

Research Article



Rethinking L2 Japanese Writing Assessment: Comparing Holistic and Analytic Rating among Persian-Speaking Learners

Elham Saberi^{1,2*} 

1. Kagamiyama, Higashi-Hiroshima, Hiroshima, Japan, Graduate School of Humanities and Social Sciences, Department of Teaching Japanese as a Second Language Program, Hiroshima University, Japan
2. Kagamiyama, Higashi- Hiroshima, Hiroshima, Network for Education and Research on Peace and Sustainability (NERPS), Hiroshima University, Japan

* **Corresponding author:** Elham Saberi, Graduate School of Humanities and Social Sciences, Hiroshima University, Japan. Email: elhamsaberi@hiroshima-u.ac.jp

ARTICLE INFO**Article History:**

Received: 27/10/2025

Revised: 18/11/2025

Accepted: 09/12/2025

Published: 26/12/2025

**Keywords:**

Analytic scoring

Holistic scoring

Japanese L2 writing

Vocabulary

Writing assessment

ABSTRACT

Introduction: Writing assessment in Japanese as a second language, has traditionally relied on holistic approaches, yet questions remain about whether analytic methods provide more diagnostic information for diverse learner populations. This study aimed to compare holistic and analytic assessment methods in evaluating Japanese compositions written by Persian-speaking learners, examining which linguistic components best predicted overall writing quality.

Methodology: Thirty-six Persian-speaking Japanese learners at intermediate proficiency level enrolled at the University of Tehran wrote 400-character compositions on assigned topics. Participants included students from both the Department of East Asian Languages and Literatures and the Japanese Language Institute Public Program. Four experienced Japanese language instructors evaluated each composition using both holistic and analytic rubrics. The holistic rubric employed a single-score approach, while the analytic rubric assessed five components: vocabulary, grammar, content, organization, and mechanics. The study conducted Pearson correlation analyses and multiple regression analyses to examine relationships between scoring methods and identify predictive components.

Results: Analysis revealed a strong correlation between holistic and analytic total scores), validating both assessment approaches. Among analytic components, vocabulary emerged as the strongest predictor of holistic scores, followed by content and grammar. Organization and mechanics demonstrated non-significant predictive power. Learner surveys indicated that 65% of participants preferred analytic assessment due to clearer diagnostic feedback for improvement.

Conclusion: Vocabulary proficiency served as the primary driver of perceived writing quality among Persian-speaking Japanese learners, challenging traditional grammar-focused pedagogies. Analytic assessment methods provided superior diagnostic value for non-Kanji countries in second language Japanese contexts, offering clearer pathways for instructional intervention and learner development.

1. Introduction

The assessment of second language (L2) writing has long been central to language education, yet significant debates persist regarding optimal

evaluation methods (Weigle, 2002). Within Japanese language education, writing assessment practices have predominantly emerged from contexts

► *Cite this paper as:* Saberi E. Rethinking L2 Japanese Writing Assessment: Comparing Holistic and Analytic Rating among Persian-Speaking Learners. 2025; 4(4):47-55. DOI: 10.58803/jclr.v4i4.141



The Author(s). Published by Rovedar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

involving learners from Chinese language backgrounds, potentially limiting their applicability to linguistically distant populations (Ishibashi, 2012; Wakita, 2016). Persian-speaking learners of Japanese represent one such understudied group, facing unique challenges stemming from fundamental typological differences between Persian (Indo-European) and Japanese (Japonic language isolate). Recent demographic shifts underscore the urgency of validating assessment practices across diverse learner populations: between 2008 and 2013, Vietnamese student enrollment in Japanese language schools increased 14-fold while Nepali enrollment increased 6-fold, as Korean student numbers, learners with a Kanji background, decreased to one-quarter of previous levels (Shimada, 2014). This dramatic shift has exposed fundamental questions about whether assessment instruments and pedagogical priorities developed for Kanji-background learners function appropriately for non-Kanji populations.

Writing assessment typically employs two primary approaches: holistic scoring, which assigns a single score based on overall impression, and analytic scoring, which evaluates specific linguistic components separately (Yamanishi et al., 2019). While holistic methods offer efficiency and practicality for large-scale testing (Lumley, 2005; Ono et al., 2019), analytic approaches promise greater diagnostic utility by identifying specific areas requiring instructional intervention (Yang et al., 2025). However, empirical comparisons of these methods within L2 Japanese contexts remain scarce, particularly for non-Asian learner populations.

Recent scholars have increasingly recognized vocabulary knowledge as a critical predictor of L2 writing quality across languages (Kyle & Crossley, 2015; Maamujav, 2021). Studies examining English L2 writing demonstrate that lexical sophistication and diversity correlate more strongly with holistic scores than traditional grammar-focused metrics. Yang et al. (2025) found that vocabulary-related features significantly outperformed grammatical accuracy in predicting writing quality among Chinese learners of English. Yet whether these patterns extend to Japanese, particularly for learners from non-Kanji backgrounds who lack orthographic support and face substantial vocabulary acquisition challenges—remains an empirical question with direct pedagogical implications. Shimada (2014) documented widespread confusion among Japanese language educators confronting non-Kanji learners, noting that traditional "grammar-focused, knowledge-injection" teaching methods often failed to produce communicatively proficient language users despite examination success. If instructional emphasis has historically overweighted grammatical accuracy while underweighting vocabulary breadth,

and if assessment rubrics reflect this imbalance, a misalignment may exist between pedagogical priorities and features that predict writing quality for non-Kanji learners.

While Japanese writing assessment research has examined various methodological approaches, including multi-trait rubric frameworks that assess purpose/content, organization/cohesion, audience awareness, and language accuracy/appropriateness (Tanaka et al., 2009), and holistic versus analytic rating comparisons (Yamanishi et al., 2019; Ono et al., 2019), systematic investigations of which specific linguistic features most strongly predict writing quality for non-Kanji learners remain limited. Tanaka et al. (2009) noted that the absence of shared evaluation criteria in Japanese language education has resulted in assessment being largely left to individual teachers' judgment. Their workshop with 20 Japanese language teachers using a five-trait rubric (rated on a 0-6 scale) demonstrated overall reliability, yet questions about linguistic predictors of quality, particularly across diverse learner populations, persist. Furthermore, Hōsei (2018) observed that while writing assessment plays a significant role in proficiency testing and instruction, research comparing assessments across rater backgrounds remains scarce, despite evidence in other foreign language contexts showing differences based on raters' native language. Shimada (2014) argued that the rapid increase in non-Kanji learners exposed not inadequacies in the learners themselves, but rather deep-seated problems in traditional Japanese language pedagogy and assessment, problems that had been masked when student populations were predominantly from Kanji-using backgrounds. This observation underscores the need for empirical validation of assessment practices across linguistically diverse populations.

The present study addresses these gaps by examining holistic and analytic assessment approaches for compositions written by Persian-speaking Japanese learners at the University of Tehran, contributing to Japanese language education scholarship in four significant ways.

First, it provides the first systematic comparison of holistic and analytic assessment methods specifically for Persian-speaking learners of Japanese, extending validation evidence beyond East Asian contexts where most research has been conducted. This addresses Shimada's (2014) call for understanding diverse learner populations by testing whether assessment instruments function equivalently across maximal linguistic distance. Traditional analytic rubrics that separately weight mechanics (including Kanji/orthography) alongside vocabulary, grammar, and content may inadvertently disadvantage non-Kanji learners whose vocabulary

knowledge and communicative competence exceed their orthographic automaticity. Understanding whether holistic and analytic approaches converge for Persian speakers, or whether systematic differences emerge, provides construct validity evidence essential for fair assessment practices.

Second, it empirically investigates which linguistic features (vocabulary, grammar, content, organization, mechanics) most strongly predict holistic writing quality judgments for learners from a linguistically distant background. This directly responds to the pedagogical challenge Shimada (2014) identified: educators report uncertainty about which features to prioritize when teaching non-Kanji learners. If vocabulary emerges as the strongest predictor despite Persian-speaking learners lacking cognate support and facing substantial lexical acquisition challenges, it suggests that vocabulary-focused instruction should receive greater curricular emphasis, even for (or especially for) linguistically distant learners. Conversely, if grammar proves to be the most predictive, it will validate traditional grammar-focused approaches. By providing empirical evidence about feature importance, this study offers data-driven guidance for curriculum design that addresses documented educator uncertainty.

Third, it examines whether vocabulary's documented primacy in English L2 writing (Yang et al., 2025) extends to Japanese for learners without orthographic or semantic transfer advantages. Persian-Japanese represents a critical test case: with no shared Kanji, no cognate vocabulary, and contrasting typological structures (SVO vs. SOV, alphabetic vs. logographic-syllabic), Persian speakers must acquire every lexical item explicitly without L1 support. If vocabulary nevertheless outperforms grammar in predicting quality, it would provide compelling cross-linguistic evidence for prioritizing lexical development across language pairs and learner backgrounds, extending theoretical understanding of L2 writing proficiency beyond European language contexts.

Fourth, it integrates learner perspectives on assessment methods and writing challenges, providing ecological validity by examining whether empirical patterns align with learner metacognitive awareness. This mixed-methods approach responds to five. Hōsei's (2018) advocacy for research incorporating learner perspectives and Shimada's (2014) emphasis on learner-centered pedagogy. If learners identify vocabulary as their primary challenge and empirical evidence confirms vocabulary as the strongest predictor, instructional alignment exists. However, if learners perceive grammar as most important while data show vocabulary as most predictive, a metacognitive

calibration gap exists, one that could inform both teaching practices and learner strategy training.

From a practical standpoint, this research directly informs curriculum development and assessment design for the growing population of non-Asian Japanese language learners. As Japanese language education expands globally beyond traditional East Asian markets, programs increasingly serve learners from Middle Eastern, African, and other linguistically distant backgrounds. Understanding whether assessment rubrics developed primarily for Chinese-background learners function equivalently when applied to Persian speakers—or whether they require modification to capture salient features of linguistically distant learners' writing—has immediate implications for fair and effective evaluation practices. Additionally, identifying which linguistic features most strongly predict writing quality for Persian speakers can guide curricular priorities, addressing the documented pedagogical confusion that Shimada (2014) observed: educators uncertain about "how to teach" and "what to prioritize" for non-Kanji learners now have empirical evidence to inform instructional decisions.

The Persian-Japanese language pair offers a particularly valuable context for examining these questions due to maximal linguistic distance across multiple dimensions. Unlike Chinese-Japanese pairs, where shared Kanji facilitate lexical acquisition, or Korean-Japanese pairs where similar grammar structures enable syntactic transfer, Persian and Japanese share virtually no cognate vocabulary, employ entirely different writing systems (alphabetic versus mixed logographic-syllabic), and exhibit contrasting syntactic structures (SVO versus SOV word order, prepositions versus postpositions). This linguistic distance potentially magnifies the role of specific competencies, such as vocabulary breadth, that cannot be supplemented through L1 transfer, making Persian-speaking learners an ideal population for isolating the relative importance of different linguistic components in writing assessment.

In this regard, this study aims to address the following five research questions:

1. What is the relationship between holistic and analytic assessment scores for L2 Japanese compositions?
2. Which analytic components (vocabulary, grammar, content, organization, mechanics) best predict holistic scores?
3. How do learners perceive the relative importance of vocabulary versus grammar for writing quality?
4. Which assessment method do learners prefer,

and why?

5. What factors do learners identify as primary obstacles to writing proficiency?

Recent scholars have increasingly recognized vocabulary knowledge as a critical predictor of L2 writing quality across languages (Kyle & Crossley, 2015; Maamuujav, 2021). Studies examining English L2 writing demonstrate that lexical sophistication and diversity correlate more strongly with holistic scores than traditional grammar-focused metrics. Yang et al. (2025) found that vocabulary-related features significantly outperformed grammatical accuracy in predicting writing quality among Chinese learners of English. Yet whether these patterns extend to Japanese, a language with distinct orthographic, morphological, and syntactic characteristics, remains an empirical question.

While Japanese writing assessment research has examined various methodological approaches, including multi-trait rubric frameworks that assess purpose/content, organization/cohesion, audience awareness, and language accuracy/appropriateness (Tanaka et al., 2009), and holistic versus analytic rating comparisons (Yamanishi et al., 2019; Ono et al., 2019), systematic investigations of which specific linguistic features most strongly predict writing ability of learners.

2. Methodology

2.1. Participants

Thirty-six Persian-speaking Japanese learners from the University of Tehran participated in this study. The sample comprised 22 undergraduate students (12 third-year, 10 fourth-year) from the Department of East Asian Languages and Literatures and 14 students from the Japanese Language Institute Public Program. All participants were native Persian speakers learning the Japanese language, aged 20-35 years, with no prior residence in Japan.

Proficiency levels ranged from JLPT N3 to JLPT N2, representing intermediate to upper-intermediate competence. The University of Tehran's curriculum targets N2-level proficiency by graduation, with participants having completed foundational grammar instruction and introductory composition courses (Composition book I and Composition book II). This proficiency range was deliberately selected to capture learners who possessed sufficient linguistic resources to produce extended discourse while still exhibiting developmental variation in their writing abilities.

2.2. Writing task

Participants completed a timed composition (50 minutes) responding to the prompt: "If you were wealthy, what would you do?" The hypothetical conditional format was chosen to elicit both linguistic complexity and personal elaboration. Compositions were limited to 400 characters on standard Japanese horizontal writing paper, with participants instructed to use plain form (da/dearu style) for stylistic consistency, though polite form (desu/masu style) was permitted.

This length was selected to balance two considerations: providing sufficient text for reliable assessment while remaining achievable within the time constraint for intermediate learners. The prompt's open-ended nature minimized content-knowledge barriers, focusing evaluation on linguistic rather than topical expertise.

2.3. Writing scoring methods

Two complementary scoring methods were employed, adapted from internationally recognized frameworks to suit the L2 Japanese context:

Writing performance was assessed using both holistic and analytic scoring approaches. For holistic scoring, a 20-point scale adapted from the TOEFL iBT writing rubrics was used, focusing on overall communicative effectiveness and raters' global impressions of writing quality; each composition received a single integrated score reflecting its overall success. For analytic scoring, a five-component rubric adapted from IELTS writing assessment criteria was employed to evaluate specific dimensions of writing, including vocabulary (20 points; range, appropriateness, and sophistication), grammar (31 points; accuracy, complexity, and structural variety), content (30 points; idea development, relevance, and depth), organization (10 points; logical structure, coherence, and paragraph unity), and mechanics (12 points; orthography, punctuation, and formatting). Component scores (totaling 103 points) were standardized to a 4-point scale per component (20 points total) for comparison with holistic scores.

2.4. Raters

Four experienced Japanese language instructors from the Department of East Asian Languages and Literatures at the University of Tehran served as raters. All four held advanced degrees in Japanese language education and had 3-8 years of teaching experience. To ensure reliability, raters participated in a calibration session, where they in

reviewed scoring guidelines and anchor compositions. Inter-rater reliability, calculated using Pearson correlations across all rater pairs, ranged from $r = 0.432$ to $r = 0.872$ ($p < .01$), indicating fair to excellent agreement (Cicchetti, 1994). Raters evaluated all 36 compositions using both methods, with holistic scoring conducted first to capture immediate impressions, followed by analytic scoring three weeks later to minimize carryover effects.

2.5. Learners' questionnaire

Following the composition submission, participants completed a researcher-developed structured questionnaire examining their perspectives on Japanese writing assessment and composition challenges. The questionnaire was developed based on an extensive literature review and consultation with Japanese language education experts and applied linguistics scholars in both Japan and Iran. Given the absence of validated instruments for Persian-speaking learners of Japanese, an understudied population, the questionnaire was designed specifically for this context.

The questionnaire consisted of both closed-ended and open-ended items assessing: perceived importance of vocabulary versus grammar for writing quality, preferred assessment method (holistic vs. analytic) with written justification, identified writing challenges using a 7-item checklist encompassing linguistic elements, instructional factors, and metacognitive aspects, and strategies for improving writing ability using a 4-item checklist.

To establish content validity, the questionnaire underwent expert review by the researcher's supervisors at the University of Tehran, Japanese language education scholars in Japan, and experienced instructors from both the Department of East Asian Languages and Literatures and the Japanese Language Institute Public Program. The instrument was pilot-tested with a small group of Persian-speaking Japanese language learners ($n = 5$) to ensure clarity and comprehensibility before administration to the full participant sample.

The questionnaire was administered in both Persian and Japanese immediately following composition submission, allowing participants to respond in their most comfortable language to maximize comprehension and response quality. While the questionnaire did not undergo psychometric validation (e.g., reliability testing via Cronbach's alpha), content validity was established through the iterative expert review and pilot testing process.

2.6. Data analysis

Quantitative data analysis was conducted using SPSS 20.0. To address the first research question, Pearson correlation coefficients were examined to determine relationships between holistic scores, analytic total scores, and individual analytic components (vocabulary, grammar, content, organization, mechanics). To address the second research question, multiple regression analysis using the enter method determined which analytic components predicted holistic scores. Descriptive statistics, including frequencies and percentages, were used to summarize learner questionnaire responses addressing research questions three through five. P-value was set at .05 for all statistical tests.

3. Results

To address the first research question examining the convergence between holistic and analytic assessment methods, Table 1 presents correlations between holistic scores, analytic total scores, and individual analytic components.

Table 1 presents correlations between holistic scores, analytic total scores, and individual analytic components.

Table 1.
Correlations between Holistic Scores and Analytic Components

Component	Holistic Score	r	p
Analytic Total		0.80	<.001
Vocabulary		0.82	<.001
Content		0.75	<.001
Grammar		0.64	<.001
Mechanics		0.58	<.001
Organization		0.36	<.01

The holistic and analytic total scores demonstrated a strong positive correlation ($r = 0.80$, $p < .001$), indicating substantial construct overlap between assessment methods. This convergent validity evidence suggests both approaches measure a largely equivalent underlying construct of writing proficiency, despite methodological differences.

Among individual components, vocabulary exhibited the strongest correlation with holistic scores ($r = 0.82$), followed by content ($r = 0.75$), grammar ($r = 0.64$), and mechanics ($r = 0.58$). The organization showed the weakest relationship ($r = 0.36$), though still statistically significant.

To address the second research question, investigating which linguistic features most strongly predict holistic writing quality, multiple regression analysis examined whether analytic components collectively predicted holistic scores. The model demonstrated excellent fit ($R = 0.875$, $R^2 = 0.765$, $p < .001$), explaining 76.5% of the variance in holistic scores.

Table 2.
Multiple Regression Analysis: Analytic Components Predicting Holistic Scores

Component	β	t	p	Tolerance
Vocabulary	0.526	8.182	<.001	0.411
Content	0.297	4.655	<.001	0.417
Grammar	0.116	2.086	.039	0.549
Mechanics	0.103	1.978	.050	0.625
Organization	-0.077	-1.606	.111	0.749

Vocabulary emerged as the dominant predictor ($\beta=0.526$), contributing more than twice the explanatory power of content ($\beta=0.297$), the second-strongest predictor. Grammar showed modest predictive value ($\beta=0.116$, $p=.039$). Notably, the organization failed to reach significance ($p=.111$), and mechanics demonstrated marginal significance ($p=.050$).

To address the third research question concerning learner perceptions of which linguistic element most influences writing quality, when asked which linguistic element most influences writing quality, 54.9% of participants ($n = 20/36$) identified vocabulary, compared to 45.1% ($n = 16/36$) selecting grammar. This distribution closely parallels the regression findings, suggesting learner intuitions align with empirical patterns.

To address the fourth research question, examining learner preferences for assessment methods, participants demonstrated a clear preference for analytic assessment: 65% ($n = 23/36$) favored analytic methods, versus 35% ($n = 13/36$) preferring holistic approaches. Qualitative justifications emphasized the analytic assessment's diagnostic clarity, enabling targeted improvement efforts.

To address the fifth research question regarding identified writing challenges and preferred improvement strategies, among seven listed challenges, "inability to correctly use linguistic elements (grammar, vocabulary, organization, content, mechanics) when expressing thoughts" received the highest endorsement (45%), followed by "inability to organize ideas" (40%) and "inability to utilize knowledge and experience accurately" (35%). For improvement strategies, 60% of participants selected "attention to the writing process (grammar, vocabulary, organization, content, mechanics)," substantially exceeding other options, including instructional methods (15%) or enrichment activities like reading and film (35%).

4. Discussion

4.1. Vocabulary as Primary Predictor of Writing Quality

The finding that vocabulary constitutes the strongest predictor of holistic scores ($\beta = 0.526$)

carries significant implications for L2 Japanese pedagogy, particularly for linguistically distant learner populations. This pattern aligns with recent cross-linguistic research demonstrating lexical knowledge's primacy in writing assessment (Kyle & Crossley, 2015; Yang et al., 2025), extending these findings to the Japanese context.

Several factors may explain the dominance of Persian-speaking learners. First, Japanese and Persian share minimal cognate vocabulary, unlike European language pairs, where lexical transfer facilitates acquisition. Persian speakers must therefore construct their Japanese lexicon largely from scratch, making lexical breadth and depth particularly salient markers of proficiency.

Second, Japanese orthography intertwines vocabulary knowledge with character recognition. Kanji usage not only signals lexical sophistication but also demonstrates orthographic competence, a dimension captured by raters' holistic impressions even without explicit orthographic criteria. This orthographic-lexical entanglement may amplify vocabulary's apparent importance relative to alphabetical languages.

Third, the IELTS-based analytic rubric employed here operationalizes vocabulary broadly, encompassing range, sophistication, and appropriateness. This multidimensional conceptualization may capture aspects of linguistic maturity that grammar rubrics—often focused narrowly on accuracy—do not. Recent computational linguistics research similarly finds lexical diversity and sophistication outperform syntactic complexity measures in predicting human writing quality judgments (Crossley, 2020; Maamujav, 2021).

These findings challenge traditional Iranian Japanese language curricula, which have historically prioritized grammatical accuracy over lexical development. While morphosyntactic precision remains important ($\beta = 0.116$, $p = .039$), its predictive power pales beside vocabulary, suggesting potential curricular realignment toward vocabulary-enriched instruction.

4.2. Content's Secondary Role

Content emerged as the second-strongest predictor ($\beta = 0.297$), indicating that ideational quality substantially influences perceived writing proficiency beyond mere linguistic accuracy. This finding complicates purely formalistic approaches to L2 writing assessment, which risk undervaluing communicative substance.

The moderate correlation between content and holistic scores suggests raters integrate both

"what" (message) and "how" (linguistic form) dimensions when forming global impressions. For learners at intermediate proficiency levels, where linguistic resources constrain expression, effective content development may depend on strategic vocabulary deployment, potentially explaining the high intercorrelation between vocabulary and content scores.

4.3. Organization's Minimal Predictive Power

The organization's failure to predict holistic scores ($\beta = -0.077$, $p = .111$) warrants careful interpretation. Two explanations merit consideration. First, intermediate-level compositions may exhibit insufficient structural variation to differentiate writers. If most learners produce similarly structured responses (e.g., simple three-paragraph essays), organization becomes a non-discriminating variable.

Second, Persian and Japanese rhetorical conventions differ substantially. Japanese writing often favors implicit, reader-responsible organization (kishōtenketsu pattern), while Persian rhetoric, influenced by Arabic traditions, employs more explicit structural markers. Raters may have unconsciously applied culturally specific organizational expectations, potentially disadvantaging learners navigating between competing rhetorical frameworks (Kubota, 2020). This interpretation aligns with contrastive rhetoric research demonstrating cultural variation in preferred discourse structures (Connor, 2002).

Alternatively, organization's non-significance may reflect measurement issues. The 10-point organizational rubric employed here may lack sufficient granularity to capture subtle structural differences, compared to the 30-point content and 31-point grammar scales.

4.4. Holistic vs. Analytic Assessment: Pedagogical Implications

The strong correlation between holistic and analytic total scores ($r = 0.80$) demonstrates both methods' construct validity, validating their use for different purposes. However, learners' preference for analytic assessment (65%) highlights a critical pedagogical consideration: formative feedback requires diagnostic specificity.

Holistic scoring excels in efficiency and reliability for summative purposes (Ono et al., 2019), making it suitable for high-stakes placement or certification testing. Yet for classroom assessment—where the goal is instructional improvement—analytic methods' component-level feedback appears more valuable. This finding aligns with Yamanishi et al. (2019), who advocate "hybrid"

approaches matching assessment methods to purpose.

Recent research in Japanese contexts similarly emphasizes assessment localization (Koizumi et al., 2016). What constitutes "good writing" reflects not only linguistic competence but also culturally situated rhetorical values. For Persian-speaking learners navigating substantial linguistic and cultural distance, analytic assessment may provide the transparent feedback necessary for developing metapragmatic awareness.

These findings suggest three pedagogical priorities for Japanese language programs serving linguistically distant learner populations:

First, vocabulary instruction should receive prominence commensurate with its demonstrated impact on writing quality. Given that vocabulary explained over 50% of the variance in holistic scores, curricula should emphasize explicit teaching of high-frequency vocabulary, academic vocabulary, and collocational patterns. Laufer and Nation (1995) demonstrated that the Lexical Frequency Profile serves as a reliable and valid measure of lexical richness in writing, correlating well with independent measures of vocabulary size. Such measures could inform curriculum design by identifying lexical gaps and tracking vocabulary development. Instruction should be supplemented by extensive reading to build incidental vocabulary knowledge and exposure to authentic Japanese texts across registers.

Second, classroom writing assessment should employ analytic rubrics providing component-level feedback, enabling learners to identify specific development areas. The strong learner preference for analytic feedback (65%) suggests such approaches align with learner expectations and support metacognitive development. Rubrics should be transparent and shared with learners before writing tasks, allowing them to internalize quality criteria and engage in meaningful self-assessment and peer review.

Third, writing instructors require professional development in analytic assessment, including calibration activities ensuring reliable, consistent scoring across raters. Training should address potential biases stemming from raters' own language learning experiences and cultural backgrounds, particularly when assessing learners from linguistically distant L1 backgrounds. The finding that non-native raters in this study privileged vocabulary over grammar suggests rater characteristics may influence which features are deemed salient, a pattern requiring empirical investigation and instructional attention.

5. Conclusion

This study compared holistic and analytic assessment methods for evaluating L2 Japanese writing among Persian-speaking learners and identified which linguistic features most strongly predict perceived quality. The findings demonstrated that both assessment approaches provide valid evaluations, supporting flexible method selection based on instructional purposes. However, the modest sample size, cross-sectional design, single-prompt task, and use of non-native raters constrain generalizability. Future research should employ longitudinal designs across proficiency levels, expand to diverse L1 backgrounds, develop Japanese-specific analytic rubrics capturing unique orthographic and sociolinguistic features, and examine vocabulary-focused instructional interventions. As Japanese language education expands beyond traditional East Asian contexts to serve linguistically distant learner populations, assessment practices must adapt accordingly.

Declarations

Competing interests

The author declares no competing interests.

Use of AI Tools

AI language assistance tool was used for proofreading, formatting, and English language editing and DeepL for translating Japanese sources. All research design, data collection, analysis,

interpretation, and original content generation were conducted by the author without AI assistance.

Authors' contribution

Author designed the study, collected and analyzed the data, interpreted the results, and wrote the manuscript. The author read and approved the final manuscript.

Funding

This research received no external funding.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable requests, subject to participant privacy protections.

Acknowledgments

The author gratefully acknowledges the students and instructors of the Department of East Asian Languages and Literatures at the University of Tehran, and those from the Japanese Language Institute Public Program, who participated in this research. The author is deeply grateful to Prof. Tomoko Watanabe, Graduate School of Humanities and Social Sciences, Teaching Japanese as a Second Language Program, Hiroshima University, Japan, for her expert guidance, continuous support, and invaluable advice on Japanese language education and assessment research. Her mentorship has been essential to the development of this study.

References

- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Connor, U. (2002). New directions in contrastive rhetoric. *TESOL Quarterly*, 36(4), 493-510. <https://doi.org/10.2307/3588238>
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415-443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Hōsei. (2018). Raitingu hyōka kenkyū no kadai to tenbō: Bogowasha kyōshi to hibogowasha kyōshi ni yoru hyōka no sōi wo chūshin ni [Issues and prospects in writing assessment research: Focusing on differences in evaluation between native and non-native speaker teachers]. *Nihon Tesuto Gakkaishi*, 14(1), 71-84. https://doi.org/10.24690/jart.14.1_71
- Ishibashi, R. (2012). Dainigengo raitingu no gengo shiyō: Gakushūsha kōpasu no kōchiku to bunseki [Language use in second language writing: Construction and analysis of learner corpus]. Kazama Shobo. <https://swu.repo.nii.ac.jp/records/5671>
- Koizumi, R., In'nami, Y., Asano, K., & Agawa, T. (2016). Validity evidence of Criterion® for assessing L2 writing proficiency in a Japanese university context. *Language Testing in Asia*, 6, 5. <https://doi.org/10.1186/s40468-016-0027-7>
- Kubota, R. (2020). Confronting epistemological racism, decolonizing scholarly knowledge: Race and gender in applied linguistics. *Applied Linguistics*, 41(5), 712-732. <https://doi.org/10.1093/applin/amz033>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786. <https://doi.org/10.1002/tesq.194>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective* (Vol. 3). P. Lang.
- Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Ono, M., Yamanishi, H., & Hijikata, Y. (2019). Holistic and analytic assessments of the TOEFL iBT® integrated writing task. *JLTA Journal*, 22, 65-88. https://doi.org/10.20622/jltajournal.22.0_65
- Shimada, K. (2014). Hi-kanji-ken gakushūsha ni tai suru nihongo shidōhō~ 「Manabu koto, oshieru koto」 no bapponteki na minaoshi [Teaching method of Japanese language for the students from non-Kanji backgrounds: Fundamental review of learning & teaching]. *Ryūgaku Kōryū*, 11, 1-16. <https://cir.nii.ac.jp/crid/1010000782237793672>
- Tanaka, M., Nagasaka, A., Narita, T., & Sugai, H. (2009). Dainigengo to shite no nihongo raitingu hyōka wākushoppu- Hyōka kijun no kentō[Second language Japanese writing assessment workshop: Examination of evaluation criteria]. Sekai no Nihongo Kyōiku: Nihongo Kyōiku Ronshū, 19, 157-176. <https://jpf.repo.nii.ac.jp/records/415>
- Wakita, S. (2016). Dainigengo to shite no nihongo sakubun no hyōka [Assessment of Japanese composition as a second language]. (Doctoral dissertation, Doshisha University, Japan). <https://doi.org/10.14988/pa.2017.0000014462>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>

17. Yamanishi, H., Ono, M., & Hijikata, Y. (2019). Developing a scoring rubric for L2 summary writing: A hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, 9, 13. <https://doi.org/10.1186/s40468-019-0087-6>
18. Yang, W., Cushing, S. T., & Yu, G. (2025). Linguistic predictors of L2 writing performance: Variations across genres. *Assessing Writing*, 66, 100985. <https://doi.org/10.1016/j.asw.2025.100985>