skor kinerja, $F(2, 31) = 12,494$; $p < 0,001$. Beban kerja tidak signifikan ($B = -0,068$; $p = 0,557$; H1 tidak didukung), sedangkan motivasi positif dan signifikan ($B = 0,693$; $p < 0,001$; H2 didukung). Keputusan bertahan pada HC3, leave-one-out, dan model 29-butir eksploratif. Temuan situs tunggal berbasis laporan diri ini bersifat asosiatif, bukan kausal.

1. Introduction

Teachers divide finite time and attention among instruction, preparation, assessment, student support, communication, documentation, and institutional duties. Reviews describe the resulting work intensification and time poverty as central features of teachers' working conditions (Creagh *et al.*, 2025; Morris *et al.*, 2026), while overload and extended hours are linked to exhaustion and poorer wellbeing (Moens *et al.*, 2026; Sandmeier *et al.*, 2022; Zhou *et al.*, 2024). These outcomes matter, but they do not establish how a broad workload composite relates to teachers' accomplishment of professional duties. Workload is therefore distinguished here from wellbeing and from performance. It represents the volume and coordination of teaching hours, classes, students, instructional administration, and additional duties. Particular demands may compete for time, yet some are integral to performing the teaching role; task-level and challenge-hindrance evidence consequently cautions against assuming one universal coefficient (Jerrim & Sims, 2021; LePine *et al.*, 2005). Gan (2025) found generally poorer performance with greater workload but also variation among workload components, and Indonesian models embed working conditions within wider organisational configurations (Ulfah *et al.*, 2024).

Work motivation is conceptually different: it concerns the willingness to initiate, direct, and sustain professional effort. The present measure covers enthusiasm, responsibility, achievement orientation, discipline, and professional development. Reviews connect teacher motivation with cognitive, emotional, and behavioural processes in teaching, while studies across settings generally report positive motivation-performance associations (Bardach *et al.*, 2022; Lazarides *et al.*, 2025; Layek & Koodamara, 2024). Motivation nevertheless remains embedded in organisational conditions and should not be reduced to an isolated personal trait (Collie, 2023; Ryan & Deci, 2020). The opportunity-capacity-willingness framework organises these distinctions (Blumberg & Pringle, 1982). Workload is treated as a situational condition related to opportunity because competing duties can alter the time and attention available for particular tasks; motivation represents willingness to invest effort. Capacity—knowledge, skill, and other personal resources—was not measured. The study therefore estimates two theory-informed correlates rather than a complete test of the framework. The outcome is self-reported professional accomplishment across work quality and quantity, timeliness, resource use, independence, and interpersonal performance. It is not student achievement, school effectiveness, a supervisory rating, or observed classroom practice. Measuring

both predictors and the outcome in the same questionnaire also creates common-source and perceptual conditions that may strengthen associations, particularly between motivation and reported performance (Podsakoff *et al.*, 2003). Prior studies differ in workload content, motivational construct, performance source, and institutional context (Eryilmaz *et al.*, 2025; Gan, 2025; Habibi *et al.*, 2025; Kholifah *et al.*, 2024; Toropova *et al.*, 2021; Ulfah *et al.*, 2024). The unresolved question is whether broad workload and work motivation show distinct simultaneous associations with self-reported performance within the bounded organisational setting of an Indonesian madrasah. Evidence from all eligible teachers at one site can answer that local question, but it cannot produce a nationally representative estimate. Accordingly, this study examines the simultaneous associations of workload and work motivation with self-reported performance among 34 teachers at one madrasah. Its contribution is a transparent, context-specific qualification of heterogeneous prior findings, including diagnostics suited to a small dataset and a clear separation between the prespecified 32-item analysis and post hoc scale refinement. The cross-sectional design supports associative, not causal, interpretation.

Because a broad workload measure may combine facilitative and obstructive responsibilities, its hypothesis is non-directional. A positive motivation hypothesis follows from its alignment with willingness to perform and the prevailing direction of empirical evidence. H1: Teacher workload is associated with self-reported teacher performance. H2: Work motivation is positively associated with self-reported teacher performance. The framework keeps the constructs analytically separate. Opportunity concerns situational features that enable or constrain work, willingness concerns motivational readiness, and capacity concerns the resources required to perform (Blumberg & Pringle, 1982). Workload is an imperfect indicator of opportunity rather than an inherently harmful exposure; motivation is an indicator of willingness rather than a substitute for opportunity or capacity. The empirical model therefore estimates the partial association of workload (X_1) and work motivation (X_2) with self-reported teacher performance (Y). It does not test mediation, moderation, interactions, indirect effects, or whether one coefficient is statistically larger than the other. Capacity and task-specific workload components remain outside the model. Figure 1 depicts this parsimonious architecture. Its arrows denote hypothesised associations, not temporal or causal effects.

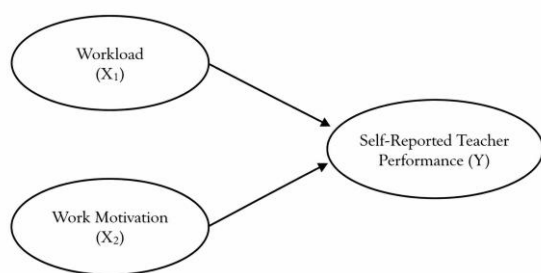


Figure 1. Conceptual Framework of the Study

2. Research Methodology

2.1 Research Design and Setting

This study used a quantitative, cross-sectional, observational questionnaire design to estimate simultaneous associations without manipulating any variable. Exposure and outcome information was collected during one period, so temporal order and causal effects cannot be established. The study was conducted at MTs Darul Abror Kedungjati, Bukateja, Purbalingga, Indonesia. The individual teacher was the unit of analysis. The site was retained explicitly as the organisational boundary of the evidence.

2.2 Participants and Data Collection

The study population comprised all 34 active teachers registered in SIMPATIKA who met the eligibility criteria. Total population sampling (a census) was used, and all 34 teachers participated and were included in the analysis, yielding 100% coverage of the eligible site population. Thus, the census fully describes this site's eligible teacher group but does not remove uncertainty for claims about teachers elsewhere. Paper questionnaires were distributed and collected in person between 16 July and 20 August 2026. All 34 forms contained complete responses to the 32 study items, yielding a 34/34 response set; no item imputation was used. Numeric row identifiers were used in the analytical matrix, and only aggregate results are reported. The university issued research-permission letter No. 119/061051.A/PH/2026 on 9 July 2026, and the madrasah issued authorisation letter No. 017/MTs.DA/C.01/VII/2026 on 15 July 2026. The questionnaire cover letter invited willingness to participate, stated that responses would be confidential, and limited their use to research. The archived materials do not document a separate signed consent form, review or exemption by a research ethics committee, or a formal protocol preventing school management from accessing completed individual forms. These documentation limits are reported rather than inferred.

2.3 Measures

The questionnaire was administered in Indonesian and measured self-reported teacher performance, workload, and work motivation. Every statement used a five-point scale from 1 (Strongly Disagree) to 5 (Strongly Agree); all were keyed in the same direction and summed without weights. Higher scores indicate higher reported performance, workload, or motivation. The English renderings in Appendix A were prepared for reporting and were not part of administration; the records do not document translation/back-translation, expert review, or a separate pilot test.

2.4 Self-Reported Teacher Performance

The performance measure comprised 12 study-developed, context-specific first-person statements covering work quality and quantity, timeliness, resource use, independence from supervision, and interpersonal impact. Wahyuni *et al.* (2024) is retained as a construct/domain source, but its article describes a larger questionnaire and does not publish the wording needed to establish item-by-item equivalence. Azzahra *et al.* (2019) provides an aligned six-domain framework, and Koopmans *et al.* (2014) provides broader measurement context; neither is claimed as the verbatim source of the administered statements.

2.5 Workload

The workload measure comprised 10 study-developed statements arranged as paired indicators of teaching hours, class load, student load, instructional administration, and additional duties. Ndruru and Yudiaatmaja (2025) reports a six-item workload measure but does not publish its wording; it is therefore treated as a construct-level source, not a direct item source. Spector and Jex (1998) is cited only as international workload context.

2.6 Work Motivation

The work-motivation measure comprised 10 study-developed paired statements on enthusiasm, responsibility, achievement orientation, discipline, and professional development. Indrastuti *et al.* (2025) does not publish the wording required to verify direct adaptation and is treated as a construct-level source. The Multidimensional Work Motivation Scale has a different regulatory structure (Gagné *et al.*, 2015) and is contextual only. No administered item is presented as a verbatim reproduction of either source. Table 1 summarises operationalisation and provenance. Appendix A supplies every administered Indonesian statement, an English rendering, and an item-level record of its contextual focus and source relationship.

Table 1. Operationalisation, Scoring, and Instrument Provenance

Construct	Operationalisation	Administered items and scoring	Published-source relationship
Self-Reported Teacher Performance (Y)	Perceived quality and quantity, timeliness, resource use, independence, and interpersonal impact	12 Indonesian statements; five-point agreement responses; unweighted sum, 12–60	Study-developed wording. Wahyuni <i>et al.</i> (2024): construct/domain source; wording not published. Azzahra <i>et al.</i> (2019) and Koopmans <i>et al.</i> (2014): framework/context only.
Workload (X_1)	Teaching hours, class and student load, instructional administration, and additional duties	10 Indonesian statements in five pairs; five-point agreement responses; unweighted sum, 10–50	Study-developed wording. Ndruru and Yudiaatmaja (2025): construct-level source; six source items reported but wording not published. Spector and Jex (1998): context only.
Work Motivation (X_2)	Enthusiasm, responsibility, achievement, discipline, and professional development	10 Indonesian statements in five pairs; five-point agreement responses; unweighted sum, 10–50	Study-developed wording. Indrastuti <i>et al.</i> (2025): construct-level source; wording not published. Gagné <i>et al.</i> (2015): context only.

Table 2. Primary and Post Hoc Exploratory Scale Specifications

Construct	Primary specification	Original corrected item–total range	Post hoc exploratory specification	Interpretation
Performance	12 items; $\alpha = 0.822$	0.162–0.666	11 items; Y9 removed; Y1 retained; $\alpha = 0.834$; range = 0.335–0.620	Primary inference uses all 12 items
Workload	10 items; $\alpha = 0.843$	0.235–0.702	8 items; X1.1 and X1.9 removed; $\alpha = 0.865$; range = 0.476–0.691	Primary inference uses all 10 items
Motivation	10 items; $\alpha = 0.869$	0.505–0.654	10 items unchanged; $\alpha = 0.869$; range = 0.505–0.654	No post hoc exclusion

2.7 Prespecified Scoring, Internal Consistency, and Exploratory Refinement

The primary specification retained all 32 administered items: 12 performance, 10 workload, and 10 motivation statements. Each composite was the unweighted item sum. In the 34 responses, Cronbach's alpha was 0.822 for performance, 0.843 for workload, and 0.869 for motivation. Original corrected item–total correlations ranged from 0.162 to 0.666, 0.235 to 0.702, and 0.505 to 0.654, respectively. These statistics describe internal coherence in this dataset; they do not establish content, structural, convergent, discriminant, or criterion validity. A post hoc diagnostic identified Y1 = 0.293, Y9 = 0.162, X1.1 = 0.257, and X1.9 = 0.235 below 0.30. Removing Y9, X1.1, and X1.9 produced 11-, 8-, and 10-item composites with alphas of 0.834, 0.865, and 0.869. Because selection and hypothesis testing used the same 34 responses and no prior expert decision justified exclusion, this 29-item model is labelled exploratory and cannot be used to claim a validated scale or reliability improvement beyond this sample. Y1 was retained in the exploratory set

because it was borderline and its corrected item–total correlation became 0.335 after Y9 was removed.

2.8 Data Analysis

The primary analysis used the original 32-item specification. Composite descriptives, corrected item–total correlations, Cronbach's alpha, and Pearson correlations were recalculated from the archived item matrix. Multiple linear regression and HC3 calculations used statsmodels 0.14.6; the original IBM SPSS Statistics 29 output was used as a cross-check. All tests were two-tailed with $\alpha = 0.05$.

The primary regression model was:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (1)$$

Where Y denotes the 12-item Self-Reported Teacher Performance sum, X_1 the 10-item Workload sum, X_2 the 10-item Work Motivation sum, β_0 the intercept, β_1 and β_2 the partial regression coefficients, and ε the error term. H1 was evaluated with β_1 and H2 with β_2 . OLS coefficients, standard errors, standardised coefficients, t statistics, two-tailed p-values, and 95% confidence intervals were examined. The model F-

test assessed the predictors jointly; R^2 and adjusted R^2 describe sample-level explanatory coverage, not causal effects. Diagnostics covered residual normality (Shapiro–Wilk), multicollinearity (VIF), heteroskedasticity (Glejser and Breusch–Pagan), functional form (Ramsey RESET and centred quadratic terms), leverage, Cook's distance, externally studentised residuals, and DFFITS. HC3 standard errors were used because the Glejser result was not uniformly consistent with homoskedasticity (Hayes & Cai, 2007). Leave-one-out models assessed coefficient and R^2 stability. Durbin–Watson is reported descriptively only because the archived rows have no meaningful temporal order; independence rests chiefly on distinct respondents, not that statistic. HC3 does not remedy common-source measurement, misspecification, or influential observations. Because the study enumerated one bounded site rather than drawing a probability sample from a wider population, reported p-values and confidence intervals are model-based quantities conditional on these 34 observations and the specified linear model. They are not design-based estimates of sampling uncertainty for Indonesian teachers or other madrasahs. The post hoc 29-item model is reported separately as exploratory. Data collection and the archived calculations were completed in 2026; the cited 2026 review articles were available before final manuscript preparation.

3. Results and Discussion

3.1 Results

3.1.1 Respondent and Composite-Score Overview

All 34 eligible teachers were analysed: 21 women (61.8%) and 13 men (38.2%). Mean age was 35.03 years ($SD = 6.07$; range = 23–48), and mean teaching tenure was 8.03 years ($SD = 4.39$; range = 1–15). Thirty-one held a bachelor's degree (91.2%), two a master's degree (5.9%), and one had completed senior secondary education (2.9%). In the primary 32-item specification, alpha was 0.822 for performance, 0.843 for workload, and 0.869 for motivation. Original corrected item–total correlations ranged from 0.162 to 0.666, 0.235 to 0.702, and 0.505 to 0.654. The values indicate internal coherence in this dataset only; they do not establish construct validity. Primary mean sum scores were 49.382 ($SD = 4.404$) for performance, 37.676 ($SD = 5.227$) for workload, and 39.941 ($SD = 4.292$) for motivation. Motivation correlated positively with performance ($r = 0.663$, $p < 0.001$). Workload correlated neither with performance ($r = 0.019$, $p = 0.916$) nor with motivation ($r = 0.146$, $p = 0.409$). These bivariate coefficients are descriptive; hypotheses were evaluated in the regression.

3.1.2 Regression Diagnostics

Residual normality was adequate (Shapiro–Wilk $W = 0.965$, $p = 0.339$), and VIF was 1.022 for both predictors. Ramsey RESET ($F = 0.071$, $p = 0.792$) and a joint test of centred quadratic terms ($F = 0.273$, $p = 0.763$) did not indicate functional-form misspecification. The Glejser test was nonsignificant for workload ($p = 0.092$) but significant for motivation ($p = 0.025$); the Breusch–Pagan F test was nonsignificant ($p = 0.099$). HC3 inference was therefore retained. Durbin–Watson was 1.860, but row order was arbitrary and the statistic is descriptive rather than proof of independence. Influence diagnostics identified observations requiring sensitivity checks rather than automatic deletion. Five cases exceeded the leverage reference $2(k + 1)/N = 0.176$ (maximum = 0.331, case 21); three exceeded Cook's $D = 4/N = 0.118$ (cases 25, 26, and 34; maximum = 0.268); four had |externally studentised residual| > 2 (cases 1, 5, 7, and 26; maximum = 2.341); and cases 25, 26, and 34 exceeded the DFFITS reference 0.594 (maximum = 0.926). In 34 leave-one-out models, the workload coefficient ranged from -0.124 to -0.023 and remained nonsignificant ($p = 0.275$ – 0.853), whereas the motivation coefficient ranged from 0.633 to 0.792 and remained positive and significant (all $p < 0.001$); R^2 ranged from 0.374 to 0.496. Case numbers denote archived analytical row order, not identities.

3.1.3 Hypothesis Tests and Robustness

The primary analysis utilizing the 32-item model yielded statistically significant results, with $F(2, 31) = 12.494$ and $p < 0.001$. The model demonstrated an R value of 0.668, indicating a moderate correlation, while the R^2 value of 0.446 suggests that 44.6% of the variance in performance scores can be explained by the two predictors. However, it is important to note that this finding does not imply a causal relationship or a population-wide estimate. When controlling for motivation, the analysis revealed that workload was not significantly associated with performance, as indicated by a regression coefficient of $B = -0.068$ (OLS $SE = 0.114$, $\beta = -0.080$, $t = -0.593$, $p = 0.557$; HC3 $p = 0.508$). Consequently, Hypothesis 1 (H1) was not supported. Conversely, when controlling for workload, motivation exhibited a positive association with performance, with a regression coefficient of $B = 0.693$ (OLS $SE = 0.139$, $\beta = 0.675$, $t = 4.997$, $p < 0.001$; HC3 $p < 0.001$). This supports Hypothesis 2 (H2), indicating that for each one-point increase in motivation, there is a corresponding increase of 0.693 points in reported performance, with workload held constant; this coefficient does not reflect an intervention effect. The post hoc exploratory model, which included 29 items, produced similar conclusions with slightly adjusted estimates: workload $B = -0.132$ (OLS $p = 0.272$; HC3 $p = 0.262$) and

motivation $B = 0.665$ (both OLS and HC3 $p < 0.001$). The R^2 value for this model was 0.460, and it achieved $F(2, 31) = 13.217$ ($p < 0.001$). However, since the same respondents were used for item selection and

estimation, the observed differences in fit should not be interpreted as an improvement in the model.

Table 3. Primary 32-Item Descriptive Statistics, Reliability, and Correlations

Variable	k	M	SD	Possible / observed range	α	1	2	3
1. Performance	12	49.382	4.404	12–60 / 40–58	0.822	—		
2. Workload	10	37.676	5.227	10–50 / 26–50	0.843	0.019	—	
3. Motivation	10	39.941	4.292	10–50 / 30–50	0.869	0.663***	0.146	—

Table 4. Primary-Model Diagnostics and Influence Analysis

Diagnostic	Statistic / criterion	Observed result	Reference	Interpretation
Marginal score distributions	Shapiro–Wilk (descriptive)	Performance $W = 0.931$, $p = 0.034$; workload $W = 0.974$, $p = 0.592$; motivation $W = 0.912$, $p = 0.009$	Not an OLS requirement	Residual distribution is the inferential focus
Residual normality	Shapiro–Wilk	$W = 0.965$; $p = 0.339$	$p > 0.05$	No detected departure
Functional form	Ramsey RESET / centred quadratics	$F = 0.071$, $p = 0.792$ / $F = 0.273$, $p = 0.763$	$p > 0.05$	No detected nonlinear misspecification
Multicollinearity	VIF	1.022 for both predictors	$VIF < 5$	No indication
Heteroskedasticity	Glejser / Breusch–Pagan F	Workload $p = 0.092$; motivation $p = 0.025$ / $p = 0.099$	Mixed	HC3 inference reported
Residual order	Durbin–Watson	1.860	No temporal order	Descriptive only
Leverage	$h > 2(k + 1)/N = 0.176$	Cases 21, 25, 30, 32, 34; max = 0.331	Flag for review	Sensitivity analysis required
Cook's distance	$D > 4/N = 0.118$	Cases 25, 26, 34; max = 0.268	Flag for review	No automatic deletion
Studentised residual	$ r_{student} > 2$	Cases 1, 5, 7, 26; max $ r = 2.341$	Flag for review	No automatic deletion
DFFITS	$ DFFITS > 2\sqrt{((k + 1)/N)} = 0.594$	Cases 25, 26, 34; max = 0.926	Flag for review	Leave-one-out analysis used
Leave-one-out	34 refitted models	Bworkload -0.124 to -0.023 ; Bmotivation 0.633 to 0.792 ; R^2 0.374 – 0.496	Decision stability	Both hypothesis decisions unchanged

Table 5. Primary 32-Item Multiple Linear Regression Results

Predictor	B	OLS SE	β	t	OLS p	HC3 SE	HC3 p	HC3 95% CI	Decision
Constant	24.258	6.514	—	3.724	< 0.001	8.293	0.006	[7.345, 41.172]	—
Workload	−0.068	0.114	−0.080	−0.593	0.557	0.101	0.508	[−0.273, 0.138]	H1 not supported
Work Motivation	0.693	0.139	0.675	4.997	< 0.001	0.155	< 0.001	[0.376, 1.010]	H2 supported

3.2 Discussion

The primary analysis did not support a partial association between the broad workload composite and self-reported teacher performance. The HC3 interval [−0.273, 0.138] indicates modest associations in either direction, and every leave-one-out estimate

remained nonsignificant, providing insufficient evidence for H1 within the model. This outcome does not prove that workload is irrelevant or that additional duties are harmless; rather, it qualifies rather than directly contradicts earlier evidence. Gan (2025) reported generally poorer performance with greater

workload, although with variation among components, and reviews emphasize heterogeneous tasks, time poverty, and nonlinear occupational relationships (Creagh *et al.*, 2025; Jerrim & Sims, 2021; Morris *et al.*, 2026). The aggregate workload assessed here may combine role-enacting duties with competing demands, allowing task-specific relationships to offset one another. Within the opportunity–capacity–willingness framework, workload serves only as a broad indicator of situational opportunity, and the model does not disaggregate demands or measure capacity, resources, timing, or interactions, which limits its ability to identify when specific tasks facilitate or obstruct performance. In contrast, motivation demonstrated a positive partial association with self-reported performance across OLS, HC3, the post hoc specification, and every leave-one-out model. Teachers who reported greater enthusiasm, responsibility, achievement orientation, discipline, and professional development also tended to report higher levels of professional accomplishment, supporting H2 as an association that aligns with the willingness to perform, though it does not validate the entire framework.

This finding is consistent with evidence linking motivation to teaching behavior and performance-related outcomes (Bardach *et al.*, 2022; Lazarides *et al.*, 2025; Layek & Koodamara, 2024), including in Indonesian vocational and professional-development settings (Habibi *et al.*, 2025; Kholifah *et al.*, 2024; Ulfah *et al.*, 2024). However, direct comparisons of coefficients remain inappropriate due to differences in instruments, settings, and model specifications. Both motivation and performance were measured by the same respondents at the same time, which may partially explain the relatively strong correlation ($r = 0.663$) due to shared method, social desirability, response style, and conceptual overlap. While HC3 and influence analyses enhance the assessment of statistical sensitivity, they do not correct for these measurement conditions or establish temporal direction. The joint model accounted for 44.6% of the observed variation in performance scores, yet only motivation had a statistically supported partial association. This contribution is deliberately bounded, illustrating distinct patterns for one situational and one motivational condition within a single madrasah while acknowledging the limitations of a small, single-source, cross-sectional dataset. Using all administered items as primary prevents data-dependent scale refinement from influencing the main inference. The post hoc 29-item analysis serves only as an exploratory robustness description, with its higher alpha and R^2 potentially being sample-bound and requiring independent replication. The regression did not evaluate recognition, supervision, feedback, professional-development programs, or task-redesign

interventions. Based on the broader literature, madrasah leaders may consider clear responsibilities, supportive management, credible recognition, constructive feedback, and feasible development opportunities as proposals for sustaining motivation (Collie, 2023; Ryan & Deci, 2020). Their effects should be assessed with independent outcomes before being claimed. Additionally, the nonsignificant workload coefficient should not justify adding duties. Literature on demands, resources, exhaustion, and well-being supports a review of tasks by type, timing, coordination, and competition for finite time (Bakker & Demerouti, 2017; Hascher & Waber, 2021; Moens *et al.*, 2026; Sandmeier *et al.*, 2022). These considerations are informed by literature rather than representing intervention effects estimated in this study. The limitations of this research include its restriction to 34 eligible teachers in one madrasah, with model-based p-values and confidence intervals being conditional on those observations and the specified model, which do not allow for national generalizability.

The cross-sectional design fails to establish temporal order or causality, and a single organizational context cannot separate individual from site-level conditions. All analytical variables were derived from a single self-administered questionnaire, leaving common-method, social-desirability, and response-style biases untested (Podsakoff *et al.*, 2003). While the records document confidentiality language and aggregate reporting, they lack separate signed consent, ethics committee review or exemption, and formal management access safeguards. Furthermore, they do not document expert item review, a separate pilot, or administration translation; the source papers do not provide wording for item-level equivalence. The study-specific scales show only internal-consistency evidence. The post hoc exclusions utilized the analytical sample, and factor analysis was not feasible; workload categories were aggregated, capacity and relevant covariates were unmeasured, and several cases were influential by common references despite stable leave-one-out decisions. Future research should preregister scoring, validate measures in an independent and larger multi-madrasah sample, gather multisource performance evidence, disaggregate workload, measure capacity and resources, and employ longitudinal or repeated-measure designs.

4. Conclusion

In the primary 32-item analysis of all 34 eligible teachers at one madrasah, workload was not significantly associated with self-reported teacher performance after motivation was included ($B =$

-0.068, $p = 0.557$); H1 was not supported. Work motivation was positively associated with reported performance ($B = 0.693$, $p < 0.001$); H2 was supported. HC3 and leave-one-out analyses preserved both decisions, and the post hoc 29-item model yielded the same pattern. The evidence is restricted to one site, 34 observations, a cross-sectional design, and measures reported by the same respondents. It does not establish causal effects, objective performance change, a complete test of opportunity–capacity–willingness theory, or generalisation to other schools. In particular, the workload result is absence of statistical support in this model—not evidence that workload is inconsequential. One restrained practical implication is to examine motivation-supportive conditions and workload configuration as locally testable management questions. Recognition, supervision, feedback, professional development, and task redesign are literature-informed proposals rather than variables tested here; any implementation should be evaluated with independently assessed outcomes.

5. References

- Azzahra, M., Ayuningtias, H. G., Anggadwita, G., & Nurbaiti, A. (2019). The effect of work discipline on employees' performance of PT Wiratanu Persada Tama Jakarta. *Review of Integrative Business and Economics Research*, 8(S1), 136–144.
- Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, 22(3), 273–285. <https://doi.org/10.1037/ocp0000056>.
- Bardach, L., Klassen, R. M., & Perry, N. E. (2022). Teachers' psychological characteristics: Do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educational Psychology Review*, 34(1), 259–300. <https://doi.org/10.1007/s10648-021-09614-9>.
- Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of Management Review*, 7(4), 560–569. <https://doi.org/10.5465/amr.1982.4285240>.
- Collie, R. J. (2023). Teachers' work motivation: Examining perceived leadership practices and salient outcomes. *Teaching and Teacher Education*, 135, Article 104348. <https://doi.org/10.1016/j.tate.2023.104348>.
- Creagh, S., Thompson, G., Mockler, N., Stacey, M., & Hogan, A. (2025). Workload, work intensification and time poverty for teachers and school leaders: A systematic research synthesis. *Educational Review*, 77(2), 661–680. <https://doi.org/10.1080/00131911.2023.2196607>.
- Eryilmaz, N., Kennedy, A. I., Strietholt, R., & Johansson, S. (2025). Teacher job satisfaction: International evidence on the role of school working conditions and teacher characteristics. *Studies in Educational Evaluation*, 86, Article 101474. <https://doi.org/10.1016/j.stueduc.2025.101474>.
- Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspel, A. K., ... & Westbye, C. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of work and organizational psychology*, 24(2), 178–196. <https://doi.org/10.1080/1359432X.2013.877892>.
- Gan, Y. (2025). Impact of workload on the performance of junior high school teachers under China's "Double Reduction" policy. *SAGE Open*, 15(3), Article 21582440251345577. <https://doi.org/10.1177/21582440251345577>.
- Habibi, A., Attar, R. W., Zhao, Q., Wijaya, T. T., Hendra, R., Alharmali, T. M., & Alhazmi, A. H. (2025). Validating the determinants of teachers' performance in vocational education: Indonesian perspectives. *Acta Psychologica*, 259, Article 105332. <https://doi.org/10.1016/j.actpsy.2025.105332>.
- Hascher, T., & Waber, J. (2021). Teacher well-being: A systematic review of the research literature from the year 2000–2019. *Educational Research Review*, 34, Article 100411. <https://doi.org/10.1016/j.edurev.2021.100411>.
- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), 709–722. <https://doi.org/10.3758/BF03192961>.

- Indrastuti, Y. A., Subijanto, S., & Kartika, C. (2025). Pengaruh beban kerja dan tuntutan pengembangan kompetensi terhadap kinerja dan motivasi guru sebagai variabel intervening di SMA Negeri 5 Tuban. *Jurnal Multidisiplin Ilmu Akademik*, 2(5), 148–156.
- Jerrim, J., & Sims, S. (2021). When is high workload bad for teacher wellbeing? Accounting for the non-linear contribution of specific teaching tasks. *Teaching and Teacher Education*, 105, Article 103395. <https://doi.org/10.1016/j.tate.2021.103395>.
- Kholifah, N., Nurtanto, M., Mutohari, F., Subakti, H., Ramadhan, M. A., & Majid, N. W. A. (2024). The mediating role of motivation and professional development in determining teacher performance in vocational schools. *Cogent Education*, 11(1), Article 2421094. <https://doi.org/10.1080/2331186X.2024.2421094>.
- Koopmans, L., Coffeng, J. K., Bernaards, C. M., Boot, C. R. L., Hildebrandt, V. H., de Vet, H. C. W., & van der Beek, A. J. (2014). Responsiveness of the Individual Work Performance Questionnaire. *BMC Public Health*, 14, Article 513. <https://doi.org/10.1186/1471-2458-14-513>.
- Layek, D., & Koodamara, N. K. (2024). Motivation, work experience, and teacher performance: A comparative study. *Acta Psychologica*, 245, Article 104217. <https://doi.org/10.1016/j.actpsy.2024.104217>.
- Lazarides, R., Schiefele, U., Daumiller, M., & Dresel, M. (2025). From teacher motivation to teaching behaviour: A systematic review and theoretical framework of the mediating processes. *Educational Research Review*, 48, Article 100703. <https://doi.org/10.1016/j.edurev.2025.100703>.
- LePine, J. A., Podsakoff, N. P., & LePine, M. A. (2005). A meta-analytic test of the challenge stressor–hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal*, 48(5), 764–775. <https://doi.org/10.5465/AMJ.2005.18803921>.
- Moens, M., Vanblaere, B., Devos, G., & Tuytens, M. (2026). Teacher wellbeing in schools: A systematic review of job demands and job resources. *Frontiers in Education*, 11, Article 1709029. <https://doi.org/10.3389/feduc.2026.1709029>.
- Morris, C., Agbo, B., Osman, M., Basketts, J., & Kyriacou, T. (2026). The components and implications of teacher workload: A review. *Educational Research*, 68(2), 173–194. <https://doi.org/10.1080/00131881.2026.2629284>.
- Ndruru, A., & Yudiaatmaja, F. (2025). Pengaruh beban kerja dan motivasi kerja terhadap kinerja guru di SMP Negeri 2 Singaraja. *Jurnal Ekonomi, Manajemen, Akuntansi Dan Keuangan*, 6(3), Article 9. <https://doi.org/10.53697/emak.v6i3.2740>.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>.
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, Article 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>.
- Sandmeier, A., Baeriswyl, S., Krause, A., & Muehlhausen, J. (2022). Work until you drop: Effects of work overload, prolonging working hours, and autonomy need satisfaction on exhaustion in teachers. *Teaching and Teacher Education*, 118, Article 103843. <https://doi.org/10.1016/j.tate.2022.103843>.
- Spector, P. E., & Jex, S. M. (1998). Development of four self-report measures of job stressors and strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. *Journal of Occupational Health Psychology*, 3(4), 356–367. <https://doi.org/10.1037/1076-8998.3.4.356>.
- Toropova, A., Myrberg, E., & Johansson, S. (2021). Teacher job satisfaction: The importance of school working conditions and teacher characteristics. *Educational Review*, 73(1), 71–97. <https://doi.org/10.1080/00131911.2019.1705247>.

- Ulfah, M., Syahrudin, H., Thoharudin, M., Wardani, S. F., & Kholifah, N. (2024). Examining the role of family sociology, managerial support, and working conditions in determining teacher performance in Indonesia: The mediating role of motivation and professional development. *Frontiers in Education*, 9, Article 1413525. <https://doi.org/10.3389/feduc.2024.1413525>.
- Wahyuni, S., Hartinah, S., & Prihatin, Y. (2024). Pengaruh kepemimpinan, motivasi, beban kerja, kompetensi dan religiusitas terhadap kinerja guru sekolah dasar. *Journal of Education Research*, 5(3), 4060–4072. <https://doi.org/10.37985/jer.v5i3.1544>.
- Zhou, S., Slemp, G. R., & Vella-Brodrick, D. A. (2024). Factors associated with teacher wellbeing: A meta-analysis. *Educational Psychology Review*, 36, Article 63. <https://doi.org/10.1007/s10648-024-09886-x>.