

Assessing Language Proficiency through Computer Technology: Potentials and challenges

Isry Laila Syathroh*, Bachrudin Musthafa, Pupung Purnawarman
Universitas Pendidikan Indonesia
Bandung, Indonesia
*islaisya@yahoo.com, din_musthafa@upi.edu, purnawarman@upi.edu

Abstract—Recent trends in the implementation of using computer in language learning in many Indonesian high schools have led to the development of Computer-Assisted Language Testing (CALT). The aim of this paper is to elaborate the result of analytical review of relevant and recent literatures about the potentials and challenges in implementing CALT. Research articles from reputable journals were carefully chosen and reviewed to achieve the aim of this research. There are two criteria of the selected journals, namely: 1) research published within the last ten years; and 2) research topics are related to the implementation of CALT in many different institutions around the world. The result of this research suggests that despite the challenges and some its negative aspects, CALT can be one of instrumentals in expansion and innovation in language assessment. CALT has also been suitable for the 21st century generations which are fast, dynamic, individualized and efficient, making it fit as the integral parts of Indonesian educational system.

Keywords—*language assessment; Computer Assisted Language Testing (CALT); Computer-Based Test (CBT)*

I. INTRODUCTION

A. Computer Assisted Language Test (CALT)

Assessing language proficiency can be administered either by *Paper-Based Test (PBT)* or *Computer-Based Test (CBT)*. The later one is also known with the term *Computer-Assisted Language Test (CALT)*. Noijons defines CALT as “an integrated procedure in which language performance is elicited and assessed with the help of computer” [1]. There are three integrated process of CALT, namely: 1) generating the test, 2) interaction with test candidate, and 3) evaluation of responses [1]. In the generating process, the computer selects a number of items from an item bank, randomly or following some selection procedure. In its second process, CALT involves the interaction process between computers and candidates (test takers). This process is most dominant difference between CALT and paper pencil. And the last process is the evaluation of responses. In many CALT programs, a candidate's data may already have been evaluated during the preceding process, but in this process all data are called up for a final evaluation of the complete response.

B. Attributes of Computer Assisted Language Test

Suvorov and Hegelheimer develop nine attributes/ characteristics of CBT, namely: *directionality, delivery format, media density, target skill, scoring mechanism, stakes, purpose, response type, and task type* [2]. Each characteristic is explained in separated sections below:

1) *Directionality*: The term directionality refers to three types of test directions: linear, adaptive and semi adaptive. Computer linear tests administer the same number of test items in the same order to all test takers. In computer adaptive test, each task is selected by the computer based on the test taker's performance on the previous task. Successful task completion results in a more complex question, while incorrect task completion results in an easier next task. Since computer adaptive test has a lot of limitations (like high cost, exposure to test item, issue with algorithm in item selection), computer adaptive test might become the solution. Computer semi-adaptive tests are adaptive at the level of a group of items called testlets or at the level of the whole test where test takers are given a version of the test that corresponds to their proficiency level as determined by a pretest [3,4].

2) *Delivery format*: Computer Assisted Language Test (CALT) can be divided into two types of delivery format: Computer Based Test (CBT) and Web Based Test (WBT). In CBT, various offline delivery formats are used such as CD, DVD & standalone software applications that can be installed on an individual computer. However, in WBT mode, test takers do their test in an online format. Regarding to this issue, predicts that due to rapid technological advances WBT will gain more popularity and witness further development in the near future [4].

3) *Media density*: One of the issues related to the media density used in CBT is the availability of different media formats and the possibility of their integration. CBT can use a single medium like audio for listening test or test-based reading test. However, some other CBT can integrate the use of multimedia within a test, for example: audio, images, videos, animation, graphics, and so on. However, claim that

this issue can result complex measurement and poses a threat to test validity [5].

4) *Target skill*: Most CBTs are designed for assessing single language skill, for example: reading test, speaking test, listening test or writing test. However, some other CBTs can be designed to assess integrated language skills, for example: listening and speaking test. Integrated skills assessment reflects the complexity of language use contexts and is believed to enhance the authenticity of language tests through interactivity provided by integrated tasks that are typically performance-based (Plakans, n.d) [4,6,7]. One of example of integrated skill test is the new TOEFL IBT [8].

5) *Scoring mechanism*: In CBT, test takers' performance can be evaluated either by human raters or by computers. Computerized scoring of the input can be done by matching exact answers or analyzing test takers' responses. Exact answer matching entails matching test takers' responses with the correct preset responses. This type of scoring is typically used for the evaluation of receptive skills (i.e., reading and listening) and, sometimes, productive skills (e.g., writing) in the form of one word or even short phrase answers provided that the test has a prepiloted list of acceptable answers, including the ones with common spelling errors [9].

6) *Stakes*: Like in Paper Based Test (PBT) CBT also can have low, medium and high stakes for test takers. A low stakes test is a kind of test which gives little influence for test takers, for example practice test, self-study test, etc. Medium stakes test can give medium impact to the test takers, for example: progress test, placement test, etc. While high stakes test give huge impacts to test takers' lives such as National Examination, certification tests, promotion, etc [10].

7) *Purposes*: Test purpose can be defined as test type and decision which is made on the basis of the test performance. There are two types of test purpose: curriculum-related and other, or non-curriculum-related [2]. Curriculum-related tests can be used for the purposes of admission to a program, placement into a specific level of the program, diagnosis of test takers' strengths and weaknesses, assessment of their progress in the program, and their achievement of the program's objectives. While non-curriculum-related tests are used for language proficiency assessment and screening for non-academic purposes, for example: to make decisions regarding employment, immigration, and so on.

8) *Response type*: In CBT, there are two types of responses which can be done by test takers [11]. The first response type is selected responses. Selected response assessment involves tasks that require a test taker to choose a correct answer from a list of options, for example: multiple choice question. And the second response type is constructed responses. In this type, test takers must develop their own answers and produce short or extended linguistic output.

9) *Task type*: There are three categories of CBT task types: selective, productive and interactive task type [12]. The examples of selective task types are multiple choice questions, yes/no questions, etc. The examples of productive are written

and oral narratives, short answer tasks, and cloze tasks. While the examples of interactive task type are matching, dragging and dropping the answers. All nine attributes elaborated above can be summarized in the table below:

TABLE I. NINE ATTRIBUTES

NO	ATTRIBUTES	CATEGORIES
1	Directionality	Linear, adaptive and semi adaptive test
2	Delivery Format	Computer Based Test (CBT) and Web Based Test (WBT)
3	Media Density	Single medium and multimedia
4	Target Skill	Single language skill and integrated language skills
5	Scoring Mechanism	Human based scoring and computer based scoring
6	Stakes	Low stakes, medium stakes and high stakes
7	Purpose	Curriculum related and non-curriculum related
8	Response Type	Selected response and constructed response
9	Task Type	Selective, productive and interactive tasks

(Table 1: Suvorov & Hegelheimer' s attributes of CBT [2])

II. METHODOLOGY

Through this paper, the writer intended to review several recent articles related to the implementation of Computer-Assisted Language Tests (CALT) in some educational institution to get the information about its potentials and challenges. Analytical review is a written synthesis of journal articles, books and other documents which summarizes and critiques the past and current state of information about a topic, and organizes the literature into subtopics, and documents the background for a study [12]. The relevant journal articles were searched and selected systematically through *Google Scholar* and *Educational Research Information Center (Eric)* data base. There are two criteria of the selected journals, namely: 1) research published within the last ten years; and 2) research topics are related to the implementation of CALT in many different institutions around the world. Through a detailed review and analysis about this topic, the possibilities and challenges of the implementation of CALT were identified. Finally, the last part of this paper points the key findings which could provide readers with wider and deeper perspectives of CALT which can be adopted in other educational contexts.

III. RESULTS AND DISCUSSION

In this part, the writer tries to review some emerging themes found in several recent articles about the potentials and challenges of CALT implementation. The elaboration is explained into several educational aspects of CALT, namely: aspect of economy, aspect of system implementation, aspect of test administration and design, and aspect of accessibility.

A. Aspect of Economy

Viewed from the economy aspects, the implementation of CALT may be more efficient than Paper Pencil Based Test (PBT). In economic perspectives, the advantages might include several factors, like: cost-effective of test in long term, reduce paper and shipping, and preparing students for a more global economy.

Kettler, Scholz, Oderman, Hixon, & Weigert, claim that the implementation of CALT may cost less than PBT [13]. In traditional test modes, schools require the process of printing and shipping test booklets. While in CBT modes, less paper is needed. Computer-based tests significantly reduce the consumption of paper [14-16]. So we can say that the implementation of CBT can promote eco-friendly environment. Like what said that computer-based assessments reduce the costs associated with entering, collecting, aggregating, verifying, and analyzing data [17,18]. Electronic delivery is less expensive than printing and mailing large quantities of testing materials. In addition, errors found in test booklets or answer sheets can be quickly and easily corrected, instead of reprinting and reshipping testing materials at considerable expense [19,20].

Moreover, students who participate in computer-based instruction and testing may be more prepared to complete in the global economy [21]. They argue that students in the 21st century need to know how to use technology to obtain good jobs, and teaching students to navigate and successfully complete online tests might help prepare them for the future. However, in the short run, the implementation of CALT often cost more than PBT because they are costly to develop and implement. Many schools might not currently have enough computers and some other related facilities, so the cost for providing these items might be big.

B. Aspect of System Implementation

Viewed from aspect of system implementation, the implementation of CALT might give several advantages, namely: 1) the administration of CALT is more efficient. In PBT, test materials need to be distributed to and collected from each administration site. So it will need additional administrators and extra rooms to store the test materials. 2) accurate data collection. CALT will ensure that data are more likely accurately collected and easier to store. Moreover, responses generally are accurately captures and scored. In PBT however, students often make some marks on the answer sheets which can result in inaccurate scoring. 3) easy to change the test if mistakes are discovered after it has been finalized [22], 4) results and other data can be stored in much less space and it is easier to retrieved, 5) test security may be improved if schools and districts do not have hard copies ahead of time 6) on demand testing and 7) potential to shift focus from assessment to instruction. It has been an issue among educators that assessment often drives instructions. With the implementation of CALT which is closely aligned with instructional methods may have the potentials to move the focus back to instruction [13].

Nevertheless, the implementation of CALT might give some challenges, namely: 1) test-day logistics, 2) same schools might not have enough computers and some other related facilities, 3) finding instructional time to teach students how to navigate the test and how to use online tools. Classroom instruction time might be needed to give students a training about computer navigations and about how to use test tools. Since each test has different platforms, so students need to be trained specific computer navigation skills, 4) practice tests/manuals need to be available far in advance of test so teachers

can teach needed computer skills. Some test vendors sometimes fail to make practice tests and manuals available far enough ahead of test day to provide sufficient instructional time [13].

C. Aspect of Test Administration and Design

Viewed from the aspect of test administration and design, the implementation of CALT might give several potentials, namely: 1) increase authenticity. Some CBTs might have the potentials to more authentically assess students learning than PBTs. CBT assesses test taker's language ability accurately by providing more efficient standardization of test administration conditions [23], 2) innovative ways to assess students (role playing, simulations, data manipulation. Al-Amri states that the most important benefit of CBT is the innovation, efficiency and productivity that can be achieved in CBT, since input materials are presented in text, graphics, audio, and video which simulate target language situations and develop the authenticity of test tasks by enhancing the interaction between test takers and test tasks [23], 3) it is preferred by students.

Some students have used computers to play games and some of them might receive the instruction through computers. Some students might prefer CBT since they can customize the assessment based on their personal preference, like colors on the screen, font types, font sizes, etc. Due to the possibility of customizing the assessment based on personal preferences, some people prefer to take CBT version of the test. For instance, all students have the option to select their own background color and font size preference on computer screen. Although some students may prefer CBT, others may prefer paper and pencil-based test [21].

Many studies have been done to examine the preference of test takers on testing administration mode [23-25]. However, some test takers prefer paper-based testing process because they are accustomed to taking notes and circling questions and/or answers for later review. 4) self-selection options, 4) immediate results. Unlike paper examinations in conventional classrooms, immediate viewing of scores on screen is provided in CBT to give test takers the instant feedback. Immediate feedback, accurate test result reports and the possibility of printing the basic testing statistics are other advantages of using computer in assessment field that enable test takers take the test at any time [26]. Teachers often need time to assess the test and to make decision about the tests. 5) shorter test duration. However, CALT might also give challenges, namely: 1) students might not always make good choices about which embedded resources they use, 2) some students might have computer anxiety.

D. Aspect of Accessibility

Viewed from the aspect of accessibility, CBT is potential for more universally designed assessment. According to Ketterlin-Geller, CBT presents an efficient tool for customizing assessments to meet individual needs within a universally designed environment [27]. For example, all students would benefit if a CBT has allowable features that make the test easier to understand and navigate. CBTs often allow many options for interacting with the assessment, which fits well with the concept of universal design. However, CBT might be a

problem in terms of accessibility if students some students do not have keyboarding skills. Young learners might lack the hand size to use mouse and keyboards. Moreover, students who do not receive instructions via computer might be awkward to use them.

IV. CONCLUSION

Assessing language proficiency using computers has been one of the recent trends in educational realm. The aim of this paper is to highlight the potentials and challenges of CBT implementation in some educational institutions. There are four aspects discussed: economy aspect of CBT, system implementation aspect of CBT, test administration and design aspect of CBT and accessibility aspect of CBT. The result of this research suggests that despite the challenges and some its negative aspects, CALT can be one of instrumentals in expansion and innovation in language assessment. CALT has also been suitable for the 21st century generations which are fast, dynamic, individualized and efficient, making it fit as the integral parts of Indonesian educational system.

REFERENCES

- [1] J. Noijons, "Testing computer assisted language testing: Towards a checklist for CALT," *CALICO Journal*, vol. 12, no. 1, 2012.
- [2] R. Suvorov, and V. Hegelheimer, "Computer-Assisted Language Testing," *The Companion to Language Assessment*, pp. 594–613, 2013.
- [3] P. Winke, and F. Fei, "Computer-assisted language assessment," In *Encyclopedia of language and education*, pp. 353–364, 2008.
- [4] G.J. Ockey, "Developments and challenges in the use of computer-based testing for assessing second language ability," *The Modern Language Journal*, vol. 93, pp. 836–847, 2009.
- [5] D. Douglas, and V. Hegelheimer, "Assessing language using computer technology," *Annual Review of Applied Linguistics*, vol. 27, pp. 115–132, 2007.
- [6] C. Chapelle, W. Grabe, and M. Berns, *Communicative language proficiency: Definition and implications for TOEFL 2000*. In *TOEFL monograph series 10*. Princeton, New Jersey: Educational Testing Service, 2000.
- [7] L. Plakans, "Integrated assessment," [Online]. Retrieved from <http://language-testing.info/video/main.html>.
- [8] Y. Sawaki, L.J. Stricker, and A.H. Oranje, "Factor structure of the TOEFL Internet-based test," *Language Testing*, vol. 26, no. 1, pp. 5–30, 2009.
- [9] J.C. Alderson, *Diagnosing foreign language proficiency: The interface between learning and assessment*. London, England: Continuum International Publishing, 2005.
- [10] C. Roever, *A web-based test of interlanguage pragmalinguistic knowledge: Speech acts, routines, and implicatures*. University of Hawai'i at Manoa, 2001.
- [11] C.G. Parshall, T. Davey, and P.J. Pashley, *Innovative item types for computerized testing*. In Kluwer (Ed.), *Computerized adaptive testing: Theory and practice* (pp. 129–148). Dordrecht, Netherlands, 2000.
- [12] J.W. Creswell, and V. Clark, *Understanding Research: A Consumer's Guide* (1st ed.). Pearson, 2010.
- [13] R. Kettler, C. Scholz, E. Oderman, N. Hixon, and S. Weigert, "Innovative uses of technology to support assessment decision-making and curricular alignment for students with disabilities," Presentation at the Council of Chief State School Officers (CCSSO), 2010.
- [14] K. Kikis-Papadakis, and A. Kollias, *Reflections on Paper-and-Pencil Tests to eAssessments: Narrow and Broadband Paths to 21st Century Challenges*. In I. F. S. & J. B. (Eds.). T (Ed.), *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*, 2009.
- [15] G. Puhan, K. Boughton, and S. Kim, "Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing," *The Journal of Technology, Learning, and Assessment*, vol. 6, no. 3, 2007.
- [16] P. Paek, "Recent Trends in Comparability Studies," 2005.
- [17] V. Buško, "Shifting from paper-and-pencil to computer-based testing: Requisites, challenges, and consequences for testing outcomes—A Croatian perspective," In *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*. European Communities, 2009.
- [18] R. Kozma, "Transforming education: Assessing and teaching 21st century skills," *The transition to computer-based assessment*, vol. 13, 2009.
- [19] B. Bridgeman, "Experiences from large-scale computer-based testing in the USA," *The transition to computer-based assessment*, vol. 39, 2009.
- [20] S.W. Choi, and T. Tinkler, *Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting*. In annual meeting of the National Council on Measurement in Education, New Orleans, LA, 2002.
- [21] K. Cater, D. Rose, C. Thille, and D. Shaffer, "Innovation in the classroom," Presentation at the Council of Chief State School Officers (CCSSO) National Conference on Student Assessment, 2010.
- [22] K. McGrory, "Glitches Delay FCAT Scores." *Miami Herald*, [Online]. Retrieved from <http://www.miamiherald.com/2010/06/05/v-print/1665809/glitches-delay-fcatescores.html.%0A%0A>
- [23] S.S. Al-Amri, "Computer-based testing vs paper-based testing: establishing the comparability of reading tests through the evolution of a new comparability model in a Saudi EFL context," Unpublished.
- [24] H. Khoshsima, and S.M.H. Toroujeni, "Transitioning to an alternative assessment: Computer-Based Testing and key factors related to testing mode," *European Journal of English Language Teaching*, 2017.
- [25] I. Yurdabakan, and C. Uzunkavak, "Primary School Students' Attitudes towards Computer Based Testing and Assessment in Turkey," *Turkish Online Journal of Distance Education*, vol. 13, no. 3, pp. 177-188, 2012.
- [26] H. Mojarrad, F. Hemmati, M. Jafari Gohar, and A. Sadeghi, "Computer-based assessment (CBA) vs. Paper/pencil-based assessment (PPBA): An investigation into the performance and attitude of Iranian EFL learners' reading comprehension," *International Journal of Language Learning and Applied Linguistics World*, vol. 4, no. 4, pp. 418-428, 2013.
- [27] L.R. Ketterlin-Geller, "Knowing what all students know: Procedures for developing universal design for assessment," *The Journal of Technology, Learning and Assessment*, vol. 4, no. 2, 2005.