

Original Article

Comparison of Performance between Paper-Based and Computer-Based Tests on the Speech-Language Pathologist National Examination

Received: 25 January, 2025. **Revised:** 17 July, 2025. **Accepted:** 17 July, 2025.

Eun Kyoung Lee¹,
Jin-Ae Bae²,
Seong Hee Choi³

¹Department of Speech-Language Pathology, Dongshin University, Naju, Republic of Korea

²Department of Speech-Language Pathology, Luther University, Yongin, Republic of Korea

³Department of Audiology & Speech-Language Pathology, Daegu Catholic University, Kyeongsan, Republic of Korea

Purpose Despite the expansion of computer-based testing (CBT) for all healthcare professions in Korea, little is known about the applicability of CBT to the National Examination for Speech-Language Pathologists. This study aimed to examine the equivalence and relationship of test performance between CBT and PBT by Korean speech-language pathologists.

Methods Ninety-three speech-language pathologists (41 first-level and 52 second-level) participated in this study, and all subjects performed both PBT and CBT. The test consisted of four periods, and students were randomly divided into Groups A and B. The two groups were crossover-designed to take PBT and CBT. All items were five-item, multiple-choice questions.

Results The total score was not significantly different between PBT and CBT in the first-level certificate test. In contrast, the total score in the second-level certificate test was significantly higher in CBT than in PBT. Item discrimination and difficulty did not differ substantially between the two test modes. CBT total scores were positively correlated with PBT ($r = 0.852$).

Conclusions In this study, there was no performance difference between the two test modes for the first-level certificate examination, and CBT was superior for the second-level certificate examination. Therefore, this study suggests that CBT can be comparable with PBT. CBT can help students reflect and improve their performance on the national examination and may provide a rationale for switching to CBT mode.

Keywords PBT, CBT, Item discrimination, Validity, Difficulty, Speech-language pathologists

INTRODUCTION

Speech-language pathologists are professionals who diagnose, evaluate, intervene, and prevent speech, language, and communication problems that can occur across the lifespan and work to improve the language skills of people with communication problems, speech and language delays, and speech and language disorders. The professional competence of speech-language pathologists is essential because it directly affects people with communication problems, and there is a national examination to evaluate the professional competence of speech-language

pathologists. Currently, the National Examination for Speech-Language Pathologists is conducted using a paper-based test (PBT), which many believe has limitations in assessing the professional competence of speech-language pathologists.

Therefore, if CBT (Computer-Based Test) problem-solving and case-oriented questions are presented in various forms in the National Examination for Speech-Language Pathologists, it can be expected that the quality control of speech-language pathologists will be carried out by evaluating field-oriented practical skills when qualifying not only second-level speech-language pathologists but

Corresponding to Seong Hee Choi

Department of Audiology & Speech-Language Pathology, Research Institute of Biomimetic Sensory Control, and Catholic Hearing Voice Speech Center, Daegu Catholic University, 13-13, Hayang-ro, Hayang-eup, Gyeongsan, 38430, Republic of Korea

TEL. +82-53-850-2542, **FAX.** +82-53-359-6780, **E-mail.** shgrace@cu.ac.kr

also first-level speech-language pathologists. It will be a basis for presenting various forms and contents of questions close to actual clinical situations that are difficult to implement in the current paper-based examination.

This change in testing methodology will improve the quality, discrimination, and diversity of the National Examination for Speech-Language Pathologists, ultimately leading to a more qualified healthcare workforce. Therefore, the introduction of CBT to the SLP examination is warranted, and it is time to discuss the appropriate CBT modality, timing, and item development for the SLP examination.

Computer-based examinations allow automated test administration, including test administration, grading, and grade management, and present problem situations that are close to real-life situations beyond the limitations of paper-based examinations while utilizing multimedia elements such as video, voice, 3D, and animation to provide multifaceted assessment and efficient question production, as well as the advantages of reduced testing errors, convenience in scoring, and security (Park, 2021; Sung, 2010; Lim, 2014; Rudland et al., 2011).

In addition, compared to traditional paper-based tests (PBTs), CBTs can more easily include embedded multimedia elements, providing a simulated experience for students to respond and interact with (Prisacari & Danielson, 2017). The built-in assistive technology options available in CBTs can even help students with disabilities complete tests independently by providing specific features such as captioning and text-to-speech (Boo & Vispoel, 2012; Kolen, 1999).

The Ministry of Employment and Labor has converted the written national technical qualification examination to a computer-based test (CBT) since 2014. It conducted all industrial engineer tests (114) in 2021 and all engineer and service tests (122) in 2022 through a computer-based test (<http://www.kukinews.com/newsView>). Replacing PBT with CBT requires interchangeable test score data based on comparison, and according to the International Guideline on CBT, a CBT method is considered reliable

when the assessment is administered in two ways and similar scores are obtained between the two methods (International Test Commission, 2006; Khoshshima et al., 2019). In this regard, the number of studies focusing on the equivalence of PBT and CBT scores has increased over the past few decades, as has the number of studies on test-takers' feedback on the two modalities. Studies comparing language performance between PBT and CBT have shown mixed results, with some showing that PBT students outperform CBT students and others showing no difference in performance between PBT and CBT (Chapelle, 2007; Jamieson, 2005; Kim et al., 2018). Some studies show no significant difference in writing scores between PBT and CBT (Mohammadi & Barzgaran, 2012; Oz & Ozturan, 2018), while a study that compared the performance of 365 Korean students in both PBT and CBT found that students scored higher on a listening task in CBT than in PBT (Choi et al., 2003). These results suggested that visual information on the CBT could explain this significant score difference. Furthermore, in a comparative study of writing tests, 73 intermediate Saudi female university students who took the PBT made more sentence-level errors than those who took the CBT, which the researchers predicted was due to the computerized features of the CBT, such as auto-correction and smart prediction, which resulted in fewer sentence-level and lexical grammar errors (AlKadi & Madini, 2019). Factors that may affect performance on PBT and CBT include computer familiarity and preference for delivery modality (Boeve et al., 2015; Jeong, 2014; Mizrachi, 2015), suggesting that participants' test results may be influenced by their familiarity with computers and their preference for delivery modality. One prior study found that CBT may be at a disadvantage for learners who were unfamiliar with the CBT interface or had limited access to a computer, with PBT test-takers outperforming CBT test-takers in a comparison of PBT and CBT on the GRE and test-takers who were more familiar with computers scoring higher than those who were not (Goldberg &

Pedulla, 2002). In contrast, a recent study found significantly higher scores in the CBT format than in PBT (Mangen et al., 2013).

In addition, middle school students were also found to have their reading test performance affected by their familiarity with computers (Odo, 2012). In contrast, some recent studies have not found any relationship between computer familiarity and test performance (Hosseini et al., 2014; Koshima et al., 2019). However, Khoshshima et al. (2019) found that Iranian university students' preference for CBT versus PBT for the TOEFL test was significantly higher. However, there was no significant difference in performance between the two modalities. Higgins et al. (2005) also investigated the testing modality preferences of 219 fourth-grade elementary school students. They preferred CBT, but the test results of both modalities were the same. A recent study on the performance and perceptions of CBT and PBT found that students' scores on the CBT version of the test were higher than those obtained on the PBT version (Washburn et al., 2017). Attitudes toward the use of computers play an essential role in the successful implementation of CBT, with some studies reporting that students have positive attitudes toward CBT (Al-Amri, 2009), while others have intense fears or concerns about using computers, and these attitudes toward computerized testing have been associated with lower performance on CBT (Balogun & Olanrewaju, 2016); however, few studies have reported the impact of attitudes or preferences toward computerized testing on test performance among health profession students.

This study aims to find an appropriate assessment method to evaluate the job competence of speech-language pathologists and analyze its validity. To this end, examining the necessity of introducing CBT to the National Examination for Speech-Language Pathologists and whether there is a difference in the performance of examinees when the existing PBT method is changed to CBT will reduce the burden on examinees before CBT is im-

plemented and reconsider the feasibility of changing the test method.

Therefore, this study aims to lay the foundation for the efficient and stable operation of the National Speech-Language Pathologist Examination by comparing the scores and time differences of examinees' performance according to the differences in test methods to identify problems that may arise from implementing CBT in advance.

RESEARCH METHODS

This study compared the effectiveness of introducing the CBT method to the National Examination for Speech-Language Pathologists and the differences between the existing PBT method and the CBT method, as well as the problems that may arise when switching from the PBT method to the CBT method, by conducting a direct comparison of the two methods.

The mock test was conducted as a paper-based test (PBT) and a computer-based test (CBT) by developing two sets of questions based on the existing national examination method of 5-item multiple-choice.

Participants

A total of 93 people (16 males, 77 females) voluntarily participated in this mock examination, with 41 passing the First-Level National Speech-Language Pathologist Examination (9th to 11th) within three years as of June 2023 and 52 passing the Second-Level National Speech-Language Pathologist Examination (9th to 11th). The mock test was held at one of the following locations: Seoul-Capital Region, Daegu Catholic University, Daegu University, and Dongshin University.

Research tools

Members of the committee with experience in taking

the national examination developed the test items. The items were the same as the existing test method, providing the five best answer choices for each subject of the first and second levels of the National Examination for Speech-Language Pathologists. One set was composed of PBT, and another was composed of CBT.

The simulation test compared the CBT format and PBT format (paired questions) for the test type used in the existing National Examination for Speech-Language Pathologists to determine the effectiveness of the test method. The stability of the CBT implementation, test-taker responses, and test administration time were checked by comparing the CBT and PBT formats of the existing five-item, best-answer multiple-choice test. The subjects and question formats of the practice tests are shown in Table 1.

Research procedures

The mock exam, which was conducted with the five best multiple-choice questions in the manner of the existing national exam, consisted of a total of 4 periods, and candidates were divided into two classes, A and B. The A class was cross-designed to take 4 PBTs of Type A in Periods 1 and 2 and 4 CBTs of Type B in Period 5, while the B class took 4 PBTs of Type B in Periods 1 and 2 and 4 CBTs of Type A in Period 5.

Among the candidates for the first-level speech-language pathologist simulation test, Class A took the test on three subjects (neurogenic language disorders, language development disorders, and fluency disorders) in Class 1 with PBT and three subjects (voice disorders, ar-

ticulatory phonological disorders, and clinical practice of speech-language pathology) in Class 2 with CBT. After a short break, students took the CBT for three subjects (neurogenic language disorders, language development disorders, and fluency disorders) in Period 3 and three subjects (voice disorders, articulatory phonological disorders, and clinical practice of speech-language pathology) in Period 4. The PBT consisted of 72 questions in 75 minutes in Period 1 and 68 in 70 minutes in Period 2. The CBT consisted of 75 minutes for 72 questions in Period 3 and 70 minutes for 68 questions in Period 4. All CBT sessions were conducted using personal laptops, and the CBT server automatically recorded the time to answer each question. Using a cross-design method, Class B took the test the same way as Class A.

The second-level speech-language pathologist practice test is designed to be identical to the first level, with the only difference being the composition of the subjects for each class (first and third: neurogenic language disorders, fluency disorders, and voice disorders; second and fourth: language developmental disorders and articulatory phonological disorders) and the number of questions for each subject.

The overall design, subjects, and question formats of

Table 2. Designing a Practice Test (Level 1, Level 2)

Fraction	1st Period	2nd Period	3rd Period	4th Period
A	PBT (Type A)	PBT (Type A)	CBT (Type A)	CBT (Type B)
B	PBT (Type B)	PBT (Type B)	CBT (Type B)	CBT (Type A)

Table 1. Practice Test Subjects and Question Formats

	Instruction	General Questions	Exam Subjects	Question Format
Level 1	1st and 3rd periods	72 questions	Neurogenic language disorders, Language development disorders, Fluency disorders	Multiple Choice
	2nd and 4th periods	68 questions	Voice disorders, Articulatory phonological disorders, Clinical practice of speech-language pathology	Traditional Questions
Level 2	1st and 3rd periods	80 questions	Neurogenic language disorders, Fluency disorders, Voice disorders	Multiple Choice
	2nd and 4th periods	70 questions	Language development disorders, Articulation disorders	Traditional Questions

these simulations are presented in Table 2.

The mock exam was conducted in the traditional 5-item best-answer format and consisted of four periods: two periods of PBT and two periods of CBT. The first and second periods of PBT and CBT consisted of five best-of-five questions and paired response questions (Table 2).

For the Level 1 Speech-Language Pathologist Mock Exam, there were two sessions (140 questions) in the traditional PBT format and two sessions (140 questions) in the CBT format, for a total of 280 questions. For the Level 2 Speech-Language Pathologist Mock Exam, there were 150 questions in the first and second periods of the PBT format and 150 questions in the first and second periods of the CBT format, for a total of 300 questions.

RESULTS

In order to evaluate the job competence of speech-language pathologists most objectively, this study compared the examinees' scores between PBT and CBT according to the examination method of the national examination and compared the scores for the overall scores and subject-specific scores of the First and Second Level Speech-Language Pathologists' Mock Examinations, respectively.

Comparisons between PBT and CBT scores

For the Level 1 Speech-Language Pathologist Mock

Exam, the mean and standard deviation of the total score of the PBT and CBT obtained by the candidates and the scores for each subject are shown in Table 3 and Figure 1. The mean total score was 81.4 ± 12.7 for PBT and 73.7 ± 12.5 for CBT, with PBT being 7.7 points higher, but there was a wide variation by subject. By subject, CBT scored slightly higher than PBT for neurogenic language disorders and language development disorders for fluency disorders, but other subjects showed similar scores within ± 0.1 points between the two tests.

A paired-sample t-test was conducted to determine whether there was a significant difference between the mean scores of the PBT and CBT for the first-level speech-language pathologist mock test candidates, and the results showed a significant difference in the first-period subjects of neurogenic language disorders, language development disorders, and fluency disorders. Specifically,

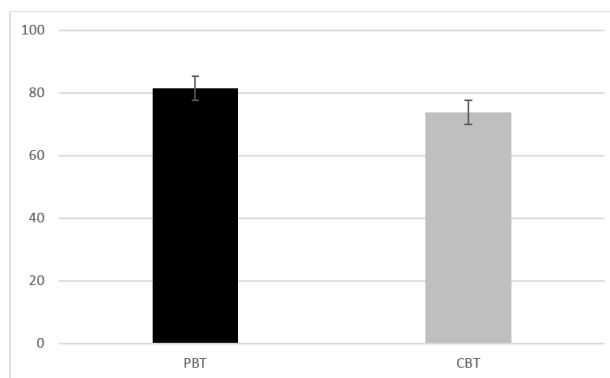


Figure 1. Comparisons of total scores between PBT and CBT scores of first-level certificate of speech-language pathologists.

Table 3. Mean Grade and Standard Deviation by Traditional Type of Course

(N=41)

Subject	PBT			CBT		
	Total Score	Average	Standard deviation	Total Score	Average	Standard deviation
Neurogenic Language Disorders	24	15.8	3.10	24	10.2	2.95
Language Development Disorders	24	17.3	3.31	24	13.5	2.48
Fluency Disorders	24	13.2	2.35	24	15.4	2.68
Voice Disorders	24	12.9	4.68	24	12.4	3.92
Articulatory Phonological Disorders	24	12.1	2.65	24	12.2	2.74
Clinical Practice of Speech-Language Pathology	20	10.1	2.54	20	9.90	2.62
All	140	81.4	12.7	140	73.7	12.5

neurogenic language disorders and language development disorders had significantly higher scores in PBT. In comparison, fluency disorders had significantly higher scores in CBT, and there was no significant difference between PBT and CBT for voice disorders, articulation

disorders, and speech rehabilitation field practice. PBT scores were significantly higher than CBT scores (Table 4), (Figure 1), (Figure 2).

Table 5 and Figure 2 show the means and standard deviations for the total score and subject scores of the

Table 4. Mean Difference between PBT and CBT (N=41)

	PBT		CBT		t-value
	Average	Standard Deviation	Average	Standard Deviation	
Neurogenic Language Disorders	15.8	3.10	10.2	2.95	11.675***
Language Development Disorders	17.3	3.31	13.5	2.48	6.554***
Fluency Disorders	13.2	2.35	15.4	2.68	-4.669***
Voice Disorders	12.9	4.68	12.4	3.92	0.846
Articulatory Phonological Disorders	12.1	2.65	12.2	2.74	-0.153
Clinical Practice of Speech-Language Pathology	10.1	2.54	9.90	2.62	0.473
All	81.4	12.7	73.7	12.5	6.227***

*** $p < .001$

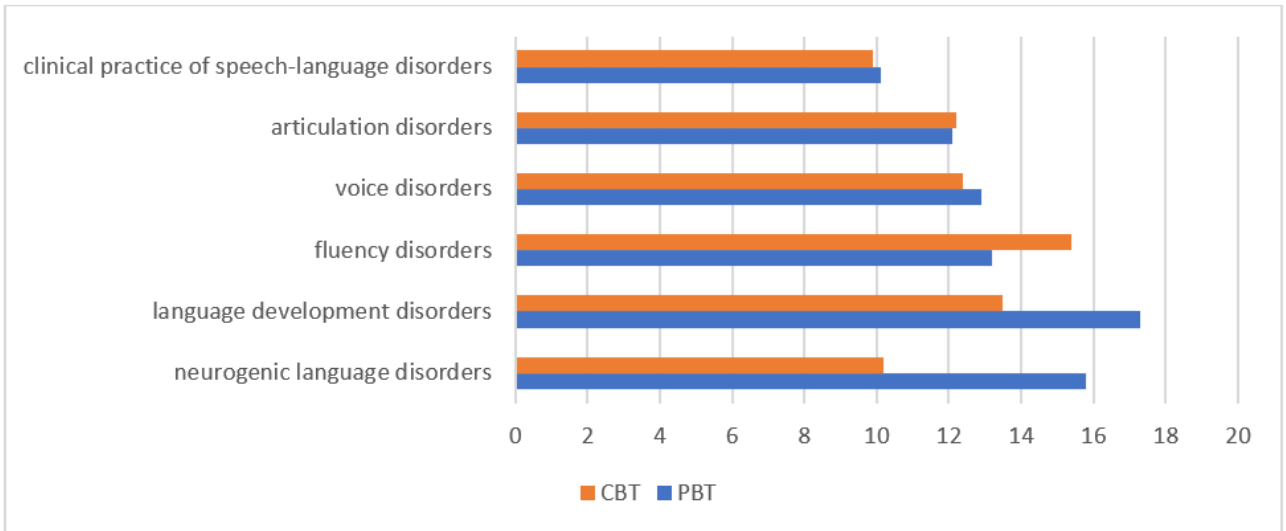


Figure 2. Comparisons of subject scores between PBT and CBT scores of first -level certificate of speech-language pathologists.

Table 5. Average Grade and Standard Deviation by Traditional Type Course (N=52)

Subject	PBT			CBT		
	Total Score	Average	Standard Deviation	Total Score	Average	Standard Deviation
Neurogenic Language Disorders	30	13.54	3.73	30	18.54	3.70
Fluency Disorders	25	16.02	3.11	25	12.52	2.78
Voice Disorders	25	14.10	3.05	25	13.25	3.59
Language Development Disorders	35	18.83	3.65	35	20.56	4.15
Articulatory Phonological Disorders	35	18.00	3.89	35	18.75	3.46
All	150	80.48	13.60	150	83.62	12.94

PBT and CBT for the second-level speech-language pathologists. The mean total score was 80.48 for PBT and 83.62 for CBT, and the mean difference between the two tests was insignificant. By subject, CBT scored slightly higher for neurolinguistic disorders and PBT for fluency disorders, but the rest of the subjects showed similar scores within ± 1 point between the two tests.

A paired-sample t-test was conducted to determine if there was a significant difference between the mean scores on the PBT and CBT for the Level 2 Speech-Language Pathologist Mock Examination candidates, and the results showed a significant difference between the two tests for all subjects except articulation and phonological disorders. Specifically, neurogenic language disorders and language development disorders scored significantly higher on the CBT, while fluency and voice disorders scored significantly higher on the PBT. The

overall score was significantly higher for CBT than PBT (Table 6), (Figure 3), (Figure 4).

Question reliability

Internal consistency analysis was conducted to examine the item reliability of the First-Level Speech-Language Pathologist Mock Test's PBT and CBT. The Cronbach's α values for each item and the overall score are shown in Table 7 and Table 8. The Cronbach's α for each First-Level Speech-Language Pathologist Mock Test was .854 for PBT and .840 for CBT.

Similarly, the subject-specific Cronbach's α for the Level 2 Speech-Language Pathologist Mock Test were slightly higher than Level 1, at .856 for the PBT and .853 for the CBT.

Table 6. Mean Difference between PBT and CBT

(N=52)

	PBT		CBT		t-value
	Average	Standard Deviation	Average	Standard Deviation	
Neurogenic Language Disorders	13.54	3.73	18.54	3.70	-10.100***
Fluency Disorders	16.02	3.11	12.52	2.78	9.258***
Voice Disorders	14.10	3.05	13.25	3.59	2.057*
Language Development Disorders	18.83	3.65	20.56	4.15	-3.294**
Articulatory Phonological Disorders	18.00	3.89	18.75	3.46	-1.659
All	80.48	13.60	83.62	12.94	-3.116**

* $p < .05$, ** $p < .01$, *** $p < .001$

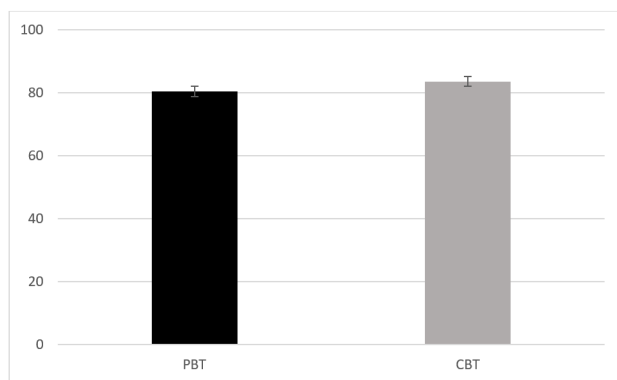


Figure 3. Comparisons of total scores between PBT and CBT scores of second-level certificate of speech-language pathologists.

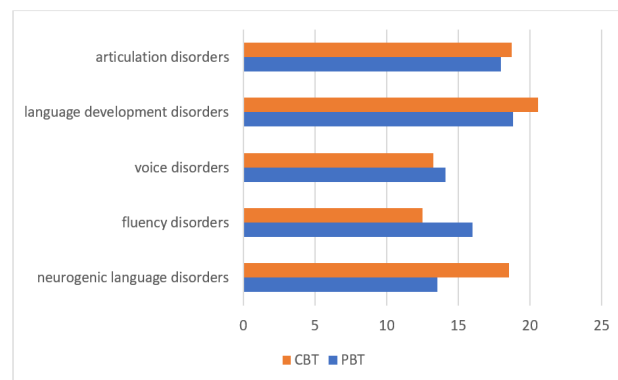


Figure 4. Comparisons of subject scores between PBT and CBT scores of second-level certificate of speech-language pathologists.

Table 7. Cronbach's α by Subject for the Level 1 Speech-Language Pathologist Mock Test

Subject	PBT	CBT
Neurogenic Language Disorders	0.596	0.526
Language Development Disorders	0.675	0.379
Fluency Disorders	0.282	0.553
Voice Disorders	0.804	0.699
Articulatory Phonological Disorders	0.444	0.584
Clinical Practice of Speech-Language Pathology	0.458	0.465
All	0.854	0.840

Table 8. Cronbach's α by Subject for the Level 2 Speech-Language Pathologist Mock Test

Subject	PBT	CBT
Neurogenic Language Disorders	0.581	0.575
Fluency Disorders	0.609	0.554
Voice Disorders	0.609	0.597
Language Development Disorders	0.609	0.652
Articulatory Phonological Disorders	0.581	0.592
All	0.856	0.853

Table 9. Correlation Coefficient between PBT and CBT by Subject (N=41)

Subject	Correlation Coefficient
Neurogenic Language Disorders	0.494***
Language Development Disorders	0.188
Fluency Disorders	0.291
Voice Disorders	0.747***
Articulatory Phonological Disorders	0.352*
Speech-Language Pathology Practice	0.599***
All	0.804***

* $p < .05$, *** $p < .001$ **Table 10.** Correlation Coefficient between PBT and CBT by Subject (N=52)

Subject	Correlation Coefficient
Neurolinguistic Disorders	0.538***
Fluency Disorders	0.576***
Voice Disorders	0.651***
Language Development Disorders	0.534***
Articulatory Phonological Disorders	0.613***
All	0.852***

*** $p < .001$

Question precision

Convergent validity analysis was conducted to examine the PBT and CBT item validity of the First-Level Speech-Language Pathologist mock Test. The convergent validity correlation analysis results between PBT and CBT for each subject are shown in Table 9 and Table 10. The overall score showed a high static correlation with a correlation coefficient of 0.804 between the two tests. Regarding individual subjects, voice disorders showed a high static correlation between PBT and CBT with a correlation coefficient of 0.747, followed by field practice with a correlation coefficient of 0.599, and neurogenic language disorders with a correlation coefficient of 0.494.

The overall score for the Level 2 Speech-Language Pathologist mock test showed a high statistic correlation of 0.852 between the two exams. Meanwhile, the individual items all showed somewhat higher statistic correlations ranging from 0.534 to 0.651.

Item difficulty and discrimination

According to the analysis of the difficulty and variance of the existing question types of PBT and CBT by certification level, the average difficulty of the questions in the First-Level Speech-Language Pathologist Practice Mock Test ranged from 50.37 to 72.26 for PBT and 45.21 to 64.13 for CBT. It can be seen that there is no significant difference in the difficulty level by question type, and the difficulty level is very similar for the second-period subjects of voice disorders, articulation and phonological disorders, and speech-language pathology field practice. The mean of item discrimination by subject ranged from 0.12 to 0.33 for PBT and 0.12 to 0.22 for CBT (Table 11). Overall, discrimination did not differ by question type.

On the other hand, the subject-specific mean of item difficulty for the Level 2 Speech-Language Pathologist Mock Test ranged from 51.18 to 64.07 for PBT and 50.55 to 61.78 for CBT, indicating that there is no significant difference in difficulty based on question type. The sub-

Table 11. PBT and CBT Difficulty and Variance by Subject

(N=41)

Subject	PBT		CBT			
	Difficulty	Level	Variability	Difficulty	Level	Variability
Neurogenic Language Disorders (24 questions)	65.75		0.18	45.21		0.20
Language Development Disorders (24 questions)	72.26		0.18	57.32		0.16
Fluency Disorders (24 questions)	55.18		0.13	64.13		0.17
Voice Disorders (24 questions)	53.56		0.33	52.24		0.22
Articulatory Phonological Disorders (24 questions)	50.41		0.12	50.61		0.22
Clinical Practice of Speech-Language Pathology (20 questions)	50.37		0.19	50.98		0.12
All (140 questions)	57.92		0.19	53.41		0.18

Table 12. PBT and CBT Difficulty and Variance by Subject

(N=52)

Subject	PBT		CBT			
	Difficulty	Level	Variability	Difficulty	Level	Variability
Neurogenic Language Disorders (30 questions)	53.89		0.32	61.78		0.31
Fluency Disorders (25 questions)	64.07		0.32	50.55		0.28
Voice Disorders (25 questions)	56.62		0.37	53.66		0.35
Language Development Disorders (35 questions)	54.19		0.26	58.42		0.31
Articulatory Phonological Disorders (35 questions)	51.18		0.27	53.99		0.27
All (150 questions)	55.99		0.31	55.68		0.30

ject-specific means for item variance ranged from 0.26 to 0.37 for PBT and 0.27 to 0.35 for CBT (Table 12). Overall, discrimination does not seem to differ by question type.

DISCUSSION & CONCLUSIONS

This study aimed to develop items for administering a computer-based national examination of speech-language pathologists and provide evidence of the feasibility of using computer-based testing in national examinations.

To develop items for the computer-based test, a workshop was held for the development committee members to familiarize them with the purpose, direction, and technical methods of the computer-based test, and then the items were developed. The developed CBT items were verified by conducting mock tests for the CBT and PBT methods for the national examination candidates. The mock test was conducted with a total of 93 test-takers.

The results of the present study showed that the total score of the First-Level Practice Test did not show a significant difference between the two delivery methods as in previous studies (Mohammadi & Barzgaran, 2012; Oz & Ozturan, 2018), but for the second-level speech-language pathologists, the total score was significantly higher in CBT than in PBT (AlKadi & Madini, 2019; Choi et al., 2003). In other words, the test scores of first-level SLPs in this study were not affected by the test delivery method, indicating the equivalence of the two test modalities. However, there was a difference between the two test modalities at Level 2, with Level 2 SLPs performing significantly better on the CBT. This may be because most participants were in their mid-20s, were highly familiar with computers, and had a positive attitude toward the change in testing methods. Therefore, it is likely that this familiarity and attitude towards computerized testing reduced their resistance to computerized testing and had a positive effect on their performance. In addition, the CBT was designed to show only one question per com-

puter screen, and this visual information may have helped test-takers focus on a given test question and recognize information on the screen more quickly. Therefore, this advantage of the computerized method may have contributed to the students' higher scores. In addition, all students took the PBT test in the first period and the comparable CBT test in the second period, with the test questions being of similar quality, so it is unlikely that there was a learning effect of the test questions. In addition, the difficulty and discrimination of the questions can affect PBT and CBT performance, but this study did not show any significant differences between PBT and CBT. For example, the overall item difficulty of the Level 1 practice test was 57.92 for PBT and 53.41 for CBT, which was not significantly different between the two test methods. In contrast, the overall discrimination rate of the Level 1 practice test was 0.19 for PBT and 0.18 for CBT, which was very similar. Likewise, the difficulty of the Level 2 Speech-Language Pathologist Practice Test was 55.99 on the PBT and 55.68 on the CBT, and the item discrimination was 0.31 on the PBT and 0.30 on the CBT. Therefore, neither item difficulty nor item discrimination appeared to affect performance on the test at either Level 1 or Level 2.

In conclusion, this study showed no difference in performance between the two modalities for the Level 1 SLP Test and that the CBT modality was superior in performance for the Level 2 SLP Test. Therefore, the results of this study suggest that CBT can help students demonstrate their abilities and improve their performance on the national examination and provide a rationale for switching to the CBT modality.

Based on the results of this study, the following conclusions were drawn from the CBT and PBT mock tests for introducing a computer-based National Examination for Speech-Language Pathologists.

First, it is necessary to analyze the item characteristics between the paper-based and computer-based tests in the National Speech-Language Pathologist Examination Practice

Tests to determine whether the test items align with the item response theory and whether any items need to be revised or supplemented. The difficulty and discrimination of the paper-based and computer-based tests were calculated, and the results showed that the difficulty of the items in the Speech-Language Pathologist Examination Practice Test was independent of the test format. This finding suggests that candidates were not challenged by the format of the test, whether paper-based or computer-based. Interestingly, in the first round of mock exams, a significant difference in scores between PBT and CBT was observed in the first period for both Level 1 and Level 2. However, the difference was reduced or became insignificant in the second period. These results suggest that for CBT to stabilize at a level similar to traditional PBT, candidates need sufficient time to familiarize themselves with CBT, which requires sufficient training and practice tests.

Second, the items were relatively easy to manipulate, and the implementation of the items was not technically challenging. However, some items may require more technical support depending on the type of item. In addition, the time required to take the CBT and PBT tests does not seem to be significantly different, so making any changes between the two tests is unnecessary. In addition, there may be differences in the types of answers between the two test types, with the PBT requiring using a single best answer type while the CBT may include questions requiring multiple answers. Therefore, differences in evaluation methods or scoring criteria may cause differences between the two types of tests, so if the scoring method is reviewed to check for consistency or fairness between the two tests and guidelines for scoring methods or criteria are provided, the level and evaluation method of the national examination for assessing speech-language pathologists' job competence can be utilized more diversely. In the future, it will be necessary to collaborate with experts in the field to evaluate and revise the test items to ensure that they fully reflect the

test-taker's expertise and experience and that they are diverse enough in terms of test-taker distribution and difficulty level to assess candidates' skills accurately.

The study's recommendations include the following:

First, although the usefulness of computer-based practice test results has been fully confirmed, the successful introduction and establishment of a computer-based national test will require prior preparation, not only in terms of item development but also in test format.

Second, CBT questions are structurally different from paper-based tests. Hence, development guidelines such as the difficulty level and allocation ratio for each question and the scoring method according to the difficulty level are necessary. In addition, one should consider gradually increasing the proportion of questions, and if one includes multimedia questions, one should specify the file extension and resolution according to their characteristics. In addition, regulations such as copyright and information protection laws should be in place for materials used to develop questions.

Third, transitioning from paper-based to CBT requires sufficient guidance and training. It is necessary to consider how to operate a familiarization training system to adapt to the CBT method and new types of questions.

This study has limitations, including the lack of a wider variety of question types in the mock exam. In addition, the number of test-takers was limited to 93, which requires pre-training and response analysis for all prospective test-takers. To overcome these limitations, further research is needed to improve the implementation technology of computerized tests.

ACKNOWLEDGMENT

This paper was funded by the research project 'Exploration of Question Types and Question Development for the Introduction of Computer-Based Speech-Language Pathologist National Examination', which was conducted

in 2023 as a service project of the Korea Healthcare Personnel Licensing Examination Institute (KHPLI) (RE02-2307-24).

REFERENCES

1. Al-Amri S. Computer-based testing vs. Paper-based testing: establishing the comparability of reading tests through the revolution of reading tests a new comparability model in a Saudi EFL context [dissertation]. Colchester: University of Essex; 2009.
2. AlKadi SZ, Madini AA. EFL learners' lexico-grammatical competence in paper-based vs. computer-based in genre writing. *Arab World English Journal*. 2019;5:154-75.
3. Balogun AG, Olanrewaju AS. Role of computer self-efficacy and gender in computer-based test anxiety among undergraduates in Nigeria. *Psychological Thought*. 2016; 9(1):58-66.
4. Boevé AJ, Meijer RR, Albers CJ, Beetsma Y, Bosker RJ. Introducing computer-based testing in high-stakes exams in higher education: results of a field experiment. *PLOS ONE*. 2015;10(12):e0143616.
5. Boo J, Vispoel. W. Computer versus paper-and-pencil assessment of educational development: a comparison of psychometric features and examinee preferences. *Psychol Rep*. 2012;111(2):443-60.
6. Carter MD. The effects of computer-based simulations on speech-language pathology student performance. *J Commun Disord*. 2019;77:44-55.
7. Chapelle CA. Technology and second language acquisition. *Annu Rev Appl Lingui*. 2007;27:98-114.
8. Choi IC, Kim KS, Boo J. Comparability of a paper-based language test and a computer-based language test. *Language Testing*. 2003;20(3):295-320.
9. Goldberg AL, Pedulla J. Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educ Psychol Meas*. 2002;62(6): 1053-67.
10. Han C, Kim M, Hwang B, Jung M. Implementation plan and item development and evaluation of computer-based national examination for oriental medicine practitioners. Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2021.
11. Han NY, Lee JY, Kwak HS, Lee BK, Lee SH, Yong CS, Kim JH, Oh JM. An expert survey for the selection of practical examination areas for the Pharmacist Board Examination. *Korean J Clin Pharm*. 2017;27(3):119-26.
12. Hassanien MA, Al-Hayani A, Abu-Kamer R, Almazrooa A. A six step approach for developing computer based assess-

- ment in medical education. *Med Teach*. 2013;35(sup1): S15-9.
13. Heo S. Computerized testing and question bank database construction in medical education. *Journal of Medical Education*. 2014;16(1):11-5.
 14. Heo S, Lee DS, Lim JH, Jeong EY, Heo Y. Measurement review, item bank construction, and administrative and financial support for the introduction of computerized adaptive testing to the Korean National Healthcare Workers' Examination. Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2018.
 15. Higgins CA, Gray G, Symeonidis P, Tsintsifas A. Automated assessment and experiences of teaching programming. *Journal on Educational Resources in Computing*. 2005; 5(3):5-es.
 16. Hochlehnert A, Brass K, Moeltner A, Juenger J. Does medical students' preference of test format (computer-based vs. paper-based) have an influence on performance? *BMC Medical Education*. 2011;11(1):1-6.
 17. Hols-Elders W, Bloemendaal P, Bos N, Quaak M, Sijstermans R, De Jong P. Twelve tips for computer-based assessment in medical education. *Med Teach*. 2008;30(7):673-8.
 18. Hosseini M, Abidin MJZ, Baghdarnia M. Comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. 2014. *Procedia – Soc Behav Sci*. 2014;98:659-67.
 19. Hwang K, Shin S, Kim N, Ryu J, Lee C, Lee S, Hwang I. A study on multimedia item development for computerized test application of the national examination for dental technicians. Seoul (South Korea): Korea Healthcare Personnel National Examination Center; 2020.
 20. International Test Commission. International guidelines on computer-based and internet-delivered testing. *Int J Test*. 2006;6(2):143-71.
 21. Jamieson J. Trends in computer-based second language assessment. *Annu Rev Appl Linguist*. 2005;25:228-42.
 22. Jeong H. A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology*. 2014;33(4):410-22.
 23. Karay Y, Schaubert SK, Stosch C, Schüttpeitz-Brauns K. Computer versus paper-does it make any difference in test performance? *Teach Learn Med*. 2015;27(1):57-62.
 24. Khoshsima H, Toroujeni SM, Thompson N, Ebrahimi MR. Computer-based (CBT) vs. paper-based (PBT) testing: Mode effect, relationship between computer familiarity, attitudes, aversion and mode preference with CBT test scores in an Asian private EFL context. *Journal of Teaching English with Technology*. 2019;1:86-101.
 25. Kim HR, Bowles M, Yan X, Chung SJ. Examining the comparability between paper- and computer-based versions of an integrated writing placement test. *Assessing Writing*. 2018;36:49-62.
 26. Kim YT, Kim JO, Jeon HS, Choi HJ, Kim MJ, Tae W, Kang MK. A job analysis study of speech-language pathologists through a Delphi survey. *Speech Therapy Research*. 2014; 23(3):147-61.
 27. Kolen MJ. Threats to score comparability with applications to performance assessments and computerized adaptive tests. *Educational Assessment*, 1999;6(2):73-96.
 28. Korea Manpower Institute. Industrial Engineer CBT test held for the first time, test takers highly satisfied. Press release. Daejeon (South Korea): Korea Manpower Institute; 2020.
 29. Lee JY, Kim KI, Kim KD, Kim MK, Kim EY, Kim JK, Kim HA, Moon Y, Ji Y, Jeong YS, Lee JY, Jung EK, Chun KH, Choi KH. A study on the development of various item types that can be implemented in computerized examinations. Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2022.
 30. Lim MK, Bo H, Jung K, Jun K. A study on the question management system of foreign healthcare licensure examinations. Seoul (South Korea): Korea Healthcare Personnel National Examination Service; 2015.
 31. Lim MK, Kim YH, Song KR, Kim HM, Lee HW, Lee SY. Basic research on computerized test administration. Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2006.
 32. Lim MK, Min H, Heo S. A study on the development of computer-based national healthcare personnel examination. Seoul (South Korea): Korea Healthcare Personnel National Examination Center; 2014.
 33. Mangen A, Walgermo BR, Brønnick K. Reading linear texts on paper versus computer screen: effects on reading comprehension. *Int J Educ Res*. 2013;58:61-8.
 34. Mizrachi D. Undergraduates' academic reading format preferences and behaviors. *J Acad Libr*. 2015;41(3):301-11.
 35. Mohammadi M, Barzgaran M. Comparability of computer-based and paper-based versions of writing section of PET in Iranian EFL context. *J Foreign Lang Teach Transl Stud*. 2012;1(2):1-20.
 36. Odo DM. Computer familiarity and test performance on a computer-based cloze ESL reading assessment. *Teaching English with Technology*. 2012;12(3):18-35.
 37. Öz H, Özturan T. Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *J Lang Linguist Stud*. 2018;14(1):67-85.
 38. Park HK, Kim DH, Kang YJ, Bae SY, Moon YJ. A study of test-taker fatigue following computer-based testing (CBT). Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2021.
 39. Priscari AA, Danielson J. Computer-based versus paper-based testing: investigating testing mode with cognitive load and scratch paper use. *Comput Human Behav*.

- 2017;77:1-10.
40. Rudland JR, Schwartz P, Ali A. Moving a formative test from a paper-based to a computer-based format. A student viewpoint. *Med Teach*. 2011;33(9):738-43.
41. Sung T. *Contemporary Educational Assessment*. 3rd ed. Seoul: Hakji; 2010.
42. Washburn S, Herman J, Stewart R. Evaluation of performance and perceptions of electronic vs. multiple-choice paper exams. *Adv Physiol Educ*. 2017;41(4):548-55.
43. Yoo SG, Won JS, Lee GJ, Choi HK, Han DK, Lee HS, Sung HJ, Jung IH, Eun J. A feasibility study on MAT item development by healthcare occupation and SBT national test system. Seoul (South Korea): Korea Healthcare Personnel National Testing Center; 2016.

Appendix 1. Traditional National Board Exam CBT (5-item best answer)

온라인 시험 응시

35 문항 35 점 만점

064 : 57

Q12 객관식 1점

전방성(anterior) 자질을 갖는 자음으로 묶인 것은?

☐ /ㅂ, ㅁ, ㄴ, ㅌ/

☐ /ㅅ, ㅆ, ㅁ, ㅋ/

☐ /ㄱ, ㄷ, ㄴ, ㅍ/

☐ /ㅁ, ㅂ, ㅅ, ㅍ/

☐ /ㅂ, ㄱ, ㅅ, ㅈ/

문제 목록

01번 | 객관식(1)

02번 | 객관식(1)

03번 | 객관식(1)

04번 | 객관식(1)

05번 | 객관식(1)

06번 | 객관식(1)

07번 | 객관식(1)

08번 | 객관식(1)

09번 | 객관식(1)

10번 | 객관식(1)

11번 | 객관식(1)

12번 | 객관식(1)

13번 | 객관식(1)

14번 | 객관식(1)

15번 | 객관식(1)

16번 | 객관식(1)

17번 | 객관식(1)

18번 | 객관식(1)

19번 | 객관식(1)

20번 | 객관식(1)

온라인 시험 응시

35 문항 35 점 만점

제한없음

Q6 객관식 1점

다음 단기목표들에 대한 장기목표로 가장 적절한 것은?

<보기>

단기목표1. 이유/원인을 나타내는 연결어미(-아/어서, -니까)를 산출할 수 있다.

단기목표2. 도구격조사(-로)를 산출할 수 있다.

단기목표3. 목적격조사(-을/-를)를 산출할 수 있다.

☐ 다양한 두 낱말의미관계 유형을 산출할 수 있다.

☐ 표현어휘를 3세 수준으로 향상시킨다.

☐ 평균발화길이(MLU-m)를 향상시킨다.

☐ 접속문과 내포문을 산출할 수 있다.

☐ 다양한 의사소통기능을 표현할 수 있다.

문제 목록

01번 | 객관식(1)

02번 | 객관식(1)

03번 | 객관식(1)

04번 | 객관식(1)

05번 | 객관식(1)

06번 | 객관식(1)

07번 | 객관식(1)

08번 | 객관식(1)

09번 | 객관식(1)

10번 | 객관식(1)

11번 | 객관식(1)

12번 | 객관식(1)

13번 | 객관식(1)

14번 | 객관식(1)

15번 | 객관식(1)

16번 | 객관식(1)

17번 | 객관식(1)

18번 | 객관식(1)

19번 | 객관식(1)

20번 | 객관식(1)