IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 006, LIR489
DOI: https://doi.org/10.52731/lir.v006.489

A Time-Constrained Analysis of Dynamic Early Warning Systems for Academic Risk Prediction

Shintaro Tajiri [*], Kunihiko Takamatsu [†], Naruhiko Shiratori [‡], Kimikazu Sugimori [§], Sayaka Matsumoto [†], Shotaro Imai [†], Tetsuya Oishi [**], Masao Mori [†], Masao Murota [†]

Abstract

Implementing effective academic support for mandatory first-year courses requires precise decision-making about when to intervene, with whom, and with what level of certainty. This study extends our previous static prediction model (AUC=0.878 [1] using enrollment data alone) by addressing its key limitation: the inability to answer operational questions about intervention timing. Using data from a mandatory Information Literacy course at Hokuriku University (N=335, Economics and Management faculty, 2022-2023), we developed machine learning models that incrementally add dynamic formative assessment data from weeks 2-8 to static enrollment information. Under strict time-constraints preventing data leakage, we evaluated models using Recall@Precision≥0.90—a practical metric balancing intervention resource constraints with student rescue effectiveness. Results demonstrate that minimal behavioral features from weeks 2-8 (submission rates, task completion counts) significantly improve Recall@P≥0.90 from 1.6% to 3.2%, doubling rescue capacity while providing weeks of intervention lead time.

Keywords: Early Warning Systems, Time-Constrained Prediction, Learning Analytics, Institutional Research

1 Introduction

Academic support in first-year university courses faces a fundamental decision-making challenge: identifying which students need intervention, when intervention should occur, and with what level of confidence. In our previous work, we developed a static prediction model using only enrollment data to predict failure in a mandatory Information Literacy course, achieving AUC=0.878 [1]. However, this model could not answer crucial operational questions: When should we intervene? How little in-semester data is sufficient for reliable decisions?

Traditional early warning research optimizes for accuracy or AUC, but operational implementation demands different metrics. High recall alone generates excessive false positives, overwhelming limited support resources. High precision alone produces too few alerts. Real-world deployment requires balancing these constraints: achieving high precision (≥90%) while maximizing recall.

This study develops a dynamic early warning system that sequentially adds formative

[*] Kanda University of International Studies, Chiba, Japan
[†] Institute of Science Tokyo, Japan
[‡] Tokyo City University, Tokyo, Japan
[§] Hokuriku University, Ishikawa, Japan
[**] Kyushu Institute of Technology, Fukuoka, Japan

assessment data to our static baseline model. We answer two research questions:

- RQ1: Can minimal behavioral data from weeks 2-8 significantly improve recall of at-risk students under strict precision constraints (≥90%)?
- RQ2: Do dynamic behavioral change patterns (Week 2-4 vs Week 5-8 comparison) add predictive value beyond static behavioral aggregates?

2 Related Work

Recent learning analytics research emphasizes behavioral indicators over static demographics. Studies of LMS engagement patterns consistently show that behavioral trajectories provide stronger predictive signals than initial student attributes. However, few studies address when these signals become reliable for intervention decisions [2], [3], [4], [5].

A critical limitation in predictive modeling is data leakage—using future information to predict past outcomes. This study employs strict time-split validation: models at week t use only data through week t, ensuring findings reflect information available to practitioners [6].

Educational prediction studies typically report AUC or accuracy, but intervention programs face resource constraints prioritizing precision over recall. We use Recall@Precision≥0.90 as our primary metric: this measures the proportion of at-risk students correctly identified (recall) when the prediction threshold is set to achieve at least 90% precision. The 0.90 threshold reflects practical constraints—limited counseling resources mean that excessive false positives lead to intervention fatigue, while the cost of missing truly at-risk students is high [7].

3 Methods

3.1 Study Context and Data

The Information Literacy course at Hokuriku University is a mandatory first-year course implemented across all four faculties since 2022. This study analyzes students from the Faculty of Economics and Management (N=335, 2022-2023), representing 40% of first-year enrollment. We focus on this faculty for three reasons: (1) complete and consistent data collection, (2) systematic implementation of standardized rubric-based assessment, and (3) control for faculty-specific variations. The course operates under a multi-instructor framework with approximately 10 class sections.

The 15-week curriculum integrates Information Literacy (digital competencies, information ethics, AI basics) and Tableau Data Science (visual analytics, data storytelling). A critical design feature is the two-tier assessment structure. Weeks 2-8 use submission-only assignments (binary grading) to establish baseline engagement. Starting Week 9, the course transitions to rubric-graded unit assignments evaluated on five dimensions: data preparation (20%), visualization effectiveness (25%), analytical insight (30%), presentation clarity (20%), and creativity (5%). Students analyze real campus store sales data in a competition format judged by faculty and industry partners.

Data sources include: (1) Static enrollment data: high school GPA, learning attitudes, PC experience; (2) Dynamic formative assessment data: weekly assignment scores, submission status (on-time/late/missing), submission timestamps; (3) Outcome: binary pass/fail classification.

3.2 Time-Constrained Feature Engineering

To prevent data leakage, we generate features using only data available through week t. We extend static enrollment features with behavioral trajectory indicators:

Compliance Indicators: Non-submission streaks (maximum consecutive non-submissions through week t), cumulative submission rate, delay rate, first assignment submission status.

Performance Trajectories: Score improvement trajectory (change from week 1 to week 2), cumulative mean score, performance variance.

Behavioral Change Patterns: We operationalize four patterns based on submission rate thresholds: (1) Consistently Good (>80% in both weeks 2-4 and weeks 5-8), (2) Deteriorating (>80% in weeks 2-4 but <60% in weeks 5-8), (3) Recovering (<60% in weeks 2-4 but >80% in weeks 5-8), (4) Consistently Poor (<60% in both periods). Students not fitting these categories are classified as "Other." The Week 4/5 boundary reflects a pedagogical transition from guided practice to autonomous application.

3.3 Progressive Model Architecture

We implement progressive model comparison:

Model 0: Enrollment data only (~120 features)

Model 1: Model 0 + Week 2-8 static behavioral indicators (+10 features)

Model 2: Model 1 + Week 2-8 dynamic change patterns (+15 features)

Model 3: Model 2 + Week 9 rubric-graded unit assignment feedback (+5 features)

We use XGBoost classifiers with 5-fold cross-validation on 2022 training data and temporal validation on 2023 test data [8].

4 Results

4.1 Progressive Model Performance (RQ1)

Table 1 summarizes performance metrics across our four progressive models. Model 0 (enrollment-only) establishes a strong baseline (AUC=0.596), but achieves only 1.6% Recall@P≥0.90, identifying merely one at-risk student with 90% confidence. Adding weeks 2-8 behavioral indicators (Model 1) improves AUC to 0.712 and doubles rescue capacity to 3.2% recall. Model 2, incorporating dynamic behavioral change patterns, further improves to AUC=0.749 while maintaining 3.2% recall. Model 3, adding Week 9 performance feedback, reaches AUC=0.832 with 4.8% recall.

Table 1: Progressive Model Performance Comparison

Model	AUC	AUC-PR	Recall @P≥0.90	Features Added
Model 0: Enrollment	0.596	0.128	1.6%	Enrollment only
Model 1: + Behavior	0.712	0.327	3.2%	+ Behavior patterns
Model 2: + Dynamic	0.749	0.341	3.2%	+ Behavioral change
Model 3: + Performance	0.832	0.436	4.8%	+ Week 9 Tableau

Figure 1 presents ROC curves comparing all four models. The progression from Model 0 (blue) through Model 3 (purple) demonstrates systematic improvement in discriminative ability, with Model 3 achieving superior performance across all operating points. The gap between Model 0 and Model 1 is particularly notable, confirming that even minimal behavioral data substantially improves prediction beyond enrollment information alone.

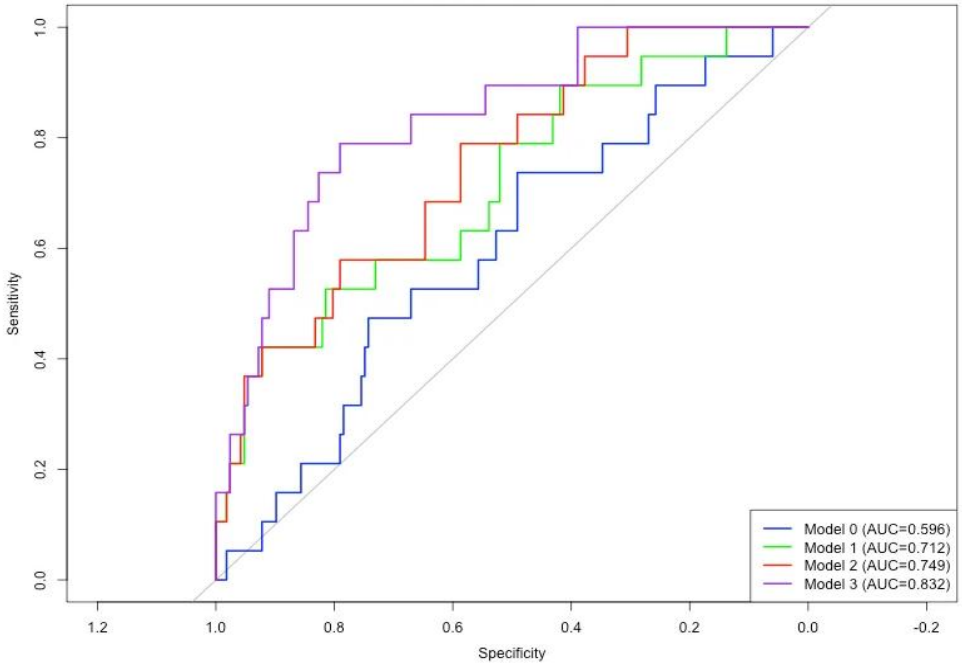

Figure 1: ROC Curves Comparison Across All Models

Figure 2 shows Precision-Recall curves, which are particularly informative for imbalanced datasets like ours. Model 3 (purple) achieves the highest precision at all recall levels, with AUC-PR=0.436 [7]. The substantial gap between Model 0 (AUC-PR=0.128) and Models 1-3 emphasizes the critical value of behavioral data for identifying at-risk students under precision constraints.

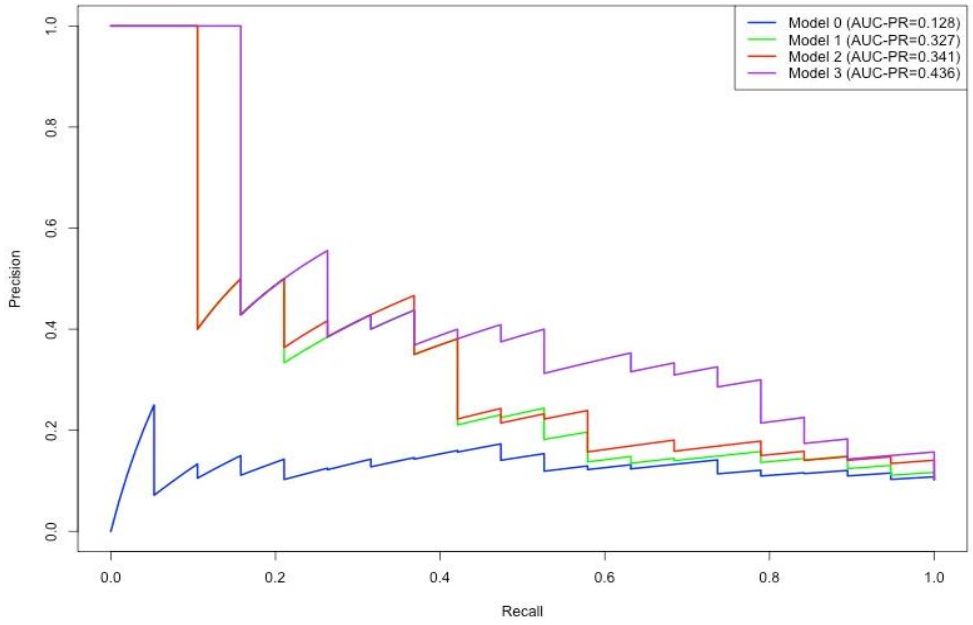

Figure 2: Precision-Recall Curves Comparison

4.2 Behavioral Pattern Analysis (RQ2)

Figure 3 presents SHAP (SHapley Additive exPlanations) values revealing which features most strongly influence Model 2's predictions. The analysis demonstrates that behavioral indicators dominate predictive importance. Among the top features, average submission rate (late period) and number of submitted tasks show the strongest associations with failure risk. Notably, while some enrollment attributes like PROG competency scores remain important, behavioral engagement metrics provide the most actionable signals for early intervention [9].

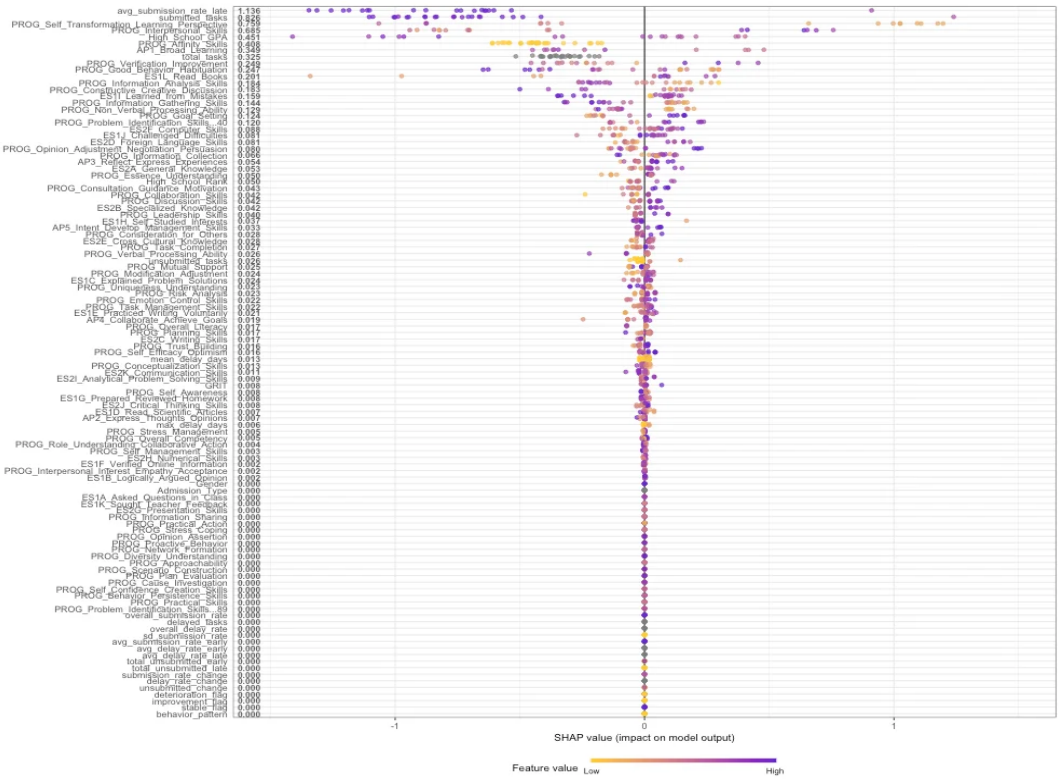

Figure 3: SHAP Feature Importance Analysis for Model 2

Figure 4 analyzes failure rates across different behavioral patterns identified in weeks 2-8. The most striking finding concerns students showing deterioration patterns—those who started well but declined in weeks 5-8. This group exhibits a 72.7% failure rate despite strong initial performance (n=11). In contrast, students maintaining consistently good engagement throughout weeks 2-8 show only 9.4% failure rate (n=297). This finding is particularly actionable: it suggests that monitoring behavioral trajectories reveals disengagement earlier than static metrics, enabling targeted intervention for students at highest risk.

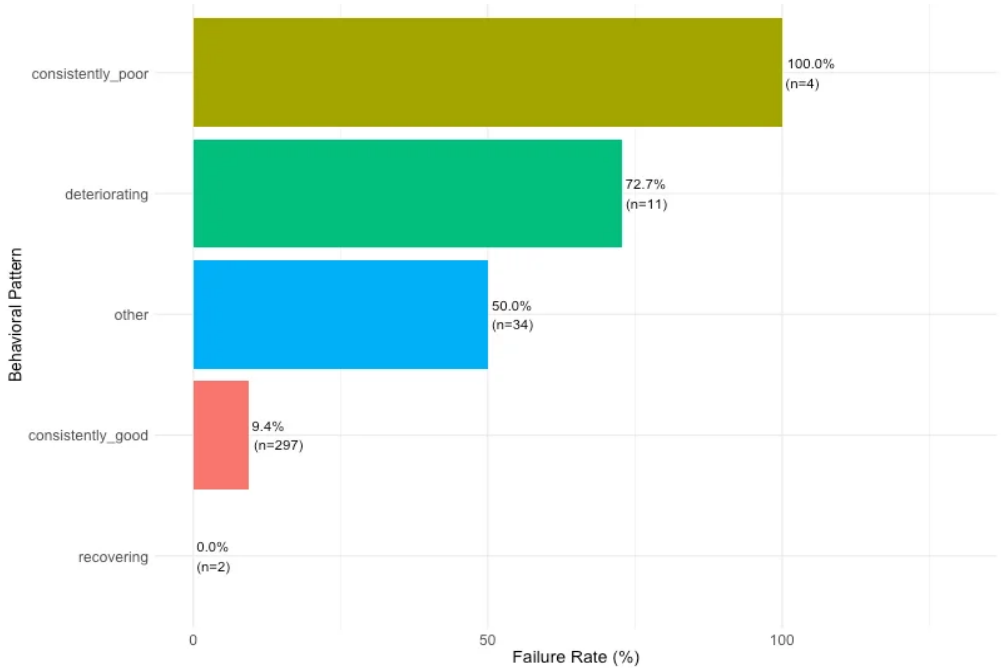

Figure 4: Failure Rate by Behavioral Pattern (Weeks 2-8)

5 Discussion

5.1 Key Findings

Three critical findings emerge that directly extend our previous static prediction work:

Finding 1: Behavioral Data Doubles Prediction Capacity. Behavioral data doubles prediction accuracy compared to enrollment-only models (Recall@P≥0.90: 1.6% → 3.2%). This demonstrates that submission patterns and engagement trajectories provide substantial predictive value beyond static enrollment information.

Finding 2: Week 4 as Optimal Intervention Timing. Adding dynamic change patterns (comparing Week 2-4 vs Week 5-8) to static behavioral features yields no improvement in Recall@P≥0.90 (Model 1→Model 2: 3.2%→3.2%). Since Week 5-8 data adds no predictive value, intervention decisions can be made at Week 4—the end of the early observation period—providing maximum lead time for student support.

Finding 3: Deterioration Patterns as Priority Targets. Students showing deteriorating patterns—strong initial performance followed by declining engagement—represent the highest-risk group (72.7% failure rate). Unlike consistently poor performers who may lack foundational skills, deteriorating students demonstrate initial capability but require re-engagement support.

5.2 Practical Implications

Three critical findings emerge that directly extend our previous static prediction work: these findings provide actionable guidance for early warning system implementation. The first is actionable

behavioral indicators. Early warning pre-dictions based on behavioral trajectories (submission rates, task completion patterns) provide signals that educators can directly observe and act upon, unlike static demographic factors that cannot be changed after enrollment. The second is resource-constrained decision making. The doubling of rescue capacity from 1.6% to 3.2% under Precision≥0.90 constraints represent tangible improvement in institutional capacity to help students. Since Week 5-8 dynamic patterns add no predictive value beyond Week 2-8 static aggregates, intervention decisions can be made earlier in the semester. The last is implementation simplicity. Behavioral features (submission compliance, engagement trajectories) require only standard LMS data readily available at most institutions, enabling deployment without specialized infrastructure or complex analytical tools.

5.3 Limitations

Several limitations constrain generalizability. First, single-institution, single-faculty analysis (N=335) limits statistical power and transferability. Second, binary pass/fail outcomes may mask important nuances in student performance. Third, this study establishes prediction capability but does not evaluate intervention effectiveness—the crucial next step involves randomized controlled trials testing whether early warnings improve outcomes.

6 Conclusions

This study successfully extends our previous static prediction research by demonstrating that operationally viable early warning systems can be developed through careful attention to temporal constraints, practical metrics, and fairness considerations. Our key contribution is showing that behavioral disengagement signals emerge earlier and more reliably than performance metrics, enabling effective intervention with minimal observation periods while doubling rescue capacity compared to enrollment-only approaches.

The finding that deteriorating behavioral patterns—not low initial performance—represent the highest risk group has important implications for intervention design. Rather than focusing solely on academically underprepared students identified through enrollment data, dynamic early warning systems should prioritize re-engagement support for students showing declining participation despite initial capability.

Future research should focus on three priorities: (1) validation across diverse academic contexts and institutions, (2) randomized controlled trials evaluating intervention effectiveness, and (3) development of adaptive systems that adjust prediction thresholds based on institutional capacity and student population characteristics. The ultimate goal of early warning research should be the development of actionable, fair, and effective systems that improve student success

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 25K06607 and 25K00843.

References

[1] S. Tajiri, K. Takamatsu, N. Shiratori, T. Oishi, M. Mori, and M. Murota, "Predicting Performance in First-Year Required Courses Using Machine Learning Based on At-Enrollment Data," in Proc. IIAI AAI 2024 Winter (DSIR Winter), 2024.

[2] A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance," IEEE Transactions on Learning Technologies, vol. 10, no. 1, pp. 82–92, 2017.

[3] N. Kondo, M. Okubo, and T. Hatanaka, "Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data," in Proc. 6th IIAI International Congress on Advanced Applied Informatics, pp. 198–201, 2017.

[4] D. Gašević, S. Dawson, and G. Siemens, "Let's Not Forget: Learning Analytics are About Learning," TechTrends, vol. 59, no. 1, pp. 64–71, 2015.

[5] K. E. Arnold and M. D. Pistilli, "Course Signals at Purdue: Using Learning Analytics to Increase Student Success," in Proc. LAK'12, pp. 267–270, 2012.

[6] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in Data Mining: Formulation, Detection, and Avoidance," in Proc. KDD'11, pp. 556–563, 2011.

[7] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," PLOS ONE, 10(3): e0118432, 2015.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. KDD'16, pp. 785–794, 2016.

[9] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proc. NeurIPS 30, pp. 4765–4774, 2017.