

Explicit Sentence-Strategy Instruction for L2 Japanese Writing Using the W/W Model

Norimichi **TAKEMURA**
Institute of Science Tokyo, Japan
takemura.n.3944@m.isct.ac.jp

Toshiki **MATSUDA**
Edogawa University, Japan
matsuda@et4te.org

Kentaro **NAGAHARA**
Institute of Science Tokyo, Japan
nagahara@ila.isct.ac.jp

For learners of Japanese as a second language (L2), revision requires selecting among alternatives under explicit task constraints, yet such decision-making is hard to observe in draft-to-draft studies. This paper examines a web-based role-play writing environment for composing 200-character self-promotion statements. In a teacher-accompanied prototype, learners repeatedly generated, compared, and selected alternatives, allowing revision to be examined as iterative cycles of alternative generation and rational judgment. Using the Warp and Woof (W/W) Model, we descriptively analyze four learners' sessions, focusing on externalized judgment criteria (C1–C4), revision-cycle occurrence, and the co-occurrence of sentence-level strategy use (S2–S4) and deterministic rule-based evaluation scores within cycles, as reusable design resources.

Keywords: L2 Japanese writing; revision cycles; game-based learning; Ways of Viewing and Thinking; Warp and Woof (W/W) Model

Introduction

For learners of Japanese as a second language (L2), revising written texts involves more than simply correcting linguistic forms. Revision requires generating multiple textual alternatives and making judgments under task-specific constraints. When learners lack explicit criteria for judgment, revision decisions tend to rely on trial and error, and their rationales are difficult to share.

To address this issue, the present study conceptualizes revision as a problem-solving process that consists of alternative generation and rational judgment. Drawing on the Warp and Woof (W/W) Model of problem solving, this study focuses on Ways of Viewing and Thinking (WVT)—decision-oriented criteria that guide learners' attention, strategy selection, and evaluation within a revision cycle.

The W/W Model is useful because it foregrounds decision criteria as well as procedural phases, making learners' revision choices understandable when task constraints are taken into account. Accordingly, in this teacher-accompanied prototype, we analyze revision cycles and examine how WVT are articulated and referenced when learners evaluate strategy options. We argue that the W/W Model suits L2 writing because it makes judgment criteria designable, observable, and discussable, enabling learners to compare options under constraints.

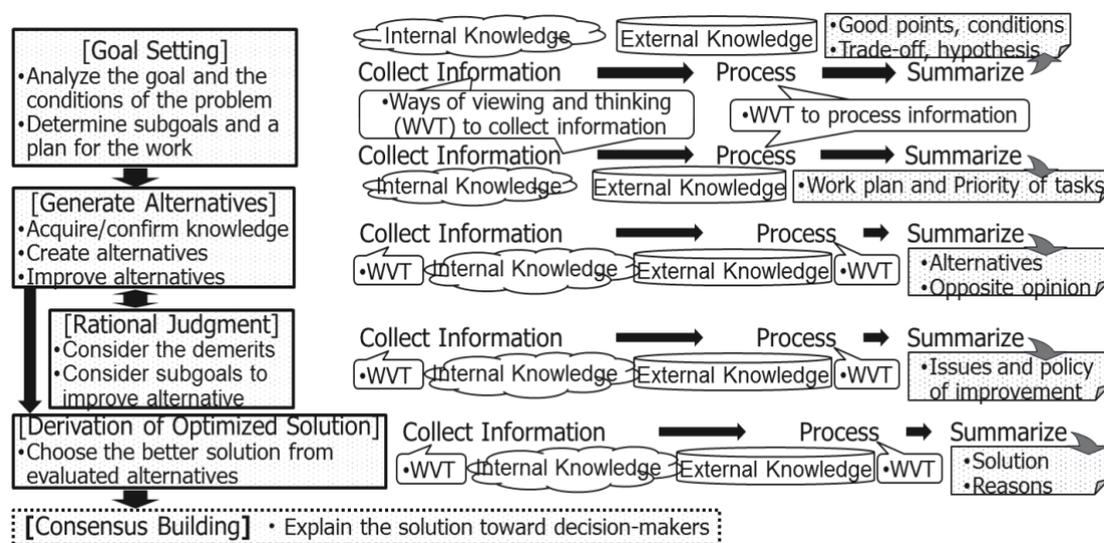
Theoretical Background

The W/W Model conceptualizes problem solving along a vertical thread (goal setting → alternative generation → rational judgment → solution selection) and a horizontal thread (collect → process → summarize) (Matsuda, 2015). A key feature of the model is its explicit emphasis on decision-oriented criteria that guide judgments in each phase, referred to here as Ways of Viewing and Thinking (WVT). Related discussions on instructional scaffolding, coaching, and game-based learning similarly emphasize supporting learners' decision making by structuring and problematizing what to attend to, rather than providing the “correct answer” (Reiser, 2004; Wood, Bruner, & Ross,

1976). This study draws on these perspectives and adopts the W/W Model as an integrative framework to examine the externalization of WVT during revision cycles.

Some scholarship suggests that, when games incorporate role enactment, constraints, and timely feedback, they can create recurring decision points—an affordance that allows researchers and teachers to revisit learners’ choices across cycles (Cornillie et al., 2012; Reinders & Wattana, 2015). Research on digital games for L2 learning also highlights the importance of designing tasks and supports that shape learners’ attention and choices during play, rather than merely adding motivational “game elements” (Reinhardt & Sykes, 2014; Peterson, 2016). Building on these discussions, the present study treats revision as recurring decision-making within a constrained role-play writing task and focuses on how judgment criteria are externalized as WVT and referenced when learners examine strategy options. This stance also aligns with learning-by-doing perspectives in goal-based scenarios (Schank, Berman, & Macpherson, 1999).

Figure 1.
The W/W Model of problem solving (Matsuda, 2015).



Focus of the Present Study

This study examines a web-based role-play writing game for composing a 200-character self-promotion statement for job-hunting in Japan, with particular attention to judgment processes within revision. At the time of implementation, some in-system judgment-support functions had not yet been implemented. Therefore, the teacher attended the sessions and temporarily substituted for these functions by verbally articulating task constraints, evaluation perspectives, and points of attention relevant to revision. Importantly, these utterances did not provide corrective rewrites or “the correct answer” but externalized criteria that learners could reference when generating, comparing, and selecting alternatives within a revision cycle. We treat this teacher-accompanied condition as a prototype form of judgment-support implementation and analyze (a) how such judgment-supporting utterances can be categorized as C1–C4 and interpreted as externalized Ways of Viewing and Thinking (WVT), and (b) how learners examined sentence-level strategies (S2–S4; claim–evidence integration, synonym replacement, and content trimming; see Table 4) within the same revision cycle under these externally articulated criteria.

Purpose and Research Questions

The purpose of this teacher-accompanied prototype design case was to externalize and organize the judgment support observed in a game-based L2 Japanese writing environment as a reusable design resource for future instructional implementation.

Specifically, teacher utterances that supported learners’ judgments were categorized as C1–C4. Revision cycles, defined as one iteration of alternative generation and rational judgment, were used as units of analysis. Strategy use (S2–S4; claim–evidence integration, synonym replacement, and content trimming) and changes in drafts were examined descriptively within the same revision cycle without making causal claims about learning outcomes.

RQ1: How were strategies S2–S4 used within the alternative generation and rational judgment phases with teacher-accompanied judgment support?

RQ2: How did strategy use and changes in rule-based writing evaluation scores co-occur within the same revision cycle?

RQ3: How did these strategies function as usable options for learners under explicit task constraints and with teacher articulation of judgment criteria?

Method

Design and Data

The instructional material examined in this study was a web-based role-play writing environment designed to support revision as a problem-solving activity. Its overall structure was informed by the W/W Model, which was used to foreground learners' judgment during revision. Learners composed a self-promotion statement for job-hunting in Japan under a strict 200-character constraint and repeatedly generated, compared, and selected textual alternatives within iterative cycles of alternative generation and rational judgment. Because some in-system judgment-support functions had not yet been implemented, a teacher-accompanied prototype condition was employed in which decision criteria were articulated verbally during revision to externalize Ways of Viewing and Thinking (WVT) (see Focus of the Present Study).

The role-play scenario asked learners to write as a job applicant describing their experience working at a convenience store in Japan. To make revision decisions observable and recordable, the system provided a bounded set of selectable topic prompts and pre-authored content elements presented as short fragments. Learners generated drafts by selecting, combining, paraphrasing, and deleting these elements under the 200-character constraint. This design created recurring decision points and explicit trade-offs among task constraints (e.g., character limits) and evaluation perspectives, thereby encouraging repeated movement between alternative generation and rational judgment.

From a game-design perspective, the task embedded (a) a role-play identity (job applicant), (b) a hard constraint as a challenge (200-character limit), and (c) immediate feedback through rule-based evaluation scores. No extrinsic rewards (e.g., points exchange) were implemented, and scores were used as informational feedback to support the comparison and reconsideration of alternatives. The bounded set of selectable content elements and the teacher's articulation of criteria served as scaffolding to keep the revision moves observable and repeatable within cycles.

The revised W/W Model used in this study is summarized in Figure 2. To support meaning confirmation while manipulating predefined content elements, the interface also provided L1 support alongside L2 prompts (Figure 3). The L1 support shown in Figure 3 was intentionally limited to meaning confirmation during information collection (collect). It was not designed to generate text on the learners' behalf or to determine strategy choices; rather, it functioned as auxiliary information that helped learners interpret and select content elements before transforming them into L2 drafts.

Figure 2.
Revised W/W Model used to structure the instructional material in this study.

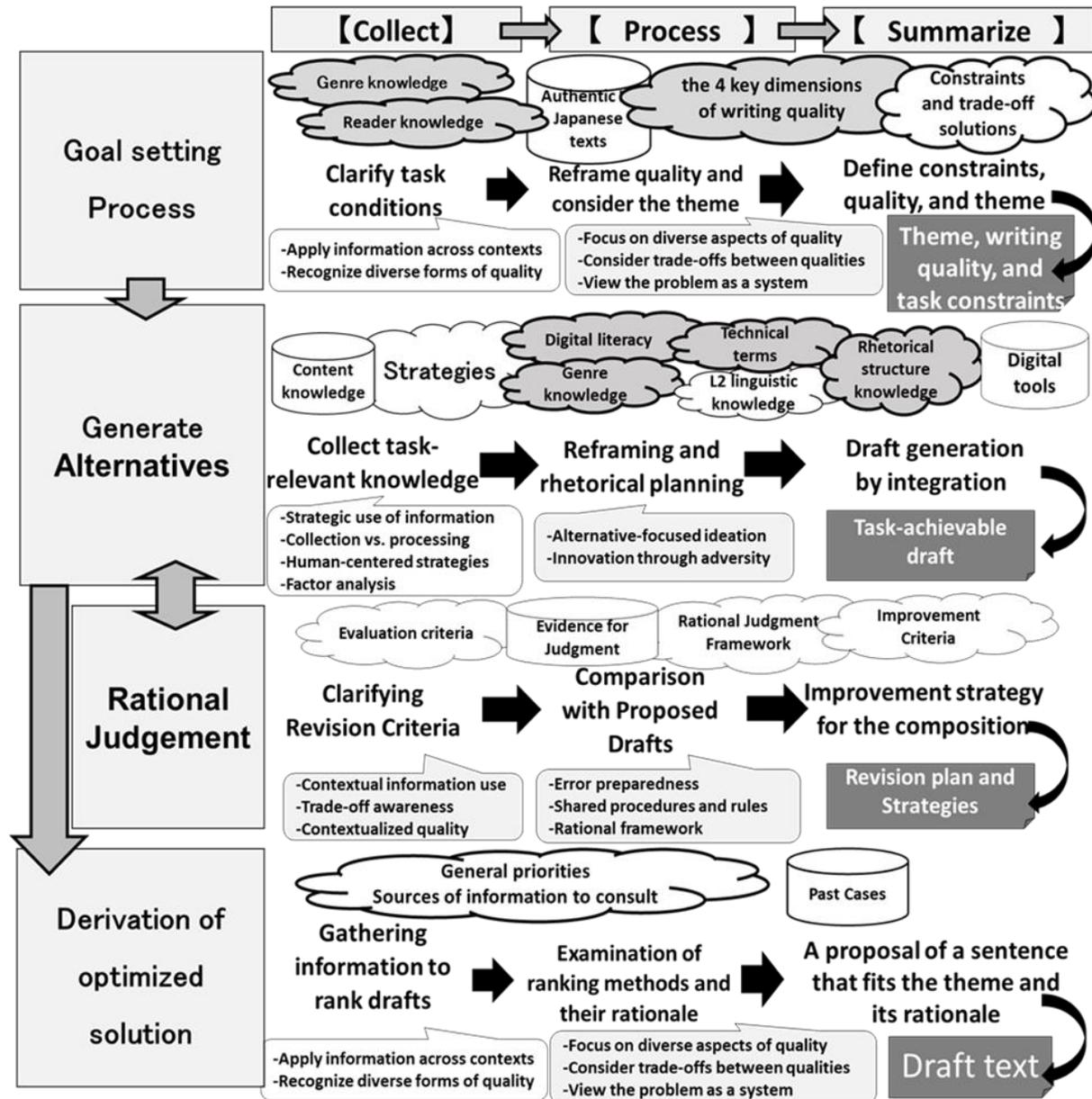
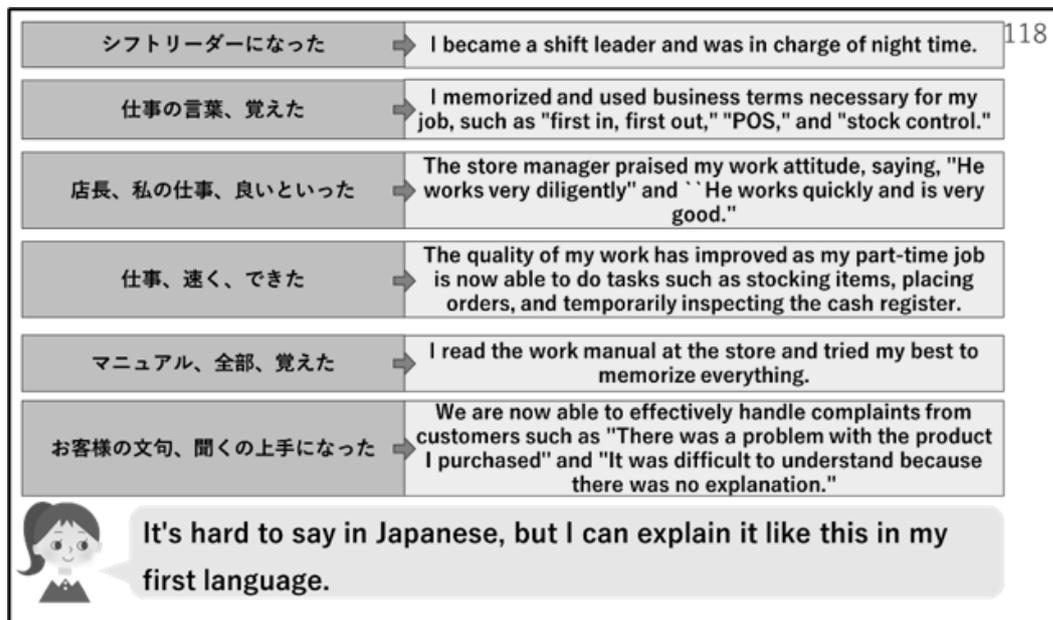


Figure 3.
Example screen presenting L2 content elements with L1 support.



Research Design and Participants

This study was conducted in July 2024 as a small-scale exploratory implementation of game-based instructional materials. The participants were four international students enrolled in a Japanese university, all of whom had studied Japanese prior to enrollment and had proficiency equivalent to JLPT N2 or N1. The implementation was designed as a process-oriented design case and was not intended for between-learner or between-group comparison.

Writing Task and Procedure

Participants engaged in the role-play task of composing a self-promotion statement for job-hunting in Japan within a 200-character limit. During the sessions, learners repeatedly generated and revised drafts while interacting with the instructional material. As explained earlier, because some judgment-support functions had not yet been implemented in the system, the teacher attended the sessions and provided limited verbal support by articulating task constraints, evaluation perspectives, and points of attention relevant to revision. Participant information is summarized in Table 1. Teacher prompts and the bounded content set were treated as part of the prototype condition rather than experimentally controlled factors.

Table 1
Study Participants

Participant	Country of Origin	Academic Year and Japanese Proficiency
A	China	Undergraduate, Year 3; JLPT N2 (passed)
B	Vietnam	Undergraduate, Year 4; JLPT N2 (passed)
C	Vietnam	Undergraduate, Year 4; JLPT N1 (passed)
D	China	Undergraduate, Year 3; JLPT N1 (passed)

Note. JLPT = Japanese Language Proficiency Test.

Unit of Analysis

The unit of analysis was a revision cycle, defined as one iteration of alternative generation and rational judgment. Cycles were counted when a feedback-check event was followed by new alternatives. Strategy use and draft changes were examined within cycles. All drafts were scored automatically using a deterministic, rule-based rubric aligned with the four writing-quality dimensions (Tables 2–3; Appendix A).

Coding of Teacher Utterances

Teacher utterances that plausibly supported learners' judgment and strategy selection were identified from the transcribed interaction data. Utterances solely related to session management or operational checks were excluded from the analysis. The remaining utterances were coded into four categories (C1–C4) based on their functional roles in supporting revision judgments.

All data were initially coded by the first author using a predefined codebook. Approximately 30% of the dataset was independently coded by a coauthor. Discrepancies were resolved through discussion with reference to the category definitions. These categories were treated as instances of WVT externalization and used as design resources rather than as indicators of instructional effects.

Of the 361 transcribed teacher utterances, 84 were excluded as session-management or operation check talk (OUT), leaving 277 utterances for C1–C4 coding.

Rule-Based Writing Evaluation (Deterministic Scoring)

Writing evaluation scores were determined using an in-system rule set aligned with the four dimensions of writing quality (Tables 2 and 3). Because identical inputs yielded identical outputs, no human raters were involved, and inter-rater reliability was not applicable. To avoid mechanically advantaging text lengths under the strict 200-character constraint, the scoring scheme distinguished between additive and normalized components: Novelty (NOV) and Effectiveness (EFF) were computed additively from fixed element/pattern triggers, whereas Credibility (content) (REL), Clarity/Organization (INT), and Credibility (form) (RELN) were normalized to a 0–20 scale based on non-blank selection events and truncated to integers (Appendix A). Appendix A summarizes the normalization procedure; full increment tables, pattern lists, and scripts are available from the authors upon reasonable request.

All drafts (from initial to final) were scored automatically by the system; the teacher did not evaluate or rate writing quality. This additive/normalized split was adopted to provide informative feedback while avoiding systematic score inflation owing to text length under the strict 200-character constraint.

The score ranges differed by design. NOV and EFF were implemented as additive indices (0–40) to reflect the accumulation of distinct content/pattern triggers, whereas REL/INT/RELN were normalized to a common 0-20 scale to make feedback comparable across drafts with different numbers of selected elements under the strict 200-character constraint.

Table 2
Original Dimensions of Writing Quality

Dimension	Content	Form
Novelty	Uniqueness of the claim	Uniqueness of writing style
Effectiveness	Claim effective for achieving the goal	Writing style effective for task achievement
Credibility	Alignment between claim and evidence	Appropriate choice of words and expressions for the task
Clarity	Clarity of the claim	Organization of the text (order of claim and evidence)

Note. Each dimension of writing quality is conceptualized in terms of both content- and form-related aspects and is used to guide learners' reflection and judgment during the writing process.

Table 3
Rearticulated Dimensions of Writing Quality for a Self-Promotion Task

	Content	Form
Novelty	An episode unique to the applicant	Uniqueness of writing style
Effectiveness	An episode that resonates with the recruiter	Writing style effective for task achievement
Credibility	The episode is supported by evidence	Appropriate choice of words and expressions for the task
Clarity	Clarity of the episode	Organization of content (order of claim and evidence)

Note. The expressions shown in this table represent task-specific paraphrases of the original writing quality dimensions, adapted to enhance learners' understanding in the context of a job-application self-promotion task.

Table 4 summarizes the writing strategies (S1–S4) introduced in the material, their related knowledge, target writing-quality dimensions, and expected length effects. Our analysis focuses on S2–S4.

Table 4
Writing Strategies Introduced in the Instructional Material and Their Alignment with Knowledge and Writing Quality Dimensions

Strategy for Applying Knowledge	Related Knowledge	Dimensions of Writing Quality Enhanced	Effect on Text Length
(S1) Selecting text drafts that possess qualities necessary to accomplish the task	Reader knowledge	Novelty, Effectiveness	No substantial change
(S2) Integrating a claim and its supporting evidence into a single sentence	Knowledge of text structure	Credibility (content), Clarity	Slight decrease
(S3) Searching for synonyms of words in use and replacing them appropriately	Genre knowledge; ICT-related knowledge	Credibility (form)	Slight increase
(S4) Retaining essential information while deleting irrelevant content	Knowledge of text structure	Credibility (content), Clarity	Substantial decrease

Note. The strategies are conceptualized as procedural operations situated within the alternative generation and rational judgment phases of the W/W Model.

Data Analysis

All analyses were descriptive and focused on identifying patterns of strategy use, judgment support, and draft changes within and across revision cycles. Strategy use (S2–S4) was identified directly from the system operation logs; each strategy button or action generated a time-stamped event record. Frequency was computed by counting these events within each revision cycle.

Ethical Considerations

Participants were informed of the study’s purpose and procedures in advance and their participation was voluntary. They were told that they could withdraw at any time, without penalty. All the data collected were anonymized and used solely for research purposes.

Results

This section reports the descriptive results observed during the implementation of the game-based writing environment structured around iterative revision cycles. The results were organized around teacher utterances (C1–C4), revision cycles, and patterns of strategy use with rule-based evaluation scores.

Teacher Utterances as Externalized WVT (C1–C4)

Across sessions, 277 teacher utterances were categorized as judgment-supporting moves (C1–C4; Table 5). C1 was most frequent, followed by C3, whereas C2 (evaluation criteria) and C4 (content expansion/reconsideration) were less frequent.

Table 5
Frequency of Teacher Judgment-Supporting Moves and Their Design Interpretation (C1–C4)

Category	Description of Teacher Utterances	Frequency	Primary Phase of the Problem-Solving Process	Corresponding Ways of Viewing and Thinking (WVT)
C1	Prompts directing attention to procedures and strategy use	112	Alternative generation (information collection)	Considering how to utilize available information; differentiating information collection from processing; reflecting on human judgment involved when using information technologies
C2	References to evaluation perspectives and judgment criteria	46	Rational judgment (information collection)	Attending to multiple dimensions of writing quality; recognizing trade-offs between different qualities; considering how required qualities vary by context and evaluator
C3	Prompts focusing on logical structure and organizational relationships	78	Alternative generation (information processing)	Decomposing problems into components and examining their relationships; generating ideas while being aware of multiple possible alternatives
C4	Suggestions encouraging content expansion or reconsideration	41	Rational judgment (information processing)	Preparing for unexpected issues or potential errors; clarifying procedures and shared rules as a framework for rational judgment
Total		277		

Note. Frequencies indicate the number of teacher utterances identified as providing judgment support during the instructional sessions.

The alignment with Ways of Viewing and Thinking (WVT) represents a design-oriented interpretation of how teacher utterances externalize judgment criteria and perspectives; it does not constitute a direct measurement of learners' internal cognitive processes.

Occurrence of Revision Cycles (Unit of Analysis)

Using revision cycles as the unit of analysis, we counted cycles for each learner. As shown in Table 6, all four learners engaged in multiple revision cycles during the instructional sessions. The total number of cycles varied across learners, indicating differences in how often they revisited alternative generation and judgment processes during revision. These results suggest that under the given instructional conditions, learners repeatedly engaged in iterative revision cycles rather than proceeding through revision in a linear manner.

Table 6
Number of Alternative Generation–Rational Judgment Cycles Repeated During the Implementation

Participant	Session 1	Session 2	Session 3	Session 4	Total Cycles
A	7	10	—	—	17
B	5	2	5	—	12
C	3	3	—	—	6
D	4	4	11	5	24

Note. A cycle refers to one iterative sequence of alternative generation and rational judgment as defined in the W/W Model. The number of cycles indicates the number of times learners revised their draft by returning to the alternative generation phase after rational judgment during the instructional sessions. Dashes indicate sessions not conducted.

Strategy Use Within the Same Revision Cycle

Strategy use was examined descriptively within the revision cycles, with a focus on S2–S4. As S1 operated as an interface-level prerequisite for constructing drafts from bounded content elements, it was not treated as a sentence-level revision operation and was excluded from the frequency analyses.

Table 7 summarizes the changes in rulebased writing evaluation scores, character count compliance, and strategy use across successive drafts for each learner. Across sessions, cycles sometimes involved repeated examination of multiple strategies alongside changes in evaluation scores.

Table 7
Changes in Writing Evaluation Scores, Strategy Use, and Character-Limit Achievement

Participant	Draft (Cycles)	Novelty	Effectiveness	Credibility (content)	Clarity/Organization	Credibility (form)	Character-limit Achievement (%)	Strategies Used (Frequency)		
								S2	S3	S4
A	Initial	10	19	15	10	4	–	–	–	–
	1st Final (7)	18	19	20	20	16	71%	0	17	0
	2nd Final (10)	33	28	20	20	20	60%	22	24	74
B	Initial	1	10	5	10	2	–	–	–	–
	1st Final (5)	15	20	10	20	17	40%	0	5	0
	2nd Final (2)	5	5	5	5	1	0%	0	0	0
	3rd Final (5)	20	28	15	15	16	60%	3	3	0
C	Initial	5	11	15	20	3	–	–	–	–
	1st Final (3)	5	20	5	10	3	33%	2	0	0
	2nd Final (3)	21	19	20	20	20	67%	4	7	0
D	Initial	16	5	15	12	5	–	–	–	–
	1st Final (4)	23	28	20	20	20	50%	0	0	15
	2nd Final (4)	18	28	20	20	20	50%	0	0	12
	3rd Final (11)	18	23	20	20	20	27%	25	11	48
	4th Final (5)	18	38	20	20	20	60%	9	7	6

Note. Writing evaluation scores were obtained using a deterministic rule-based scoring procedure embedded in the instructional material. NOV and EFF were computed additively, whereas REL/INT/RELN were normalized to a 0–20 scale and truncated to integers (see Appendix A for the full-rule specification). Character limit achievement indicates the proportion of drafts within the 200-character limit in each draft block.

Co-occurrence of Strategy Use and Rule-Based Evaluation Scores

To further examine patterns within the same revision cycle, the co-occurrence of strategy use and rule-based evaluation scores was analyzed descriptively. Evaluation scores were generated using identical scoring rules that were consistently applied across all drafts, allowing for within-case comparisons across cycles.

As shown in Tables 7 and 8, learners often considered evaluation perspectives and task constraints (e.g., the 200-character limit) when comparing options within cycles. Conversely, when only a few strategies were examined within a cycle, revision activity appeared more dispersed across drafts.

Transcripts also included instances in which learners explicitly referenced trade-offs between the 200-character constraint and evaluation perspectives when deciding whether to adopt a strategy, suggesting that externally articulated criteria were used as a frame for comparing options.

These patterns are presented as descriptive process characteristics, rather than as evidence of instructional effectiveness or learning gains.

Table 8
Frequency of Strategy Use

Strategy	A	B	C	D	Total
(S2) Integrating claims and supporting reasons into a single sentence	22	3	6	34	65
(S3) Searching for synonyms of words in use and replacing them accordingly	41	8	7	18	74
(S4) Retaining necessary information while deleting irrelevant content	74	0	0	81	155
Total frequency	137	11	13	133	294

Note. Table 7 aggregates strategy frequencies by draft blocks, whereas Table 8 reports totals across cycles; therefore, the totals are not expected to match the number of drafts.

The frequencies indicate the total number of times each strategy was applied by each learner across all alternative generation–rational judgment cycles. Multiple uses of the same strategy within counted as independent occurrences.

Summary of Results

In summary, the results indicate that (a) various types of judgment supporting teacher utterances were externalized as WVT during revision, (b) all learners engaged in multiple revision cycles defined as alternative generation and rational judgment, and (c) strategy use and rule-based evaluation scores co-occurred within these cycles in observable patterns. These findings provide a descriptive account of revision processes under a teacher-accompanied prototype condition for externalizing judgment support.

Discussion

This paper reports an exploratory, process-oriented design case of revision in a game-based L2 Japanese writing environment. With four participants, no control condition, and a teacher-accompanied prototype setting, the analyses were descriptive and did not support causal claims regarding effectiveness or learning gains. Accordingly, changes in strategy use and rule-based evaluation scores are interpreted as process indicators of decision making within revision cycles rather than outcomes attributable to the material.

First, the analysis of teacher utterances categorized as C1–C4 illustrates how different forms of Ways of Viewing and Thinking (WVT) were externalized through interaction during revision. These utterances did not function as corrective feedback or prescriptions for revision outcomes. Instead, they articulated criteria, perspectives, and points of attention that oriented learners' judgment during alternative generation and rational judgment. From a design perspective, the repeated appearance of these utterances indicates which judgment-oriented perspectives tended to become salient during revision under the given task constraints.

Second, treating the same revision cycle, defined as one iteration of alternative generation and rational judgment, as the unit of analysis proved useful for capturing revision as a recurrent decision-making process rather than as a linear sequence of drafts. Learners repeatedly returned to earlier phases to generate, compare, and evaluate alternatives under

stable constraints, such as the 200-character limit. This finding highlights the analytical value of focusing on cycles to examine how judgment processes unfold over time.

Third, sentence-level strategies (S2–S4) functioned as selectable options within these revision cycles, rather than as isolated techniques applied once per draft. The results show that multiple strategies were often examined within a single revision cycle, as learners considered trade-offs between writing quality dimensions and task constraints. These patterns illustrate how strategies were positioned within judgment-oriented revision processes and how learners navigated decision points during revision.

Taken together, these findings suggest that the design significance of sentence-level strategies lies in their availability as options that learners can select, compare, and reconsider within recurring revision cycles. The teacher-accompanied setting examined in this study functioned as a prototype environment in which judgment support was externalized through verbal articulation rather than being embedded within the instructional system. From an instructional design perspective, the results highlight the importance of embedding judgment-support functions that foreground evaluation perspectives, prompt strategy consideration, and support the comparison of alternatives without directing learners toward predetermined revisions.

In summary, this study provides design-relevant insights into how WVT can be articulated and structured in L2 writing environments. By externalizing judgment criteria aligned with the phases of the same revision cycle, instructional systems may better support learners' autonomous decision making during revision.

Conclusion

This study examined revision processes in L2 Japanese writing in a game-based instructional environment structured around the W/W Model. Focusing on revision as an iterative process of alternative generation and rational judgment, the study analyzed how judgment-oriented criteria were externalized during revision and how sentence-level strategies were examined within the same revision cycle.

The primary contribution of this study lies in identifying judgment support externalized through teacher utterances, categorized as C1–C4, and interpreting them as Ways of Viewing and Thinking (WVT). These categories capture how criteria, perspectives, and points of attention are articulated during revision to support learners' judgment rather than to prescribe specific revisions.

A second contribution is demonstrating the analytical utility of the same revision cycle as a unit of analysis. By examining strategy use and evaluation indicators within recurring cycles under stable task constraints, the study provides a descriptive account of judgment-oriented revision processes without making causal claims about learning outcomes.

Taken together, these findings suggest that the design value of sentence-level strategies lies not in their isolated use but in their availability as selectable options within recurring revision cycles. By articulating judgment criteria aligned with the phases of the revision cycle, instructional systems may better enable learners' autonomous decision making during revision.

Limitations and Future Directions

This study has several limitations. Factors shaping the observed processes include the teacher's real-time prompts (teacher-as-proxy), participants' prior familiarity with job-application genres, differences in session length, and the bounded set of selectable content elements that constrained available options. Because this study is descriptive and process-oriented, we treated these factors as part of the prototype condition rather than attempting to isolate causal effects.

The study was conducted as an exploratory process-oriented design case involving a small number of participants. The limited sample size precludes statistical analysis and does not allow generalization of the findings to L2 Japanese learners as a population. Accordingly, the results should be interpreted as descriptive accounts of revision processes observed under specific instructional conditions and not as evidence of instructional effectiveness.

In addition, the instructional environment examined in this study included teacher-accompanied judgment support. Because some judgment-support functions had not yet been implemented in the system, teacher utterances served to externalize Ways of Viewing and Thinking (WVT) during revision. Consequently, the findings do not represent the

effects of instructional materials alone, nor do they isolate the influence of individual types of support. Instead, this study documents a teacher-accompanied condition in which judgment support was externalized through interaction, with the aim of informing future system designs.

A control condition was not included because the primary goal was to elicit task requirements for judgment support functions at a prototype stage. Future work should compare teacher-as-proxy support, embedded in-system prompts based on C1–C4, and minimal-support conditions to examine how different support configurations shape strategy examination within revision cycles.

Finally, writing evaluation scores were generated using a rule-based procedure and treated as process indicators rather than as measures of learning outcomes. Differences in scoring ranges across the evaluation dimensions reflect the design choices made to accommodate strict task constraints such as the character limit. Therefore, these scores should not be interpreted as indicators of overall writing proficiency or improvement.

Given these limitations, the scope for transferability in this study lies not in learner outcomes but in the identification of design resources for judgment support. Specifically, the categories of externalized judgment support (C1–C4), their alignment with the phases of the same revision cycle, and their interpretation as WVT constitute transferable insights for the design of instructional systems.

Future research should implement these judgment-support functions directly within instructional systems and examine how learners engage with them without teacher accompaniment. Studies with larger participant groups and varied writing tasks are necessary to further explore how judgment-oriented revision processes operate across contexts.

References

- Cornillie, F., Clarebout, G., & Desmet, P. (2012). The role of feedback in foreign language learning through digital role-playing games. *Procedia - Social and Behavioral Sciences*, 34, 49–53. <https://doi.org/10.1016/j.sbspro.2012.02.011>
- Matsuda, T. (2015, January). Design framework of gaming materials to cultivate problem-solving abilities: Differences and commonalities among STEM educations. In *Proceedings of the 13th Hawaii International Conference on Education* (pp. 2147–2159).
- Peterson, M. (2016). The use of massively multiplayer online role-playing games in CALL: An analysis of research. *Computer Assisted Language Learning*, 29(7), 1181–1194. <https://doi.org/10.1080/09588221.2016.1197949>
- Reinders, H., & Wattana, S. (2015). Affect and willingness to communicate in digital game-based learning. *ReCALL*, 27(1), 38–57. <https://doi.org/10.1017/S0958344014000226>
- Reinhardt, J., & Sykes, J. M. (2014). Guest editor commentary. *Language Learning & Technology*, 18(2), 2–8. Retrieved from <https://www.lltjournal.org/item/10125-44362/>
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2
- Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 161–181). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

Appendix A. Deterministic Rule-Based Writing Evaluation

Scoring is deterministic: identical inputs yield identical outputs; no human raters were involved (IRR not applicable). For each non-blank selection event, R-CNT += 5 (thus R-CNT = 5 × non-blank events; if R-CNT = 0, REL/INT/RELN = 0). Raw accumulators (REL2/INT2/RELN2) are updated per non-blank event (0–5): REL2 and INT2 add 5 when the ordered element list matches predefined claim–evidence or organization patterns (else 0); RELN2 adds 5/3/1 depending on best/standard/other paraphrase-group matches. Bounded subscores are REL = [(REL2/R-CNT)×20], INT = [(INT2/R-CNT)×20], RELN = [(RELN2/R-CNT)×20]. NOV/EFF are additive sums of fixed increments triggered by derived element/pattern codes; full tables/lists/scripts are available from the authors upon reasonable request.