Koh, E., et al. (Eds.) (2026). Proceedings of the 2nd International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education.

# Identifying Signals of Continuance and Discontinuance in Peer-Supported JSL Learning: A Human-LLM Qualitative Approach

**Yu HIRAYAMA[a*], Yoshiko GODA[b] & Masayoshi ARITSUGI[c]**
[a]*Graduate School of Science and Technology, Kumamoto University, Japan*
[b]*Research and Education Institute for Semiconductors and Informatics, Kumamoto University, Japan*
[c]*Faculty of Advanced Science and Technology, Kumamoto University, Japan*
*hirayamataekwo@gmail.com

**Abstract:** Improving the continuity of Japanese language learning among international graduate students remains a critical challenge in Japan, particularly for those enrolled in English-taught programs. Peer-based collaborative learning has been widely implemented to support these learners, yet little is known about the early-stage signals related to participation being maintained or discontinued. This study attempts to identify "process-level signals" (i.e., qualitative indicators or markers) in participants' reflection texts that relate to later partnership sustainability through an exploratory qualitative approach.

Using 36 reflection entries collected from 10 pairs at a Japanese university, human qualitative analysis identified four major categories and ten sub-categories of interactional signals. Furthermore, we evaluated the feasibility of automating signal detection using a Large Language Model (LLM) under two prompting conditions: naïve zero-shot (Prompt A) using raw definitions, and devised zero-shot (Prompt B) with refined rules. The results reveal clear differences between the two prompting strategies. While Prompt B achieved perfect recall (1.00) and substantial agreement (κ=0.78) in identifying logistical signals such as time constraints, the stricter instructions induced an over-conservative bias. Consequently, Prompt B overlooked nuanced relational and value-oriented signals captured in Prompt A. These findings suggest that different prompting strategies are required depending on signal types, with stricter rules being effective for logistical signals and more flexible settings being necessary for relational nuances.

**Keywords:** Collaborative learning, Japanese language learning, qualitative analysis, process-level signals, LLM

## 1. Introduction

Improving the domestic employment outcomes of international students has become a key national priority in Japan. The Japanese government has announced a target of increasing the domestic employment rate of international students to 60% by 2033 (Council for the Creation of Future Education, 2023). In response, universities have implemented various initiatives, including Japanese language education and career support. Previous research has suggested that the continuity of Japanese language learning, especially for beginners, is significantly influenced by the starting timing and initial proficiency (Ryan & Hakamata, 2023). Therefore, encouraging international students to start and sustain their Japanese as a Second Language (JSL) learning remains a critical challenge.

Peer-based collaborative learning with Japanese students is a promising approach to support this challenge, as it enhances motivation and provides authentic communication opportunities (Sueshige, 2017; Nishio et al., 2020). However, while prior studies have primarily examined outcomes such as language improvement or learner satisfaction, less attention has been paid to process-level variations in the early stages of collaboration. Crucially, not all collaborative learning activities are sustained; while some thrive, others discontinue participation, often without formal intervention. Despite this variability, little is known about the process-level characteristics that relate to the persistence or vulnerability of these collaborative learning activities.

Accordingly, this study examines reflection texts from a peer-based collaborative learning program using an exploratory qualitative approach. In this study, the term “process-level signals” refers to analytically derived, qualitative indicators or markers observed in participants’ early-stage interactions, rather than statistically predictive or causal variables. First, through human qualitative analysis, the study aims to identify "process-level signals" that distinguish between potential continuance and discontinuance in early-stage collaboration. Second, the study explores the feasibility of using a large language model (LLM) to extract these signals automatically. If process-level signals can be reliably identified, such approaches could support automated early-stage screening, adaptive intervention, and improved pair matching. The research questions are:

RQ1: What process-level signals related to continuance and discontinuance can be identified in early-stage collaborative JSL learning based on participants’ reflections?

RQ2: To what extent can a large language model (LLM) identify these process-level signals from reflection texts compared with human qualitative analysis?

## 2. Related Work

Peer-based collaborative learning and tutoring activities between Japanese and international students have been implemented to support Japanese language learning and intercultural exchange in higher education. Prior studies have documented both the potential benefits and the challenges of such initiatives.

Studies on language exchange programs have pointed out some challenges, such as providing appropriate linguistic feedback, communication difficulties in online settings, and logistical problems related to scheduling and time coordination (Sueshige, 2017).

Research on Japanese student tutors emphasizes reciprocal interaction, where balanced participation leads to more positive experiences and rapport formation (Soeda, 2011). Qualitative studies using interview data also reported that international students gained self-efficacy and found meaning in their learning through sustained tutoring relationships. The study emphasized that building rapport is vital for the success of learning support programs (Iwasaki, 2011).

While these studies provide valuable insights into collaborative learning and tutoring, they primarily focus on overall outcomes, participant perceptions, or retrospective accounts. Less attention has been paid to the process-level variations observable in the early stages of collaboration. Furthermore, recent research suggests that large language models (LLMs) can analyze educational qualitative data with a level of consistency comparable to human coding (Long et al., 2024). This raises the possibility of detecting early-stage signals of continuance or discontinuance in peer-supported JSL learning contexts.

## 3. Research Setting and Data Profile

### *3.1 Program and Participants*

The data for this study were collected from a peer-based collaborative Japanese language learning program at a Japanese university. This program was designed to support international students in their acquisition of JSL and their adaptation to Japanese society through regular interaction with Japanese student supporters.

A total of 29 pairs participated in the program. Among them, 10 pairs (N=20) provided formal consent from both international and Japanese students to have their data utilized for research purposes. The international students in this study were graduate students who belonged to the Graduate School of Science and Technology (GSST). The Japanese student supporters were recruited from all departments. Participation was voluntary for both international students and Japanese student supporters.

## 3.2 Data Collection and Duration

The program ran from May 2025 to July 2025, with pairs expected to meet more than once a month. Following each session, participants submitted individual reflections via a Learning Management System (LMS) every month. For the qualitative analysis, this study focused on the first two reflection entries (R1 and R2) to identify early-stage signals related to the sustainability of the partnership.

Not all participants submitted reflections for every session, as shown in Table 1. In total, 36 reflection entries (19 from Japanese student supporters (JP supporters) and 17 from international students (IS)) were used for human coding and LLM-based detection.

Table 1. *Profile of Participating Pairs and Analyzed Data*

| Pair ID | JP Supporter (Faculty/Year/English level) | International Student (Faculty/JLPT) | Reflection (JP) *1 | Reflection (IS) *2 |
|---|---|---|---|---|
| Pair01 | Letters / 2nd yr./ B1 | GSST / - | R1, R2 | R1, R2 |
| Pair02 | Education / 1st yr./ A2 | GSST / N5 | R1, R2 | R1, R2 |
| Pair03 | Law / 2nd yr./ B1 | GSST / N4 | R1, R2 | R1 |
| Pair04 | Letters / 2nd yr./ A2 | GSST / N4 | R1 | R1 |
| Pair05 | Letters / 1st yr./ A2 | GSST / N4 | R1, R2 | R1, R2 |
| Pair06 | Letters / 1st yr./ A2 | GSST / N4 | R1, R2 | R1, R2 |
| Pair07 | Law / 1st yr./ A2 | GSST / N5 | R1, R2 | R1, R2 |
| Pair08 | Education / 1st yr./ A2 | GSST / N3 | R1, R2 | R1*3, R2 |
| Pair09 | Informatics / 1st yr./ A2 | GSST / N2 | R1, R2 | R1, R2 |
| Pair10 | Engineering / 1st yr./ A2 | GSST / N1 | R1, R2 | R1 |

*1 Reflections (JP) were written in Japanese. *2 Reflections (IS) were written in English.
*3 International student's R1 of Pair08 was only written in Japanese.

# 4. Methodology

## 4.1 Qualitative Analysis Procedure by Human Coding (RQ1)

The qualitative analysis followed an inductive process to ensure categories remained grounded in participants' experiences. First, the first author performed open coding on the 36 reflection segments, assigning descriptive labels. Second, the open codes were clustered using the KJ method to generate categories that represent early-stage signals related to continuance and potential discontinuance. To enhance the validity of the coding scheme, the second author reviewed the code definitions and the coded data.

## 4.2 Automated Analysis using LLM (RQ2)

To evaluate the feasibility of automated signal detection, we conducted an analysis using an LLM. We utilized Llama-3.1-8B-Instruct (4-bit quantized), a local LLM, to ensure data privacy and cost-effectiveness. We adopted a two-step approach:

1. Naïve zero-shot (Prompt A): Prompt based strictly on raw sub-category definitions.
2. Devised zero-shot (Prompt B): Refined prompt using specific pedagogical keywords and rules to address the "interpretative gaps" identified in Prompt A.

We evaluated the LLM coding compared with human coding using observed agreement, Cohen's kappa coefficient (κ), and recall. These measures were selected to capture not only surface-level consistency but also diagnostic reliability and coverage of human-coded signals.

# 5. Results

## 5.1 Identified Interactional Signals (RQ1)

The manual qualitative analysis identified four major categories comprising 10 sub-categories of signals (Table 2). These signals were not predefined but inductively generated from participants' reflections, capturing variations in early-stage collaborative processes. Some signals (e.g., time constraints, relational comfort, and self-efficacy-related perceptions) are consistent with prior studies (Sueshige, 2017; Soeda, 2011; Iwasaki, 2011).

As shown in Table 3, while all pairs recognized difficulties and engaged in some form of repair (B-1, B-2), these factors alone did not differentiate pairs by their eventual number of meetings. Instead, pairs that had a lot of meetings tended to exhibit signals related to relational comfort (C-2), personal interaction (C-3), and a clear orientation toward regular scheduling (A-2) from the initial stage. This suggests that while problem-solving is a necessary foundation, the emergence of relational and temporal alignment functions as a key predictor for sustained participation. Notably, the only pair that reported time constraints without expressing any intention to adjust their schedule (A-1 = 1, A-2 = 0) had the fewest meetings. This result suggests that unaddressed temporal constraints may serve as a strong early signal of disengagement, providing a clear target for automated detection and intervention.

Table 2. *Interactional Signals identified by Manual Qualitative Analysis*

| Category | Sub-category | Definition | Representative Excerpts |
|---|---|---|---|
| Engagement with Temporal Conditions | A-1 Expressions of Time Constraints | Reporting challenges related to scheduling or finding time to meet | It's difficult to schedule a time. |
| | A-2 Regular Scheduling | Descriptions of activities are becoming a fixed, regular, or routine part of the weekly schedule. | Every Monday during third period at the cafeteria |
| Awareness of Challenges and Repair Practices | B-1 Recognition of Difficulties | Identifying specific linguistic or pedagogical difficulties encountered during interaction. | It was difficult to teach pronunciation. |
| | B-2 Attempts to Address Difficulties | Using concrete strategies (translation by app, rephrasing, gestures, seeking advice from others) to mitigate identified issues. | using relatively simple Japanese or sentences. |
| Relational Deepening and Psychological Safety | C-1 Feelings of Anxiety or Tension | Expressing nervousness or interpersonal tension during the early phase of collaboration. | At first, I was worried about making mistakes. |
| | C-2 Increased Comfort and Openness | Describing increased comfort, relaxation, or ease in communicating with the partner. | The atmosphere was relaxed |
| | C-3 Personal Interaction Beyond Task | Descriptions of personal exchanges that extend beyond task-oriented learning activities, such as sharing meals or engaging in informal conversations. | we shared our personalities, likes and dislikes, hobbies, and favourite movies. |
| Reciprocal Learning and Meaning Making | D-1 General Value Recognition | Acknowledging the general value or meaningfulness of the collaborative activity. | it gives real opportunities to practice speaking in a supportive environment. |
| | D-2 Awareness of Reciprocal Learning | Recognition that learning occurs bidirectionally, with both partners gaining knowledge, understanding, or insights through the collaborative process. | The appeal lies in the fact that we can learn from each other. |
| | D-3 Self-efficacy Related Perceptions | Statements indicating participants' perceptions of having contributed effectively, helped their partner, or developed confidence in their own ability. | I gained more confidence |

Table 3. *Occurrence of R1 Signals and Total Number of Meetings*

| Pair ID | The number of meetings * | A-1 | A-2 | B-1 | B-2 | C-1 | C-2 | C-3 | D-1 | D-2 | D-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair01 | 7 | - | - | ✓ | ✓ | - | ✓ | - | ✓ | - | ✓ |
| Pair02 | 7 | - | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - |
| Pair03 | 4 | - | - | ✓ | ✓ | - | ✓ | ✓ | - | - | - |
| Pair04 | 1 | ✓ | - | ✓ | ✓ | - | - | ✓ | ✓ | ✓ | ✓ |
| Pair05 | 4 | - | - | ✓ | ✓ | ✓ | - | ✓ | - | - | - |
| Pair06 | 4 | - | ✓ | ✓ | ✓ | ✓ | - | - | ✓ | ✓ | - |
| Pair07 | 2 | - | - | ✓ | ✓ | - | ✓ | ✓ | - | ✓ | - |
| Pair08 | 4 | - | - | ✓ | ✓ | - | - | ✓ | ✓ | - | - |
| Pair09 | 9 | - | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | ✓ | - |
| Pair10 | 2 | - | - | ✓ | ✓ | ✓ | ✓ | ✓ | - | - | - |

* They were counted from the Reflection texts (R1 and R2).

## *5.2 Evaluation of LLM Performance (RQ2)*

As shown in Table 4, we evaluated the performance of the LLM under two prompting conditions, Prompt A and Prompt B. In Prompt B, the same definitions were retained, but additional keywords, boundary rules, and instructions against implicit inference were added based on issues identified in Prompt A. The excerpt of both prompts is shown in Figure 1.

Table 4. *Comparison of Detection Performance between Prompt A and Prompt B*

| Sub-Category | Agreement [%] | | Cohen’s k | | Recall | |
|---|---|---|---|---|---|---|
| | Prompt A | Prompt B | Prompt A | Prompt B | Prompt A | Prompt B |
| A-1 (Time) | 44.4 | 97.2 | 0.07 | 0.78 | 1.00 | 1.00 |
| A-2 (Regular) | 66.6 | 94.4 | 0.23 | 0.83 | 0.62 | 0.87 |
| B-1 (Difficulty) | 38.8 | 36.1 | 0.05 | -0.02 | 0.35 | 0.35 |
| B-2 (Attempts to Address) | 83.3 | 61.1 | 0.52 | 0.24 | 0.86 | 0.55 |
| C-1 (Anxiety) | 86.1 | 88.8 | -0.04 | 0.28 | 0.00 | 0.25 |
| C-2 (Comfort) | 55.5 | 63.8 | 0.20 | 0.23 | 1.00 | 0.46 |
| C-3 (Personal) | 77.7 | 50.0 | 0.56 | 0.05 | 0.70 | 0.25 |
| D-1 (Value) | 58.3 | 50.0 | 0.00 | -0.06 | 1.00 | 0.66 |
| D-2 (Reciprocity) | 55.5 | 61.1 | 0.11 | 0.22 | 0.44 | 0.33 |
| D-3 (Self-efficacy) | 55.5 | 63.8 | -0.14 | 0.11 | 0.16 | 0.50 |

**【Prompt A】**
You are a qualitative data analyst. Identify interactional signals based ONLY on the explicit definitions provided.
### DEFINITIONS:
A. Engagement with Temporal Conditions
- A-1 (Expressions of Time Constraints): Reporting challenges related to scheduling or finding time to meet.
. . .

**【Prompt B】**
You are a qualitative data analyst. Identify interactional signals based ONLY on the explicit definitions provided.
Your task is to detect whether specific interactional signals are explicitly present in a single reflection text.
Mark a signal as present (1) ONLY when there is clear, direct textual evidence.
Do NOT infer intentions, motivations, emotions, or outcomes beyond what is explicitly written.
If the evidence is implicit, vague, or ambiguous, mark the signal as absent (0).
Do NOT summarize or evaluate the activity.
Base your judgment strictly on the surface content of the text.
### SIGNAL DEFINITIONS AND CUES
A. Engagement with Temporal Conditions
- A-1: Expressions of Time Constraints
Reporting challenges related to scheduling or finding time to meet.
e.g., "busy", "could not find time", "schedules did not match", "during exams"
. . .

*Figure 1. Excerpt of Prompt A and Prompt B*

The transition from Prompt A to Prompt B phase revealed a significant shift in the model's behavior. The most remarkable improvement occurred in Category A (Engagement with Temporal Conditions). By explicitly defining exclusion rules—such as distinguishing recurring routines from isolated date reports—the LLM achieved a substantial agreement of $\kappa=0.78$ for A-1 and an almost perfect agreement of $\kappa=0.83$ for A-2. Notably, the Recall for A-1 remained at 1.00, demonstrating that the LLM could capture 100% of the students' reported time constraints without exception. This confirms that for logical and structural risk detection, a strict, human-refined prompt is highly effective.

Conversely, several categories, such as C-3 (Personal Interaction) and B-2 (Attempts to Address Difficulties), showed a decline in performance in Prompt B. This decline suggests an "over-conservative bias" triggered by the strict definitions. As the boundaries became narrower to eliminate noise, the LLM became more hesitant to assign codes, frequently missing subtle interpersonal nuances that human coders identified through contextual immersion. This result indicates that for high-context emotional or relational signals, a rigid zero-shot approach may be counterproductive, as it sacrifices sensitivity for precision.

# 6. Discussion

## *6.1 Practical Implications of the Signal Framework and LLM Use*

The primary contribution of this study is the formalization of process-level signals in peer-supported JSL learning contexts. Unlike traditional outcome-based assessments, our framework allows for the qualitative diagnosis of a pair's "health" at an early stage. Specifically, identifying the lack of Regular Scheduling (A-2) despite reported Time Constraints (A-1) provides a useful indicator of potential disengagement.

Regarding the role of the LLM, the findings demonstrate a clear functional divide. The LLM can effectively automate the detection of logistical signals, such as time constraints (Category A), which are critical indicators of potential disengagement. Given its perfect Recall for A-1, the LLM can reliably identify pairs requiring immediate structural support. However, the detection of rapport formation and value recognition (Categories C and D) remains a challenge, as strict prompting causes the LLM to overlook subtle relational cues.

These results indicate that human qualitative analysis remains essential for high-context relational signals, while the LLM is better suited for detecting logistical risks. In practice, reflection texts can be automatically analyzed after submission to identify early-stage signals and flag potentially at-risk pairs. This approach reduces analytical workload and enables more consistent and scalable detection, allowing the LLM to function as an early-warning tool that supports human-led interpretation.

### *6.2 Limitations of the Study*

Several limitations must be acknowledged in these findings. First, the qualitative analysis and the creation of baseline signals were conducted by a single researcher, although the coding scheme and coded data were reviewed by the second author to enhance validity. Second, the sample size was limited to 10 pairs at a university, restricting generalizability. Finally, this study utilized only one specific model (Llama-3.1-8B). Therefore, future studies should compare multiple models to improve detection performance.

## 7. Conclusion and Future Work

This study represents an exploratory step toward identifying and automating the detection of early-stage interactional signals in peer-based JSL learning. Future research should expand the dataset, validate signals across multiple coders, and examine the relationship between early signals and long-term outcomes. In addition, more advanced prompting techniques and model comparisons are needed to improve the detection of nuanced relational signals.

## References


Council for the Creation of Future Education. (2023, April 27). Future initiative for promoting study abroad to cultivate the next generation (Second proposal). Cabinet Secretariat of Japan. https://www.cas.go.jp/jp/seisaku/kyouikumirai/pdf/230427honbun.pdf (in Japanese)

Ryan, Y., & Hakamata, M. (2023). Career support for science and engineering graduate international students at beginner to intermediate Japanese levels: Toward the utilization of highly skilled foreign professionals (KAKENHI Grant No. 17K04680). Shizuoka University. (in Japanese)

Sueshige, M. (2017). A Language Exchange Project between International Students and Japanese Students. *Bulletin of the Institute for Education and Student Support Services*, Okayama University, 2, 77–86. (in Japanese)

Nishio, T., Fujikake, C., & Osawa, M. (2020). Language learning motivation in collaborative online international learning: An activity theory analysis. *Journal of Virtual Exchange*, 3, 27-47.

Soeda, E. (2011). An investigation into the Japanese language learning support provided by tutor activities: Perspectives of foreign students. *Journal of Language, Culture, and Japanese Language Education*, 40, 95–112. (in Japanese)

Iwasaki, Y. (2011). The meaning of learning support for exchange students : from talking about learning support by japanese tutors. *言語文化と日本語教育*, *40*, 20–29. (in Japanese)

Long, Y., Luo, H., & Zhang, Y. (2024). Evaluating large language models in analysing classroom dialogue. *npj Science of Learning*, 9, Article 60. https://doi.org/10.1038/s41539-024-00273-3