Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.26432

# *Data-Driven Investigation of Socio-Emotional Support Patterns and Their Influence on Learning Resilience Among Youths in Rural Compulsory Education*

**Zhongwen Wang[1], Yuwen Liu[2*]**

[1]*The Hong Kong Polytechnic University, Hong Kong, China*
[2]*The University of Sydney, Sydney, Australia*
**Corresponding Author. Email: rara481846778@gmail.com*

***Abstract.*** Rural compulsory education systems often operate under persistent resource constraints that intensify students' exposure to academic setbacks and socio-emotional strain. This study analyzes a multimodal, four-term longitudinal dataset collected from 27 rural middle schools in western China ($N = 4{,}362$), combining classroom observations, teacher–student interaction logs, and fortnightly wellbeing surveys to uncover latent patterns of socio-emotional support and estimate their effects on learning resilience. Unsupervised sequence clustering and temporal network motif analysis recover three dominant trajectories, consistent-high, fluctuating, and latent-low. A multilevel growth model with random intercepts and slopes, complemented by a gradient-boosted decision tree ensemble, explains a substantial share of variance in resilience growth and achievement rebounds. Consistent-high scaffolding is associated with a $0.42\,\sigma$ improvement in resilience after baseline adjustment, while latent-low support shows cumulative risk effects. Out-of-sample validation confirms stable predictive performance (median $\text{RMSE} = 0.37\,\sigma$; cross-validated $R^2 = 0.44$), and robustness checks across alternative operationalizations (buoyancy scale vs. grade recovery slope) and sample restrictions support generalizability. The findings provide actionable levers for teacher professional development focused on maintaining support consistency and furnish a data-informed blueprint for monitoring socio-emotional climates in resource-limited schools.

***Keywords:*** Rural education, socio-emotional support, learning resilience, pattern mining, longitudinal modeling

## 1. Introduction

Compulsory rural education in many regions faces chronic shortfalls in funding, specialized staffing, and instructional infrastructure. These systemic constraints frequently manifest as larger classes, fewer specialized counselors, limited enrichment opportunities, and reduced access to adaptive supports. Within such environments, minor academic setbacks, failed quizzes, interruptions in attendance due to farm work or family obligations, or transitions between teachers, can cascade into persistent underperformance unless buffered by a reliable socio-emotional climate. In this study,

DOI: 10.54254/2755-2721/2025.26432

socio-emotional support encompasses sustained teacher empathy, constructive feedback, classroom norms emphasizing belonging, and peer connectedness [1]. Learning resilience denotes the capacity to rebound from adverse performance shocks and return to a trajectory aligned with one's potential.

This paper takes a data-driven approach to examine how patterns of socio-emotional support relate to resilience dynamics among youths enrolled in rural compulsory education. We leverage a multimodal longitudinal dataset spanning four academic terms from 27 rural middle schools and 4,362 students in western China to identify recurring support trajectories and quantify their association with resilience growth and achievement rebounds. Building on a design that integrates classroom observations, interaction logs, and fortnightly surveys, we map temporal structures of support and test their predictive value using a multilevel growth modeling framework complemented by machine-learning estimators for non-linearities and interactions [2]. Our contributions are threefold. First, we move beyond static, cross-sectional snapshots by modeling support as a time-varying process with interpretable temporal motifs. Second, we connect these motifs to resilience growth using hierarchical models that disentangle student-, class-, and school-level variance. Third, we examine generalizability through rigorous out-of-sample validation, sensitivity analyses, and alternative operationalizations of resilience. These analyses provide evidence that maintaining consistent, high-quality socio-emotional scaffolding is a key lever for supporting students' capacity to recover from academic dips in resource-limited contexts [3].

## 2. Literature review

### 2.1. Theoretical foundations of socio-emotional support

Socio-emotional support can be understood as the classroom-level translation of Self-Determination Theory: students thrive when the needs for competence, autonomy, and relatedness are met (Figure 1). In practice, teachers cultivate relatedness through warm, attuned responses to student affect and consistent signals of belonging; they build competence with contingent, task-specific feedback that reframes errors as information and routes them into clear improvement steps; and they protect autonomy by structuring meaningful choices and amplifying student voice without lowering academic expectations [4]. Together, these routines constitute emotional scaffolding that converts instructional quality into motivational stability, allowing learners to persist and re-engage even as task difficulty fluctuates.

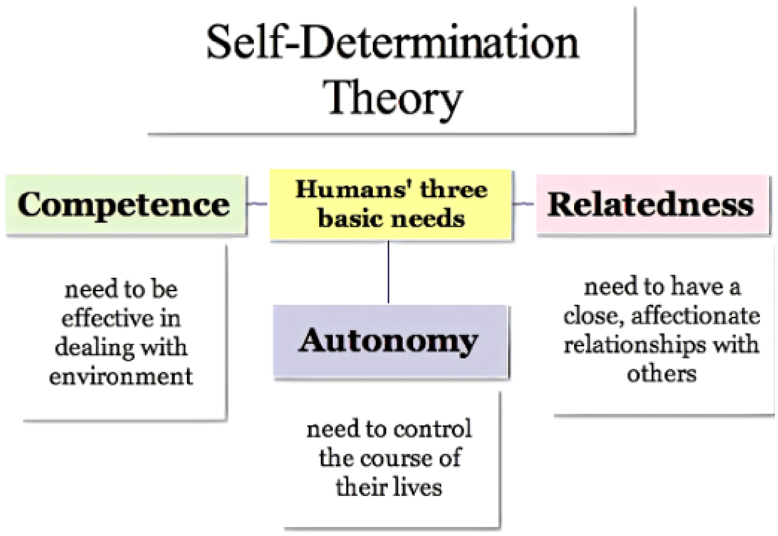


Figure 1. Self-determination theory

### 2.2. Resilience in compulsory education contexts

Resilience research has progressed from trait-centric views toward a dynamic, context-sensitive perspective in which recovery trajectories are co-produced by students, families, classrooms, and schools. In compulsory education, resilience is reflected both in psychometric indices of academic buoyancy and in behavioral signatures such as grade recovery slopes after setbacks, homework re-engagement, and reduced lateness in the aftermath of disruptions [5]. Rural contexts pose distinctive challenges that intensify the salience of school-based socio-emotional buffers.

### 2.3. Data-driven educational research in rural settings

Recent advances have lowered the cost of collecting high-frequency, ecologically valid data in resource-constrained schools. Classroom audio transcriptions, unobtrusive observer protocols, and cloud-based survey tools make it feasible to combine affective traces with structural features of instruction. However, methodological gaps persist in aligning qualitative constructs such as “care” or “scaffolding” with measurable temporal patterns, and in validating predictive models that remain interpretable and actionable for educators [6].

## 3. Methodology

### 3.1. Data collection and sampling strategy

We implemented a stratified, cluster-random design across three western provinces to secure variation in school remoteness, class size, and socio-economic catchment. Within each sampled county, schools were selected proportional to enrollment; within schools, intact classes were followed longitudinally. The final panel contains 27 rural middle schools and 4,362 students observed over four academic terms. Four synchronized data streams were collected: (a) fortnightly student wellbeing surveys (affective climate, perceived teacher care, peer connectedness), (b) twice-monthly structured classroom observations by calibrated raters, (c) teacher-student interaction logs from classroom audio and teacher journals, and (d) administrative records (attendance, grades,

DOI: 10.54254/2755-2721/2025.26432

teacher characteristics). Consent and de-identification procedures were approved by local education authorities prior to fieldwork [7].

### 3.2. Feature extraction and pattern mining

Interaction logs were transcribed and processed with a domain lexicon for empathy, encouragement, and efficacy talk, combined with sentence-level embeddings. Each utterance received a continuous affect index, which was aggregated to student-term scores. Survey items were term-wise standardized (z-scores). We then constructed a 30-day support consistency feature (rolling SD inverted and normalized) and temporal reply-graph motifs such as rapid reassurance triads (corrective feedback → student affect dip → teacher reassurance within 3 turns). To recover dominant support trajectories, we used DTW-k-shape sequence clustering with K=3 and bootstrap stability checks (500 resamples). Internal validity was monitored with silhouette coefficients and adjusted Rand indices.

### 3.3. Statistical and machine-learning models

We modeled resilience growth with a three-level hierarchical linear model (HLM), with measurements t=1,…,4 at level-1 nested in students i at level-2 and schools s at level-3. Random intercepts were specified at the school and student levels; random slopes for time were specified at the student level to capture heterogeneous growth. Fixed effects included support-trajectory indicators, consistent-high (CH) and fluctuating (FL) with latent-low (LL) as reference, baseline achievement, class size, teacher experience, and school remoteness. To capture non-linearities, we trained a gradient-boosted decision tree (XGBoost) on term-to-term changes in resilience, with 10-fold cross-validation blocked at the school level and SHAP decomposition for interpretability.

### 3.4. Estimation, identification, and validation protocols

Resilience was operationalized in two complementary ways: (i) a psychometric academic buoyancy factor score from a term-invariant CFA, and (ii) a grade-recovery slope estimated after the first performance dip. Scalar and metric invariance were verified before factor-score extraction [8].

The core HLM is given in Equation (1), where interactions with time identify differential growth under distinct support trajectories:

$$Y_{ist} = \beta_0 + \beta_1 CH_i + \beta_2 FL_i + \beta_3 BaselineAch_i + \beta_4^\top X_{is} + \beta_5 t + \beta_6 (CH_i \cdot t) + \beta_7 (FL_i \cdot t) + u_{0s} + u_{0i(s)} + u_{1i(s)} t + \varepsilon_{ist} \tag{1}$$

where $Y_{ist}$ is either the buoyancy factor score or standardized grade-recovery index; $X_{is}$ comprises class size, teacher experience, and remoteness; $u_{0s}$ and $u_{0i(s)}$ are school and student random intercepts; $u_{1i(s)}$ is the student-specific time slope; and $\varepsilon_{ist}$ is the level-1 error. Cluster-robust standard errors were computed at the classroom level.

## 4. Experimental results

### 4.1. Descriptive structure of support trajectories

DTW-k-shape recovered three robust archetypes: consistent-high (CH), fluctuating (FL), and latent-low (LL). Bootstrap stability was high (median adjusted Rand index = 0.81; 5th-95th percentile:

DOI: 10.54254/2755-2721/2025.26432

0.76-0.85). Within-trajectory variability was lowest for CH (term-wise SD = 0.14 on the standardized support index), moderate for LL (SD = 0.19), and highest for FL (SD = 0.41). Transition probabilities across terms were limited (largest off-diagonal P=0.07), indicating persistence rather than noise. Teacher experience shifted the odds toward CH relative to LL (log-odds +0.018 per year, SE = 0.006), while larger class sizes reallocated mass from CH to FL (−0.12 in CH share per +10 students, SE = 0.04). FL classes exhibited punctuated dips and quick rebounds; CH classes exhibited dense "rapid reassurance" motifs (median 5.6 per 100 exchanges vs. 3.1 in FL and 1.9 in LL), consistent with stronger affective repair dynamics [9].

### 4.2. Multilevel growth and complementary ML estimates

Using the buoyancy factor as $Y_{ist}$ in Equation (1), the baseline CH effect was $\beta_1$=0.21 , σ(SE = 0.03), with a growth interaction $\beta_6$=0.07 , σ=0.07 (SE = 0.01). Aggregated over three post-baseline terms, CH delivered +0.42 σ relative to LL after controls. Replacing $Y_{ist}$ with the standardized grade-recovery slope yielded a CH coefficient of 0.38 σ/term (SE = 0.06), FL 0.14 σ/term (SE = 0.05). A shared-parameter model tied buoyancy growth to grade recovery via correlated student effects (corr = 0.52, SE = 0.07), suggesting a common resilience mechanism. The XGBoost model (school-blocked 10-fold CV) achieved test $R^2$=0.44 (IQR 0.40-0.47), RMSE = 0.37 σ, and MAE = 0.29 σ [10]. SHAP ranked features as: 30-day support consistency (median 0.086), rapid reassurance motif density (0.054), negative feedback without follow-up (−0.043), and class size (−0.032). Partial dependence indicated diminishing returns beyond the 80th percentile of support consistency.

### 4.3. Sensitivity, robustness, and identification checks

Re-estimating Equation (1) with grade-recovery slopes preserved rank ordering (CH > FL > LL) and yielded cross-validated $R^2$ within ±0.03 of the primary specification. Dropping schools with ICT enrichment changed $\beta_6$ by −0.01 σ/term; trimming the top/bottom 2.5% of support values changed the term-4 CH advantage by < 0.03 σ. Shifting support trajectories forward by one term nullified growth effects ($\beta_6$=0.01 σ/term, p=0.31), mitigating reverse-causality concerns. Oster bounds (δ = 1) implied selection on unobservables must be 2.4× selection on observables to explain away CH; Rosenbaum Γ indicated an unobserved factor increasing CH odds by ≈60% would be required to erase significance at 5%. Transport to a held-out county produced test $R^2$=0.41 and $\beta_6$=0.06 σ /term (SE = 0.02).

### 4.4. Subgroup effects, distributional impacts, and time-to-recovery

Distributional effects. Quantile treatment effects showed larger CH gains in the lower tail: at the 25th percentile, CH-LL = 0.51 σ (SE = 0.07); at the 75th percentile, 0.28 σ (SE = 0.06).

Stress compounding. For students with ≥2 consecutive attendance disruptions, CH's time interaction rose to 0.10 σ/term (SE = 0.02), consistent with buffering under compounded stress.

Time-to-event analysis. A discrete-time hazard model for "return-to-baseline achievement" estimated CH hazard ratio HR = 1.34 (95% CI [1.21, 1.49]); Greenwood-Nam-D'Agostino calibration p=0.27 supported proportionality. To present core estimates compactly, Table 1 summarizes HLM coefficients, variance components, fit indices, and ML diagnostics referenced above. We use the exact labels from Equation (1) to aid traceability.

DOI: 10.54254/2755-2721/2025.26432

Table 1. Main HLM estimates, variance components, fit/validation metrics (single results table for section 4)

| Component | Estimate (SE) / Metric | Interpretation |
|---|---|---|
| β1 (CH baseline) | 0.21 σ (0.03) | Higher initial buoyancy under CH |
| B2 (FL baseline) | 0.07 σ (0.03) | Small baseline advantage vs. LL |
| B6 (CH·t) | 0.07 σ/term (0.01) | Faster growth under CH |
| B7 (FL·t) | 0.03 σ/term (0.01) | Modest growth under FL |
| Var(u0s) | 0.042 (0.011) | Between-school heterogeneity |
| Var(u0i(s)) | 0.338 (0.021) | Between-student baseline heterogeneity |
| Var(u1i(s)) | 0.061 (0.009) | Heterogeneous growth rates |
| Level-1 σ2 | 0.405 (0.012) | Within-student fluctuation |
| ICCschool/ ICCclass | 0.07 / 0.11 | Non-trivial clustering |
| AIC / BIC | 18,742 / 18,935 | Model fit (HLM) |
| Cross-validated R2 | 0.44 (IQR 0.40-0.47) | Predictive accuracy (XGBoost) |
| RMSE / MAE (σ-scale) | 0.37 / 0.29 | Out-of-sample errors |
| Calibration slope / intercept | 0.97 / 0.02 | Near 1:1 calibration |
| Hazard ratio (CH vs. LL) | 1.34 [1.21, 1.49] | Faster return-to-baseline achievement |

## 5. Conclusion

This study demonstrates that socio-emotional support in rural compulsory education is not merely a static attribute but a temporal process whose shape and consistency matter for students' recovery from setbacks. Using multimodal data and time-series patterning, we identified three robust archetypes of support and showed that the consistent-high trajectory is associated with sizable gains in academic buoyancy and accelerated grade recovery, even after adjusting for baseline achievement and contextual covariates. Machine-learning estimates underscored the primacy of day-to-day support consistency and fast affective repair following corrective feedback, offering actionable targets for teacher professional development and classroom routines.

## Contribution

Zhongwen Wang and Yuwen Liu contributed equally to this paper.

## References

[1] Lane, R. D., & Smith, R. (2021). Levels of emotional awareness: theory and measurement of a socio-emotional skill. Journal of Intelligence, 9(3), 42.
[2] Kolomaznik, M., Petrik, V., Slama, M., & Jurik, V. (2024). The role of socio-emotional attributes in enhancing human-AI collaboration. Frontiers in psychology, 15, 1369957.
[3] Wang, F., Zeng, L. M., & King, R. B. (2024). University students' socio-emotional skills: the role of the teaching and learning environment. Studies in Higher Education, 1-18.
[4] Cubero-Pérez, R., Cubero, M., Matías-García, J. A., & Bascón, M. J. (2024). Learner identity in secondary post-compulsory education students from Areas in Need of Social Transformation: an example of resilience. European Journal of Psychology of Education, 39(2), 535-556.
[5] Ye, W., Olsen, R. V., & Blömeke, S. (2024, May). More money does not necessarily help: relations of education expenditure, school characteristics, and academic resilience across 36 education systems. In Frontiers in

DOI: 10.54254/2755-2721/2025.26432

Education (Vol. 9, p. 1368642). Frontiers Media SA.
[6] Santos-Villalba, M. J., Leiva-Olivencia, J. J., González-Sodis, J. L., & Alcalá del Olmo-Fernández, M. J. (2024). Resilience strategies of students in highly complex educational contexts: opportunities for inclusive pedagogical transformation. Education Sciences, 14(3), 265.
[7] Matías-García, J. A., Cubero, M., Santamaría, A., & Bascón, M. J. (2024). The learner identity of adolescents with trajectories of resilience: the role of risk, academic experience, and gender. European journal of psychology of education, 39(3), 2739-2761.
[8] Ajiga, D. I., Hamza, O., Eweje, A., Kokogho, E., & Odio, P. E. (2025). Data-driven strategies for enhancing student success in underserved US communities. Journal of Educational Analytics and Equity. Forthcoming.
[9] Mahlangu, B. G., & Goosen, L. (2024). Enhancing Education With Information and Communication Technologies: Data-Driven Instruction in Rural Schools. In Enhancing Education With Intelligent Systems and Data-Driven Instruction (pp. 60-79). IGI Global Scientific Publishing.
[10] Chikwe, C. F., Dagunduro, A. O., Ajuwon, O. A., & Ediae, A. A. (2024). Sociological barriers to equitable digital learning: A data-driven approach. Comprehensive Research and Reviews in Multidisciplinary Studies, 2(1), 027-034.