Overview of School Readiness to Face the Computer-Based National Examination in Gowa District, South Sulawesi: Analysis of Test Results of Computer-Based Tests (CBT) and Paper-Based Tests (PBT)

Syahrul[1,*] Iwan Suhardi[1] Alifya N.F.H.[1]

[1]*Faculty of Engineering, Universitas Negeri Makassar, Makassar, Indonesia*
Corresponding author. Email: syahrul@unm.ac.id

ABSTRACT

At present, the Computer-Based National Examination (CBNE) model to replace the Paper and Pencil-Based National Examination (PBNE). The National Examination Organizer assumes that the items displayed on the monitor screen have the same difficulty index when presented on paper. Based on this assumption, CBNE and PBNE are considered equal. However, judging from the context and the situation, the paper pencil-based test (PBT) and computer-based test (CBT) model has differences, namely the number of items in the range of views, test aid, the types of question that can be displayed, the color of the test items, how to do the test, the need for basic knowledge of operating a computer, and the habit factor. The differences between the two test models can be psychologically influential when working on a computer-based test. This research is to review how schools prepare their students for CBNE. This research uses a combination of development methods and quantitative methods. The development method is used to develop (1) a set of question item packages and (2) Computer-Based Testing (CBT) software. The test was conducted on 452 students in class XII, from 3 (three) schools in Gowa Regency, South Sulawesi, namely SMK N 1 Gowa, SMK N 2 Gowa, and SMK N 4 Gowa. The statistical test results showed the level of school readiness in preparing students for CBNE. If the results of the statistical test do not show any differences in the results of the test scores, then the school students are considered ready to implement CBNE. If the results of the statistical test show that there are differences in the results of the test scores, then the magnitude of the average results will be seen. If the average size of the PBT group test scores is lower than the CBT group test scores, then the school students are considered ready for CBNE. However, if the average amount of PBT group test scores is higher than the CBT group test scores, then the school students are considered not ready for CBNE. The results showed that, in general, students at the schools are considered ready for CBNE.

Keywords: School readiness level indicator, CBNE, CBT, PBT, score results

1. INTRODUCTION

This research has been done because there is an increasing number of implementations of the Computer-Based Testing (CBT) model as a way to analysis the ability of test-takers. CBT began to gradually replace the function of the Paper-Based Test (PBT) [1] [2]. A concrete example of CBT penetration is in the development of the National Examination in Indonesia. The trend of using the CBT testing model in an educational sector is predicted to increase, replacing the PBT model continually.

The implementation of the National Examination in Indonesia currently uses 2 (two) testing models, namely using paper and computer test and it has been changed to the CBT model. The National Examination, which uses paper, is known as PBNE, or in Indonesia, it is termed as UNKP. The computer-delivered national examination is known as CBNE, or in Indonesia, it is termed as UNBK (Ujian Nasional Berbasis Komputer). The National Examination Institutional assumes that the items displayed on the UNBK software monitor screen will have the same level of difficulty when delivered at UNKP, which uses paper. With this assumption, the results-scores of the UNBK and UNKP test models are

considered equivalent. The difference in the context and the situation between the PBT and CBT testing models, as well as the habit factor in exam-taking, can psychologically affect the students during their computer-based examination. On the other hand, the anxiety factors could influence the examination process, which then can negatively affect the focus of the students during the computer-based exam.

Psychometrically, there are almost no advantages between CBT compared to PBT. The testing models use the same number of items for each participant or fix-length test. In scoring, the PBT and CBT models use Classical Test Theory (CTT) with the formula:

$$\text{Test Scores} = \frac{\text{Question items answered correctly}}{\text{The Total of Question Items}}$$

Generally, the test scores are converted to a score of 0 to 100. The greater the test scores obtained, then the estimated level of students' ability is considered to be higher (in other words, students are considered smarter).

As it utilizes a more modern media tools, CBT has advantages compared to PBT, namely (1) increase standardization, (2) increase test security, (3) enhance test display capabilities, (4) minimize the error of measurement, and (5) accelerate the provision of score and interpretation [3]. Although they have the same paradigm for measuring the ability of test-takers, the PBT and CBT models have striking differences in terms of context and situation. The difference in the context and situational aspects between PBT and CBT is possible to influence the results of the test scores and determine the estimation of students' ability. Psychometrics experts, such as Rudner [4] and Grist [5], argue that the item parameters used on PBT may not match the appearance on the computer monitor screen.

Table 1. Comparison of the Context and Situational Aspects of the PBT and CBT Models

Context and Situation	PBT Model	CBT Model
The number of items in the range of views	Consist of many question items	Generally, it only has 1 (one) question item. I need to scroll the question if it is too long.
Test Aid	Paper and Pencil	Monitor screen, CPU, keyboard, mouse, and speaker
Types of question that can be displayed	Text and Pictures	Text, picture, audio, and video
The way question items can be answered	Marking the selected answer using a pencil	Select the correct answer using a mouse or keyboard
Basic knowledge of information technology	Not needed	Needed
The color of question items given	Mostly black	Possible for all colors
The habit factors in addressing the questions	Students are used to it	Students are still unfamiliar with it

The objective of this paper is comparing the CBT model and the PBT model in student readiness in Gowa Regency.

2. METHOD

This research used a combination of development methods and quantitative methods. The development method was used to develop the test package of questions set and CBT software. The developed questions were chosen from general subjects. They were taught to all high school students without the difference in Core Competencies and Basic Competencies. The question package consisted of 50 (fifty) items with material taken from Indonesian Language Class XII in accordance with the latest 2013 Curriculum. The data was obtained by forming 2 (two) equal groups of respondents, and each group worked on the same questions package with the PBT and CBT testing models.

The CBT model was developed using a web-based client-server system that can be accessed via the LAN network. Similar to the PBT testing model, CBT software was designed to enable the respondents to choose the desired item number. Furthermore, they could review their response if they wanted to replace it. The way to answer the question items on the PBT model is by circling the chosen answers on the answer sheet using a pencil. In contrast, the way to answer the question items on CBT is by selecting the answer that is considered correct with the mouse or keyboard.

In general, the flowchart in this study is presented in Figure 1. Statistical analysis is used after collecting of test scores of the groups of respondents. It aims to find out that there are any significant differences between the two test models. The results of the analytical output from SPSS were then evaluated and used as an indicator to determine whether the students will be ready for a computer-based test or not.
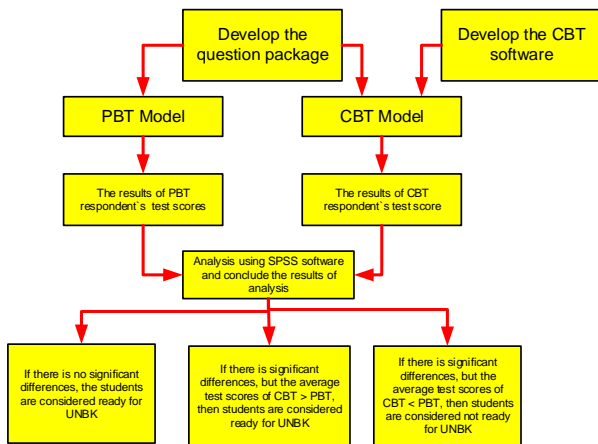
Figure 1 The Flowchart of The Research

3. RESULTS AND DISCUSSION

Respondents were drawn from students in class XII in schools in Gowa Regency, South Sulawesi, namely in SMK N 1 Gowa, SMK N 2 Gowa, and SMK N 4 Gowa. The respondents are 452 students, consisting of the PBT model with 230 respondents and 222 respondents used CBT test model. After the test was conducted, the results of the PBT group and CBT group are presented in Table 2 and Table 3.

Table 2. The Results of the PBT

38	68	54	34	48	56	46	46	56	80	36
44	68	54	20	44	38	28	38	50	56	38
44	76	54	52	44	70	44	50	46	34	50
36	74	50	40	34	62	44	54	54	40	42
36	64	44	40	32	38	38	58	54	56	34
50	76	58	58	64	50	26	56	48	54	48
58	66	48	42	34	62	50	66	34	54	42
46	74	44	50	50	38	28	46	66	56	40
58	68	44	56	28	58	52	36	68	50	50
30	74	64	42	34	36	30	76	54	50	38
58	70	64	48	42	38	50	18	54	48	32
58	78	64	56	24	52	28	22	26	22	70
58	48	68	48	60	48	54	36	50	54	58
58	64	74	48	26	54	38	48	58	40	34
60	74	74	44	46	56	46	48	52	50	22
54	76	64	50	44	36	46	60	40	56	30
60	76	64	48	50	52	50	38	58	48	36
44	78	34	56	50	42	56	32	38	48	38
56	82	34	54	48	28	66	50	64	24	24
50	40	54	68	74	70	70	66	67	16	50
46	74	64	58	20	52	48	52	42	48	
Number of Respondents									230	
Minimum Score									16	
Maximum Score									82	
Average Score									49.72	
Deviation Standard									13.83	

Table 3. The Results of the CBT Test Model

50	60	42	58	42	34	52	40	46	66	60
30	46	44	58	52	32	46	60	52	68	66
34	46	40	46	24	58	46	34	36	68	68
36	60	46	52	52	50	28	64	48	70	56
30	60	46	46	60	44	36	32	48	74	56
28	68	46	44	66	28	62	60	42	72	48
42	66	42	48	52	26	18	52	38	48	42
50	56	54	42	36	54	54	32	46	26	42
38	68	52	40	38	38	24	30	46	32	44
50	58	36	26	44	30	86	22	40	48	50
22	66	50	40	50	50	62	44	50	46	58
28	60	40	44	38	60	54	60	26	46	30
52	66	36	48	54	48	58	32	58	48	36
50	62	34	58	54	62	64	44	18	62	30
50	70	56	42	46	50	54	68	52	28	62
50	40	56	60	46	42	50	50	60	72	40
52	56	56	34	50	50	44	54	36	68	44
62	62	40	18	46	40	38	28	60	38	58
52	42	50	52	58	28	44	38	42	62	44
50	36	36	34	60	60	30	38	44	40	46
60	36									
Number of Respondents									222	
Minimum Score									18	
Maximum Score									86	
Average Score									47.41	
Deviation Standard									12.42	

Table 4. Tests of Normality

Students		Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	*df*	*Sig.*	*Statistic*	*df*	*Sig.*
Score	PBT	.059	230	.049	.990	230	.093
	CBT	.055	222	.096	.992	222	.264
a. Lilliefors Significance Correction							

Table 4 shows the results of the Lilliefors and Shapiro Wilk tests. The score of P-value (Sig) Lilliefors was 0.049 in the PBT group and 0.096 in the CBT group. The P-value of the Shapiro Wilk test in the PBT group was 0.93, and in the CBT group was 0.264. Table 3 shows that the scores of the PBT group for the Lilliefors test obtained p-value (Sig) of 0.049, all the results of the test scores > 0.05. The scores of the PBT group for the Lilliefors test obtained p-value (Sig) of 0.049 is generally assumed that both the PBT and CBT groups are normally distributed. Assumptions that the two groups are normally distributed can be seen in the visual appearance of the Q-Q Plot and P-P Plot as follows. The display Q-Q Plot and P-P plot, it is found that the data or point spreads around the diagonal line and follows the direction of the diagonal line.
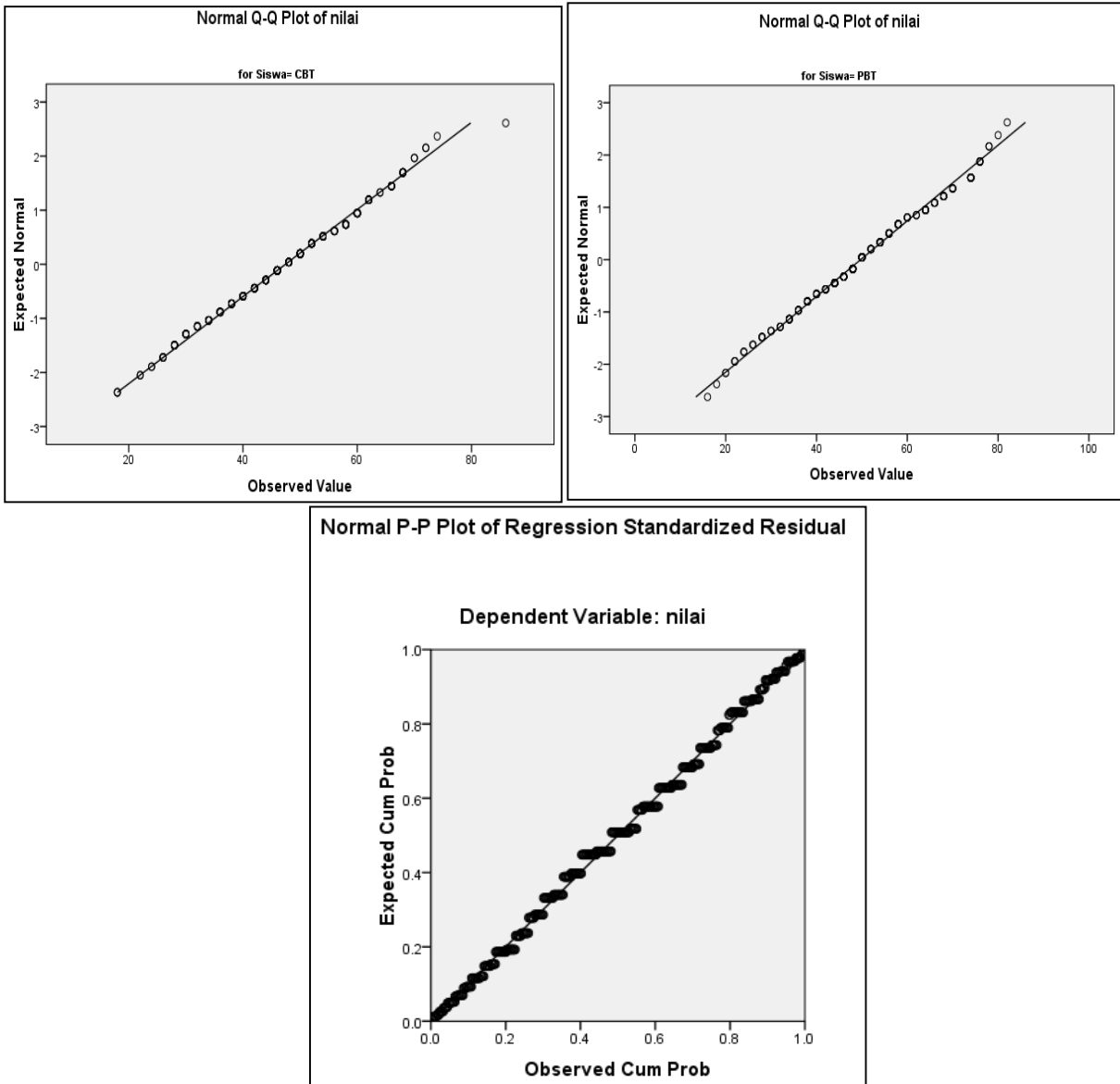
Figure 2 The Visual Display of Q-Q Plot dan P-P Plot

Homogeneity independent t-test results for the entire respondent's data are shown in Table 5. Levene values are shown in the line of values based on mean, which is 1.313 with a p-value (sig) of 0.252, which is > 0.05, which means there is a similarity of variances between groups or which means homogeneous.

Stem-leaf diagrams (Figure 3) and box-plot diagrams (Figure 4) were used to detect outliers. In the stem-leaf diagram, the presence of an outlier is marked based on whether there are extreme values above and below the stem-leaf. Based on the results, there are no extreme values, which means there are no outliers in the two groups. Outlier detection can also be assessed using the following box-plot.

Table 5 Homogeneity Test

		Levene Statistic	df1	df2	Sig.
Value	Based on Mean	1.313	1	450	.252
	Based on Median	1.268	1	450	.261
	Based on Median and with adjusted df	1.268	1	439.984	.261
	Based on trimmed mean	1.305	1	450	.254

The box-plot does not indicate that there are plots above and / or below the box-plot, which means there are no outliers in the data of the two groups. Because all assumptions are fulfilled, it can proceed to the next test, namely the Independent T-Test.

Comparing the average test results, the difference between PBT group and the CBT group is conducted by

using the t-test number. The analysis assumes both variances are similar (equal variances are assumed). This figure was used because the results of the analysis using the F test showed no difference in variance between the two groups.
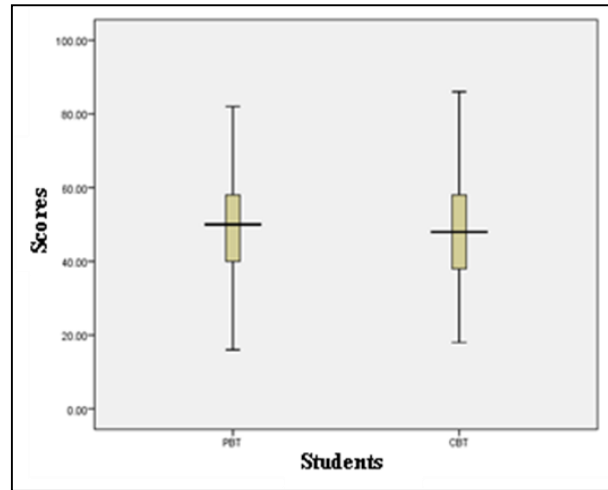

Figure 3 Detection of Outliers with Box-plot

Table 6 SPSS Outputs Assessment Results of PBT and CBT Testing Models

Group Statistics					
	Students	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Scores	PBT	230	49.7217	13.82710	.91173
	CBT	222	47.4144	12.41658	.83335

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	*Sig.*	*t*	*Df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	*95% Confidence Interval of the Difference*
									Lower	*Upper*
scores	Equal variances assumed	1.313	.252	1.864	450	.063	2.30732	1.23755	-.12477	4.73942
	Equal variances not assumed			1.868	447.686	.062	2.30732	1.23520	-.12019	4.73484

The steps are taken to test the average P&P Test model, and the CBT model is:

1. Formulating research hypotheses
   There are differences in the average assessment results between groups of PBT and CBT testing models
2. Formulating operational hypotheses (null and alternative hypotheses)
   Ho: The average assessment results using the PBT and CBT methods are the same
   Ha: The average results of the assessment using the PBT and CBT methods are not the same

3. Determining the level of trust used
   The confidence level used is 95% or using alpha 5%.
4. Determining the decision-making rules
   The rule in making decisions is to accept H0 if the t count is smaller than the t table and reject H0 if the t count is greater than t table. Based on the t table with alpha, 5% 2-way test or 2.5%, and the degree of freedom df = 450, the value of the t table is 1.96525. So the decision taken is to accept H0 if t arithmetic smaller than 1.96525 and reject H0 if t arithmetic greater than 1.96525
5. Calculating t count or t statistics

The SPSS 16.0 assistive program is used to calculate t count is 1.864

6. Decision making and interpretation of results

After calculating t statistics, the final step is to make a decision on the results of the analysis and interpretation of these results. The average difference in the assessment results between the PBT and CBT methods is 2.30732, with a standard deviation of 1.23755. The results of the t statistical calculation resulted in a value of 1.864 and a significance of 0.063.
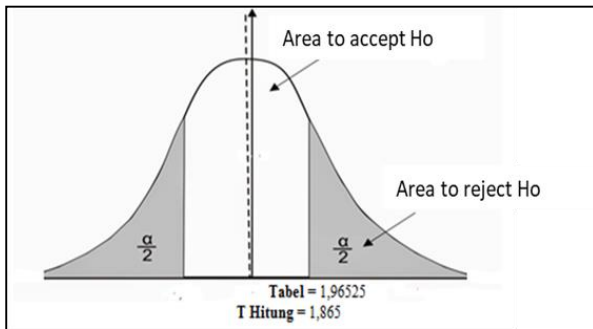

Figure 4 Two Tail Hypothesis Test Results Assessment

With a significance result of 0.063, the decision to accept H0 can be taken because the significance level is greater than alpha (0.025). The results of the calculation of the t value (1.864) turned out to fall in the reception area, then H0 is accepted. Therefore, H0 is accepted, and Ha is rejected. This means, from the results of the analysis, it can be stated that statistically, the values of the two methods (PBT and CBT models) are stated to be no significant difference.

In a deeper analysis, there is an average difference (mean difference) of 2.30732, which indicates the difference in the average scores of PBT and CBT, which is 49.7217 for PBT and 47.4144 for CBT. This indicates that there is no significant difference between two groups. However the average scores CBT model are smaller than the PBT model. So that, it can be concluded that the same package of questions, when done with the CBT method, is still felt more difficult by test-takers compared to if it is done with the PBT method.

The analysis shows that in classical theory, there are differences in the average results of the exam scores if the same package of questions is displayed with the PBT model and the CBT model. The package of questions displayed on the CBT monitor screen using a mouse and keyboard is felt to have a different level of difficulty when presented on a paper and using a pencil.

In addition to the different aspects of context and situation, the factor of the unfamiliarity of students working on problems with a computer-based test model can contribute to the differences of average scores between the PBT and CBT models. In general, not many schools have applied computer-based testing models to classroom learning practices. The number of computer laboratories is not proportional to the overall number of students. With the UNBK policy implemented by the government, it was responded by conducting socialization on the use of computer-based exams for final-level students several months before the National Examination. However, such a short time does not necessarily result in students becoming accustomed to taking computer-based exams.

In the use of CBT, it is important to consider the aspects of computer self-efficacy or the students' confidence as being successful in working on a computer-based test. Computer self-efficacy helps to reduce student anxiety levels in taking computer-based exams [6] [7] [8]. With the reduced level of anxiety, students can focus more on working on problems and can show their best abilities. On the other hand, one way to reduce test anxiety using a computer is to improve students' computer experience and confidence in taking computer-based exams [9] [10]. It is important to provide the opportunities for students to become familiar with the CBT model [11]. Making students familiar with reproducing CBT model trials before test day can reduce anxiety factors. It will be beneficial for students who are economically disadvantaged and do not have computers. Furthermore, the habituation of this exam model is expected to increase students' confidence in being able to succeed at this UNBK so that they do not feel significant obstacles compared to other students who take the UNKP.

Although there are indicators that students are declared ready to face UNBK in the use of computer-based testing media. It does not mean that they will be successful and get high test scores if a computer-based testing model is conducted. This strategy is only an indicator of readiness to use a computer that is related to the level of student habits towards computer-based testing. To get high scores on UNBK/CBNE, of course, it remains focused on the mastery of the material contained in the learning process.

4. CONCLUSION

It can be concluded that there were no significant differences in the score results if the same question items were displayed with the Paper Based Test (PBT) model and the Computer-Based Testing (CBT) model. This indicates that, in general, students are considered ready to do computer-based exams such as CBNE (UNBK). However, there is a difference in the average (mean difference) of 2.30732. It indicates a difference in the average scores of PBT and CBT, which is 49.7217 for PBT and 47.4144 for CBT. Although statistically stated there is no significant difference between groups working on the PBT and CBT testing models, but when viewed from the average scores, it can also be stated that the same question package, when worked with the CBT method,

will be felt as more difficult for test-takers compared to if it is done by the PBT method. In this case, students need more opportunities to get used to using a computer-based testing model so that the level of readiness to face CBNE (UNBK) may increase.

REFERENCES

[1] A. C. Bugbee Jr, "The equivalence of paper-and-pencil and computer-based testing," J. Res. Comput. Educ., vol. 28, no. 3, pp. 282–299, 1996.

[2] P. for I. S. Assessment, PISA Computer-based Assessment of Student Skills in Science. OECD, 2010.

[3] C. V. Bunderson, D. K. Inouye, and J. B. Olsen, "The four generations of computerized educational measurement," ETS Res. Rep. Ser., vol. 1988, no. 1, pp. i–148, 1988.

[4] L. M. Rudner, "An on-line, interactive, computer adaptive testing tutorial," ERIC Clear. Assess. Eval., 1998.

[5] S. Grist, Computerized adaptive tests. ERIC Clearinghouse, 1989.

[6] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," MIS Q., pp. 189–211, 1995.

[7] S. P. John, "Influence of computer self-efficacy on information technology adoption," Int. J. Inf. Technol., vol. 19, no. 1, pp. 1–13, 2013.

[8] H. K. Sam, A. E. A. Othman, and Z. S. Nordin, "Computer self-efficacy, computer anxiety, and attitudes toward the Internet: A study among undergraduates in Unimas," J. Educ. Technol. Soc., vol. 8, no. 4, pp. 205–219, 2005.

[9] M. Zeidner, "Test anxiety," Corsini Encycl. Psychol., pp. 1–3, 2010.

[10] R. M. Liebert and L. W. Morris, "Cognitive and emotional components of test anxiety: A distinction and some initial data," Psychol. Rep., vol. 20, no. 3, pp. 975–978, 1967.

[11] M. Russell, "Testing On Computers," Educ. Policy Anal. Arch., vol. 7, p. 20, 1999.