www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
Budapest Institute

Education Evidence Approach: Type Validity of Computer Based Test (CBT) In Formative and Summative Assessment for Vocational High Schools

Rais Amin[1], Sri Rahayu[2]
[1,2]*Universitas Pendidikan Indonesia, Indonesia*
raisamin03@student.upi.edu, *srirahayu@upi.edu*

Abstract

This study aims to provide some guidelines for SMK to choose the right type of CBT. Advances in adult technology that are now increasingly rapidly have led to new breakthroughs in the national education system in Indonesia. Pandemic outbreak current Covid-19, the government and institutions engaged in education are now starting to take steps again to develop and improve the concept of distance assessment. One of them is the Computer Based Test (CBT) which is a technology in assisting the implementation of the assessment using an online. Handling and developing tests carried out remotely using a digital system at Vocational High Schools (SMK). This study distinguishes CBT into three different types, namely linear tests, automated tests, and computed adaptive tests. So that in the search and consideration in selecting CBT it is necessary to be guided by the purpose of the test in question. The test approach can determine which goals can be achieved using a particular test. The discussion of this study focuses on two different test approaches, namely the formative test assessment approach and the summative test assessment approach. The suitability of each type of CBT in measuring the performance of different test approaches was evaluated based on three test characteristics, namely the purpose of the test, the duration of the test, and the computerized adaptive test. This study aims to provide some guidelines in selecting the appropriate type of CBT for use in SMK.

Keywords
computer based test; assessment; formative and summative

I. Introduction

Indonesian education in 2021 can now be considered as *online*, due to the *Sars-Cov-2* (*Covid-19*) outbreak. The epidemic forced many educational institutions and universities to close temporarily, but distance learning was still being carried out. Several areas of the world are beginning to be affected by this and there is a fear of missing the entire current or future semester.

According to Astuti et al (2019) Education is an obligation of every human being that must be pursued to hold responsibilities and try to produce progress in knowledge and experience for the lives of every individual. Education is one of the efforts to improve the ability of human intelligence, thus he is able to improve the quality of his life (Saleh and Mujahiddin, 2020). Education is expected to be able to answer all the challenges of the times and be able to foster national generations, so that people become reliable and of high quality, with strong characteristics, clear identities and able to deal with current and future problems (Azhar, 2018).

Vocational High Schools (SMK) which are part of one of the vocational education institutions with practical learning needs are constrained by this. As per the researchers' judgment, it is not certain to return to full normal teaching in the near future, as social

distancing is highly favored at this stage. This will have a negative impact on learning opportunities. Education units are struggling to find options in dealing with this challenging situation, these conditions make us realize that scenario planning is an urgent need for academic institutions (Dhawan, 2020).

Educators and school administration staff need to think of a renewal by focusing on educational unit level institutions in creating quality and competent human resources (Amin & Rahayu, 2021) in order to create active, effective, and productive schools. One of them is the use of *online* and *software* that continues to grow, because instructors in traditional classrooms complement face-to-face interactions with *online*. As this trend continues, it is important for instructors to determine the effectiveness and usefulness of combining traditional classroom teaching with *online* (Wooten, 2016).

learning *Online* can be referred to as a tool that can make the teaching and learning process more student-centered, more innovative, and even more flexible.learning *Online* defined as "a learning experience in a synchronous or asynchronous environment on the use of different devices, such as mobile phones, laptops, and others with internet access of course. In this scope, students can be anywhere (independently) to learn and interact with instructors or other students" (Singh & Thurman, 2019).

Measurement of test assessment in measuring science both in theory and practice in order to achieve consensus by showing the ability of students to create contributors (Burke et al., 2013) in learning is also done *online*. Assessment is another point of consideration in distance education and also has many challenges, such as the use of software, hardware, internet connectivity so that it affects the learning process and occurs when the processing demands generated exceed the processing capacity of the cognitive system (Junco & Cotten, 2012). In addition, this creates a problem in providing an objective assessment. Assessment *Online* allows for quick, timely and responsive assessment, but issues such as assessment of descriptive questions, strategies for different subjects need to be addressed (Sarrayrih & Ilyas, 2013).

Digital test developers are one of the first choices that need to be made by the development of test equipment with the type of *Computer Based Test* (CBT) developed. The usefulness and feasibility of different types of CBT-based tests need to be evaluated to select the optimal type of test for a particular testing situation (Becker & Bergstrom, 2013).

The test approach determines which goals can be achieved. There are four different test approaches, but in this study the assessment used was formative and summative assessment. Therefore, the purpose of this study is to discuss three different types of CBT such as linear tests, automated tests, and computerized adaptive tests. The suitability of the CBT types for different testing approaches was assessed based on the characteristics of the test.

II. Research Method

The method used in this study is a *literature review*. This method is a literature search both on an international and national basis which is carried out using data collection from various sources. In the early stages of searching for journal articles, 17,500 articles were obtained in the range of article years, namely 2011 to 2021 using the keyword "*Computer Based Test* (CBT) *Based Exam For Assessment*" which was identified and has not been explored for suitability of articles for compilation. The journal has only about 30 articles that are worked on and relevant. The number of articles taken was only 10 with full criteria, eight articles of medium quality, and two articles of low quality.

III. Results and Discussion

3.1 Educational Test Approaches

Despite the various test approaches and definitions for these approaches, we prefer formative assessment and summative assessment. Before and during test development, the main objective is to determine which approach is best for the test.

a. Formative Assessment

The first approach is formative assessment, which focuses on supporting and improving the learning process in learning facilities by making decisions at the learning level in the classroom where individual characteristics such as performance or knowledge are measured. In addition, there are three types of formative assessment in the form of data-based decision making, assessment for learning, and diagnostic testing. Using this data, teachers may be able to set learning goals based on students' current level of knowledge (Van der Kleij et al., 2015).

Assessment by learning differs from teaching, usually focusing on measurement combined with improving learning, whereas pure instruction focuses on improving learning. The size of the exam reports should be sufficiently precise at the individual level, so that instruction can be adapted (assessment for learning), teachers get a precise picture of the learners' developmental stages (diagnostic testing) or can be combined to provide an overview of the current status of curriculum and school performance (decision making). data-based) (van Groen et al., 2014).

b. Summative Assessment Summative

Test testing is used to make conclusions about individual students (Haertel, 2013) based on measurements. Test results play a role in decision making about mastery of content mostly by students or classes (Van der Kleij et al., 2015), guide decisions about grouping students' abilities, determine entry and exit from education, assistance in vocational schools in accepting decisions (Haertel, 2007). 2013) and affects whether a student passes or not. Two uses of the summative assessment instrument must be designed in such a way that it supports learning and that the teacher must be held accountable for serving the purpose of summation". The first refers to the use of "learning assessment" test results, while the latter has implications for vocational teacher training needs in *Language Assessment Literacy* (LAL) (Black & William, 2018).

Four proposed components are important in any class-based test assessment procedure, namely by setting goals (formally and informally), assigning assignments, and grading (Ahmad, 2020). This implies that these benchmarks are valid, reliable, practical, and the teachers are sufficiently trained to not only design these benchmarks but also administer and assess them according to the standards. Summative assessment from this view is expected to produce most of the *washback* (assessment that affects teaching materials) hindering the adaptation of instructions to support the learning of vocational students (Abdullateef & Mohammed Zain, 2021).

Table 1. Characteristics of the Test Approach

	Test administration purpose	Test length	Level	Report		
				Scope	Report Measure	Precision
Formative assessment	Assessment: Enhance learning and instruction	Preferably short tests because testing is often frequent	Individual or class	One or multiple narrow domains or skills	Ability estimate, score, or indicator for each domain	Low at the individual level
Summative assessment	Assessment: Make a decision about mastery of a domain or admission	Long test acceptable for high-stakes testing, short tests acceptable for low-stakes testing	Individual	One or multiple broad domains	Requires a mastery decision, ability estimate, score, or indicator for each domain or for the entire test	Low for low-stakes testing, high for high-stakes testing at the individual level

(van Groen et al., 2014)

3.2 Types of CBT

Test characteristics that depend on the type of CBT several types CBT. The three types of discussion here are linear tests, automated tests, and computerized adaptive tests.

a. Linear Tests

The first type of CBT is linear testing, where the content, *items*, order of *items*, and the length of the test are the same for everyone. Some forms of linear tests may allow parallelization. Nonetheless, all test forms were collected prior to test administration taking place. *The items* for each test are selected manually by the test developer before the test is administered, often *pre-test* about *the items*.

This type of test is inefficient in terms of measurement accuracy, because the test does not adapt to individual learners (Yan et al., 2016) for test results, total percentile scores, distributions, classifications, or reported ability estimates

b. Tests that Automatically Generated

The second type of CBT is an automatically generated test that targets micro-scale applications, due to the small dimensions and relatively large forces involved (Silva et al., 2021). Therefore, it is also known as auto-assembled testing. Adaptive test assembly is discussed separately. The disadvantage of this test is that the resulting test is not in accordance with the abilities of each student.result, *item* may be too easy or too difficult and may be inefficient in terms of testing. Automated tests have the same reporting possibilities as linear tests.

c. Computerized Adaptive Tests

The third type of CBT is computerized adaptive testing, in which *items* are selected according to the examinee. Computerized adaptive testing produces more precise than linear performance estimates with the same test length. After each response, the examinee's ability is estimated and an item is selected automatically which has *the property that* optimal measurement selection methods have been developed, for example in selecting *items* that are measured optimally at the cut-off point when making classification decisions (van Groen et al., 2014).

Another computerized adaptive test is using a *multi-segment*, the test consists of various *items* adaptiveThis allows test developments to combine multiple subjects into a single test by adaptively within and between subjects.

IV. Conclusion

When developing an educational test, the development of the test should define these objectives. Based on this objective, the development selected relevant test approaches such as formative assessment and summative assessment. After the development by selecting the objectives and testing approach, the next must choose the appropriate type of test. This article aims to provide a test development of an overview of the suitability of different types of CBT for these two testing approaches. Three types of CBT are discussed namely linear tests, automated tests, and computerized adaptive tests.

Automatically generated and computerized linear adaptive tests are best suited for formative assessment and summative assessment. Computerized adaptive tests can also be used given their strong focus on measurement and the wide range of possible adaptations of their test designs.

The focus of this article is to determine which type of CBT is appropriate for each testing approach based solely on four characteristics, namely the purpose of the test administration, the length of the test, the level of interest in the measurement, and the size of the report. The purpose of a test is the importance of a characteristic in developing the test. Other criteria, including the complexity and cost of test administration software, *item*, experience with this type of CBT, acceptance of the type of CBT by relevant stakeholders, and so on are also important in deciding which type of CBT should be used for a particular testing situation.

One of the challenges when defining test objectives is testing that has multiple intended objectives. The author assumes that the test has one most important purpose. If a test has multiple objectives, the development must check whether the type of CBT selected is suitable for all relevant and important purposes. Test programs may also combine formative and summative tests. In these situations, test development must ensure that each test conforms to the intended test knowledge and in the program.

References

Abdullateef, ST, & Muhammedzein, F. (2021). Dynamic Assessment: A Complementary Method to Promote EFL Learning Shifan. *Arab World English Journal*, *12*(2), 279–293.

Ahmad, Z. (2020). Summative assessment, test scores and text quality: A study of cohesion as an unspecified descriptor in the assessment scale. *European Journal of Educational Research*, *9*(2), 523–535. https://doi.org/10.12973/eu-jer.9.2.523

Amin, R., & Rahayu, S. (2021). Application of DRILL Learning Method in Vocational Secondary Education. *VOCATECH: Vocational Education and Technology Journal*, *2*(2), 100–107. https://doi.org/10.38038/vocatech.v2i2.52

Astuti, R.W., Waluyo, H.J., and Rohmadi, M. (2019). Character Education Values in Animation Movie of Nussa and Rarra. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*. P. 215-219.

Azhar,A. (2018). Students' Trends in Islamic Communication Postgraduate in2010-2016 State Islamic University of North Sumatera (UINSU). *Budapest International Research and Critics Institute (BIRCI-Journal)*, P.206-214.

Becker, KA, & Bergstrom, BA (2013). Test administration models. *Practical Assessment, Research and Evaluation*, *18*(14), 1–7. https://doi.org/10.7275/pntr-yz21

Black, P., & William, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Burke, MG, Carter, JD, & Hughey, AW (2013). The Use of Case Study Competitions to Prepare Students for the World of Work. *Industry and Higher Education*, *27*(3), 157–162. https://doi.org/10.5367/ihe.2013.0156

Dhawan, S. (2020). Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22. https://doi.org/10.1177/0047239520934018

Eggen, TJHM (2018). Multi-Segment Computerized Adaptive Testing for Educational Testing Purposes. *Frontiers in Education*, *3*(December), 1–8. https://doi.org/10.3389/feduc.2018.00111

Haertel, E. (2013). How Is Testing Supposed to Improve Schooling? *Measurement*, *11*(1–2), 1–18. https://doi.org/10.1080/15366367.2013.783752

Junco, R., & Cotten, SR (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers and Education*, *59*(2), 505–514. https://doi.org/10.1016/j.compedu.2011.12.023

Linden, WJ van der, & Glas, CAW (2010). *Elements of Advertisements Elements of Advertisements*. Springer New york.

Saleh, A., Mujahiddin. (2020). Challenges and Opportunities for Community Empowerment Practices in Indonesia during the Covid-19 Pandemic through Strengthening the Role of Higher Education. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*. Volume 3, No 2, Page: 1105-1113

Sarrayrih, MA, & Ilyas, M. (2013). Challenges of Online Exam, Performance and Problems for Online University Exam. *International Journal of Computer Science Issues*, *10*(1), 439–443.

Silva, PF, Jouda, M., & Korvink, JG (2021). Magnetostatic reciprocity for MR magnet design. *Magnetic Resonance*, *2*(2), 607–617. https://doi.org/https://doi.org/10.5194/mr-2-607-2021

Singh, V., & Thurman, A. (2019). How Many Ways Can We Define Online Learning? A Systematic Literature Review of Definitions of Online Learning (1988-2018). *American Journal of Distance Education*, *33*(4), 289–306. https://doi.org/10.1080/08923647.2019.1663082

Van der Kleij, FM, Vermeulen, JA, Schildkamp, K., & Eggen, TJHM (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy and Practice*, *22*(3), 324–343. https://doi.org/10.1080/0969594X.2014.999024

Van Groen, MM, Eggen, TJHM, & Veldkamp, BP (2014). Item Selection Methods Based on Multiple Objective Approaches for Classifying Respondents Into Multiple Levels. *Applied Psychological Measurement*, *38*(3), 187–200. https://doi.org/10.1177/0146621613509723

Wooten, T. (2016). Leveraging online testing to enhance student learning. *Advances in Accounting Education: Teaching and Curriculum Innovations*, *19*, 141–163. https://doi.org/10.1108/S1085-462220160000019006

Yan, D., Davier, AA Von, & Lewis, C. (2016). *Computerized Multistage Testing: Theory and Applications*. CRC Press.