An Analysis on the Characteristics of the Items in Paper-Based Test (PBT) and Smart Device-Based Test (SBT) - Focused on the Optician National Board Examination

Eun Joo Kim [1] and Jung Un Jang [2]

[1]*Professor Kim Eun Joo, Faculty of Liberal Arts, South Korea*
[2]*Professor Jung Un Jang, Department of Optometry, Eulji University, South Korea*
[1]*kej70@eulji.ac.kr,* [2]*jju@eulji.ac.kr*

Abstract

This study is aimed to analyze the characteristics of the items in the optician national board examination according to item response theory and to find if the test items are adequate or need to be modified for the evaluation of learners' ability level. Through the item analysis based on item response theory, it is possible to estimate an error of measurement accurately according to individual learner's ability level. In a design step, paper-based and smart device-based trial tests for the optician national board examination were planned to show multimedia type, data-suggestion type, and text type items to test takers. As a result, item difficulty and item discrimination were not greatly different between the paper-based and smart device-based trial tests for the optician national board examination. In particular, to collect opinions of the persons concerned for the introduction of smart device-based tests and to stabilize the introduction of the smart device-based national board examination perfectly, it is necessary to conduct various pilot projects to support experiences of the smart device-based test. To develop multimedia items to encourage the problem-solving ability of people with different kinds of occupations, it is necessary to establish information use guidelines that help to access the object and standardized data which do not infringe an information use policy, among internet-sharable video needed to develop items.

Keywords: Paper-Based Test (PBT), Smart device-based test (SBT), National board examination, Item response theory

1. Introduction

In particular, with respect to the ophthalmological roles, the public demands an improvement in the quality level of opticians. The most effective way to improve and verify opticians' quality level is a national licensing examination system. [1][2][3] For this reason, a national licensing examination should guarantee validity and reliability and accurately assess the learning process and ability of opticians. In other words, it is necessary to develop the evaluation method of inferring learners' understanding and thinking processes for objects based on their active responses, and of finding their cognitive level. In the US where the computer-based national evaluation of achievements is performed, the US medical licensing examination and the US-registered nurse licensing examination are based on computer. [4][5][6][7][8][10][11][12] In Korea, some schools carry out a computer-based evaluation for national scholastic aptitude

---

examination on a trial basis (Seoul Metropolitan Office of Education and Research Information Service, 2008). Nevertheless, the Smart Device-Based Test (SBT) can have a variety of multimedia items that are not supported by the PBT. Compared to the PBT, the SBT delivers information in the form of images and videos, rather than texts only. The SBT requires learners to have the minimum linguistic understanding so that it is possible to reduce the difference between item responses in the same test items. [12][13][14][15][16][17][18]

This study developed the PBT and SBT items for the equal evaluation items of a trial test for the optician national licensing examination and analyzed the items based on Item Response Theory (IRT). Through the item response theory is not based on classical test theory in which an error of measurement is estimated with the use of test reliability, but makes it possible to calculate an error of measurement accurately in line with individual learners' ability level. Also tried to analyze the characteristics of the items in the optician national licensing examination according to item response theory and to find if the test items are adequate or need to be modified for the evaluation of learners' ability level. In a design step, paper-based and smart device-based trial tests for the optician national licensing examination were planned to show multimedia type, data-suggestion type, and text type items to test takers. It tried to look into how these types of items influence the item difficulty.

2. Method

2.1. Study subjects

Paper-based and smart device-based trial tests for the optician national licensing examination were conducted, respectively. In March 2018, sixty third-year students majoring in optical science of the four-year E university situated in Seongnam, Gyeonggido, took the trial tests.

Table 1. General characteristics of study subjects (N=60)

Observed variables		Frequency (persons)	Percentage (%)
Gender	Male	26	43.3
	Female	34	56.7
GPA in previous semester	Less than 2.0	0	.0
	Between 2.1 and 3.5	40	66.7
	Between 3.6 and 4.0	13	21.7
	Between 4.1 and 4.5	7	11.7

2.2. Data analysis

The characteristics of the items of the optician national licensing examination were analyzed in Item Response Theory (IRT). Unlike classical test theory based on the total score of a test, the IRT tries to analyze the attributes of items according to the unique Item Characteristic Curve of each item that is incorporated into a test. For the parameter estimation by the IRT, Bayesi An Program was applied to analyze items according to Bayesian statistics. The program can analyze a one-parameter item response model and a two-parameter item response model. This study selected a two-parameter item response model to estimate item difficulty, item discrimination, and Cronbach'a reliability. Also, as a method of analyzing test takers' abilities, EAP (Bayesian 'Expected a Posteriori' Estimation) and MAP (Bayesian Modal Estimation) were applied. Thus, based on the IRT, this study calculated and compared the item difficulty

and item discrimination of the items in the paper-based and smart device-based trial tests for optician national licensing examination. As shown in [Table 2], according to the IRT, the linguistic terms of item difficulty can be classified into five types: 'very easy', 'easy', 'neither difficult nor easy', 'difficult', and 'very difficult'. The classification is not an absolute standard. The data analysis used the Bayesian 1.0 for Windows program. The detailed analysis method automatically proceeds when the subject enters the BayesiAn program with the data file answered the question. When the analysis is conducted by the program, a total of four files are created. A data file should have an extension of *.dat or *.txt as a data file for question analysis. The input data files include the number of questions, the number of ID columns and the correct answers, the number of answers, the choice code whether to include the analysis of the questions or not and the response data.

Table 2. The scope of item difficulty in linguistic expressions

Linguistic expression	Item difficulty index
Very easy	-2.0 or less
Easy	-2.0 to -.5
Neither difficult nor easy	-.5 to +.5
Difficult	+.5 to +2.0
Very difficult	+2.0 or more

In item response theory, item difficulty means the point of one's ability level in which the probability of answering an item correctly is 0.5, and it is denoted by β or b. An item answered correctly by a student with a high level of ability is difficult, whereas an item answered correctly by a student with a low level of ability is easy. Just as the difficulty of classical test theory, item difficulty represents a level of item difficulty, which ranges from -2 to +2.

Item discrimination represents the slope of an item characteristic curve that shows item difficulty, and it is denoted by α or a. Generally, the slope of an item characteristic curve is the steepest and straight at ±0.5 of item difficulty. Therefore, the slope of the point on an item characteristic curve that presents item difficulty is estimated to be item discrimination. An item discrimination value is mostly a positive number but sometimes is found to be a negative number. An item whose item discrimination is a negative number reduces the probability of a correct answer with a rise in a test taker's ability level. Therefore, such an item needs to be removed. Table 3 presents the linguistic expressions according to the item discrimination index in the logistic model.

Table 3. The scope of item discrimination in linguistic expressions

Linguistic expressions	Item discrimination index
None	0.00
Almost none	0.01 to 0.34
Low	0.35 to 0.64
Appropriate	0.65 to 1.34
High	1.35 to 1.69
Very high	1.70 or more
Perfect	$+\infty$

3. Results

3.1. Difficulty and discrimination of the items in paper-based and a smart device-based trial test

Table 4. Item difficulty and item discrimination

Paper-Based Test				Smart Device-Based Test			
Difficulty	Discrimination	Ability	Reliability	Difficulty	Discrimination	Ability	Reliability
M (SD)	M (SD)	M (SD)	Cronbach'α	M (SD)	M (SD)	M (SD)	Cronbach'α
.621 (1.77)	.979 (.348)	-.148 (.644)	.790	.714 (2.60)	.951 (.307)	-.134 (.546)	.695

Regarding the item difficulty of the paper-based and smart device-based trial tests for the optician national licensing examination, the item difficulty of the PBT was .621, and that of the SBT was .714, both of which were 'difficult'. Therefore, item difficulty was not different between the two trial tests. About their item discrimination, the PBT had. 979, and the SBT had. 951, both of which were 'appropriate' in the scope of item discrimination. Therefore, item discrimination was not different between the two trial tests.

3.2 The difficulty and discrimination of the smart device-based test items in item response theory

The item difficulty and item discrimination values of the smart device-based test items for the optician national licensing examination were calculated. The results are presented in Table 5. In Bayesian statistics, the difficulty, discrimination, and estimation error of the estimated items, and the maximum item information of each item were analyzed. As a result, the no. 2 items were found to be the most difficult (15.066), and the no. 71 item was the easiest (–2.853). The item with the best discrimination was the no. 66 item (1.926).

Table 5. The item difficulty and item discrimination of the smart device-based test items

Item no.	item discrimination	Error of estimation for item discrimination	item difficulty	Error of estimation for item difficult	Max item info.
1	.832	.376	-.059	.321	.173
2	1.000	.500	15.066	192.777	.250
3	.554	.448	2.617	1.326	.077
4	.747	.394	-.712	.417	.139
5	.650	.473	4.497	2.282	.105
6	.820	.378	-.144	.324	.168
7	.728	.409	1.428	.722	.133
8	.864	.371	-.144	.308	.187
9	.634	.412	-.471	.436	.100
10	.768	.400	-1.299	.567	.147

11	.816	.382	-.579	.363	.166
12	.909	.464	3.718	1.823	.207
13	.731	.409	1.422	.719	.134
14	.528	.438	1.498	.871	.070
15	.649	.409	.393	.457	.105
16	.748	.427	-2.380	1.005	.140
17	.825	.415	2.060	.941	.170
18	.900	.410	2.053	.913	.202
19	.638	.411	-.360	.422	.102
20	.574	.434	1.666	.902	.082
21	.568	.425	.466	.526	.081
22	.673	.406	-.663	.444	.113
23	1.017	.374	-1.403	.516	.258
24	1.012	.350	.142	.281	.256
25	.823	.427	2.430	1.116	.169
26	1.155	.380	-1.779	.621	.333
27	1.043	.425	2.551	1.125	272
28	1.535	.300	-.403	.193	.589
29	.527	.455	2.975	1.511	.069
30	.603	.427	1.441	.797	.091
31	.896	.452	3.289	1.563	.201
32	1.001	.352	-.217	.269	.251
33	.711	.399	.451	.435	.126
34	1.321	.344	1.124	.437	.436
35	.539	.431	.628	.586	.073
36	1.049	.346	-.283	.261	.275
37	.867	.374	-.556	.341	.188
38	.689	.408	1.016	.599	.119
39	.695	.400	.258	.409	.121
40	1.000	.500	15.066	192.777	.250
41	.833	.376	-.144	.319	.174
42	.650	.420	1.467	.778	.105
43	1.255	.324	.343	.259	.394
44	.869	.387	1.195	.580	.189
45	.820	.380	-.489	.348	.168
46	.963	.379	1.283	.578	.232
47	.766	.447	3.101	1.478	.147
48	1.003	.373	1.233	.549	.252
49	.912	.364	-.222	.294	.208
50	.770	.440	2.822	1.330	.148
51	.829	.440	2.874	1.342	.172
52	.752	.414	1.778	.849	.141
53	.747	.391	-.331	.361	.140
54	.646	.421	1.474	.783	.104
55	1.198	.373	1.569	.624	.359
56	1.366	.418	-2.508	.937	.466

57	.716	.396	-.241	.370	.128
58	.656	.436	2.389	1.163	.108
59	.892	.370	-.546	.331	.199
60	.607	.447	2.791	1.378	.092
61	1.048	.352	-.711	.317	.275
62	1.231	.386	-1.945	.675	.379
63	1.075	.345	.410	.306	.289
64	.757	.398	1.028	.569	.143
65	1.083	.380	1.588	.659	.293
66	1.926	.286	.563	.218	.927
67	1.338	.320	-.609	.244	.447
68	1.738	.374	-1.907	.584	.755
69	.990	.362	.778	.412	.245
70	1.891	.403	-2.222	.730	.894
71	1.164	.435	-2.853	1.153	.339
72	1.059	.386	-1.780	.644	.281
73	.774	.463	3.788	1.864	.150
74	.961	.389	-1.682	.633	.231
75	1.175	.330	.108	.243	.345
76	1.086	.418	-2.506	.973	.295
77	1.786	.387	-2.059	.652	.797
78	1.060	.397	-2.040	.754	.281
79	1.277	.362	-1.542	.504	.408
80	1.112	.359	-1.222	.432	.309
81	.869	.389	1.304	.618	.189
82	1.539	.308	-.729	.237	.592
83	1.216	.337	-.848	.310	.369
84	.884	.453	3.334	1.590	.195
85	.790	.388	-.776	.412	.156
86	1.624	.295	.360	.213	.660
87	.963	.368	-1.001	.408	.232
88	1.087	.348	.628	.349	.295
89	.721	.404	1.082	.604	.130
90	1.118	.346	-.824	.325	.313
91	.776	.386	-.054	.342	.151
92	1.482	.301	.064	.195	.549
93	.661	.420	1.576	.813	.109
94	1.168	.368	-1.544	.527	.341
95	.925	.373	.926	.473	.214
96	1.460	.304	.175	.208	.533
97	.641	.461	3.694	1.829	.103
98	1.257	.321	-.205	.218	.395
99	.575	.435	1.818	.962	.083
100	.938	.365	.562	.373	.220

4. Discussion and Conclusion

In this study, the characteristics of the items in the trial tests for the optician national licensing examination were analyzed according to item response theory. In the analysis, whether there are any items to be modified in each type of trial test to evaluate learners' ability level was analyzed. As a result, item difficulty and item discrimination were not greatly different between the paper-based and smart device-based trial tests for the optician national licensing examination. Given that the items developed with multimedia were found to be good in terms of discrimination, it was possible to relatively find well learners' ability level. As such, an optician national licensing examination should be able to evaluate one's ability to solve a problem in a real job context. The current paper-based test with selection type can mainly evaluate how much test-takers memorize or understand relevant knowledge of jobs, rather than how much they solve a problem. To apply the practicability or reality of relevant jobs to the examination, it is necessary to seek a policy to introduce the smart device-based national licensing examination system. In particular, to collect opinions of the persons concerned for the introduction of smart device-based tests and to stabilize the introduction of the smart device-based national licensing examination perfectly, it is necessary to conduct various pilot projects to support experiences of smart device-based tests. It is required to make many efforts to improve functions in terms of smart device technologies, instead of prematurely establishing a system to introduce the smart-based national licensing examination. Some persons are still unfamiliar with table PC-type smart devices. Moreover, test takers' ability and assurance of using smart devices can work as interference factors of test performance. Therefore, in the follow-up research, it will be necessary to analyze any performance difference depending on smart device function improvements and function factors. With the expansion of smartphones and tablet PCs to various fields, smart devices are highly related to people's life. Given the sharp rise in the use of new smart devices, it is natural to be interested in why and how the devices are utilized. According to the report of the Korea Internet & Security Agency, the penetration rate of students' smart devices increased to 64.5% in the second half of 2012.[19] In the circumstance of the change to smart device-based tests in the smart era, it is necessary to look into the actual conditions, education environments, and other general situations of smart devices used to provide smart device-based smart education effectively in the education field. The use of the internet and mobile devices becomes part of students' everyday life. Seol et al., reported that school education is forced to change to smart device-based education to live up to the changes of times, beyond its traditional learning venue.[20] In the circumstance of the change to smart device-based tests in the smart era, it is necessary to look into the actual conditions, education environments, and other general situations of smart devices used to provide smart device-based smart education effectively in the education field.

Acknowledgments

References

[1] Yoou S. K., Won J. S., Lee K.J., Han D.K., Sung H. Jm Lee H. S., Lim H. C., Kim E. J., Choi H. Y., "Study on validity of the national examination based on smart device-based test (SBT) by 7 types of healthcare occupation," The Korean Journal of Emergency Medical Services, vol.23, no.3, pp.7-34, **(2018)**

[2] Kim D. P., Lakshimi A. S., Lee C. N., Shehzad N., Debbie F. S., "Analysis of contact lens education in Korea on the base of IACLE students trial exam (STE)," The Korean Society of Vision Science, vol.13, no.2, pp.149-155, **(2011)**

[3] Ye K. H., Cho S. A., "A study on the scales of life of ophthalmic optics students," The Korean Society of Vision Science, vol.22, no.1, pp.41-49, **(2020)**

[4] Donal E. M., Gerard F. D., David B. S., "Medical licensing examinations in the United States," Journal of dental education, vol.66, no.5, pp.610-611, **(2002)**

[5] Steven A. H., Peter J. K., Gerard F. D., "The evolution of the united states medical licensing examination (USMLE): Enhancing assessment of practice-related competencies," The Journal of the American Medical Association, vol.310, no.21, pp.2245-2246, **(2013)**

[6] Josef G. T., Renee F. M., Salvatore S., Linda P., "Do United States medical licensing examination (USMLE) scores predict in-training test performance for emergency medicine residents?" J Emerg Med, vol.38, no.1, pp. 65-69, **(2010)**

[7] Kathleen E. K., Dawn Y., Kevien R. W., Gavin C. R., Terrence E. G., Shawn M. Q., Charles C. W., Andre J. B., Adam L. L., Isamu E. Y., Stephen E. D., Bryan G. K., "Correlation between United States medical licensing examination and comprehensive osteopathic medical licensing examination scores for applicants to a dually approved emergency medicine residency," J Emerg Med, vol.52, no.2, pp.216-222, **(2016)**

[8] Hasty R. T., Snyder S., Suciu G.P., Moskow J.M., "Graduating osteopathic medical students' perceptions and recommendations on the decisionto take the United States medical licensing examination," J Am Osteopath Assoc, vol.112, pp.83-89, **(2012)**

[9] Weizberg M., Kass D., Husain A., Cohen J., Hahn B., "Should osteopathic students applying to allopathic emergency medicine programs take the USMLE exam?" Department of Emergency Medicine (UCI), vol.7, no.1, pp.101-106, **(2014)**

[10] Gimpel J. R., "New COMLEX-USA-to-USMLE conversion formula needed," J Am Osteopath Assoc, vol. no.10, pp.577-578, **(2010)**

[11] McMahon G. T., Tallia A. F., "Perspective: Anticipating the challenges of reforming the United States medical licensing examination," Acad Med., vol.85, no.3, pp.453-456, **(2010)**

[12] https://www.korea.kr/archive/expDocView.do?docId=33027 **(2012)**

[13] Kim Y. R., Chung M. H., Kim J. H., "A study on the actual condition and utilization plan of smart devices for educational purpose," J Korean Society for Internet Information, vol.14, no.3, pp.47-55, **(2013)**

[14] Huh S., "Can computerized tests be introduced to the Korean medical licensing examination," J Korean Med. Assoc., vol.55, no.2, pp.124-130, **(2012)**

[15] Brian K. F., Louedes F., Richard T., "Student performance and attitudes during the transition from paper to computer-based examinations," The Journal of the Association of Schools and Colleges of Optometry, vol.15, no.1, pp.1-10, **(2019)**

[16] Jeong H., "A comparative study of scores on computer-based tests and paper-based tests," Beh Inform Technol, vol.33, no.4, pp.410-422, **(2014)**

[17] Karay Y., Schauber S. K., Stosch C., Schuttpelz B. K., "Computer versus paper—does it make any difference in test performance?" Teach Learn Med., vol.27, no.1, pp.57-62, **(2015)**

[18] Boevé A. J., Meijer R. R., Albers C. J., Beetsma Y., Bosker R. J., "Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment," PLoS ONE, vol.10, no.12, **(2015)**

[19] https://www.korea.kr/archive/expDocView.do?docId=33027 **(2012)**

[20] Seol M. G., Sohn C. I., "A survey on teacher's perceptions about the current state of using smart learning in elementary schools," Journal of The Korean Association of Information Education, vol.16, no.3, pp.309-318, **(2012)**

Authors

Eun Joo Kim, PhD

Professor of Liberal Art at Eulji University, Korea. She is a graduate of the Department of Education at Kookmin University in Korea, and the areas of research that are interested are managing the quality of education such as learners' learning experience, learning performance and education evaluation, creativity education.

Jung Un Jang, PhD

Professor, Deaprtment of Optometry, Eulji University, Seongnam

This page is empty by intention.

Eun Joo Kim and Jung Un Jang