Measuring teacher collaboration in Swedish schools:

A validation study using survey data from primary school teachers

Rebecka Persson
Center for Educational Leadership & Excellence
Stockholm School of Economics
PO Box 6501
11383 Stockholm
Rebecka.Persson@hhs.se

Abstract

This working paper evaluates the measurement properties of teacher collaboration items from a survey distributed by LegiLexi to Swedish primary school teachers (N = 2,124). Nine items capturing the frequency and quality of professional interactions with teacher teams, school leaders, and broader school climate were examined using descriptive statistics, principal component analysis, confirmatory factor analysis, internal consistency estimates, and intraclass correlations. A two-factor confirmatory model distinguishing horizontal collaboration (teacher team) from vertical collaboration (principal/school leaders), with correlated uniquenesses for parallel item content, provided excellent fit to the data ($\chi^2(5)$ = 7.37, p = .195, CFI = .999, RMSEA = .016). The two factors were moderately correlated (r = .36), confirming empirically distinct dimensions. Preliminary school-level correlations with student literacy outcomes from LegiLexi reading assessments were explored. The results establish a measurement foundation for future research on teacher collaboration and student reading outcomes in Swedish primary schools.

Introduction

Teacher collaboration is recognized as an important factor in school effectiveness and student achievement. Empirical research has linked collaborative teacher practices to improvements in teaching quality and pupil outcomes (Goddard et al., 2007; Ronfeldt et al., 2015; Vangrieken et al., 2015). Teacher collaboration encompasses diverse practices, including exchanging teaching materials, jointly planning instruction, engaging in cross-curricular projects, and providing feedback on colleagues' teaching (OECD, 2020). The Teaching and Learning International Survey (TALIS) distinguishes between surface-level collaboration, such as exchanging materials and discussing students, and deeper professional collaboration, such as co-teaching and peer observation (OECD, 2020). This distinction matters, as research suggests that deeper forms of collaboration may have a more substantial impact on instructional quality and student learning (Daly et al., 2021; Ronfeldt et al., 2015).

A growing body of research also connects teacher collaboration to school leadership. Principals play a central role in creating the organizational structures and conditions that facilitate meaningful teacher interaction (García-Martínez et al., 2021; Spillane et al., 2015). In a recent longitudinal study of Swedish schools, Persson, Demir, and Wennberg (2025) demonstrated that teacher collaboration partly mediates the relationship between principal instructional leadership and pupil educational achievement, as measured by final-year grade point averages. However, this mediating effect was not present when achievement was measured by standardized test scores, where principal instructional leadership had a direct relationship that superseded the influence of teacher collaboration. The study further showed that the observed relationships were driven by stable between-school differences rather than within-school changes over time, suggesting that school culture and organizational practices are deeply embedded. These findings underscore the importance of measuring teacher collaboration accurately: the construct appears to function differently depending on school context, team structure, and the type of outcome under study.

In the Swedish context, school leadership research has been described as "important but neglected" (Ärlestig et al., 2016), and the same can be said for the systematic measurement of teacher collaboration. While the Swedish School Inspectorate has for over a decade collected teacher survey data that includes items on collaboration and school climate, and while these items have been used in research (e.g. Persson et al., 2025), the measurement properties of teacher collaboration scales have received less attention than those of leadership scales. A recent validation study of principal leadership measures from two Swedish government agencies found that existing scales meet conventional reliability and validity criteria (Persson, 2024), establishing a precedent for the kind of psychometric evaluation needed also for collaboration measures.

An opportunity to study teacher collaboration in relation to student outcomes has emerged through a teacher survey distributed by LegiLexi, a Swedish foundation that provides reading assessments for primary school students. The survey, administered in 2026, collected responses from over 2,000 teachers across Swedish primary schools. Alongside the survey, LegiLexi's student-level reading assessment data make it possible to link teacher-

reported collaboration to student literacy outcomes at the school level. However, before such substantive analyses can be undertaken, the measurement properties of the collaboration items must be established. The items in the survey were partly drawn from the Swedish School Inspectorate's teacher survey, providing continuity with the measures used in previous research, and partly drawn from previous research.

The purpose of this working paper is to evaluate the measurement properties of the current set of teacher collaboration items; partly comparing the scale from the Swedish School Inspectorate to scales used in research and partly preparing for future research. Specifically, the study examines: (a) descriptive statistics and distributional properties of individual items, (b) internal consistency of proposed scales, (c) the factor structure through principal component analysis and confirmatory factor analysis, (d) intraclass correlations to assess school-level clustering, and (e) preliminary school-level correlations between teacher-reported collaboration and student literacy outcomes. The analyses serve to establish a measurement foundation for future research using these data.

Method

Sample

The sample consisted of $N = 2{,}124$ teacher respondents from a survey distributed by LegiLexi to Swedish primary school teachers in 2026. No demographic selection criteria were applied; all teachers using the LegiLexi platform were invited to participate. The responding teachers were distributed across 858 schools, with a median of 1 teacher per school ($M = 2.5$, $SD = 1.3$). Approximately one-third of respondents ($n = 695$) lacked a valid school identifier, reducing the effective sample for school-level analyses to 1,429 teachers in 858 schools.

Measures

Nine items from two survey sections were examined. *Work Context and Collegial Collaboration* contained six items capturing three facets of professional interaction, conversation frequency, perceived value of discussions, and relational closeness, directed at two targets: the teacher team (items 5.6, 5.7, 5.8) and the principal/school leaders (items 5.9, 5.9b, 5.10). The measurement was derived from Pil and Leana (2009) and further developed and validated by Demir (2024).

School Climate and Collaboration contained three items capturing agreement with statements about cross-disciplinary work (7.1), usefulness of school conferences (7.2), and collegial collaboration in planning (7.3). The items originate from the Swedish School Inspectorate's teacher survey (Skolenkäten) and have been used in previous research on teacher collaboration and school leadership (Persson et al., 2025). All items were recoded to numeric scales with higher values indicating more frequent interaction, greater perceived value, closer relationships, or stronger agreement.

Student literacy outcomes were obtained from LegiLexi's reading assessment data, covering six domains: decoding (*avkodning*), language comprehension (*språkförståelse*), reading

comprehension (*läsförståelse* and *läsförståelse 2*), listening comprehension (*hörförståelse*), and letter understanding (*bokförståelse*). Because the two datasets could not be linked at the individual level, both were aggregated to school means using the school unit identifier for the correlational analyses.

Analytical approach

Descriptive statistics, inter-item correlations, and Cronbach's alpha were computed for all items and proposed scales. Principal component analysis (PCA) with varimax rotation was used to explore the factor structure. Confirmatory factor analysis (CFA) was conducted to test the hypothesized measurement model, estimated using maximum likelihood with missing values (MLMV). Intraclass correlations (ICCs) were estimated from empty two-level models. Subgroup analyses were conducted by teacher team type (grade-level teams vs. cross-grade teams). School-level correlations between teacher collaboration composites and student literacy outcomes were examined for the merged sample. All analyses were conducted in Stata.

Results

Descriptive statistics

Table 1 presents summary statistics for all items. Effective sample sizes per item ranged from $N = 1{,}071$ (item 5.9b) to $N = 1{,}836$ (item 5.3), reflecting varying amounts of missing data. The lower sample size for item 5.9b likely reflects respondents who reported no conversations with school leaders and consequently skipped the follow-up item on the value of those conversations.

Table 1. Descriptive statistics for individual items.

Item	N	M	Mdn	SD	Skew	Kurt
5.2 Time with students	1,654	2.30	2	1.58	2.08	6.81
5.3 Time at school	1,836	5.05	6	1.89	−0.75	2.33
5.5 Time in team	1,741	3.92	3	1.95	0.02	1.75
5.6 Discussion freq. (team)	1,754	3.37	4	1.22	−0.36	2.06
5.7 Value of disc. (team)	1,742	3.79	4	0.93	−0.42	2.65
5.8 Relation (team)	1,723	3.95	4	0.83	−0.48	2.83
5.9 Discussion freq. (principal)	1,724	1.97	2	1.05	0.86	2.82
5.9b Value of disc. (principal)	1,071	2.79	3	1.00	0.17	2.49
5.10 Relation (principal)	1,689	3.42	3	0.95	−0.27	2.69
7.1 Cross-disciplinary work	1,580	3.28	3	1.11	−0.43	2.43

7.2 Conferences useful	1,560	3.64	4	0.92	−0.76	3.51
7.3 Collaboration in planning	1,552	3.89	4	1.00	−0.98	3.69

Note. Total N = 2,124. Items 5.2, 5.3, 5.5 coded 1–7; items 5.6–5.10 coded 1–5; items 7.1–7.3 Likert 1–5.

Internal consistency

Cronbach's alpha was computed for three proposed scales (Table 2). All values fell below the conventional threshold of $\alpha \geq .70$, which may reflect the brevity of the scales (three items each) and the heterogeneity of facets within each scale.

Table 2. Internal consistency.

Scale	Items	α
Teacher team interaction	5.6, 5.7, 5.8	.63
Principal interaction	5.9, 5.9b, 5.10	.69
School climate	7.1, 7.2, 7.3	.51

Inter-item and inter-scale correlations

Within-scale inter-item correlations were moderate: .35–.42 for the teacher team scale, .41–.46 for the principal scale, and .23–.28 for the climate scale. Between scales, teacher team and principal items correlated most strongly on matching facets (e.g. frequency–frequency $r = .39$, relationship–relationship $r = .30$), with weaker cross-facet associations (.11–.19). Item 7.3 (collaboration in planning) correlated .33–.34 with all teacher team items, while item 7.2 (conferences useful) correlated more strongly with principal items (.36–.37). Item 7.1 (cross-disciplinary work) had the weakest associations overall. Composite scale correlations were .44 (team–climate), .35 (team–principal), and .28 (principal–climate). The full inter-item correlation matrix is presented in Table 3.

Table 3. Inter-item correlation matrix.

	5.6	5.7	5.8	5.9	5.9b	5.10	7.1	7.2	7.3
5.6	—								
5.7	.37	—							
5.8	.35	.42	—						
5.9	.39	.18	.11	—					
5.9b	.12	.25	.07	.45	—				
5.10	.13	.18	.30	.41	.46	—			
7.1	.15	.21	.18	.06	.06	.12	—		
7.2	.16	.30	.22	.19	.36	.37	.27	—	
7.3	.33	.34	.34	.09	.13	.13	.23	.28	—

Note. All $r \geq .09$ significant at $p < .05$. N ranges from 970 to 1,754 due to pairwise deletion.

Principal component analysis

PCA was conducted on the nine items (listwise $N = 970$). The unrotated solution yielded three components with eigenvalues exceeding 1.0 (2.77, 1.56, 1.04), jointly accounting for 59.7% of the total variance. A varimax-rotated three-component solution produced an interpretable structure largely consistent with the proposed scales (Table 4). The teacher team items (5.6, 5.7, 5.8) loaded on Component 1, the principal items (5.9, 5.9b, 5.10) loaded on Component 2, and two of three climate items (7.1, 7.2) loaded on Component 3. Item 7.3 (collaboration in planning) cross-loaded on both Component 1 (.36) and Component 3 (.31), consistent with its content overlapping with collegial team interaction.

Table 4. Varimax-rotated component loadings.

Item	Comp 1 (Team)	Comp 2 (Principal)	Comp 3 (Climate)	Unexplained
5.6 Discussion freq. (team)	.62			.29
5.7 Value of disc. (team)	.44			.49
5.8 Relation (team)	.50			.45
5.9 Discussion freq. (principal)		.57		.30
5.9b Value of disc. (principal)		.55		.33
5.10 Relation (principal)		.53		.40
7.1 Cross-disciplinary work			.60	.50
7.2 Conferences useful			.58	.36
7.3 Collaboration in planning	.36		.31	.50

Note. Loadings < .30 suppressed. Rotated variance per component: 1.96, 1.94, 1.47.

Confirmatory factor analysis

To complement the exploratory analyses, a series of confirmatory factor models was estimated for the six items measuring teacher–team interaction (items 5.6, 5.7, 5.8) and principal interaction (items 5.9, 5.9b, 5.10). All models were estimated using maximum likelihood with missing values (MLMV), which uses all available data under the assumption that data are missing at random. The effective sample size was $N = 1{,}769$ after exclusion of 355 observations with missing data on all six items.

Model 1: Single-factor model. As a baseline, all six items were specified as indicators of a single latent Collaboration factor. The model fit was poor: $\chi^2(9) = 716.02$, $p < .001$, RMSEA = .211 [90% CI: .198, .224], CFI = .635, TLI = .391. Standardized factor loadings ranged from .45 to .63. The low CFI and TLI values indicate substantial misspecification, suggesting that a single dimension does not adequately represent these items.

Model 2: Two correlated factors. A two-factor model was specified, distinguishing between a Team factor (items 5.6, 5.7, 5.8) and a Principal factor (items 5.9, 5.9b, 5.10), with the two factors freely correlated. This model showed substantially improved fit: $\chi^2(8) = 361.16$, $p < .001$, RMSEA = .158 [.144, .172], CFI = .818, TLI = .658. A likelihood-ratio test confirmed the improvement: $\Delta\chi^2(1) = 354.86$, $p < .001$. Standardized loadings ranged from .58 to .73, and the estimated factor correlation was .50. However, absolute fit remained below conventional thresholds, motivating further inspection.

Modification indices revealed large expected parameter changes for correlated residuals between parallel item pairs, items measuring the same facet (frequency, perceived value, relational closeness) but directed at different targets. The largest modification index was for the frequency pair (MI = 219.1), followed by relational closeness (MI = 102.5) and perceived value (MI = 52.9). These correlated residuals are theoretically expected: the items share content beyond what the latent factors capture, consistent with method effects documented in the multitrait-multimethod literature.

Model 3: Two factors with correlated uniquenesses. The final model retained the two correlated factors and added three correlated residuals between the parallel item pairs: frequency (5.6 ↔ 5.9), perceived value (5.7 ↔ 5.9b), and relational closeness (5.8 ↔ 5.10). This model fit the data well: $\chi^2(5) = 7.37$, $p = .195$, RMSEA = .016 [.000, .040], CFI = .999, TLI = .996. The non-significant chi-square indicates that the model-implied covariance matrix is not distinguishable from the observed data. The likelihood-ratio test against Model 2 was highly significant: $\Delta\chi^2(3) = 353.79$, $p < .001$.

Standardized factor loadings in Model 3 ranged from .56 to .74 (Table 6). On the Team factor, the value of discussions item (5.7) loaded highest ($\lambda = .67$), followed by relational closeness (5.8; $\lambda = .63$) and conversation frequency (5.6; $\lambda = .56$). On the Principal factor, the value of discussions (5.9b) loaded highest ($\lambda = .74$), followed by relational closeness (5.10; $\lambda = .67$) and conversation frequency (5.9; $\lambda = .63$). The factor correlation was .36, lower than in Model 2 (.50), as expected once method variance is partialled out. All three correlated residuals were significant ($p < .001$), with the frequency pair strongest ($r = .41$).

Table 5. Fit indices across CFA models.

Model	χ^2	df	RMSEA	CFI	TLI	AIC	$\Delta\chi^2$ (Δdf)
M1: One factor	716.02	9	.211	.635	.391	26154	—
M2: Two factors	361.16	8	.158	.818	.658	25801	354.86*** (1)
M3: Two factors + CU	7.37	5	.016	.999	.996	25453	353.79*** (3)

Note. N = 1,769. MLMV estimation. CU = correlated uniquenesses. *** $p < .001$.

Table 6. Standardised factor loadings and residual correlations (Model 3).

Item	Team	Principal	CU (r)	R^2
5.6 Conversation freq. (team)	.56	—	.41	.31
5.7 Value of disc. (team)	.67	—	.25	.45

5.8 Relational closeness (team)	.63	—	.29	.40
5.9 Conversation freq. (principal)	—	.63	.41	.40
5.9b Value of disc. (principal)	—	.74	.25	.55
5.10 Relational closeness (principal)	—	.67	.29	.45
Factor correlation (Team ↔ Principal)		.36		

Note. CU = correlated uniqueness between parallel items. R^2 = proportion of variance explained by the latent factor. All loadings and CUs significant at $p < .001$.

Summary. The confirmatory factor analyses support a two-dimensional structure of teacher collaboration. A model distinguishing horizontal collaboration (teacher team) from vertical collaboration (principal/school leaders), with correlated uniquenesses for method effects, provided excellent fit. The moderate factor correlation ($r = .36$) indicates that the two dimensions share approximately 13% of their variance, confirming related but distinct aspects of teachers' professional interactions.

Intraclass correlations

To assess the extent to which teacher responses cluster at the school level, intraclass correlations were estimated from empty two-level models on the subsample with valid school codes (Table 7).

Table 7. Intraclass correlations.

Scale	ICC	95% CI	LR test p
Teacher team	.073	[.026, .188]	.012
Principal	.109	[.051, .220]	.001
School climate	.036	[.005, .229]	.136

Subgroup analyses by team type

Teachers were grouped by team structure: grade-level teams ($n = 551$) and cross-grade teams ($n = 1{,}190$). In cross-grade teams, teacher team and principal interaction were substantially more correlated (.42 vs. .25), suggesting greater overlap between horizontal and vertical communication in these structures. In grade-level teams, the team–climate association was somewhat stronger (.46 vs. .43). ICC estimates differed markedly: in grade-level teams, substantial school-level variance was observed for both teacher team interaction (16%) and principal interaction (26%), while in cross-grade teams only the principal scale showed significant clustering (12%). School climate showed no significant school-level variance in either group.

Preliminary correlational analysis: teacher collaboration and student literacy outcomes

The merged school-level dataset contained 756 matched schools. Teacher coverage per school was thin ($M = 1.33$ teachers per school). In the pooled sample, no statistically

significant associations were observed between any teacher-reported scale and any student achievement outcome (r ranging from approximately –.07 to .07). When examined by team type, cross-grade team schools (n = 88) showed several modest associations: school climate was positively associated with decoding (r = .23, p < .05), and principal interaction correlated with reading comprehension 2 (r = .22, p < .05) and listening comprehension (r = .22, p < .05). Same-grade team schools (n = 273) showed no significant associations.

Discussion

This working paper evaluated the measurement properties of teacher collaboration items from a survey distributed by LegiLexi to Swedish primary school teachers. The results support a two-dimensional structure distinguishing horizontal collaboration (within teacher teams) from vertical collaboration (with principals and school leaders). This distinction is consistent with the broader literature on teacher professional interactions, which recognizes that relationships with peers and relationships with school leaders serve different functions in teachers' professional lives (Honingh & Hooge, 2014; Meyer et al., 2022).

The confirmatory factor analysis demonstrated that a model with correlated uniquenesses for parallel item content provided excellent fit, while simpler models did not. The method effects captured by the correlated uniquenesses, reflecting shared response tendencies across items measuring the same facet (frequency, value, closeness) for different targets, were substantial, particularly for the frequency items. This finding has practical implications: composite scale scores computed as simple averages conflate substantive factor variance with method variance, which may attenuate or distort correlations with external criteria. Latent variable approaches that model measurement error explicitly are preferable when the data permit.

Internal consistency estimates were below conventional thresholds for all three proposed scales ($\alpha = .51–.69$). This is not unexpected for short scales where items capture conceptually distinct facets of a construct (e.g. how often teachers talk, how useful they find it, and how close they feel) rather than serving as interchangeable indicators. The CFA results confirm that these items do form coherent latent factors despite their heterogeneity, reinforcing that Cronbach's alpha underestimates the quality of measurement in such cases.

The preliminary correlational analyses at the school level yielded no significant associations between teacher collaboration and student literacy outcomes in the pooled sample. Several factors likely contribute to this null finding. First, the thin teacher coverage (median of 1 teacher per school) means that school-level means are largely based on individual teacher responses rather than reliable aggregates, introducing substantial measurement error that attenuates correlations. The confirmatory factor analyses demonstrated that horizontal and vertical collaboration are empirically distinct dimensions ($r = .36$), with substantial method variance shared across parallel items. Aggregating these items into simple composites at the school level compounds measurement error from both thin coverage and unmodelled method effects. Multilevel structural equation modelling, which preserves the latent factor structure and accounts for measurement error, would yield more credible school-level estimates, but requires substantially greater teacher coverage per school than the present data afford. The confirmatory factor analyses establish that the measurement model is sound at the individual teacher level, providing a foundation for such analyses in future data collections with denser school-level sampling.

Second, the differential findings across team types suggest that team type may function as a moderator of the relationship between teacher collaboration and student outcomes. The emergence of significant associations in cross-grade team schools, despite a

smaller sample, points to potential structural differences in how these team types operate that warrant further investigation.

Last, and most fundamentally, there is likely meaningful within-school variation in teacher experiences; some teachers within the same school may perceive stronger leadership or a better climate than others. Aggregating to school means eliminates this within-school variance, which may be where important relationships reside. Persson et al. (2025) found that the relationships between principal instructional leadership, teacher collaboration, and pupil achievement were driven by stable between-school differences rather than within-school changes over time. Future analyses should preserve the multilevel structure of the teacher data and model individual teacher perceptions within schools, with school-level student outcomes as the dependent variable.

In sum, this study establishes that the teacher collaboration items in the LegiLexi survey form a measurement structure consistent with the constructs they are intended to capture. The two-dimensional factor structure, horizontal and vertical collaboration, is well-supported, and the individual-level measurement model provides a solid foundation for future substantive research linking teacher collaboration to student reading outcomes.

References

Daly, A. J., Liou, Y.-H., & Der-Martirosian, C. (2021). A capital idea: Exploring the relationship between human and social capital and student achievement in schools. *Journal of Professional Capital and Community, 6*(1), 7–28.

Demir, E. K. (2024). *The role of social capital for teacher professional learning and student achievement* (Doctoral dissertation, University of Cambridge).

García-Martínez, I., Montenegro-Rueda, M., Molina-Fernandez, E., & Fernández-Batanero, J. M. (2021). Mapping teacher collaboration for school success. *School Effectiveness and School Improvement, 32*(4), 631–649.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

Honingh, M., & Hooge, E. (2014). The effect of school-leader support and participation in decision making on teacher collaboration in Dutch primary and secondary schools. *Educational Management Administration & Leadership, 42*(1), 75–98.

Meyer, A., Richter, D., & Hartung-Beck, V. (2022). The relationship between principal leadership and teacher collaboration: Investigating the mediating effect of teachers' collective efficacy. *Educational Management Administration & Leadership, 50*(4), 593–612.

OECD (2020). *TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals.* OECD Publishing.

Persson, R. (2024). Ledarskapsmätning i skolan – en explorativ studie om svenska myndigheters mätningar av rektors pedagogiska ledarskap. *CELE Working Paper.* Stockholm School of Economics.

Persson, R., Demir, E. K., & Wennberg, K. (2025). Principal instructional leadership and teacher collaboration: A longitudinal study of the influence on pupil achievement. *Educational Management Administration & Leadership.* https://doi.org/10.1177/17411432251350793

Pil, F. K., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *Academy of Management Journal, 52*(6), 1101–1124.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475–514.

Spillane, J. P., Hopkins, M., & Sweet, T. M. (2015). Intra- and interschool interactions about instruction: Exploring the conditions for social capital development. *American Journal of Education, 122*(1), 71–110.

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review, 15*, 17–40.

Ärlestig, H., Johansson, O., & Nihlfors, E. (2016). Sweden: Swedish school leadership research – An important but neglected area. In H. Ärlestig, O. Johansson, & E. Nihlfors (Eds.), *A Decade of Research on School Principals* (pp. 103–122). Springer.