Preprints.org

Article

Not peer-reviewed version

# Quiet and Passive Quitting Predict Work Engagement in a Machine Learning-Enhanced SEM Analysis

Marcin Nowak *, Marta Pawłowska-Nowak, Piotr Osmański

Posted Date: 13 March 2026

doi: 10.20944/preprints202603.1086.v1

Keywords: quiet quitting; passive quitting; work engagement; SEM; machine learning

*Article*

# Quiet and Passive Quitting Predict Work Engagement in a Machine Learning-Enhanced SEM Analysis

**Marcin Nowak *, Marta Pawłowska-Nowak and Piotr Osmański**

Poznan University of Technology, Jacka Rychlewskiego 1, 61-131 Poznań, Poland
* Correspondence: marcin.nowak@put.poznan.pl

**Abstract**

In structural equation modeling (SEM), model fit is commonly evaluated on a single sample, which limits evidence about out-of-sample generalisability in HR research. This study adapts repeated cross-validation to SEM and introduces a normalised generalisation gap (NG) index to quantify fit deterioration on unseen data relative to training fit. Using employee survey data (N = 1,040), we estimated a model in which quiet quitting and passive quitting predict work engagement and evaluated RMSEA across repeated 10×10 folds. In-sample fit indices were stable across training subsamples, yet test-set fit was consistently weaker. The NG index operationalises this decline as a standardised ratio, enabling transparent reporting of overfitting risk and model transportability. The procedure supports more robust evaluation of measurement models and structural relations in people analytics applications.

**Keywords:** quiet quitting; passive quitting; work engagement; SEM; machine learning

---

## Introduction

The development of people analytics and the broader stream of algorithmic management has led organisational research to increasingly position itself at the intersection of management science, psychometrics, and data science: organisations possess large volumes of employee data, and statistical models (including latent models) are used not only to test hypotheses but also to support HR decision-making [1]. In this context, the question of the credibility of out-of-sample generalisation becomes crucial, as diagnostic applications in HR require robustness to sample variability rather than merely good fit to data "here and now" [2].

Assessing the risk of declining employee engagement is particularly important, as engagement constitutes a central construct in the organisational behavior literature and is measured, among others, by the short version of the UWES-9 scale [3]. At the same time, meta-analytic evidence indicates that engagement at the business-unit level is associated with a range of organisational outcomes (e.g., productivity, turnover), which justifies interest in this construct from the perspective of HR analytics [4]. Theoretical frameworks such as the Job Demands–Resources model further systematise the mechanisms through which job resources and demands influence engagement and exhaustion, creating a natural bridge between psychological research and predictive modeling in organisational settings [5]

Against this background, phenomena related to effort reduction and employee withdrawal are gaining importance. Quiet quitting is increasingly being operationalised and empirically examined- for example, measurement instruments are being developed (including scales with distinct dimensions of withdrawal and lack of initiative), demonstrating that the construct is no longer merely a media label but is entering scientific discourse [6]. Research also suggests that quiet quitting may be associated with key HR variables such as job satisfaction, affective commitment, and turnover intentions, particularly under conditions of digital transformation and varying levels of psychological safety [7]. Moreover, quiet quitting has been embedded in mechanisms explaining the consequences of exhaustion: for instance, in occupational samples (nurses), a positive association between burnout and quiet quitting has been demonstrated, as well as the mediating role of job satisfaction, with direct implications for HR wellbeing and retention policies [8].

At the same time, both the literature and practice highlight the need to distinguish between intentional effort reduction and more "apathetic" withdrawal resulting from exhaustion and loss of meaning at work. In this context, the conceptual and measurement proposal of passive quitting has emerged and is being developed in research [9]. Passive quitting is defined as unintentional withdrawal and a decline in engagement that may be more strongly associated with the psychological costs of work than quiet quitting understood as boundary-setting [10]. Within this same stream, measurement instruments for QQ and PQ have been developed, along with approaches integrating SEM and ML-inspired procedures to assess the psychometric and predictive utility of scales (including in relation to engagement) [11,12].

From the perspective of HR analytics, it is therefore crucial not only to demonstrate associations between QQ/PQ and engagement, but also to assess the extent to which a latent model retains its properties out-of-sample when intended to serve as a component of risk diagnostics across different employee samples (e.g., in subsequent survey waves or across organisational segments). Structural equation modeling remains the standard in organisational research because it integrates measurement models (CFA) with the structural component and enables estimation of relationships between latent constructs. However, reporting practices in SEM largely rely on single-sample fit indices and their interpretive thresholds (e.g., CFI/TLI/RMSEA), which by definition refer to in-sample fit [13].

Additionally, the classical process of model "refinement" (e.g., through post-hoc modifications) may lead to capitalisation on chance-a form of overfitting in which improved fit results from tailoring the model to the idiosyncrasies of a specific sample rather than from a more accurate representation of the data-generating mechanism [14]. In data science, the standard response to this problem is to separate model evaluation into "training" and "test" components and to apply cross-validation in order to estimate predictive error and model stability. The distinction between explanatory and predictive modeling, emphasised in methodological and psychological literature, indicates that good fit/explanation does not necessarily translate into good generalisation [2].

Importantly, the idea of cross-validation is not foreign to SEM: classical works already justified the use of cross-validation for covariance structures, and methodological reviews have highlighted its usefulness for moment-structure models (including SEM) [15]. In recent years, direct proposals for out-of-sample prediction tools within SEM have also emerged, reinforcing the argument that SEM can be evaluated through the lens of generalisation rather than exclusively through single-sample fit [16]. Despite the existence of methodological foundations, organisational and HR research practice continues to be dominated by the evaluation of SEM models based on single-sample fit indices, without formal, quantitative assessment of fit deterioration outside the estimation sample [13]. In the domain of people analytics-where a model may constitute part of a diagnostic tool-this gap is particularly problematic, as it hinders the distinction between "good in-sample fit" and the model's actual ability to generalise to new employee data [1].

The aim of the present article is therefore to (1) adapt to SEM a repeated cross-validation procedure known from machine learning and (2) propose the normalised generalisation gap (NG) index as a simple, standardised measure of out-of-sample fit deterioration relative to in-sample fit. The approach is illustrated using a model in which quiet quitting and passive quitting explain work engagement, thereby combining a methodological contribution with an application in HR diagnostics. The remainder of the article presents: the Methodology (SEM specification, cross-validation procedure, and construction of the NG index), followed by the Results (single-sample fit, in-sample stability, out-of-sample fit, and NG values), and finally the Discussion and implications for HR analytics practice (how to report the generalisation of latent models and how to interpret fit deterioration in diagnostic applications).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2026 doi:10.20944/preprints202603.1086.v1



# Methodology

Step 1. Specification of the SEM Model

The study employed previously validated research instruments designed to measure the phenomena referred to as quiet quitting and passive quitting [9,10]. The constructed instrument takes the following form:

- QQS1. At work, I consciously limit myself to performing only those tasks that fall within my formal duties and for which I am compensated.
- QQS2. I deliberately do only what is necessary because maintaining work–life balance is more important to me than going beyond my responsibilities.
- QQS3. When I know that I am overexerting myself at work, I consciously reduce my effort to what is strictly necessary in order to retain my job.
- QQS4. Because additional effort is not recognised in my company, I have decided to restrict myself to my basic duties.
- QQS5. I consciously refrain from undertaking tasks beyond my responsibilities, even if they are interesting.
- QQS6. Because I do not feel supported by the organisation, I have decided not to engage beyond what my duties require.
- QQS7. To maintain psychological balance, I consciously distance myself from work and do not engage more than necessary.
- PQS1. I am often too tired or overwhelmed to put more than the minimum effort into my work.
- PQS2. I care less and less about my job and its outcomes-it is difficult for me to become engaged.
- PQS3. I feel so exhausted by work that it is difficult for me to maintain the quality of what I do.
- PQS4. At times, I perform my duties mechanically, without engagement or initiative.
- PQS5. I feel that a lack of appreciation discourages me from investing effort in my work.
- PQS6. My work no longer gives me satisfaction-I do only what is necessary to get through the day.
- PQS7. I feel that the lack of development and meaning at work has led me to care less and less about what I do.

The scales were evaluated in terms of face and content validity-the CVI reached 0.929 for content validity and 0.914 for clarity, while the CVR achieved a value of 1.00. In the subsequent stage, an exploratory factor analysis (EFA) was conducted. The observed floor effect ranged from 3.46% to 13.75%, and the ceiling effect ranged from 6.25% to 13.65%. Slight skewness and negative kurtosis indicated distributions approximating normality. The suitability of factor analysis was supported by the results: KMO = 0.918 and a statistically significant Bartlett's test ($\chi^2 = 9026.24$, $p < 0.001$). Both the Kaiser criterion and the scree plot suggested the presence of two factors corresponding to the dimensions of quiet and passive quitting. Principal Axis Factoring with oblimin oblique rotation was applied, confirming a two-factor structure consistent with the adopted theoretical assumptions. The extracted factors jointly explained more than 66% of the variance. Scale reliability assessed using Cronbach's alpha coefficients amounted to 0.910 for quiet quitting and 0.916 for passive quitting, indicating high internal consistency and robust psychometric properties of both instruments. An extended description of the validation procedures is presented in [9]. Organisational engagement was measured using the UWES-9 scale [3].

The developed model comprises a measurement component (CFA) for QQ and PQ, as well as a structural component in which WE is explained by the latent variables QQ and PQ. The measurement component was specified as:

$$x_{QQ} = \Lambda_{QQ} \cdot \eta_{QQ} + \varepsilon_{QQ} \tag{1}$$

$$x_{PQ} = \Lambda_{PQ} \cdot \eta_{PQ} + \varepsilon_{PQ} \tag{2}$$

 doi:10.20944/preprints202603.1086.v1



where $\eta_{QQ}$ and $\eta_{PQ}$ are latent variables, $\Lambda$ denotes the loading matrices, and $\varepsilon$ represents measurement errors. Additionally, a covariance between the latent predictors was assumed:

$$\mathrm{cov}(\eta_{QQ}, \eta_{PQ}) \neq 0 \tag{3}$$

The structural component takes the following form:

$$WE = \beta_{QQ} \cdot \eta_{QQ} + \beta_{PQ} \cdot \eta_{PQ} + \zeta \tag{4}$$

where $\beta_{QQ}$ and $\beta_{PQ}$ are regression coefficients, and $\zeta$ is the residual term.

The parameters $\theta$ were estimated using the ML method by minimising the discrepancy function:

$$F_{ML}\big(S, \Sigma(\theta)\big) = \log|\Sigma(\theta)| + \mathrm{tr}(S\Sigma(\theta)^{-1}) - \log|S| - p \tag{5}$$

where $S$ is the covariance matrix derived from the data, $p$ denotes the number of observed variables, and $\Sigma(\theta)$is the model-implied covariance matrix.

Step 2. Baseline SEM Estimation on the Full Sample (Single-Run Result)

In the first step, the SEM model was fitted to the full dataset in order to obtain:

- parameter estimates ($\hat{\theta}$),
- standard fit indices (including $\chi^2$, CFI, TLI, RMSEA).

RMSEA was calculated in accordance with the standard definition:

$$RMSEA = \sqrt{\max\left(\frac{\chi^2 - df}{df(N-1)}, 0\right)} \tag{6}$$

where $df$ denotes the model degrees of freedom.

Step 3. Cross-Validation of Model Fit (In-Sample Stability)

In the subsequent step, repeated $K$ -fold cross-validation was applied to assess the stability of model fit. For each repetition $r = 1, \ldots, R$ and each fold $k = 1, \ldots, K$, a training set $D_{train}^{(r,k)}$ of size $N_{train} \approx \frac{K-1}{K} N$ was created. Subsequently:

$$\hat{\theta}^{(r,k)} = \underset{\theta}{\mathrm{argmin}} F_{ML}(S_{train}^{(r,k)}, \Sigma(\theta)) \tag{7}$$

After estimation on the training set, fit indices were computed for the training data, in particular:

$$RMSEA_{train}^{(r,k)} = \sqrt{\max\left(\frac{\chi^2 - df}{df(N_{train} - 1)}, 0\right)} \tag{8}$$

The final outcome of Step 2 consisted of the distributions and descriptive statistics of the fit indices (mean, standard deviation, min–max) obtained from $R \times K$ model fits.

Step 4. Validation of Generalisability (Out-of-Sample) and the NG Index

To assess model generalisability, in each fold the test set $D_{test}^{(r,k)}$ and the corresponding covariance matrix $S_{test}^{(r,k)}$ were used. The model fitted on the training data yields the implied covariance matrix:

$$\Sigma\big(\hat{\theta}^{(r,k)}\big) \tag{9}$$

Subsequently, the test fit statistic was computed within the covariance-based framework:

$$\chi_{test}^{2(r,k)} = (N_{test} - 1) F_{ML}\left(S_{test}^{(r,k)}, \Sigma\big(\hat{\theta}^{(r,k)}\big)\right) \tag{10}$$

as well as the corresponding RMSEA for the test set:

 doi:10.20944/preprints202603.1086.v1



$$RMSEA_{test}^{(r,k)} = \sqrt{max(\frac{\chi_{test}^2 - df}{df(N_{test} - 1)}, 0)} \tag{11}$$

The out-of-sample deterioration in fit was parameterised by the normalised gap index (NG):

$$NG_{RMSEA}^{(r,k)} = \frac{RMSEA_{test}^{(r,k)} - RMSEA_{train}^{(r,k)}}{RMSEA_{train}^{(r,k)}} \tag{12}$$

Values of $NG_{RMSEA} < 1$ were interpreted as indicating acceptable generalisability in relative terms, that is, the out-of-sample deterioration in fit does not exceed twice the fit obtained on the training data.

Step 5. Aggregation of Results

The final measures of stability and generalisability were obtained by aggregating the results from $R \times K$ model fits:

$$\overline{RMSEA}_{train} = \frac{1}{RK} \sum_{r=1}^{R} \sum_{k=1}^{K} RMSEA_{train}^{(r,k)} \tag{13}$$

$$\overline{RMSEA}_{test} = \frac{1}{RK} \sum_{r=1}^{R} \sum_{k=1}^{K} RMSEA_{test}^{(r,k)} \tag{14}$$

$$\overline{NG}_{RMSEA} = \frac{1}{RK} \sum_{r=1}^{R} \sum_{k=1}^{K} NG_{RMSEA}^{(r,k)} \tag{15}$$

Additionally, the distributions of $RMSEA_{train}$, $RMSEA_{test}$, and $NG_{RMSEA}$ (histograms) were analyzed as an illustration of fit stability and generalisation variability. Complementing the aggregation of results is the interpretative classification of the normalised generalisation gap NG_RMSEA, which enables a qualitative assessment of the magnitude of fit deterioration between the training and test samples. For this purpose, the interpretative thresholds presented in Table 1 were adopted.

**Table 1.** Interpretation of NG_RMSEA Values.

| NG_RMSEA | Interpretation | Generalisation |
|---|---|---|
| < 0.25 | very good | stable model |
| 0.25 – 0.50 | good | minor overfitting |
| 0.50 – 0.75 | moderate | noticeable deterioration |
| 0.75 – 1.00 | acceptable (borderline) | substantial overfitting |
| 1.00 – 1.50 | poor | large decline in generalisability |
| > 1.50 | very poor | unstable model |

The table defines the interpretation of the relative decline in model fit outside the estimation sample. Lower NG_RMSEA values indicate high concordance between training and test fit, and thus greater model stability and lower risk of overfitting. As the index increases, the magnitude of fit deterioration also increases, which is interpreted as limited transferability of the model to new data. The adopted classification allows NG_RMSEA to be treated as a synthetic measure of SEM model generalisability: it enables model comparison, assessment of the impact of specification modifications, and reporting of model quality from a data-science perspective, that is, not only in terms of in-sample fit but also predictive stability.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2026 doi:10.20944/preprints202603.1086.v1



## Results

Step 1. Specification of the SEM Model

The empirical part of the study was conducted based on an online CAWI survey carried out between 23–29 May 2025 on a sample of 1,040 employees from Poland (Ariadna panel). The research questionnaire included demographic information, proprietary instruments for measuring Quiet Quitting (QQ) and Passive Quitting (PQ), as well as the UWES-9 organisational engagement scale. The collected data were subsequently analyzed using SEM modeling in accordance with the measurement and structural components presented in the methodological section.

Step 2. Baseline SEM Estimation on the Full Sample (Single Estimation Result)

The SEM model was fitted to the full dataset (N = 1,040). Estimation of the measurement component demonstrated high and statistically significant factor loadings for all QQ and PQ indicators ($p < 0.001$), confirming the correctness of the measurement model specifications. In the structural component, a significant negative effect of Quiet Quitting ($\beta = -0.60$, $p < 0.001$) and Passive Quitting ($\beta = -0.92$, $p < 0.001$) on work engagement (WE) was identified. Additionally, a small yet significant covariance between the latent factors QQ and PQ was observed. The fit indices indicate a good model fit to the data: $\chi^2(88) = 608.67$, CFI = 0.946, TLI = 0.936, RMSEA = 0.075, GFI = 0.938, AGFI = 0.926. The obtained values meet the adopted criteria for acceptable model fit.

Step 3. Cross-Validation of Model Fit (In-Sample Stability)

The stability of the model fit was assessed using repeated 10×10 cross-validation. In each fold, the model was estimated on a training set of N_train = 936, and the fit indices were subsequently calculated. The distribution of RMSEA_train indicates high stability of model fit: mean RMSEA_train = 0.076 (SD = 0.002), ranging from 0.071 to 0.079. In all 100 model estimations, the RMSEA < 0.08 criterion was met, confirming that the acceptable fit is not the result of a single sample. Similarly, the remaining fit indices showed minimal variability across folds (CFI: M = 0.946, SD = 0.002; TLI: M = 0.936, SD = 0.003; GFI: M = 0.937, SD = 0.002). The low variance of the fit indices indicates parameter stability and robustness of model fit with respect to sample partitioning. The in-sample validation results suggest that the SEM model maintains good fit across different data subsamples, providing grounds for further assessment of its generalisability beyond the estimation sample. The histogram in Figure 1 presents the distribution of RMSEA values obtained on the training sets (RMSEA_train) within the repeated 10×10 cross-validation procedure, that is, across 100 independent SEM model estimations.

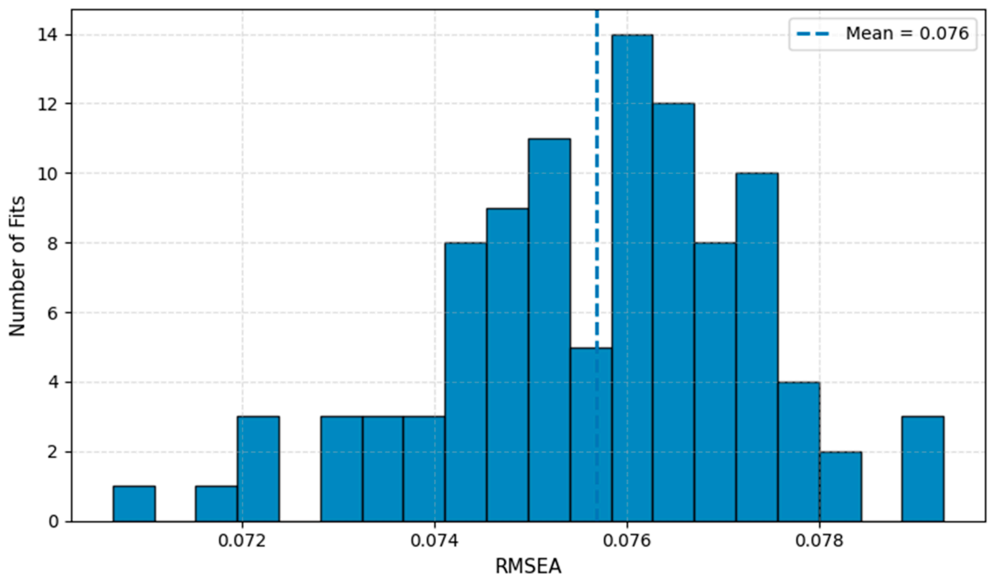


**Figure 1.** Histogram of RMSEA from SEM Cross-Validation (In-Sample Stability).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2026 doi:10.20944/preprints202603.1086.v1



The distribution of RMSEA_train is clearly concentrated around 0.076, with minimal dispersion (SD = 0.002) and a range of 0.071–0.079, indicating high stability of model fit across successive folds. The absence of values exceeding 0.08 signifies that in each of the 100 model estimations the model maintained an acceptable level of fit, and thus the result is not an artifact of a single estimation. The shape of the histogram (narrow, without "long tails") further suggests that the model fit is robust to variations in the composition of the training sample (N_train = 936), thereby reinforcing the conclusion regarding parameter stability and providing coherent grounds for proceeding to the assessment of out-of-sample generalisability.

Step 4. Validation of Generalisability (Out-of-Sample) and the NG Index

The generalisability of the model was assessed by analyzing out-of-sample fit within a repeated 10×10 cross-validation framework. In each fold, the model fitted on the training set was used to compute fit indices on the test set, allowing for the estimation of RMSEA_test and the normalised generalisation gap (NG). The mean RMSEA_test value was 0.143 (SD = 0.014), ranging from 0.110 to 0.180, indicating a systematic deterioration of fit outside the estimation sample relative to the training fit (RMSEA_train ≈ 0.076). At the same time, the distribution of RMSEA_test remains relatively stable across folds, suggesting a replicable generalisation effect. The normalised generalisation gap reached a mean value of NG = 0.89 (SD = 0.20), with a range from 0.43 to 1.52. This indicates that out-of-sample model fit is, on average, approximately 89% weaker than the training fit; however, in relative terms, it satisfies the adopted criterion of acceptable generalisability (NG < 1). The results suggest that the SEM model retains a stable structure, yet-consistent with expectations-its out-of-sample fit is markedly poorer than its in-sample fit. The histograms of RMSEA_test (Figure 2) and NG (Figure 3) illustrate the distribution of the decline in fit and demonstrate that the generalisation effect is systematic rather than incidental.

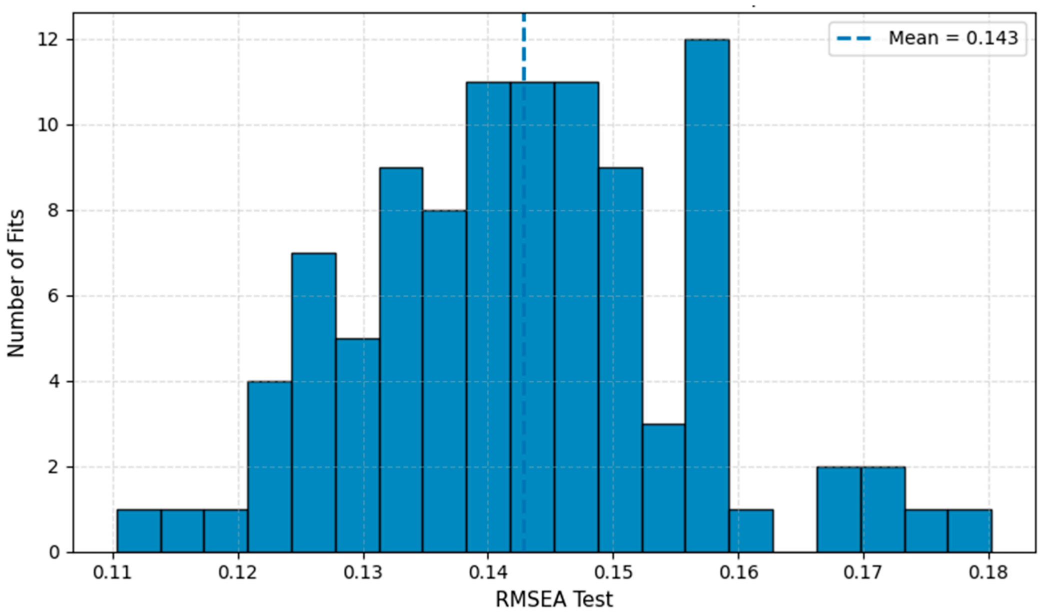


**Figure 2.** Histogram of RMSEA_test from SEM Cross-Validation (Out-of-Sample Stability).

The histogram of RMSEA_test demonstrates that model fit on unseen data is consistently weaker than during training: the distribution is centered around 0.143 with SD = 0.014 and a range of 0.110–0.180, confirming a systematic (rather than incidental) deterioration in fit relative to RMSEA_train ≈ 0.076. At the same time, the moderate width of the distribution suggests that the effect is replicable across folds, that is, it does not strongly depend on a specific sample partition but constitutes a stable property of the model within this dataset.

 doi:10.20944/preprints202603.1086.v1



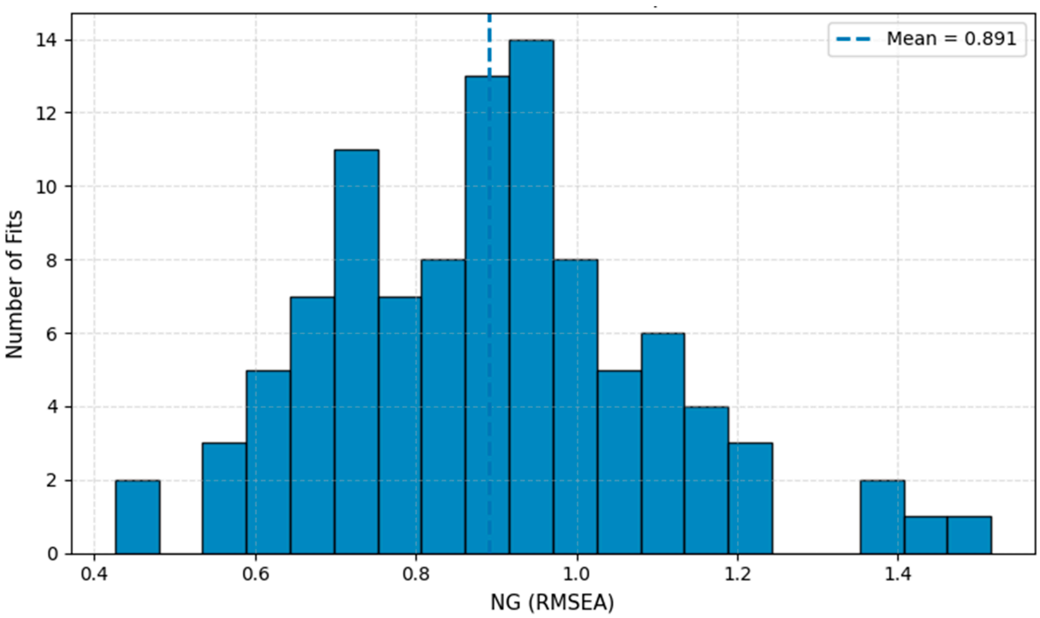


**Figure 3.** Histogram of NG from SEM Cross-Validation (Out-of-Sample Stability).

The NG_RMSEA histogram specifies the magnitude of this deterioration in relative terms: the values are concentrated around 0.89 (SD = 0.20), with a range of 0.43–1.52, indicating that test fit is, on average, approximately 89% weaker than training fit. The predominance of observations within the interval of approximately 0.75–1.00 points to borderline acceptable generalisability (a pronounced yet predictable decline), whereas the tail of the distribution above 1.00 signals that in some partitions weaker generalisation (stronger overfitting) occurs-although this is not the prevailing case. Overall, the histograms support the conclusion that the model is structurally stable (the decline is consistent) but reveals a substantial generalisation gap, the magnitude of which should be reported alongside classical in-sample fit indices.

Step 5. Aggregation of Results

In the final step, the results obtained across all cross-validation iterations (R × K = 100 model estimations) were aggregated. The aggregation included mean values, standard deviations, and ranges of variability for the key fit and generalisability indices. The mean model fit in the training samples was RMSEA_train = 0.076 (SD = 0.002), confirming stable and acceptable fit across different data subsamples. In the test samples, the mean fit was markedly weaker (RMSEA_test = 0.143; SD = 0.014), indicating a systematic decline in fit outside the estimation sample. Aggregation of the generalisation index yielded a mean value of NG_RMSEA = 0.89 (SD = 0.20), reflecting a moderate reduction in fit while maintaining relatively acceptable model generalisability.

The distributions of RMSEA_train, RMSEA_test, and NG_RMSEA indicate low variability across sample partitions and confirm the stable character of the obtained results. Taken together, the aggregated findings demonstrate that the SEM model is characterised by stable in-sample fit and a predictable, systematic decline in out-of-sample fit, thereby enabling a more rigorous assessment of its generalisation properties than is possible in a conventional single-sample analysis.

## Conclusions

This article proposes a shift in the logic of SEM model evaluation from the classical "in-sample fit" perspective toward a generalisation-oriented framework consistent with the data science paradigm, in which the primary criterion of model quality is its performance on unseen data. The theoretical contribution lies in the formal transfer of repeated cross-validation procedures to SEM, as well as in the introduction of the normalised generalisation gap (NG) as a standardised measure of the relative deterioration of model fit outside the estimation sample. NG enables the operationalisation of “hidden” overfitting in SEM: a situation in which standard fit indices remain stable across subsamples, while test fit systematically deteriorates. An empirical illustration based on employee data (N = 1,040) demonstrates that RMSEA obtained on training sets remains at an acceptable level with low variance, whereas test RMSEA is substantially higher, and NG reveals variability in the generalisation gap across folds. Methodologically, this implies that merely meeting conventional fit thresholds is insufficient for inferring the transportability of latent models, particularly when they are intended to serve diagnostic or predictive functions.

The practical implications are particularly relevant for HR analytics and people analytics. In organisational applications, SEM models are used to validate measurement instruments (e.g., attitude scales), construct risk indicators (e.g., declining engagement), and support decisions regarding wellbeing or retention interventions. The proposed CV-SEM procedure combined with NG provides HR departments and data analysts with a straightforward quality-control mechanism: it enables reporting not only “how well the model fits,” but also “how stably it will perform in a new employee sample.” In the context of quiet quitting and passive quitting, this has applied significance, as these constructs may function as early warning signals of declining engagement; a generalisation-oriented assessment reduces the risk of implementing models that are well fitted to a single survey wave but fail in subsequent organisational segments or measurement waves.

The study’s limitations stem primarily from the nature of the data and the adopted operationalisation of generalisation. First, the analyzed sample originates from a single country and a single data collection mode (CAWI), which limits conclusions regarding cross-population stability and model behavior in other cultural or industry contexts. Second, NG was demonstrated using RMSEA, which enhances interpretive transparency but does not exhaust the issue: other fit indices may respond differently to the train–test split, and RMSEA itself has known properties dependent on model complexity and sample size. Third, cross-validation measures generalisation in terms of “internal replicability” within the same dataset; it does not replace external validation on an independent dataset, which constitutes the strongest test of transportability.

Future research should extend this approach in several directions. First, it is advisable to develop a family of NG indicators for multiple criteria (e.g., CFI/TLI, the ML discrepancy function, predictive measures for endogenous variables) and to examine their properties in simulation studies (sensitivity to misspecification, number of indicators, degrees of freedom, sample size, and assumption violations). Second, CV-SEM could be integrated with regularisation techniques (e.g., regularised SEM), treating NG as a criterion for selecting the penalty–complexity trade-off, analogously to hyperparameter tuning in machine learning. Third, cross-organisational and longitudinal replications are needed: particularly in HR analytics, it is crucial to examine whether NG remains stable across measurement waves and whether QQ/PQ models retain predictive utility for changes in engagement. Finally, a promising direction involves applying the procedure to multigroup models and measurement invariance analyses, where generalisation has not only a statistical but also a comparative dimension (across departments, industries, or countries). Overall, the study indicates that integrating SEM with evaluation tools known from data science constitutes a viable pathway toward enhancing the robustness of latent models in organisational research and developing more reliable diagnostic tools in HR practice.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Polzer, J.T.: The rise of people analytics and the future of organizational research. Res. Organ. Behav. 42, 100181 (2022)
2. Shmueli, G.: To explain or to predict? Statistical Science, 289–310 (2010)
3. Schaufeli, W.B., Shimazu, A., Hakanen, J., Salanova, M., De Witte, H.: An ultra-short measure for work engagement. European Journal of Psychological Assessment (2017)

 doi:10.20944/preprints202603.1086.v1




4. Harter, J.K., Schmidt, F.L., Hayes, T.L.: Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. Journal of Applied Psychology 87(2), 268 (2002)
5. Bakker, A.B., Demerouti, E.: The job demands-resources model: State of the art. Journal of Managerial Psychology 22(3), 309–328 (2007)
6. Galanis, P. et al.: The quiet quitting scale: Development and initial validation. AIMS Public Health 10(4), 828 (2023)
7. Kim, K.T., Sohn, Y.W.: The impact of quiet quitting on turnover intentions in the era of digital transformation: The mediating roles of job satisfaction and affective commitment, and the moderating role of psychological safety. Systems 12(11), 460 (2024)
8. Galanis, P. et al.: The influence of job burnout on quiet quitting among nurses: The mediating effect of job satisfaction. International Journal of Nursing Practice 31(5), e70057 (2025)
9. Nowak, M.: Spadek zaangażowania organizacyjnego - ciche i bierne odchodzenie (Decline in organizational engagement - quiet and passive quitting). Polskie Towarzystwo Ekonomiczne, Poznań (2026)
10. Nowak, M., Wiecek-Janka, E., Zajkowski, R.: The Impact of Quiet Quitting and Passive Quitting Phenomena on Occupational Burnout (2026)
11. Nowak, M., Gigol, T.: Evaluation of the Psychometric Scale of Quiet and Passive Quitting Using an Integrated Approach Combining Structural Equation Modeling and Machine Learning (2026)
12. Nowak, M., Zajkowski, R.: An Integrated Structural Equation Modelling and Machine Learning Framework for Measurement Scale Evaluation - Application to Voluntary Turnover Intentions. AppliedMath 5(3), 105 (2025)
13. Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6(1), 1–55 (1999)
14. MacCallum, R.C., Roznowski, M., Necowitz, L.B.: Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin 111(3), 490 (1992)
15. Cudeck, R., Browne, M.W.: Cross-validation of covariance structures. Multivariate Behavioral Research 18(2), 147–167 (1983)
16. de Rooij, M., Karch, J.D., Fokkema, M., Bakk, Z., Pratiwi, B.C., Kelderman, H.: SEM-based out-of-sample predictions. Structural Equation Modeling 30(1), 132–148 (2023)