

## ARTICLE

# Comparison of Test Scores Using Paper and Computer Media as an Indicator of Computer Self-Efficacy and Test Anxiety in Facing Computer-Based-Testing

---

Nusantara Journal of  
Social Sciences and  
Humanities

Issue 1: 2020

<https://lekantara.com/index.php/njsh>

**Suhardi Iwan**

Universitas Negeri Makassar  
Makassar, Indonesia

Email: [iwan.suhardi@unm.ac.id](mailto:iwan.suhardi@unm.ac.id)



### Abstract

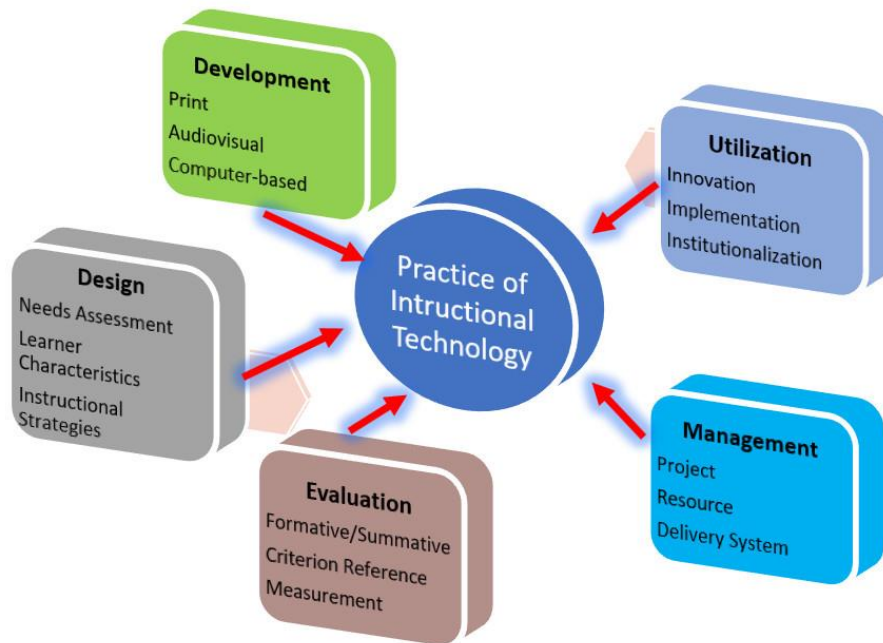
Due to unfamiliarity, students often feel that they are not ready to face exams using computers. Students often feel insecure and anxious when facing exams using computers. This study aims to measure the readiness of the group of test-takers in facing the exam using computer media. This research is a combination of development methods and quantitative studies. The research begins with a study of whether there is a difference in the test taker's assessment scores if the same question package is tested on the Paper-Based Testing (PBT) and Computer-Based Testing (CBT) models. Furthermore, a statistical test was carried out to determine whether there was a significant difference between the results of the assessment score obtained from the results of the PBT and CBT testing. From these statistical tests, it can be determined the level of readiness of the test-takers in facing the test using computer media. If the statistical test showed no difference in the assessment scores of the PBT and CBT groups, in general, the test participant group was declared ready to face the test with computer media. If the statistical test shows that there is a difference in the assessment score, then the mean score is seen. If the mean score for the PBT group assessment is lower than the mean score for the CBT group assessment, in general, the test participant group is declared to be ready for the exam using computer media. However, if it is found that the mean score for the PBT group is higher than the mean score for the CBT group, then in general the test participant group is declared not ready to face the test using a computer.

**Keywords:** Paper-Based Testing, Computer-Based Testing, computer self-efficacy, test anxiety

## Introduction

Unstoppable, information technology has touched various fields, including education (Sudibyo, 2012: 177). According to Seels & Richey (1994: 28-29), learning technology has 5 (five) domains, namely: design, development, utilization, management, and assessment of processes and resources for learning. The five domains are closely related and synergistic with the relationship shown in Figure 1 below.

**Figure 1. Learning Technology Domain (Adapted from Seels & Richey, 1994)**



The domain of development and assessment does not escape the penetration of information technology developments. The role of the Computer-Based Testing (CBT) testing model began to gradually replace the function of the Paper Based Test (PBT) testing model. Most test administrators assume that a question item written on paper media has the same difficulty level when the item is displayed on a computer monitor screen. However, the equivalence between the difficulty level of the items on paper and computer media still lacks an in-depth study. From a psychometric aspect, the working principle of CBT is considered to move the PBT paradigm from paper media to computer screens. However, from the context and atmosphere aspects, the appearance of the two test models has many differences.

In general, students in Indonesia are accustomed to carrying out the learning process and taking exams using paper and are not yet used to learning and testing using computers. The habit of dealing with new things can have a psychological effect when working on computer-based exam questions. The difference in context and atmosphere between the PBT and CBT testing models can cause awkwardness and anxiety, especially for students who are dealing with the CBT testing model for the first time. The influence of anxiety and anxiety factors before and on the exam process can cause students to be unable to focus on doing the test well when using a computer-based exam model. This computer self-efficacy factor is important in helping reduce students' anxiety levels in taking computer-based exams so that they are expected to show their best abilities.

## Literature Review

The role of Computer-Based Testing (CBT) has gradually replaced the function of the Paper-Based Test (PBT) (Bugbee, 1996; OECD, 2010). Likewise in Indonesia, the trend of using CBT in the educational environment is predicted to continue to increase to replace the PBT model (Syahrul, 2018).

There is almost no psychometric advantage to CBT compared to PBT. PBT and CBT use the same number of items for each participant or the fix-length test. Even though they have the same paradigm of measuring the ability of test-takers, the PBT and CBT models have striking differences from the context and atmosphere (Suhardi, 2017; Suhardi, 2018).

Psychometric experts, such as Rudner and Grist, argue that the parameter values of items written on paper may no longer be appropriate when displayed on a computer monitor screen (Rudner, 1998; Grist, 1989). The difference in context and atmosphere between the PBT and CBT testing models, as well as the habitual factors in taking the exam, can affect psychologically when working on computer-based exam questions. The influence of anxiety factors before and on the exam process can cause students to be unable to focus on doing the test when using CBT.

Comparison of context and atmosphere aspects between PBT and CBT models faced by test takers is presented in Table 1.

**Table 1. Comparison of Context and Atmosphere Aspects of the PBT and CBT Models (Adapted from Suhardi, 2018)**

<b>Context and Atmosphere</b>	<b>PBT Model</b>	<b>CBT Model</b>
Number of items within eye reach	Consists of many items	Usually, there is only 1 (one) item, even long items have to be scrolled
Exam aids	Paper and pencil	Monitor screen, CPU, keyboard, mouse, and speakers
How to do the items	Marking the answers that are considered correct with a pencil	Select the correct answer with the mouse or keyboard
Aspects of basic knowledge about information technology	Not required	Required
The test-taking habit factor	It is common	Not yet a habit

Along with the increasing penetration of CBT replacing PBT in the scope of education, the level of equivalence of the PBT and CBT testing models needs to be further analyzed. It is also necessary to study a simple measurement and easy to apply so that it can be used as an indicator of the readiness of the test-takers in facing the testing model using computer media.

## Research Method

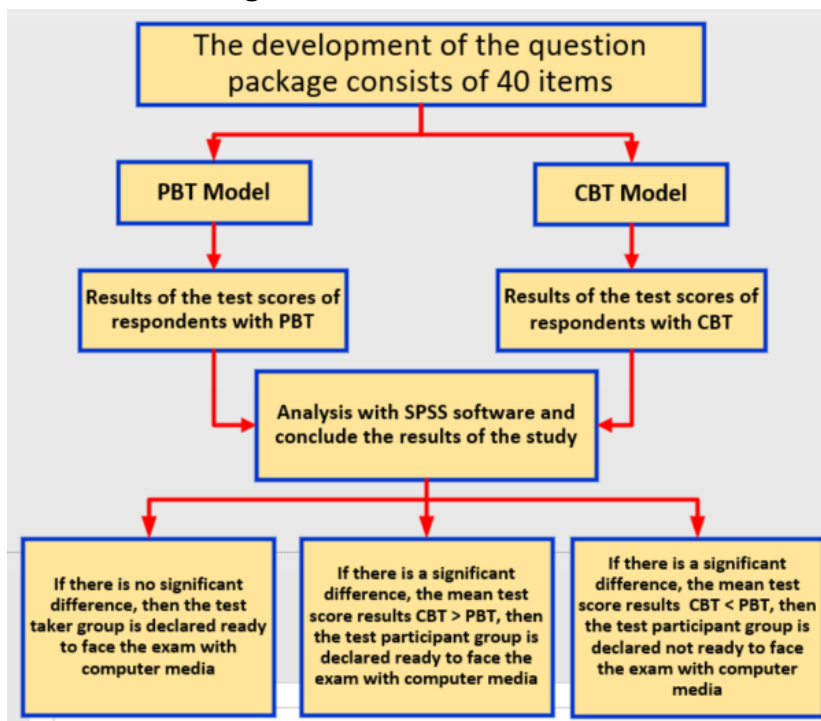
This research uses a combination of development methods and quantitative studies. The development method was used to develop a test question package and CBT software. The question package consists of 40 (forty) items with text display. The material for the exam is taken from Indonesian for Class X Senior High Schools. The characteristics of the items on PBT and CBT were made as similar as possible. As with the technicality of working on the PBT testing model, CBT software is designed so that test-takers can choose the desired item numbers and can review the responses if they want to replace them. The only difference is

how to respond to the items and their appearance. In PBT, how to respond to questions is done by choosing the correct answer on the exam sheet using a pencil. While in CBT, this is done by selecting the correct answer on the computer monitor screen using a keyboard or mouse.

Experiments were carried out by forming 2 (two) groups of respondents. The number of each group is 100 (one hundred) respondents. Respondents have been made to have equal abilities. One group worked on problems using the PBT model and the other group worked on problems using the CBT model. In the PBT model, question packages are printed and distributed as many as the number of respondents. Whereas in the CBT model, the developed question packages are entered into the CBT software database. Each respondent uses a set of computer equipment.

The test results score was then analyzed using the SPSS tool, to test whether there was a statistically significant difference between the two test models. The results of the SPSS output were then analyzed to determine the level of readiness of the group of test-takers for the exam using computer media. In general, the flowchart of this research is presented in Figure 2.

**Figure 2. Research Flowchart**



**Results and Discussion**

After the testing process was carried out using the same question package, the test scores obtained from the PBT and CBT groups were presented in Table 2 below.

**Table 2. Scores on PBT and CBT Models**

PBT SCORE (100 participants)			CBT SCORE (100 participants)		
65	47,5	92,5	20	67,5	30
62,5	62,5	92,5	15	55	45

60	50	60	25	47,5	60
55	62,5	97,5	50	70	50
55	62,5	67,5	75	60	52,5
67,5	87,5	95	42,5	67,5	30
70	65	57,5	55	65	30
92,5	52,5	77,5	60	67,5	55
72,5	75	70	57,5	55	40
85	32,5	92,5	52,5	55	40
67,5	32,5	97,5	55	62,5	47,5
60	92,5	62,5	47,5	70	37,5
65	82,5	62,5	32,5	65	45
72,5	82,5	55	52,5	60	57,5
60	87,5	62,5	55	57,5	47,5
65	95	75	45	55	42,5
100	72,5	80	37,5	55	52,5
70	77,5	57,5	72,5	55	45
65	65	87,5	52,5	55	42,5
62,5	47,5	75	37,5	55	35
60	67,5	97,5	37,5	77,5	45
60	97,5	77,5	40	55	30
77,5	100	65	50	47,5	42,5
70	67,5	52,5	37,5	65	40
65	60	50	37,5	72,5	52,5
70	67,5	70	50	60	47,5
70	67,5	67,5	52,5	60	37,5
62,5	67,5	82,5	55	47,5	45
57,5	70	57,5	52,5	47,5	55
60	95	60	65	55	47,5
52,5	70	67,5	52,5	15	87,5
47,5	60	70	52,5	20	42,5
55	95		22,5	10	
52,5	87,5		25	37,5	

The statistical test of the comparison of scores from PBT and CBT was carried out with the SPSS tool. The method used is the two-sample free t-test. The results of the SPSS output are presented in Figure 3 below.

**Figure 3. SPSS Output**

**T-Test**

<b>Group Statistics</b>					
	<b>Siswa</b>	<b>N</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>Std. Error Mean</b>
<b>Score PBT</b>		100	69.4500	14.84422	1.48442
<b>CBT</b>		100	49.2600	13.66158	1.36616

		<b>Levene's Test for Equality of Variances</b>		<b>t-test for Equality of Means</b>				<b>95% Confidence Interval of the Difference</b>		
		<b>F</b>	<b>Sig.</b>	<b>t</b>	<b>df</b>	<b>Sig. (2-tailed)</b>	<b>Mean Difference</b>	<b>Std. Error Difference</b>	<b>Lower</b>	<b>Upper</b>
<b>Score</b>	<b>Equal variances assumed</b>	.940	.334	10.008	198	.000	20.19000	2.01740	16.21165	24.16835
	<b>Equal variances not assumed</b>			10.008	196.651	.000	20.19000	2.01740	16.21149	24.16851

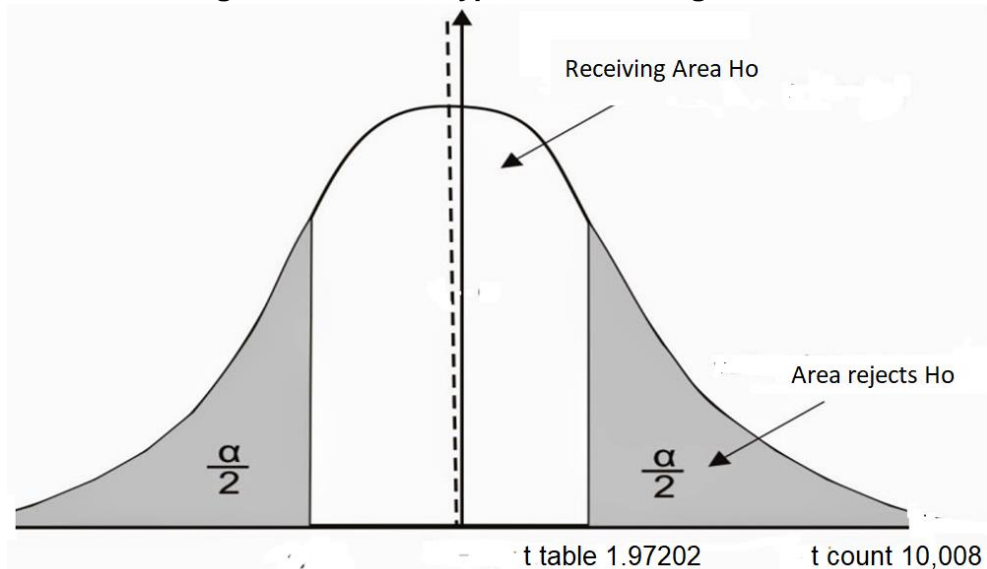
To test whether there were differences in the PBT and CBT scores, the variance similarity test was performed. The variance similarity test was carried out using the F test or Lavene test. The F test value which assumes the two variants are the same is 0.940 with a probability (sig) of 0.334. Because the probability value is  $> 0.05$ , then  $H_0$  is accepted and  $H_a$  is rejected. So it means that there is no difference in score variance between PBT and CBT. In the absence of differences in the variance of the PBT score and the CBT model, then the mean test was carried out between the two test models.

Comparing the mean PBT and CBT scores was done by using the t-test value which assumes that the two variances are the same (equal variances assumed). This value is used because the results of the analysis using the F test show that there is no difference in variance between the two groups.

The steps taken to test the mean score of PBT and CBT are:

1. Formulate a research hypothesis.  
There was a difference in the mean score between PBT and CBT model groups.
2. Formulate operational hypotheses (null and alternative hypotheses).  
 $H_0$ : The mean scores of PBT and CBT models are the same.  
 $H_a$ : The mean scores of PBT and CBT models are not the same.
3. Determine the level of confidence used.  
The confidence level used is 95% or by using an alpha of 5%.
4. Determine the rules for decision making.  
The rule in decision making is to accept  $H_0$  if the t count is smaller than the t table and reject  $H_0$  if the t count is greater than the t table. Based on t table with alpha, 5% 2-way test or 2.5%; and the degree of freedom  $df = 198$  obtained t table value of 1.97202. So the decision taken is to accept  $H_0$  if the t count is smaller than 1.97202 and reject  $H_0$  if the t count is greater than 1.97202.
5. Calculating t count or t statistic.  
To calculate the statistical value, the SPSS tool is used, the t count is 10,008.
6. Decision making and interpretation of results  
The mean difference in scores between the PBT and CBT testing models was 20.19 with a standard deviation of 2.017. The result of t statistical calculation yields a value of 10.008 and the significance is 0.0001.

Figure 4. Two Tail Hypothesis Testing Scores



With a significant result of 0.001, a decision can be made to reject  $H_0$  because the significance level is smaller than alpha (0.025). The results of the calculation of the t value (10.008) also fell in the rejection area, so  $H_0$  was rejected. Therefore  $H_a$  was accepted. That is, from the results of statistical analysis it can be stated that the results of the test scores of the two methods (PBT and CBT) differ significantly.

The mean difference figure of 20.19 indicates a difference in the average score between PBT and CBT, namely 69.45 for PBT and 49.269 for CBT. So in general, in this case, it can be concluded that the same items when done with CBT will feel more difficult for the test-taker than if done with PBT. The results of this research study indicate that the question packages displayed on the monitor screen using a mouse and keyboard have different levels of difficulty when presented on a sheet of paper using a pencil.

In addition to differences in context and atmosphere aspects, students are not accustomed to working on questions using computer media which can result in differences in test scores between PBT and CBT. In general, there are not many schools in Indonesia that implement testing using computer media in learning practices in classrooms. The number of computers in the school computer laboratory is generally not proportional to the needs of the total number of students in the school.

Having the basic ability to operate computer equipment is not a guarantee that students are familiar with testing using computer media. This is because the habit of using paper-based testing has been going on for years before. There may be psychological barriers that influence the difference between PBT and CBT. The unfamiliarity of doing exams using computer media makes students unable to show their best abilities when doing exams. The students' habit of working on the items using PBT without realizing it has a less supportive effect when using CBT.

In the use of CBT, it is necessary to pay attention to aspects of computer self-efficacy, namely how confident a student sees himself as being successful in taking tests using computer media. Computer self-efficacy helps reduce students' anxiety levels in taking exams using computer media (Compeau & Higgins, 1995; Surej, 2013; Sam, Othman & Nordin, 2005). With

reduced levels of anxiety, students can focus more on doing questions and can show their best abilities.

One way to reduce test anxiety in facing exams using computer media is to increase the experience of using computers and increase self-confidence in taking exams using computer media (Zeidner & Matthews; 2003; Liebert & Morris, 1967). The best way is to provide opportunities for students to become more familiar with CBT (Russell, 1999). Familiarizing students with more CBT trials before the test day can reduce anxiety factors. Familiarizing students with CBT will benefit students who are less well off economically and do not have computers at home to improve their computer operating experience. The hope is that the test results can be maximized, such as using test media using paper media.

From the results of this study, practically it can be used as an indicator to determine the level of student habits using computer equipment as a testing medium. The level of students' habit of using computer equipment determines the degree to which students are prepared to use computer media as a testing aid. So far, there is no practical and simple method to determine whether students of a school are ready to use computer equipment to replace testing using paper media.

This research study provides a theoretical basis that one of the indicators of student readiness to use computer media as testing equipment to replace paper media is to compare the test scores between PBT and CBT. If the comparison of the test scores produced by PBT and CBT does not show a significant difference, then it means that the test taker feels the same atmosphere even though the test is done with different exam media. Test takers were unaffected when they encountered PBT or CBT. Test participants are stated to have a sufficient level of the habit of using computers. Test takers already have high self-confidence and do not feel anxious when facing exams using computer media. In this criterion, test participants do not feel any psychological burden due to differences in the testing model from using paper media to computer media.

If the comparison of scores on PBT and CBT is significant, it means that the test taker feels a different atmosphere when working on the same questions between PBT and CBT. If it is found that the mean score on PBT is greater than the mean score on CBT, then the test participant is declared not to have a sufficient level of habit in using computers. Test participants do not have self-confidence and still feel anxiety when facing exams using computer media. In this criterion, test participants still feel the psychological burden of the difference in testing from using paper media to computer media.

However, if it is found that the mean score on PBT is lower than the mean score on CBT, it means that the test taker is declared to have a high level of familiarity in using computers. In this criterion, the test participants do not have the psychological burden of the difference in the testing model from using paper media to computer media. Test takers already have high self-confidence and do not feel anxious when facing exams using computer media. The test takers have even felt the convenience of testing using computer media compared to using paper media because computers have more profitable facilities and make the process of taking the exam easier.



Although it can be an indicator of the test taker's readiness in facing testing using computer media, it does not mean that test-takers will be successful and get high test scores if the test is carried out using computer media. This strategy is only an indicator to determine the readiness to use computers as a means of testing media. To get high score results on testing using computer media, of course, it is still focused on the aspects of understanding the material contained in the learning process.

## **Conclusion**

From the results of the study and analysis, it is concluded that in classical theory it is possible to have a statistical difference in scores if the same item is displayed with a Paper-Based Test and Computer-Based Testing. The difference in scores between PBT and CBT can be used as an indicator to see the level of readiness of the test takers, namely computer self-efficacy and test anxiety in facing exams using a computer

By comparing the scores on PBT and CBT statistically obtained 3 (three) possibilities that can be used as an indicator of the test taker's readiness level, namely: (1) If there is no significant difference in scores, then the average item difficulty level on PBT is equivalent to the average level of difficulty items on the CBT. The test participant is declared to be ready to face the exam using computer media, (2), If there is a significant difference in scores and the mean score on PBT is lower than CBT, then the test participant is declared very ready to face the exam using computer media, and (3) If there is a difference The score is significant and the mean score on PBT is higher than CBT, then the test participant is declared not ready to face the exam using computer media. In this case, test takers need more opportunities to get used to facing exams using computer media so that their level of readiness increases.

## **Limitation of the Study**

The need for further studies to analyze the relationship between the level of basic knowledge and the level of habitability of test-takers to computer operation with the item difficulty index on PBT and CBT to become the basis for the psychometric field to understand the tendency of differences in test scores on PBT and CBT.

## **Declaration of conflicting interest**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding**

An earlier version of this paper was presented as a background paper to the Commonwealth Local Government Conference in Gaborone, Botswana and the authors duly acknowledge the support of the CLGF authorities.

## **References**

1. Bugbee, A. C. (1996) The equivalent of paper-and-pencil and computer-based testings. *Journal of Research on Computing in Education*, 28(3), pp. 282–299. <https://doi.org/10.1080/08886504.1996.10782166>
2. Compeau, D.R., & Higgins, C. A. (1995) Computer self - efficacy: development of a measure and initial test. *MIS Quarterly*. 19 (2), pp. 189–211. Available at: <https://pdfs.semanticscholar.org/391d/d4ca2ee3b2995260425d09dd97d4d9aaac29.pdf>

3. Grist, S. (1989) *Computer adaptive tests*. ERIC Digest No. 107 [Versi elektronik]. ERIC Clearinghouse on Tests Measurement and Evaluation Washington DC, American Institutes for Research Washington DC. Available at: <http://files.eric.ed.gov/fulltext/ED315425.pdf>
4. Liebert, R.M., & Morris, L.W.(1967) *Cognitive and emotional components of test anxiety: a distinction and some initial data*. Psychological Reports 20975–978. <http://dx.doi.org/10.2466/pr0.1967.20.3.975>
5. OECD. (2010) *PISA Computer-Based Assessment Of Student Skills In Science*. Available at: <http://www.oecd.org/publishing/corrigenda>. <https://doi.org/10.1787/9789264082038-en>.
6. Rudner, M. L. (1998) *An On-Line, Interactive, Computer Adaptive Testing Tutorial*. Available at: <http://EdRes.org/scripts/cat>.
7. Russell, M. (1999) Testing on computers: a follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7, 20. <https://doi.org/10.14507/epaa.v7n20.1999>.
8. Sam, H. K., Othman, A. E. A., & Nordin, Z. S. (2005) Computer self-efficacy, computer anxiety, and attitudes toward the internet: a study among undergraduates in Unimas. *Educational Technology & Society*. 8 (4), pp. 205-219. Available at: [https://www.j-ets.net/ETS/journals/8\\_4/19.pdf](https://www.j-ets.net/ETS/journals/8_4/19.pdf)
9. Seels, B. R. & Richey R. (1994) *Intructional technology: the definition and domains of the field*. Washington DC: *Association for Educational Communications and Technology (AECT)*. p. 28-29.
10. Sudibyo, L. (2011) Peranan dan dampak teknologi informasi dalam dunia pendidikan di Indonesia. *WIDYATAMA*. No.2. Volume 20, pp. 175-185. Available at: <http://download.portalgaruda.org/article.php?article=268282&val=7107&title=Peranan>
11. Suhardi, Iwan. (2017) Kajian deskriptif perbandingan model pengujian paper based test dan computer based test: tinjauan dari aspek psikometrik, konteks dan suasana, serta psikologi pengguna. *Jurnal Media Komunikasi Pendidikan Teknologi dan Kejuruan*. Vol 5, No 2 (2018). Available at: <https://doi.org/10.26858/mekom.v5i2.7437>; <https://ojs.unm.ac.id/mkpk/article/view/7437/4320>
12. Suhardi, Iwan. (2017) Pengembangan model bank soal mata pelajaran bahasa Indonesia di sekolah lanjutan indeks atas dengan tampilan multi media untuk penilaian berbasis computerized adaptive testing. Unpublished Doctoral Dissertation, Yogyakarta State University (UNY) Postgraduate Program.
13. Surej, P.J. (2013) Influence of computer self-efficacy on information technology adoption. *International Journal of Information Technology*, Vol. 19, No. 1, 2013. Available at: [http://intjit.org/cms/journal/volume/19/1/191\\_2.pdf](http://intjit.org/cms/journal/volume/19/1/191_2.pdf)
14. Syahrul. (2018) *Penyelenggaraan pengujian berbasis komputer dalam lingkup dunia pendidikan di indonesia: perkembangan, tantangan, dan arah inovasinya di masa depan*. Professorship Inauguration Speech. Presented at the UNM Senate Extraordinary Open Session, April 25, 2018.
15. Zeidner, M. & Matthews, G. (2003) Test anxiety. *Encyclopedia of Psychological Assessment*. SAGE Publications Ltd. pp: 965-969. Print ISBN: 9780761954941. Online ISBN: 9780857025753. <http://dx.doi.org/10.4135/9780857025753.n202>.