Acitya: Journal of Teaching & Education, Vol. 2 No. 2 2020
Acitya: Journal of Teaching & Education
Website: http://journals.umkt.ac.id/index.php/acitya
Research Papers, Review Papers, and Research Report
Kampus 1 UMKT Jl. Ir. H. Juanda No 15, Samarinda, Indonesia 75123

DIGITAL-BASED LANGUAGE TESTING IMPLEMENTATION DESIGNED FOR EFL LEARNERS

Magdalena Kartikasari Tandy Rerung, S.S., M.Hum[1], Junita, S.Hum[2]
Universitas Bunda Mulia
mrerung@bundamulia.ac.id, junita@bundamulia.ac.id

ABSTRACT
The technology-based test has been a controversial issue during the last decade. On the other hand, it can offer solutions for an effective long-distance learning or e-Learning program which can be applicable in the situations where learning process is limited either in physically or time-constraint. Despite of monitoring students' progress assessment, tests are needed to evaluate whether the learning process is effective or not. With the intervention of internet and technology in the education field, this offers many ways to apply assessment for the learning progress. Web-based-test (WBT) offers a great varieties and features which is easily accessed. On the other hands, technology based tests seem able to gain students interest and motivations dealing with the language tasks. Therefore the following research intends to find out which types of web based test is more effective to be applied for language evaluation. Furthermore, this research used two kinds of WBT, namely *Google Form* and *Quizizz* to look further of some benefits from these types of tests also to investigate its effectiveness. The following study applied the quantitative method where scoring result is gathered and calculated to measure the average score and the most preferable test-method. There were around 66 university students from two different majors. They were taking Business English at their second semester in Bunda Mulia University. The tests were conducted during the online program due to the pandemic situation. The finding reveals that both tests have its strongest and weakest points whereas the better result shows that *Quizizz* is higher than in *Google Form*.
Keyword: ***digital testing, language learning, assessment***

1. INTRODUCTION

Assessment procedures in informal education and training mostly still on paper-based while technology is rapidly developing – computers, software, and the internet. This development of technology unconsciously affects education. Such as teaching and learning have lifted the standard from paper-based to a computer-based system of assessment, and it also can be a media in the teaching and learning process. Griffin (2003 qtd in Summak et al., 2010) pointed out that technological and educational improvements are to be expected to modify the way that many schools look and work. Students are expected to be technology literate to outshine in future jobs and to be creative citizens. Besides, according to Holznogel (2005 qtd in Summak et al., 2010), technology is a tool that teachers can use to deliver materials and implement practices in better ways. Students seem to be more motivated to learn through it, and it is proved by Idio (2000) who reported if technology, which is a computer, increases student motivation to learn and is particularly imperative for teachers as instructional tools in the classroom.

The correlation between scores Paper-Based Testing (PBT) and Computer-Based Testing (CBT) has been a controversial issue during the last decade. According to Wolfe and Manalo (as quoted in (Greiff & Funke, 2009), "A serious shortcoming of most research concerning score differences attributable to test delivery medium is the fact that most of these studies examine group differences rather than individual differences. These studies have suggested that, on average, there are only small differences between scores on computer-based and pencil and-paper tests." In addition, Ridgway and McCusker (as quoted in Lent, G. V.) stated:

> "They list several uncomfortable truths that impact the success of computer-based assessment including, working with ICT across the educational sector is particularly difficult because of the wide range of hardware and software platforms that are used; ICT has had minimal impact on classroom practices – let alone on attainment; optimistic claims for the likely effectiveness of e-assessment [especially e-portfolio work] are rarely grounded in evidence, such evidence as we have about the benefits of e-portfolios is weak, and discouraging; we know far too little about how to design assessment to support learning." (P.78)

Besides, the research result of Computer-Based (CBT) VS. Paper-based (PBT) Teaching: Mode Effect, Relationship between Computer Familiarity, Attitudes, Aversion and Mode Preference with CBT Test Score in an Asian Private EFL Context by Khoshsima et al., some of the respondents who preferred to take PBT stated that they could write down or underline some key-words or phrases for future returning. In PBT, they could put a bullet next to the questions they did not know their answers for future review. They claimed that CBT required more technical knowledge and also expressed their concern about the system breaking down and crash. They were afraid of computers not working as they expect during the test. Some well-known authoring software that can be since early 90s for instance Pinpoint, Vocab, Wordstore, Storyboard, Gapmaster, Choicemaster, Testmaster, CoMIL, Dasher, CALIS, and Gapkit. At the end of 90s, there are some well-developed updated version such as LIBRA, WinCALIS, CoMIL, RealEnglish, Course Builder, Multimedia Testmaster and Teleste Partner Tools (Supyan, 2004 qtd in Nordin, Arshad, Razak, & Jusoff, 2010)

Assessement and Technology

According to Brown, "assessment is an integral part of the teaching learning cycle. In an interactive, communicative curriculum, assessment is almost constant. Tests, which are a subset of assessment, can provide authenticity, motivation, and feedback to the learner. Tests are essential components of a successful curriculum and one of several partners in the learning process." There are seven fundamental principles (Fairbairn & Brown, 2005):

1. Periodic Assessments, both formal and informal, can increase motivation by serving as milestones of student progress.
2. Appropriate assessments aid in the reinforcement and retention of information.
3. Assessments can confirm areas of strength and pinpoint areas needing further work.
4. Assessments can provide a sense of periodic closure to modules within a curriculum.
5. Assessments can promote student autonomy by encouraging students' self-evaluation of their progress.
6. Assessments can spur learners to set goals for themselves.
7. Assessments can aid in evaluating teaching effectiveness.

One of the assessment types is computer-based testing. Recent years have seen a growth of assessment in which the test-taker performs responses on a computer. Brown (2004, in (Fairbairn & Brown, 2005) acknowledged some computer-based tests (also known as "computer assisted" or "web-based" tests) are small-scale "home-grown" tests available on websites. A specific type of computer-based test, a computer-adaptive test, has been available for many years but has recently used by both teachers and students. Computer-based testing, with or without CAT technology, offers these advantages:

- Classroom based testing
- Self-directed testing on various aspects of a language (vocabulary, grammar, discourse, one or all of the four skills, etc.)
- Practice for upcoming high-stakes standardized tests
- Some individualization, in the case of CATs
- Large-scale standardized tests that can be administered quickly to thousands of test-takers at many different stations, then scored electronically for rapid reporting of results

Some disadvantages are present in our current weakness for computerized testing. For example:

- Lack of security and the possibility of cheating are inherent in classroom based, unsupervised computerized tests
- Occasional "home-grown" quizzes that appear on unofficial websites. May be mistaken for validated assessments.
- The multiple-choice format preferred for most computer-based tests contains the usual potential for flawed item design.
- Open-ended responses are less likely to appear because of the ·need for human scorers, with all the attendant issues of cost, reliability, and turnaround time.
- The human interactive element (especially in vocal production) is absent.

Evolution of Language Testing in the Digital Era

First introduced in the late ‘70s and early ‘80s, computer-based tests (CBTs) started being well-developed and implemented in order to have a simpler and more practical of testing comparing with the paper and pencil tests. CBTs was effectively helped the test takers to take an individual test relying to a computer or a closed network. Even though students prefer to use CBTs rather than the traditional ones, this preference does not prove their performance better (O’Malley et al.,2005 qtd in (Álvarez, 2016).

The next development in technology-based testing is the use of Computer adaptive testing (CAT) where the types of tests, the computer can make the necessary calculations needed to estimate a person’s proficiency and to choose the questions to present. It further argued that computer adaptive testing “offers a potentially more efficient way of collecting on people’s ability” (Hughes, 2003 qtd in (Álvarez, 2016). Some benefits showed in CATs are like “self-pacing; challenging; immediate feedback; multimedia presentation” (Dunkel, 1999 qtd in (Álvarez, 2016). The later invention as the development of internet, the Web-Based Testing (WBT) is introduced. In WBTs the test is located as a website on the tester’s server where it be accessed by the test-taker’s computer. The client’s browser software displays the test, the test-taker completes it and some can directly show the score once it is done. Some advantages shown in WBT such as “flexibility in time and space, easy to write, and affordability” (Álvarez, 2016)

Looking the diversity of computer devices, the latest invention is accessible not only from computer but also from smartphones or tablets. It is called as Computer-assisted language testing (CALT). Pathan (2012, qtd in (Álvarez, 2016) identifies three domains in CALT: “(1) the use of computers for generating tests automatically, (2) the interaction between test-takers and the

computer (in the form of online interaction), and (3) the use of computers for the evaluation of test takers' responses". Based on the development of testing in the digital era, various of application occurred such as Quizziz that is going to be applied as a test tool in this research. Quizziz is categorized as WBTs and the integration in CALT where the students can both access the test from either computer or other devices that has the internet connection.

Web-based Formative Assessment

Google Forms are generally used to do surveys without difficulty and quickly since the users can assess them anywhere and whenever. It helps users to plan events, ask questions to their workforces or customers, and gather the varied type of information and efficiently. Nowadays, educators also use Google Forms to assess their students. They believe that it is easy and straightforward to use. It allows the users to include different types of questions such as short answers, paragraphs, multiple selections, verification boxes, pull-down, linear scale, a grid of several options, among others.

Below are some advantages of using Google Forms (as quoted from DataScope):

- It is a free online tool that allows its users to collect information quickly and efficiently.
- With Google Forms, users can create surveys in a few minutes to ask your clients or collaborators about your products or service.
- To start using this tool, users only need a Google account
- The interface is straightforward to use. Any user with common Internet knowledge can create forms using this tool.
- The assistant is simple to use. The What-You-See-Is-What-You-Get interface makes it easy to drag and drop form elements and organize them based on actions or events.
- At the design level, it is possible to choose between a palette of colors and one's images as a background.
- Google forms stores the feedback received so we can analyze it in detail.
- The forms are integrated with Google spreadsheets; therefore, it is accessible a spreadsheet view of the collected data.
- The general configuration of forms or surveys allows you to collect the recipient's email address and limit the answers.
- For advanced users, the type of data that can be inserted into a field can be customized using regular expressions. It helps customize the form even more.

Moreover, the disadvantages of using Google forms are (as quoted from DataScope):

- It is necessary to have the internet to be able to use this tool.
- Design customization is minimal. Advanced users can change the design to use the tool for a higher number of purposes.
- There are some security concerns. The user has to create the right password and protects it to increase the level of security.
- There are certain limitations regarding the capabilities of this tool. It accepts texts up to 500 Kb; images up to 2 Mb; and for spreadsheets, the limit is 256 cells or 40 sheets.

On the other hand, Quizizz is a self-paced learning tool that helps every student celebrate their accomplishments. Teachers integrate Quizizz into instruction, review, and assessment to sustenance students in Pre-K through College. It is designed for educators, which allows them to conduct student-paced formative assessments in a fun and engaging way for students of all ages. (taken from Quizizz help center) The prominent features contain:

- Student-paced: Questions appear on each student's screen, so they can answer questions at their own pace and then review their answers at the end.
- BYOD: It can be played by students using any device with a browser, including PCs, laptops, tablets, and smartphones. Besides,
- Thousands of free quizzes: Amazing teachers around the world create thousands of great questions on Quizizz every day! This community effort generates excellent content that everyone can use.
- Quiz Editor: It allows the users to pluck questions from any quiz, easily add images from the internet, auto-save your progress, and tons of other features.
- Reports: The reports give users detailed class-level and student-level insights for every quiz the user conduct. The user can also download the reports as an Excel spreadsheet.
- Quiz Customization: The user has multiple options to customize their quiz session to toggle the level of competition, speed, and other factors.

Besides, Quizizz provides a feature where the teachers can assess the students through games, and a game joint using a teacher account can have up to 500 players in a single game. There are eight settings that the user can control before starting a quiz where they can see the settings menu after selecting their preferred game type. According to Vogel et al. (2006), collaboration with computers can advance learning motivation, and the design, which makes the learning process similar to playing a game, can boost learning motivation. Game is believed can be influenced by learning motivation since it can engage the students in the learning process.

Assessment on the other hand raises competitive skill among students, if the teachers view tests as part of challenging activities then it is possible to use tools that can combined both excitement and evaluation as well. Here the research was designed to find out which of the web-based test selected are more effective to be applied in the classroom and also to investigate which language skills are affected by these tests.

2. METHOD

Research design

To achieve the research goals, the study employed a quantitative approach including descriptive statistics also t-test to measure the significance result based on the web-test used. Results analyzed based on the tables and mean scores.

Participants

The participants were sixty-six students studied in Bunda Mulia University and taking Business English subject during their second semester. The participants come from two different majors those are Management and Design and Visual Communication program. Both of the test were instructed differently according to the major.

Research instrument

The test itself will be based on WBT and the name is "Quizziz", it is a wellknown platform that can be accessed freely. What makes it interesting is how it is visually displayed differently from the test takers and test-boards. There are two different interfaces showing. Below is the comparison of visual displaying for each tests.

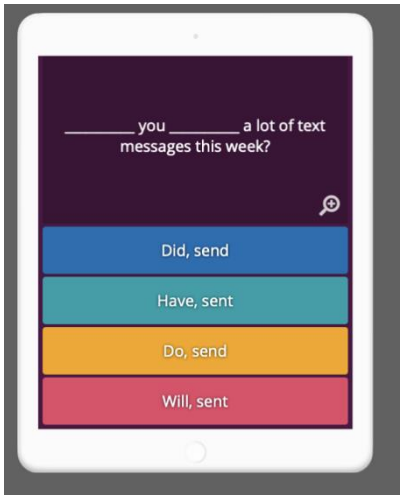

Image 1. Quizizz Display

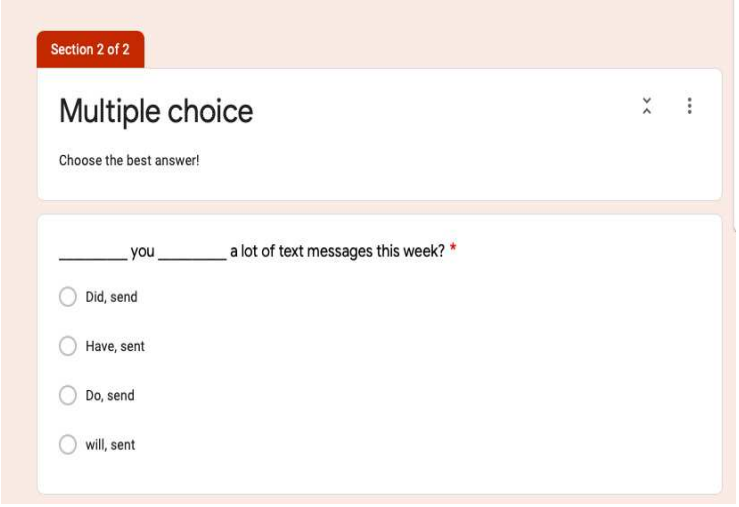

Image 2. Google Form Display

There are 20 items asked in the test and those items were divided into three categories, those are *structures* (5 items), *vocabulary* (11 items), *reading comprehension* (4 items). Topics discussed were about situations in Business English which they had learned during the semester.

Data collection procedures

The tests were distributed at the end of the semester to evaluate their learning progress based on the topics learned. Since at the end of the semester the students were instructed to study at home, therefore the tests were taken outside the classroom and conducted at the same time. The duration for both tests were practically two-hours and closed once it reached the time limit. As the result, both web-tests provide statistical data in Microsoft Excel format where the researchers could directly classify and calculate based on the needs.

Data analysis

Once the result gathered, the data analyzed quantitatively following the statistical procedures. The average score from each distributed test will be used to observe the progress of students' language learning related to the lesson plan during the semester. The overall result also discussed comparing between Google Form-based test and Quizizz-based test to investigate the differences in both of the result. The method used is the statistical approach and t-test value.

3.FINDINGS AND DISCUSSION

RQ 1. Are there any different result based on the two tests?

As explained earlier that the test was conducted based on two different groups on the same subjects. The time allocation was given at the same time which was around two-hours per session. However due to the students' various competencies, some of them had successfully conducted the test less than an hour, and some of them require more time. Furthermore, couple of cases were taking longer because of the network or technical issues so these students had to retake the test. Yet these problems were actually have been anticipated earlier, since they had used the same web-based test during the quiz, therefore the problems were not becoming major issues. To answer the result questions, below is the score gathered based on *Google-Form* and *Quizziz* web/mobile-based test. In the first table, it shows the test is homogeneous based on the significance point *0.2* which is higher than *0.05* and the test is acceptable.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Score_GoogleForm	.095	66	.200*	.971	66	.119
Score_Quizziz	.093	66	.200*	.978	66	.282

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 1. Normality Test

Statistically, the result shown in table 2 both groups have different result in average. The group that has higher mean score is the group that used Quizziz which 60 points in average and Google-Form group got around 56 points in average, with mean the lowest score is 15 and the highest score is 95 both for Google Form while in Quizziz the lowest score is 20 and the highest is 100 points which a perfect score.

Statistics

		Score_Google Form	Score_Quizziz
N	Valid	66	66
	Missing	0	0
Mean		56.74	60.00
Median		60.00	60.00
Mode		65	70
Std. Deviation		19.952	19.155
Minimum		15	20
Maximum		95	100

Table 2. Statistics Result

Furthermore, to show the evidence whether both of the test got a significant result, the independent t-test is used with the test value of 51 and below is the result.

One-Sample Test

Test Value = 51

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Score_GoogleForm	2.338	65	.022	5.742	.84	10.65
Score_Quizziz	3.817	65	.000	9.000	4.29	13.71

Table 3. Significance Result

Based on the result, we can see there is a significant difference in mean score ($p < .001$) , especially in Quizziz score where it can support that the test is slightly effective than Google Form test-based. In addition to see how the score distributed based on the category of the questions the results below is to explain based on the students correct answers and also to answer the second research question.

RQ 2. How do the result difference based on the questions categories?

To answer the second question, the data was analyzed based on its question categories, those are *language structure, vocabulary,* and *reading comprehension*. The numbers were gathered based on the correct answers for each category. Each categories were included 3 to 5 items of questions and varied based on topics of the lessons. Below is the table of the overall result

Question_Categories		Question_Types_Quizziz	Question_Types_GoogleForm
Structure	Mean	29.40	31.80
	N	5	5
	Std. Deviation	9.633	3.114
	Minimum	13	27
	Maximum	37	35
	Range	24	8
Vocabulary	Mean	37.91	38.73
	N	11	11
	Std. Deviation	10.368	9.242
	Minimum	17	26
	Maximum	53	54
	Range	36	28
Reading	Mean	38.75	45.50
	N	4	4
	Std. Deviation	10.689	13.229
	Minimum	24	29
	Maximum	48	60
	Range	24	31
Total	Mean	35.95	38.35
	N	20	20
	Std. Deviation	10.455	9.837
	Minimum	13	26
	Maximum	53	60

Table 4. Question-Categories Result

Based on the table, it can be seen that the average of correct answers are found almost equally in each category of both tests. However, the most correct answers are found in *reading comprehension* with the total around 45 corrects from total of 66 students answered in Google Form based test which is above higher than in Quizzis. On the other hand, *vocabulary* questions just almost got number of correct answers for both test which are around 37 to 38 correct answers. Similarly in *structure* questions, both tests got in between 29 to 31 for the answers. However, the overall results show that Google Form test-based gathered more correct answers comparing to Quizziz even though the differences among scores are not quiet far difference.

Based on the results, there are some interesting facts that can be discussed further. First, it can be seen that the score result based on the two tests showing a significant different. The reason may be relied on the features that Quizziz has for example it is more colorful and attractive plus it also provide some hints for the students to help them answering the difficult questions, on the other hand Google Form might provide a plain visual and does not offer any hints or clues to guide the students. On the contrary, Quizziz can provide a limited-time for each question this probably can be such a challenging task for students who are better-performed yet can be some burdens for lesser-performed students. Google Form on the other hand does not limit the time for each item,

however the teacher can still control when the test is accessed or not. This type of time limitation can be beneficiary for some students who needs longer time in thinking and focusing.

Secondly, for each category of questions showed different result where most correct answers were found in Google Form. From all the categories, *reading comprehension* got the highest correct answers yet there is a quiet different range in both tests. This might explain that students are probably more comfortable in answering text-based test without time limited and in the form of passage like in Google Form, unlike Quizzes each item in reading test there were 2 minutes duration and also students have to scroll up and down from their phone or laptop to find the answers. This technique might find them quite difficult and ineffective.

4. CONCLUSIONS

The objective of the present study was to find the significant different between the two web-based tests, which are Google Form and Quizzz. The result and also the discussion showed that both are showing different results and it is significant. Quizziz which showed the higher result is suggested to be implemented as a means of assessment that is not only challenging yet entertaining. In this case the students are not merely focused on the test item yet with their device they can feel more comfortable in accessing the test. Comparing to Google Form in which students might still feel like paper-based test where they have to focus on each item of the test. This might trigger for those students who are easily pressured by any kind of tests or assessment. On the other hand, time allocation might bring disadvantage for Quizziz since each item should have its own duration. Students who are not good in time-management could feel uncomfortable whereas students who like some challenges would find this feature as an advantage for them.

For the categories of questions, *reading comprehension* showed the strongest result among other skills like *structure*, and *vocabulary*. Looking at the differences, it is suggested that reading comprehension should be considered once the test used Quizziz, since viewing the passage from any devices could make students' visual ability easily tired. Google Form is strongly suggested in this case, it offers comfortability to meet the students' eyes and more effective in answering each item of the questions, so they do not need to scroll up and down like in Quizziz

More researches on how the Web-based test forms can attract students to actively participate and how these web-based test affect the students' motivations an e-Learning effectiveness as a whole are also necessary.

Acknowledgments

Most thankful for the University of Bunda Mulia that has provided financial support for this paper, and we are also grateful for the insightful comments offered by the internal reviewers team at Bunda Mulia University. The generosity and knowledge of one and all have enhanced this study in immeasurable ways and make this study becomes better and avoided by many errors.

Reference:

Álvarez, M. F. (2016). Language testing in the digital era. *Technology-Enhanced Language Learning for Specialized Domains: Practical Applications and Mobility*, *December*, 61–72. https://doi.org/10.4324/9781315651729

Fairbairn, S., & Brown, H. D. (2005). Language Assessment: Principles and Classroom Practices. H. Douglas Brown. *TESOL Quarterly*. https://doi.org/10.2307/3588320

Greiff, S., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In *Main* (Issue September). https://doi.org/10.2788/60083

Idio, I. E. (2000). *A Study of Teachers Perceptions about Their Ability to Integrate Computer Technology into the Instructional Process: A case study, Dogwood Elementary*. University of Sarasota.

Summak, M. S., Samancioğlu, M., & Bağlibel, M. (2010). Technology integration and assesment in educational settings. *Procedia - Social and Behavioral Sciences*, *2*(2), 1725–1729. https://doi.org/10.1016/j.sbspro.2010.03.973