

RESEARCH ARTICLE

Comparison of Biological Concept Assessment Scores by Handwriting and Computer Writing

Jeong-hwa Seong¹ · Sarah Meilani Fadillah² · Minsu Ha^{2*}¹Kangwon National University High School²Kangwon National University

손 글쓰기와 컴퓨터 쓰기에 따른 생물개념평가 점수 비교

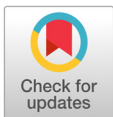
성정화¹ · 파딜라 사라 메이라니² · 하민수^{2*}¹강원대학교 부설고등학교 · ²강원대학교

*Corresponding Author: msha@kangwon.ac.kr

ABSTRACT

As students become accustomed to the digital environment, various discussions are taking place on the comparison of handwriting assessment and computer-based assessment methods. This study attempts to examine whether there are significant differences in the number of words, biological concept level, and biological concept confidence for each assessment when handwriting assessment or computer writing assessment is performed to confirm biological concepts. A descriptive assessment was conducted by selecting 13 questions related to life science and dividing 361 students from general high school into handwriting and computer writing groups. As a result of propensity score matching and linear regression analysis, the handwriting group showed significantly higher number of characters in the descriptive assessment answer and the concept score of the descriptive assessment than the computer writing group. There was no significant difference in the level of confidence of descriptive assessment. These results confirmed that students' concepts were accurately expressed when handwriting assessment was performed rather than computer writing assessment with an assessment tool.

Key words: Biological concepts, handwriting assessments, computer writing assessments, descriptive assessments, propensity score matching



OPEN ACCESS

Brain, Digital, & Learning
2022, Vol. 12, No. 4, 663-674.

<https://doi.org/10.31216/BDL.20220041>

Received: November 23, 2022

Revised: December 21, 2022

Accepted: December 21, 2022

© 2022. Institute of Brain based Education,
Korea National University of Education



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

The 21st century society emphasizes the convergence and interaction between knowledge rather than absolute accumulation of knowledge due to globalization and informatization. As the times change, the human resources required by society also change. The public education system has shifted from an academic curriculum to a competency-based curriculum to cultivate creative convergence talents needed by the 21st century. The competency-centered curriculum aims to educate the necessary competencies in social life away from the knowledge curriculum. Especially after identifying core competencies in the OECD's DeSeCo project, many countries around the world reorganized their curricula into competency-oriented curriculum (So, 2017).

The modification to the curriculum also altered the assessment method. The major change is the expansion of descriptive and process-oriented assessments. Beyond the assessment's current purpose, which focuses on objectively assessing and quantifying the degree of achievement presented in the subject, assessment is considered an integral part of the class and should be used to enhance the teaching-learning process (Ha et al., 2019). Ha et al. (2019) suggested formative assessment that was equipped with feedback as well as assessment that stressed on process over results should be implemented often times more rather than general assessment. Additionally, formative assessment emphasizes the activities that focus on the areas for improvement through identifying strengths and weaknesses as information for teachers for instructional planning and for students for their further understanding (Cizek, 2010). Essay and descriptive assessment, especially if they were equipped with feedback are able to give more information about students' learning process, since descriptive tasks required students to explain their actual understandings on certain concepts. As a result of these shifts in the paradigm of assessment, the use of various assessment tools, descriptive assessments, and essay-type assessments have expanded, and a new assessment method which focus on process-oriented assessment has emerged. However, the descriptive question task might burden in terms of efficiency compared to optional task such as multiple-choice test, especially when it is used in a large population. To this issue, prior studies began to identify how digitalization of descriptive task would benefit the efficiency in implementing formative assessment.

Recently, the digitalization of learning processes and the cultivation of digital talent have become national goals in Korea (Ministry of Education, 2022). Besides, UNESCO (Cornu, 2011) suggested the educational practitioners to develop a digital environment in which they can shift from the paper-based pedagogy to digital-based pedagogy, which for example can be done by performing computer-based assessment activities. Another reason for students to be familiar with computer-based test is that more writing assignment and official assessment are being conducted in a digital setting (Alderson, 2000; Im et al., 2008; Mogey & Fluck, 2015). As a result, it is essential to understand the potential impact of computer typing on students' learning and performance.

Furthermore, there are various research results on the effects of learning and assessment media. According to prior studies, learning on a computer can generate interest and save time, although it does not have a better advantage in terms of cognitive process, as computer typing is inferior to handwriting in concept structuring and conceptual understanding (Kim et al, 2018; Noyes & Garland, 2008; Wollscheid et al., 2016). In computer-based assessment, it is possible to present questions according to the level of the evaluators and respondents, quickly score, and efficiently manage assessment results; however, it can be affected by visual limitations on computer screens, such as difficulties in note-

taking and computer familiarity (Kim et al, 2018; Wollscheid et al., 2016). Various assessment methods are necessary to be introduced as the need for remote education beyond time and space emerges due to the expansion of the descriptive assessment ratio, the establishment of process-based assessment, and the introduction of lifelong education and high school credit system. In this study, we aim to identify how each student's handwritten and computer-submitted answers might affect the result obtained from descriptive assessment that confirm the concept of living things.

A Preliminary Study on the Change of Writing Method

Previous studies have reported a significant proportion of students who have experience with computer-based writing, for instance, Jang (2013) surveyed high school students and enquired about the frequency of pen-based writing and computer-based writing, stating that more than 23.3% had more experience in computer-based writing. In a survey of teenagers on whether they prefer internet writing or paper writing, Moon (2014) suggested that 60.8% of middle school students and 59.1% of high school students prefer internet writing. Additionally, there have been a number of studies examining the impact of different writing media on learning attitudes such as motivation (Chua & Don, 2013), confidence in using computer (Garra & Hassan, 2018), and mood (Fouladi et al., 2002). Adding confidence level in a test performance can further be helpful for study of metacognition, since confidence judgment can be thought as the ability to accurately assess one's own ability in a particular area of task (Tiffin & Paton, 2019). The confidence level in conceptual test has been widely performed in both paper-based and computer-based tests, but few studies have directly compared confidence level across these two mediums. One study by Lee (2001) reported that students felt their confidence improved when internet writing was used instead of traditional writing.

More research has been conducted abroad. Berninger et al. (2009) suggested that students who face difficulty in writing exhibited a positive effect on the amount and content of their writing when writing using a computer. From a different perspective, Kim (2005) pointed out the issue on paragraphs that are not well distinguished because, when writing using a computer, fragmentary contents are frequently listed without a clear idea. Mueller and Oppenheimer (2014) confirmed that students who wrote class contents using electronic media received lower scores in the conceptual assessment than students who wrote them by hand. Additionally, in a letter memorization method study conducted on pre-school children, Longcamp et al. (2005) confirmed that handwritten letters were better remembered than computer-written letters. Recent studies which compares handwriting and computer typing in science related concepts suggested that typing modality might influence the cognitive achievement. Aberšek et al. (2018) studied the effect of typing notes in the process of learning science subjects where understanding the content of the subject is important. According to an analysis of their performance, it was observed that students who used computers to take notes exhibited cognitive overload, low achievement levels, poor accuracy of technical terms, and a lack of understanding of the interrelationships between concepts. Although the results of learning according to the writing medium are not consistent, it can be observed that changes in the writing medium might have a positive or negative effect on the learning process and results of students. Depending on the writing medium, there is a difference between visual attention and tactile input, which can also affect the process and outcome of learning (Mangen et al., 2016). Therefore, it is necessary to understand the characteristics of the media and continuously study the effect on learning and cognition depending on the type of media used for learning.

A Preliminary Study on the Change of Assessment Media

Assessment using computers is spreading more widely in the digital environment. The advantages of computer-based assessment are: first, assessment can be conducted regardless of the time and place; second, accurate and rapid grading, effective analysis, and application of results; third, simulation and animation, which are difficult to present in paper-based assessment, thus, it facilitates the presentation of various types of questions (Alderson, 2000; Im et al., 2008; Kim et al., 2018). In addition, it is possible to present questions tailored to the corresponding levels based on respondents' responses, as well as to increase the efficiency of learning through immediate feedback (Rudland et al., 2011, Wise & Flake, 1990). However, the assignment or test takers should have appropriate familiarity of computers, such as the ability in using mouse and keyboard including capability in reading text on computers, if they do not want to be disadvantaged by computer-based assessment (Alderson, 2000). Additionally, cheating may occur while taking tests, and problems related to security may appear. Finally, it is expensive to build computer-based assessment (Ockey, 2009).

The assessment process affects learning results according to the assessment medium. Spray, Ackerman, Reckase, and Carlson (1989) allowed students to skip questions regardless of the order of the questions, such as handwriting, or come back at the end to correct their answers, showing no significant difference from the assessment results of the handwriting group. Mazzeo and Harvey (1988) noted that when comparing the results according to the assessment medium based on the student's intelligence, attitude, personality, and achievement level, there was no significant difference in the power test, but there was a significant difference in the speed test.

There is also a study on the effect of the assessment medium on the affective domain of students. Computer familiarity can be presented in the affective domain, however, researchers have observed that computer familiarity was not suggested as significant influence on computer-typing assessment. Odo (2012) observed that computer familiarity affects the assessment results, but it is not a key factor in determining the assessment score. Previous studies also suggested that computer familiarity does not affect the results as a result of taking the TOEFL test (Alderson, 2000; Al-amri, 2008). In addition, Do (2013) stated that computer familiarity and computer-based assessment were not correlated. Research on computer familiarity generally demonstrated that computer familiarity did not affect the assessment results and it was found that anxiety about the test itself is greater than anxiety about using computers to perform the test. Weerakon et al. (2001) noted that neither computer experience nor the anxiety was significantly correlated with the performance resulted from computer-based assessment.

According to the background and the literature, this study aims to compare the students' responses on descriptive science-related questions that were performed from two different mediums, namely handwriting and computer typing. Specifically, the research questions are as follows:

1. How is the difference between the amount of characters produced by handwriting and typing in a science-related descriptive questions?
2. How does the conceptual outcome from handwriting and computer typing differ?
3. How does the confidence level for handwriting differ from that for computer typing?

Methods

Development of Biological Conceptual Descriptive Assessment Questions

The concept of biology in integrated science, which is a common subject in the 2015 revised curriculum, was confirmed in order to conduct tests on first and second-year high school students. By analyzing the content elements, achievement standards, and achievement levels presented in the 2015 revised curriculum, the core concepts of biological concepts, life system, biological diversity and maintenance, as well as ecosystem and environment were selected. Further, test questions were produced according to the concepts.

The questions that were developed by the AAAS were used to secure the validity of the test paper questions. Questions on evolution and natural selection were translated by Yoon (2015) and the ones whose reliability and validity were verified through internal consistency reliability analysis and Rasch analysis were used. Lee and Ha (2015) translated the questions on reproduction and genetics and the reliability and validity of the instruments were verified. The question regarding the interdependence of the ecosystem was translated by Lee et al. (2021) and its validity was confirmed after analyzing student responses. The mean-square (MNSQ) value, which indicates the item response fit value that could produce productive measurement, fell within the range between 0.5-1.5. This cut-off considered the items as appropriate according to Linacre (2002). Among the contents of the integrated science, AAAS assessment questions were reviewed based on the achievement criteria for bio-related concepts, and a total of 13 questions were selected, including six questions regarding “life system”, four regarding “biological diversity and maintenance”, and three regarding “ecosystem and environment”. Since the questions developed in AAAS were produced in an multiple choice form, they were modified and supplemented in a descriptive form according to the purpose of this study. The example of the adopted items were presented in Table 1.

Table 1. Example of conceptual assessment questions

Multiple choice test (original)	Open-ended test (revised)
A mouse's tail was caught in a mouse trap. When he was later released, his tail was cut off. If the mouse had a baby, how would it affect its tail length and what is the most valid explanation for the reason? ① It will affect the tail length of all babies due to the influence of their parent. ② Changes occurred only in one of the parents, so it will somewhat affect the tail length of the baby. ③ It does not affect the tail length of the baby because only the traits of the parents have changed. ④ The traits may vary depending on the age of the parent, so it depends on the age of the mouse when its tail was cut off.	A mouse's tail was caught in a mouse trap. When he was later released, his tail was cut off. If the mouse had a baby, explain how it would affect the tail length of the baby mouse and why.

A measure of confidence used by Woo et al. (2017) was performed to confirm the degree of biological concept confidence (self-assessment). A measure of confidence was presented below each optional and descriptive assessment question. Students were asked to select a feeling after answering the question, with prompts such as "How sure are you in the answers you provided?" A five-points Likert's scale ranging from “strongly not sure” to “strongly sure” was utilized. For handwriting assessment, a paper test paper containing 13 descriptive questions and a measure of confidence

for each question was produced, and the same Google Form test paper was produced for the computer writing assessment. Since the students who took the paper test and the Google Form test were different individuals, an optional questionnaire containing similar content of multiple-choice questions was used as the equalization.

Measurement of the Validity and Reliability of the Biometrics Verification Questions

Rasch analysis was performed to measure the validity of the student test paper questions, and Cronbach's alpha analysis was performed to measure the reliability. As a result of the Rasch analysis for each item, it can be confirmed that the fit of each item is appropriate for evaluating the biological concept as the Infit MNSQ and Outfit MNSQ values of all items from 1 to 13 for both optional and descriptive items are in the range of 0.5 to 1.5. The reliability of internal consistency was also suitable as 0.6 or higher in both the descriptive and selective types.

Participants and Data Collection

The subjects of this study were natural science students in the first and second grades of Public High School A in Gangwon-do and Public High School B in Gyeonggi-do. These participants consist of 97 first-year students (50 male students, 47 female students) and 105 second-year students (60 male students, 55 female students) from High School A participated in the study; 96 first-year students (51 male students, 45 female students) and 63 second-year students (46 male students, 17 female students) participated in the study from High School B. The assessment was conducted after two tests in the second semester when the students who participated in the study learned all the core concepts presented in integrated science. For each group, one descriptive assessment and one multiple choice assessment were performed, respectively. First, the students answered the test paper consisting of 13 descriptive questions for 30 minutes. The handwriting group responded to a paper test that could check the concept of living things in the classroom, and the computer writing group responded to a test paper made of Google Form in the school computer room using a computer. Ten days after completing the descriptive assessment, the students who participated in the study completed an multiple-choice questionnaire consisting of the same questions and order as the descriptive assessment questionnaire so that their level of knowledge could be assessed.

In the descriptive assessment, the two public schools performed both handwriting and computer writing methods, however the participants of the group were assigned randomly. The number of students who participated in the test that required them to write answers in handwriting or the Google Form test that required them to write answers in computer writing is as follows. 175 out of a total of 361 students participated in the handwriting test; among them, 97 were first-year students (48 male and 49 female students) and 78 were second-year students (41 male and 37 female students). In this study, 186 out of 361 students participated in the Google test paper; among them, 96 were first-year students (53 male and 43 female students) and 90 were second-year students (55 male and 35 female students).

Data Analysis

A total of 284 students answered the multiple choice test and descriptive assessment questions and this reflected in the analysis, of which 141 students participated in the handwriting assessment and 143 students participated in the computer writing assessment. The number of male students who participated in the study was 146 and the number of female students was 138. ANOVA was conducted in SPSS to analyze the statistical significance of differences between groups on the gender and input method of the handwriting group and the computer writing group, the number of characters in the descriptive assessment answer, the descriptive assessment concept score, and the level of confidence of the descriptive assessment answer. The computer writing group had higher result when they perform the items in multiple-choice format, and in terms of gender, female students scored higher than male students. Since this indicates that there is a difference in knowledge level according to group and gender, it is necessary to analyze the number of characters, descriptive assessment scores, and the level of confidence in descriptive assessment answers according to group and gender after controlling the knowledge level. To this end, Propensity Score Matching (PSM) analysis and linear regression analysis were conducted to control knowledge and gender that could affect the descriptive assessment of the handwriting and computer writing groups. Propensity Score Matching uses a logistic regression model to reduce covariates to a one-dimensional “propensity score” (Yoo, 2013); to match and compare people with similar propensity scores based on gender, class, school, and multiple choice scores.

After generating one-dimensional propensity scores based on gender, class, school, and multiple choice assessment scores, there were 102 cases in which the difference in propensity scores was less than 0.1 were matched and analyzed as a paired sample t-test. For 102 cases in which the difference in propensity score is less than 0.1, the paired sample t-test results for the number of characters in the descriptive assessment answers of the handwriting group and the computer writing group were performed. After controlling the knowledge level of the handwriting group and the computer writing group, this study further analyzed the concept scores of descriptive assessment, the number of characters in descriptive assessment answers, and the level of confidence of descriptive assessment answers. Furthermore, Linear regression analysis was also performed to confirm a more reliable basis. In linear regression analysis, significant differences between the two groups (handwriting and computer writing) were confirmed by adding variables to be controlled, such as scores as dependent variables and multiple choice assessment scores as independent variables.

Result and Discussion

After creating one-dimensional propensity scores based on factors such as gender, class, school, and multiple choice assessment scores, the 102 cases were selected and analyzed using a paired sample t-test. As a result of the t-test of the corresponding sample for the number of characters in the descriptive answer, the average of the handwriting group was 36.226 and the average of the computer writing group was 30.090. It can be observed that the number of descriptive characters in the handwriting group is significantly higher than that of the computer writing group ($p < 0.05$). According to the result, the number of characters in the descriptive assessment responses of the handwriting group was significantly higher than the number of characters of the computer writing group, as determined by the paired-sample t-test and linear regression analysis with Propensity Score Matching.

When students use a computer for writing, fragmentary contents are often listed (Kim, 2005). Based on the results of previous studies, the assessment predicted that the computer writing group would have more characters in the descriptive assessment answer than the handwriting group, however the results of this study were reversed since in our study, the handwriting group significantly had more characters than the computer writing group. The length of the writing has been found to be related to writing fluency, which refers to the number of correctly formed letters or handwriting outcomes produced in a given time period (Feng et al., 2019). According to Feng et al. (2019), when students perform computer-based tests, typing requires them to remember the locations of letters on the keyboard, while handwriting requires them to accurately and efficiently form each letter. They also noted that these different physical requirements may affect writing fluency differently depending on the method of transcription used. However, previous research was conducted with the purpose of solving tasks using different media in the writing learning process; this study focuses on the characteristics of various media used for assessment in science-related question. In this study, students answer the questions according to their knowledge level regardless of the assessment medium. Since the assessment questions did not require an essay-style responses that necessitated a large number of characters, but required the concept of science to be written according to the problem situation, the handwriting assessment similar to the test situation is thought to have aided in the retrieval of scientific concepts.

In terms of descriptive assessment concept score, the paired sample t-test gave outcome that the average of the handwriting group was -0.336, and that of the computer writing group was -0.802. It is also noted that the handwriting group has a significantly higher descriptive concept score ($p < 0.05$) than the computer writing group. As a result of the t-test of the paired sample for the level of confidence of the descriptive answer, the average of the handwriting group was -0.175, and that of the computer writing group was -0.392. The handwriting group demonstrated a higher level of confidence in descriptive answers than the computer writing group, but it was not significant ($p > 0.05$).

Linear regression analysis was conducted with the number of characters in the descriptive assessment answer, the concept score of the descriptive assessment, and the level of confidence of the descriptive assessment answer as dependent variables (Table 2) to identify the influence of the handwriting group. Linear regression analysis was conducted to identify whether the influence of handwriting assessment was related to the number of characters in the descriptive assessment answer. Since the handwriting assessment was $\beta = 0.125$ ($p < 0.05$), the handwriting assessment had a significant effect on the number of characters in the descriptive assessment answer. As a result of analyzing the effect of handwriting assessment, it was confirmed that the number of characters in the descriptive assessment answer was higher in the handwriting assessment. Next, since $\beta = 0.175$ ($p < 0.05$) of handwriting assessment, handwriting assessment had a significant effect on descriptive assessment concept scores. Since the handwriting assessment was $\beta = 0.019$ ($p > 0.05$), the handwriting assessment did not have a significant effect on the confidence level of the descriptive assessment answer.

Table 2. Results of linear regression analysis

Dependent variable	Independent variable	Non-standardized coefficient		Standardized coefficient	t	p	R ²
		β	Std. Error	β			
The number of characters in the answer	Constant	28.439	2.147		13.243	0.000	0.208
	MCT score	5.397	0.751	0.394	7.182	0.000	
	Handwriting	4.789	2.078	0.125	2.305	0.022	
	Male	-6.390	2.083	-0.167	-3.068	0.002	
Descriptive Assessment Score	Constant	-1.107	0.101		-10.960	0.000	0.436
	MCT score	0.485	0.035	0.636	13.726	0.000	
	Handwriting	0.374	0.098	0.175	3.826	0.000	
	Male	-0.231	0.098	-0.108	-2.354	0.019	
Confidence in descriptive assessment	Constant	-0.690	0.150		-4.602	0.000	0.049
	MCT score	0.117	0.052	0.134	2.233	0.026	
	Handwriting	0.046	0.145	0.019	0.317	0.752	
	Male	0.499	0.145	0.204	3.430	0.001	

In both the paired-sample t-test after propensity score matching and the linear regression analysis, the descriptive assessment concept scores of the handwriting group and the computer writing group were significantly higher in the handwriting group than in the computer writing group. Based on the results of a previous study, the students who took notes in class using electronic media received lower scores in conceptual assessment than students who took notes by handwriting (Mueller & Oppenheimer, 2014). In the assessment, the handwriting group was expected to obtain a higher score in the descriptive assessment concept than the computer writing group, and the results of this study aligned with the literature. Previous studies compared students' conceptual assessment scores according to the method of taking notes on learning contents, but this study relates to the characteristics of different media used for assessment. The reason why the descriptive assessment concept score was significantly higher in the handwriting group seems to help students' cognitive recall and structure of concepts rather than computer writing assessment. The cognitive advance in terms of recalling the scientific concepts also favoured the handwriting groups compared to computer typing group (Smoker et al., 2009). Moreover, in the case of computer writing assessment, there are visual limitations in viewing the screen, and the inability to take notes may be the cause, and because it is not a high-pressure test, a somewhat unfamiliar computer test environment for students may have affected the score (Kim et al, 2018; Wollscheid et al., 2016). Noh (2013) suggested that the score was higher in paper-based assessment than in computer-based assessment in factual comprehension measurement, and there was no statistically significant difference between the two groups in inference and comprehensive comprehension measurement. Since the questions presented in the assessment of this study were questions that measure factual understanding, it is judged that the handwriting group was able to obtain higher conceptual scores than the computer writing group.

There was no significant difference in the level of confidence in the descriptive assessment answer between both mediums tested by paired sample t-test after propensity score matching and linear regression analysis. This means that the level of confidence in the answer is not affected by the input medium. One research have identified performance as well as psychological tests acquired from computer-based and paper-based test, and it was found that the test effects of self-efficacy variable (students' belief in their abilities to succeed the learning) was found to be significant in both test

mediums, indicating that there is no difference in psychological test (including self-efficacy variable) when the tests are taken using either handwriting or computer typing (Chua & Don, 2013). This study supports the finding of Chua and Don (2013), that test medium through which tests are taken does not have a significant impact on the confidence level of tests taker.

Conclusions

Various discussions are taking place on the comparison between handwriting and computer writing in terms of learning and assessment as students become familiar with the digital environment. Due to the recent expansion of the descriptive assessment uses, the introduction of process-based assessment, lifelong education, and the introduction of high school credit systems, the need for remote education beyond time and space has emerged. In this study, when handwriting and computer writing assessments were conducted to confirm whether there were significant differences in the number of characters, biological concept levels, and degree of confidence of answers written by students for each assessment mediums. This study confirmed that the handwriting assessment had more characters in the answer than the computer writing assessment and the concept score was higher in the handwriting descriptive assessment to confirm the concept of living things. Through these studies, it is necessary to create an environment for computer writing assessment in the form of cognitive recall and structure of concepts in consideration of question types, question placement, and problem-solving methods. In the future, school education practitioners should indeed prepare not only handwriting assessment, but also assessment methods that utilize various media. Further research is required to determine the most effective way of each assessment medium for helping students to retain and recall concepts as well as to develop their cognitive structures.

References

- Aberšek, M. K., Aberšek, B., & Flogie, A. (2018). Writing versus typing during science teaching: Case study in Slovenia. *Journal of Baltic Science Education*, 17, 84-96.
- Al-Amri, S. (2008). Computer-based testing vs. paper-based testing: A comprehensive approach to examining the comparability of testing modes. *Essex Graduate Student Papers in Language & Linguistics*, 10, 22-44.
- Alderson, C. (2000). Technology in testing: The present and the future. *System*, 28, 593-601.
- Berninger, V. W., Garcia, N. P., & Abbott, R. D. (2009). Multiple processes that matter in writing instruction and assessment. In G. A. Troia (Ed.), *Instruction and Assessment for Struggling Writers: Evidence-based Practices* (pp. 15–50). The Guilford Press.
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, 29, 1889-1895. doi:10.1016/j.chb.2013.03.008
- Cizek, G. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade, & G. Cizek (Eds.), *Handbook of Formative Assessment* (pp. 3-17). New York, NY: Routledge.
- Cornu, B. (2011). Digital natives: How do they learn? How to teach them. *UNESCO Institute for Information Technology in Education*, 52, 2-11.
- Do, S. (2013). *The Effect of Computer-based Tests on Test Anxiety* (Unpublished Master Thesis). Hankuk University, Seoul, Republic of Korea.

- Feng, L., Lindner, A., Ji, X. R., & Malatesha Joshi, R. (2019). The roles of handwriting and keyboarding in writing: A meta-analytic review. *Reading and Writing*, 32, 33-63.
- Fouladi, R. T., McCarthy, C. J., & Moller, N. (2002). and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9, 204-215.
- Garas, S., & Hassan, M. (2018). Student performance on computer-based tests versus paper-based tests in introductory financial accounting: UAE evidence. *Academy of Accounting and Financial Studies Journal*, 22, 1-14.
- Ha, M., Lee, G. G., Shin, S., Lee, J. K., Choi, S., Choo, J., ... Park, J. (2019). Assessment as a learning tool and utilization of artificial intelligence: WA3I project case. *School Science Journal*, 13, 271-282.
- Im, E. J., Lee, W. K., Lee, Y. C., Choe, B. H., Chung, S. K., Lee, T. H., ... Lee, J. M. (2008). Development of computer-based test (CBT) and student recognition survey on CBT. *Korean Journal of Medical Education*, 20, 145-154.
- Jang, E. (2013). A Case Study on Internet Writing Process of Secondary High School Students (Unpublished Doctoral Dissertation). Korea University, Seoul, Republic of Korea.
- Kim, H. R., Bowles, M., Yan, X., & Chung, S. J. (2018). Examining the comparability between paper-and computer-based versions of an integrated writing placement test. *Assessing Writing*, 36, 49-62.
- Kim, S. (2005). A study on aspects of paragraph in word processor writing. *Journal of CheongRam Korean Language Education*, 32, 1-20.
- Lee, G. (2001). A Study on the Internet of Writing Education (Unpublished Master's Thesis). Sookmyung Women's University, Seoul, Republic of Korea.
- Lee, J. K. & Ha, M. (2015). Development of two equivalent genetics concept assessments using AAAS genetics concept test items and Rasch model. *School Science Journal*, 9, 165-174.
- Lee, S., Hwang H., Ha, M., & Park, E. (2021). Study on reliability of translation of aaas ecological concept assessment and student's misconception. *Biology Education*, 49, 240-250.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16, 878.
- Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta psychologica*, 119, 67-79.
- Mangen, A., & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, 5, 99-106.
- Mazzeo, J., & Harvey, A. L. (1988). The Equivalence of Scores from Automated and Conventional Educational and Psycholological Tests: A Review of the Literature. 88-8. New York: College Entrance Examination Board.
- Ministry of Education. (2022, March). Education in Korea 2021. Retrieved from <http://moe.go.kr>.
- Mogey, N., & Fluck, A. (2015). Factors influencing student preference when comparing handwriting and typing for essay style examinations. *British Journal of Educational Technology*, 46, 793-802.
- Moon, M. (2014). A study on differences by gender and school level in attitude and recognition on internet writing (Unpublished Master's Thesis). Korea National University of Education, Cheongju, Republic of Korea.
- Mueller, Pam A., & Daniel M. O. (2014). Study on Differences by Gender and School Level in Attitude and Recognition on Internet Writing. *Psychological Science*, 25, 1159-1168.
- Noh, Y. O. (2013). Comparison of computer-based tests and paper-based tests for assessing high school students' reading abilities. *Secondary English Education*, 6, 83-103.
- Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51, 1352-1375.

- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, 93, 836-847.
- Odo, D. M. (2012). Computer familiarity and test performance on a computer-based cloze ESL reading assessment. *Teaching English with Technology*, 12, 18-35.
- Rudland, J. R., Schwartz, P., & Ali, A. (2011). Moving a formative test from a paper-based to a computer-based format. A student viewpoint. *Medical Teacher*, 33, 738-743.
- Smoker, T. J., Murphy, & Rockwell, A., K. (2009). Comparing memory for handwriting versus typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53, 1744-1747.
- So, K. (2017). *Understanding Curriculum*. Seoul: Gyoyookgwahaksa
- Spray, J. A., Ackerman, T. A., Reckase, M. D., & Carlson, J. E. (1989). Effect of the medium of item presentation on examinee performance and item characteristics. *Journal of Educational Measurement*, 26, 261-271.
- Tiffin, P. A., & Paton, L. W. (2019). Does 'online confidence' predict application success and later academic performance in medical school? A UK-based national cohort study. *BMJ Open*, 9, e034437.
- Weerakoon, G. L. P. (2001). The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *Medical teacher*, 23, 152-157.
- Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development*, 23, 3-10.
- Wollscheid, S., Sjaastad, J., Tømte, C., & Løver, N. (2016). The effect of pen and paper or tablet computer on early writing—A pilot study. *Computers & Education*, 98, 70-80.
- Woo, M., Rachmatullah, A. & Ha, M. (2017). Validating overconfidence biases measurement using genetics and evolution concept test and exploring gender and academic track differences. *The Journal of Learner-Centered Curriculum and Instruction*, 17, 27-51.
- Yoo, J. E. (2013). Misunderstandings and mis-applications on ANCOVA in educational studies. *The Journal of Learner-Centered Curriculum and Instruction*, 13, 27-49.
- Yoon, Y. (2015). *Evolution Concept in Accordance with The Level of Understanding of The Evolution Misconceptions Analysis of Middle · High School Students: Use The Evolving Concept Aaas Item* (Unpublished Master's Thesis). Korea National University of Education, Cheongju, Republic of Korea.

Authors Information

Seong, Jeong-hwa: Kangwon National University High School, Teacher, First Author

Fadillah, Sarah Meilani: Kangwon National University, Graduate student, Co Author

Ha, Minsu: Kangwon National University, Professor, Corresponding Author