JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 19, No. 1 (2023), pp. 13-18

Development of online system checkable for Japanese writing tasks

Tsutomu Sekine[a,1], Kohki Takahashi[b]

[a]*Seikei University, Faculty of Science and Technology – Tokyo (Japan)*

[b]*Seikei University, Graduate School of Science and Technology – Tokyo (Japan)*

(submitted: 01/06/2022; accepted: 22/04/2023; published: 27/4/2023)

Abstract

Online learning environments have attracted attention of many educators especially in recent years since COVID-19 is still ongoing situation. Meanwhile, the various resources are becoming more and more available in online. In this study, some available online resources were used to create the system checkable for some writing abilities and the depth of understanding for Japanese writing tasks. The system was also made to provide some evaluation scores without depending the number of characters. The demonstration of system were given after the integration and implementation of some modules customized using online resources. The data sheet in the system finally saved the written content for 67 students. The writing task was given as the writing of summarization for what a student understand in a class. The following features were demonstrated from the analytical findings of online system developed in this study. The effectiveness of some available online resources was indicated through the demonstration of system checkable for some writing abilities and the depth of understanding for Japanese writing tasks. It was definite that the system was also made to provide some evaluation scores without depending the number of characters.

KEYWORDS: Online Evaluation System, Writing Task, Key Word, Key Sentence, Latent Semantic Analysis.

DOI
https://doi.org/10.20368/1971-8829/1135687

CITE AS
Sekine, T., & Takahashi, K. (2023). Development of online system checkable for Japanese writing tasks. *Journal of e-Learning and Knowledge Society*, *19*(1), 13-18.
https://doi.org/10.20368/1971-8829/1135687

1. Introduction

Education environment with information and communication technology (ICT) becomes more and more important than ever in the ongoing COVID-19 situation. The systems related to online learning environment such as e-learning have attracted attention of many educators especially in recent years. Moreover, the systems have been developed with the derived variation of educational terms such as distance learning, online training, distance education and so on (Ouadoud et al., 2021). To enhance students' learning effectiveness and experience, multimedia technologies are implemented to the systems in some learning objectives (Lau et al., 2014). The systems are generally created by integrating various modules. Slavuj et al. (2016) described several modules to create the system for language learning. With technological advances of ICT, the various resources have been gradually available in online. Even if a focus is given to the one module in the system research and development, it is possible to lead to the development of an advanced system.

The evaluation module of writing tasks is one of the modules in the online learning system. The module has been studied so far for the automation of evaluation. Descriptive type questions currently appear on the written assignment or examination. These questions are typically classified into two types. The one is commonly referred to essay question, and the other is known as short answer question (Ishioka, 2016). In essay question, the scoring process takes a lot of time if a teacher or human grader has to check all the answers written by test-takers. Automated essay scoring has been studied to enhance the scoring efficiency and to overcome the coupled difficulties since the 1960s (Ramesh & Sanampudi, 2022).

The various related methods have been proposed as some automated essay scoring systems. The work by Ajay et al. (1973) is known as Project Essay Grader

[1] corresponding author - email: ts_s@outlook.com – address: 3-3-1 Kichijoji-kitamachi, Musashino-shi, Tokyo 180-8633 (JP)

(PEG) in which an essay is graded based on the writing characteristics such as grammar, diction, construction, and so on. The PEG was modified by Shermis et al. (2001), and a high correlation was obtained between human graders and the system. Intelligent Essay Assessor (IEA) using latent semantic analysis was developed to estimate a holistic score (Foltz et al., 1999). E-rater was introduced to evaluate the analytical writing assessment in Graduate Management Admission Test (GMAT) (Powers et al., 2002). Since then, the other systems were also created using natural language processing techniques (Rudner et al., 2006). Recently, the systems with deep learning techniques were proposed for essay scoring (Dong et al., 2017).

Some essay scoring systems have been also developed so far to deal with Japanese language. One of the systems was Japanese essay scoring system (Jess) (Ishioka & Kameda, 2006). Jess is similar to E-rater, and the system focuses especially on evaluating rhetoric, organization, and content. Mainichi Daily News data was used in the system development to collect linguistic features such as sentence length and "kanji"/"kana" ratio. JWriter (Lee & Hasebe, 2017) is also one of the Japanese essay scoring systems. The readability score was proposed based on a regression analysis of data collected from 100 language textbooks and the balanced corpus of contemporary written Japanese (Hasebe & Lee, 2015). There also exist several studies to develop Japanese essay scoring systems with machine learning techniques (Tanaka & Tsubone, 2011; Hirao et al., 2020). However, few application cases have been reported on these Japanese essay scoring systems, and the available Japanese language data is not necessarily sufficient to make these systems more sophisticated.

The related technologies of automated essay scoring have matured gradually with the times. Especially, natural language processing and its related technologies are developing rapidly with the advancement of information and communication technologies. In recent years, the various resources are available in online. The background conditions would enable us to create an automated essay scoring system more easily through organizing some resources.

The objective of this study is to assess one of the possibilities in developing an online learning system. Some available online resources are used to create the system checkable for some writing abilities and the depth of understanding for Japanese writing tasks. The system is also made to provide some evaluation scores without depending the number of characters. The demonstration of system will be given after the integration and implementation of some modules customized using online resources.

2. Online system checkable for writing tasks

Important criteria were considered before the creation of an online system. This study focuses on writing tasks in a lecture for undergraduate students in our university. The lecture is related to production engineering and managed by one of the authors. The writing task in each class was set to be essay-style content in Japanese language. It is well known that an evaluation of writing quality and dexterity contain many complex factors. The cause of complexity results from the existence of various criteria to evaluate writing elements (Steendam et al., 2012). It means that the criteria depend on the subject of writing task, and the common, uniform ones are indeterminate in practice. In this study, some criteria were thus arranged to check the writing quality and dexterity in written content submitted by a student taking the lecture. These were vocabulary, Chinese characters, formality, sentence length, and descriptive validity.

The four criteria except for descriptive validity were evaluated through "Yasashii Nihongo Checker" which was an online support system of Japanese official document (Iori, 2016). The system has been developed with incorporating various tools in Japanese natural language processing such as Japanese dictionary and dependency parser. An entered text can be analyzed with the system. The evaluation of four criteria are obtained on a 5-point scale. The 3rd grade vocabulary of Japanese-Language Proficiency Test was a measure of scoring the vocabulary in a target sentence. Chinese characters was graded using the content rate of Chinese characters in a target sentence. Formality was determined based on the density of nouns in a target sentence. The average sentence length was used to evaluate sentence length in a target sentence.

Descriptive validity expresses the essential conformity of writing contents and the depth of understanding for a writing task. This study employed latent semantic analysis (LSA) to reveal the criterion based on the linguistic features of a target sentence. The analyzing algorithm includes singular value decomposition and can statistically compute the amount of words' or sentences' information. It is known that the scoring results of humans and LSA are in agreement to some extent (Landauer et al., 1998). In addition, the other related study reported the results with extracting keywords and key sentences from Japanese sentences (Tsubakimoto et al., 2007). The results indicated that the key sentences were well accorded with the sentences on which human graders focused during the scoring. On the basis of these results, this study applied LSA to the estimation of descriptive validity. A proofreading tool was also used to check grammatical correctness and typographical errors.

An online system was created in this study. Figure 1 shows the architecture of online system checkable for the above-mentioned criteria of writing tasks. The system was organized with some sub-functions implemented using Google Apps Script. The criteria can be checked using the sub-functions after an essay-style content is submitted by a student taking a lecture. The online form for submission was developed with Google

forms. Student's identification numbers were also collected through the form. Then, the collected information was individually accumulated to a data sheet created using Google spreadsheet. The results of Yasashii Nihongo Checker were extracted to the data sheet after some analyses in the system. Application programming interface (API) of external tools was also used to obtain the data related to morphological analysis and proofreading. We took advantage of Japanese morphological analysis API and proofreading API provided by Yahoo! Japan, respectively.

Two kinds of text analysis results were prepared to compare them in the developed system. The one was a result of select sentences whose selection was implemented after closing students' submissions, and the other was a result of target sentences written by each student. Here, what should be careful is that model, thoughtful sentences are determined by a lecturer as the select sentences. Figure 2 illustrates the brief flowchart of computational processes. These processes are as follows. Firstly, extraction of writing contents is executed after each student's submission. Then, the processing results in vocabulary, Chinese characters, formality, and sentence length can be acquired through the sub-function designed specially to communicate data with Yasashii Nihongo Checker. After that, Japanese morphological analysis is performed before LSA. Information contents of keywords and key sentences are obtained as a result of LSA. Key sentences are selected with reference to the result, and resultant keywords are extracted from the selected key sentences. The above-mentioned processes are accomplished in almost the same manner.

The conformity of two results can be checked through the comparison of select and target sentences. The following formula was used in the comparison:

$$CF = \frac{10 \cdot KW_{match}^{\ 2} \cdot rnc}{KW_{target} \cdot KW_{select}} \qquad (1)$$

where CF is the conformity of two results; KW_{match} is the number of matches for resultant keywords; rnc is Ratio of the number of characters in each sentence to be compared; KW_{target} is the number of keywords in a target sentence; KW_{select} is the number of keywords in a select sentence. In the calculation of a certain target sentence, the above formula is used in combination of all select sentences. Then, a certain target sentence is graded using the total score of calculated results. In addition, Japanese proofreading can be performed through proofreading API provided by Yahoo! Japan. In the proofreading, deduction points are calculated based on the number of errors.

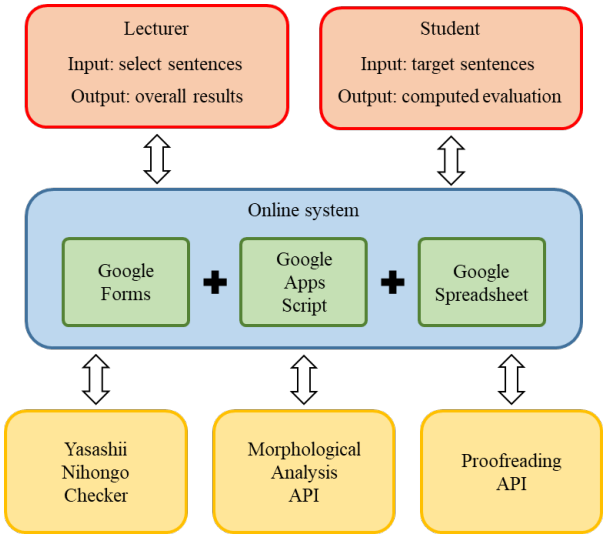

Figure 1 - The architecture of online system.

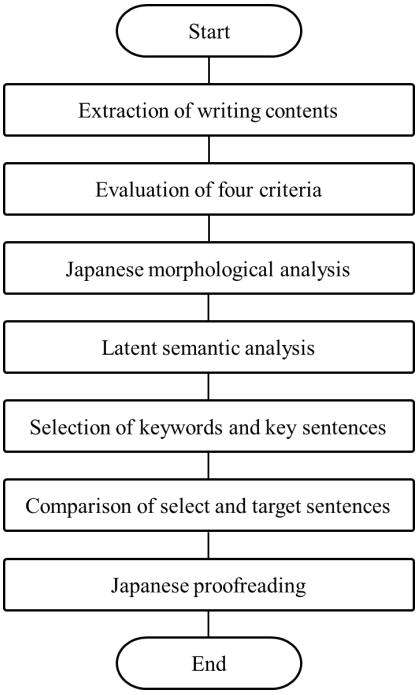

Figure 2 - The flowchart of computational processes.

3. Results and discussion

The writing task was given as the writing of summarization for what a student understand in a class. The all writing tasks were collected through the online system. The series of processes mentioned in section 2 were executed through a customized algorithm operable on Google Apps Script. The data sheet in the system finally saved the written content for 67 students.

Figure 3 shows the relationship between the number of morphemes and the number of characters. From the figure, it was found that two data had a proportional relationship. The correlation coefficient was 0.92, and the result represented a strong correlation between the two data. The fact demonstrated that morphological analyses were properly implemented through Japanese morphological analysis API provided by Yahoo! Japan. Moreover, we confirmed that there was no undoubted problem in some analysis results.

Figure 4 displays the details of scores in each criterion. In the figure, legends express each evaluation score, respectively. Figures 5 to 8 also express the relationships between the number of characters and evaluation score in each criterion. The plotted data on the figures were obtained as the results of Yasashii Nihongo Checker, and the external tool returned integer values only.

In vocabulary, Figures 4 and 5 illustrated that about 95% of the students acquired the score of 1 and 2. Two causes were considered to have an effect on the result. The one was that this class dealt with some specialized content for third-year university students, and the other was that Yasashii Nihongo Checker was an online support system of Japanese official document. It was also clear from the two figures that there was scarcely correlation between the number of characters and evaluation score, and the correlation coefficient was -0.36.

In Chinese characters, Figures 4 and 6 denoted that more than 80% of the students obtained the score of 2 and 3. From the result, it was revealed that a lot of the students used Chinese characters found commonly in Japanese official documents. In addition, the two figures clarified that there was no correlation between the number of characters and evaluation score, and the correlation coefficient was calculated as -0.14.

In formality, Figures 4 and 7 displayed that there dispersedly existed the students with the score from 2 to 5. This result was considered to imply an improvement in the writing task of this class. In fact, there was a wide range of formality in sentences written by the students. It was also definite from the two figures that there was no correlation between the number of characters and evaluation score, and the correlation coefficient was computed as -0.19.

In sentence length, Figures 4 and 8 represented that about 85% of the students achieved the score of 2 and 3. The result indicated that most of the students provided the length of a sentence found commonly in Japanese official documents. Moreover, the two figures demonstrated that there was no correlation between the number of characters and evaluation score. The correlation coefficient was 0.015, and it was confirmed to be extremely small.

In the results explained above, we could not detect any relationship between the number of characters and evaluation score in each criterion. From this fact, the scores were almost unaffected from the number of characters.

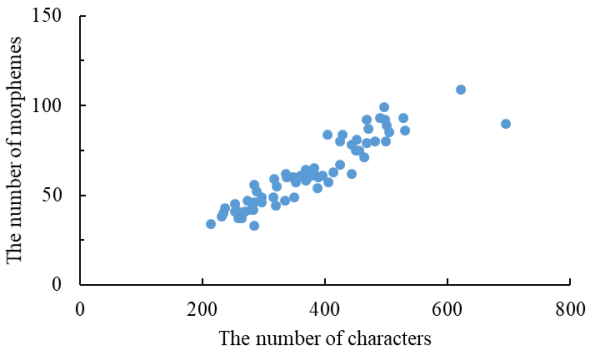

Figure 3 - Relationship between the number of characters and morphemes.

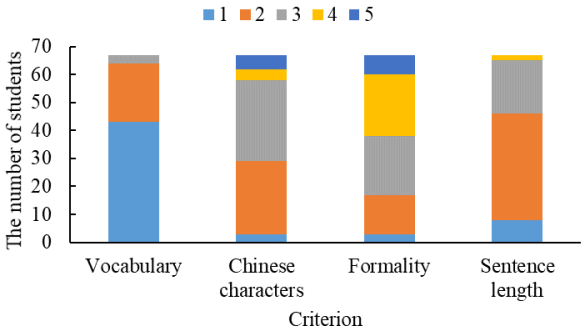

Figure 4 - Details of scores in four criteria obtained from an external tool used in the developed system.

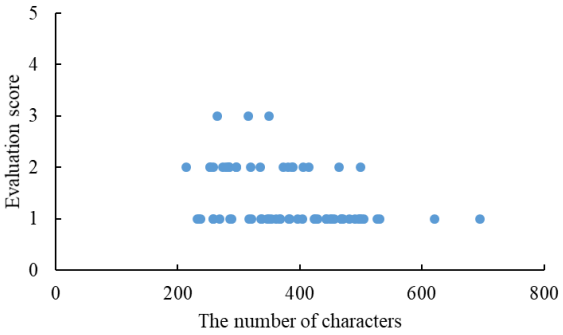

Figure 5 - Relationship between the number of characters and evaluation score in vocabulary.

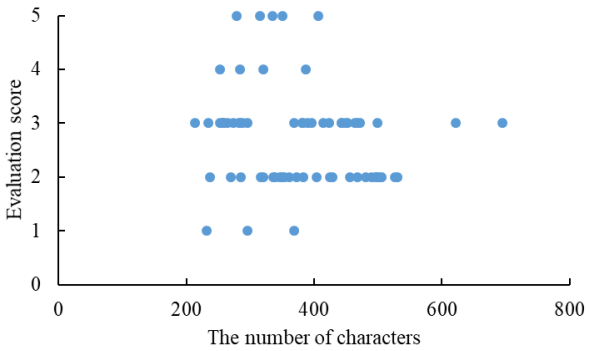

Figure 6 - Relationship between the number of characters and evaluation score in Chinese characters.

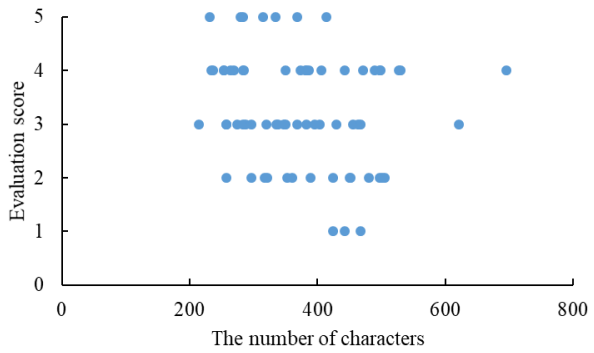

Figure 7 - Relationship between the number of characters and evaluation score in formality.

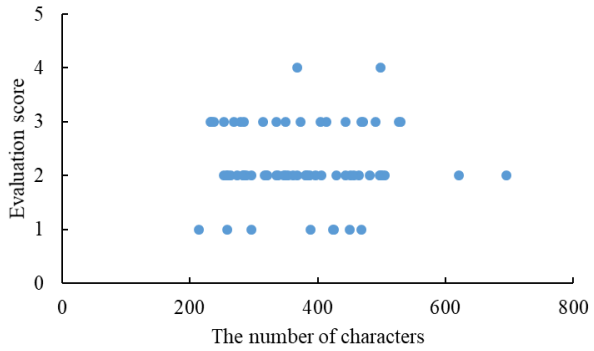

Figure 8 - Relationship between the number of characters and evaluation score in sentence length.

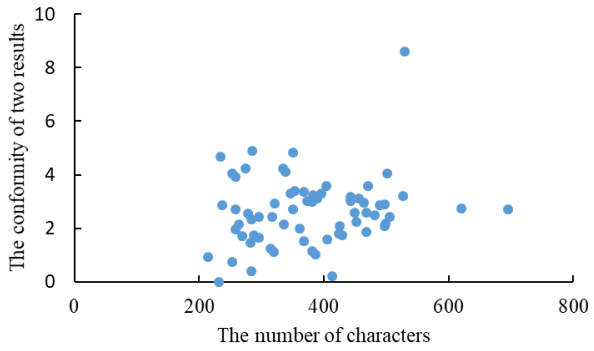

Figure 9 - Relationship between the number of characters and the conformity of two results.

Figure 9 shows the conformity of two results after the comparison of select and target sentences. The scores plotted on the figure were evaluated on a scale of 10 points, and double-precision real number was used in the calculation. The evaluation approach for this criterion was rather strict seemingly since almost all of the students acquired the score lower than 5. The figure also indicated that there was no correlation between the number of characters and evaluation score; furthermore, the correlation coefficient was 0.19.

What should be emphasized was that there was no evidence indicating a relationship between the number of characters and evaluation score in each criterion even if the last result was added to the other results. According to these results, the developed system would enable us to check an evaluation score based on each evaluation approach. In this writing task, we had students check whether unconscious errors are in the written content before submitting. The several cautionary instructions were displayed on Google forms, so that the proofreading only pointed out a few mistakes such as typographical error and so on.

4. Conclusion

An online system checkable for a writing task was developed in this study. The system had the functionalities of external tools such as Google forms, Yasashii Nihongo Checker, Japanese morphological analysis API, and so on. Five criteria were managed to check the writing quality and dexterity in written content submitted by a student taking the lecture. These were vocabulary, Chinese characters, formality, sentence length, and descriptive validity. The all writing tasks were collected through the online system. The series of processes mentioned in section 2 were executed through a customized algorithm operable on Google Apps Script. The data sheet in the system finally saved the written content for 67 students. The following features were demonstrated from the analytical findings of online system developed in this study. The number of morphemes had a proportional relationship between the number of characters. Evaluation scores using the results of Yasashii Nihongo Checker had little relations between the number of characters. There was no correlation between the number of characters and evaluation score obtained as the conformity of two results after the comparison of select and target sentences. From these results, the developed system could provide the reasonable criterion without depending the number of characters. The effectiveness of some available online resources was indicated through the demonstration of system checkable for some writing abilities and the depth of understanding for Japanese writing tasks. It was definite that the system was also made to provide some evaluation scores without depending the number of characters. In the future work, the further sophistication would be required in scoring approach and algorism for the detailed evaluation of Japanese writing tasks.

Acknowledgements

This study was partially supported by JSPS KAKENHI Grant Number JP19K03063. The authors sincerely acknowledge the supports.

References

Ajay HB, Tillett PI, Page EB. (1973). Analysis of essays by computer (AEC-II) (No. 8-0102). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, National Center for Educational Research and Development.

Dong F, Zhang Y, Yang J. (2017). Attention-based recurrent convolutional neural network for automatic essay scoring. In: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), pp. 153–162.

Foltz PW, Laham D, Landauer TK. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 1, 2, http://imej.wfu.edu/articles/1999/2/04/index.asp

Hasebe, Y., Lee, J. (2015). Introducing a Readability Evaluation System for Japanese Language Education. The 6th International Conference on Computer Assisted Systems For Teaching & Learning Japanese (CASTEL/J).

Hirao, R., Arai, M., Shimanaka, H., Katsumata, S., Komachi, M. (2020). Automated Essay Scoring System for Nonnative Japanese Learners, the 12th Language Resources and Evaluation Conference, pp. 1250–1257.

Iori, I. (2016). The Enterprise of Yasashii Nihongo : For a Sustainable Multicultural Society in Japan, Hitotsubashi review of arts and sciences, 10, pp. 4-19.

Ishioka, T., Kameda, M. (2006). Automated Japanese Essay Scoring System based on Articles Written by Experts, the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp. 233–240.

Ishioka, T., (2016). Computer-based Writing Tests, The journal of the Institute of Electronics, Information and Communication Engineers, 99(10), pp. 1005-1011. (in Japanese)

Landauer, T.K., Foltz, P.W., Laham, D. (1998) An introduction to latent semantic analysis, Discourse Processes, 25, pp. 259-284.

Lau, R.W.H., Yen, N.Y., Li, F., Wah, B. (2014). Recent development in multimedia e-learning technologies, World Wide Web, 17, pp. 189–198.

Lee, J., Hasebe, Y. (2017). jWriter Learner Text Evaluator, URL: https://jreadability.net/jwriter/.

Ouadoud, M., Rida, N., Chafiq, T. (2021). Overview of E-learning Platforms for Teaching and Learning, International Journal of Recent Contributions from Engineering, Science & IT, 9(1), pp. 50–70.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: challenging the validity of automated essay scoring. Computers in Human Behavior, 18(2), pp. 103–134.

Ramesh, D., Sanampudi, S.K. (2022). An automated essay scoring systems: a systematic literature review. Artificial Intelligence Review, 55, pp. 2495–2527.

Shermis MD, Mzumara HR, Olson J, Harrington S. (2001). On-line grading of student essays: PEG goes on the World Wide Web. Assess Eval High Educ 26(3), pp. 247–259.

Slavuj, V., Kovačić, B., Jugo, I. (2016). Adaptive E-learning system for language learning: Architecture overview, 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 951-955.

Tanaka, M., Tsubone, Y., (2019). New Writing Feedback: A Collaborative Approach Utilizing Human and Machine Evaluation, CAJLE 2019 Proceedings, pp. 315-324. (in Japanese)

Tsubakimoto, M., Nakamura, M., Kishi, M. (2007). A Fundamental Examination of the Report Assessment Method Focusing on Key sentences and Key words, Journal of Science Education in Japan, 31, pp. 210-219.

Van Steendam, E., Tillema, M., Rijlaarsdam, G., Van den Bergh, H. (Eds.) (2012). Measuring Writing: Recent Insights into Theory, Methodology and Practices.