European Journal of Educational Research

The Development of Computerized Economics Item Banking for Classroom and School-Based Assessment

Friyatmi[*]	**Djemari Mardapi**	**Haryanto**	**Elvi Rahmi**
Universitas Negeri Yogyakarta/ Universitas Negeri Padang, INDONESIA	Universitas Negeri Yogyakarta, INDONESIA	Universitas Negeri Yogyakarta, INDONESIA	Universitas Negeri Padang, INDONESIA

Abstract: The advancement of information and technology resulted in the change in conventional test methods. The weaknesses of the paper-based test can be minimized using the computer-based test (CBT). The development of a CBT desperately needs a computerized item bank. This study aimed to develop a computerized item bank for classroom and school-based assessments. A research and development method is used in this study, which consisted of four phases, i.e., planning, item development, system development, and field testing. Data is collected through documentation, expert judgment, and field testing. The data were analyzed using descriptive statistics and item response theory. The sample of this study was teachers and high school students in West Sumatera province selected using purposive random sampling techniques. The results of the study are as follows. 1) The computerized item bank has excellent quality based on expert validation. 2) There are 120 items inputted into the item bank system that has a moderate difficulty and good discriminant index based on item response theory. 3) The field testing indicated the computerized economics item banking has high effectiveness of usability, usefulness for the teachers, and feasible for classroom and school-based assessment.

Keywords: *Item banking, CBT, assessment, economics.*

Introduction

The assessment of students' achievement is one of the essential aspects of the education process. A high-quality assessment serves several purposes: for students, it serves as the basis for continuing their study to a higher level of education and also to find a job later. For the teachers and the school, the assessment result becomes evidence of individual and institutional success. In the assessment system, there are various techniques to measure the student's achievement. The assessment model chosen would influence the effectiveness of the education being conducted.

So far, teachers in Indonesia preferred to use the test to measure students' competence, especially by using the paper-and-pencil based test (PPT). It is usually limited to assessing students' cognitive achievement. Teachers are very familiar with the PPT because of its simple administration. However, there are several weaknesses in the implementation of PPT, mainly when it is used for large-scale testing, such as high costs of test production, administration, distribution, and scoring (Wang & Shin, 2009). There is also a chance for cheating in the test (Cizek, 2001).

Cheating is a crucial problem that frequently occurred in the national examination of Indonesia. The occurrence of fraud in the form of illegal peddling of the answer key before or during large-scale tests such as the national examination (Novitasari, 2012). Not only happening in the large-scale examination but also cheating occurs in the small-scale test like the final examination in the classroom (Friyatmi, 2011). The weakness of this paper and pencil test based need to be addressed in order to improve the quality of students' assessment.

The development in technology necessitates the modification and change to conventional testing methods (Abubakar & Adebayo, 2014; Quellmalz & Pellegrino, 2009). The implementation of computer-based testing (CBT) could become an alternative solution to overcome the weakness of PPT. The use of CBT is believed to be more effective for measuring

---

[*] **Corresponding author:**
Friyatmi, Universitas Negeri Yogyakarta, Indonesia. ✉ fri.yatmi@fe.unp.ac.id

students' skills, knowledge, understanding, and that is the embodiment of the authentic assessment (Ockey, 2009). For that reason, the Minister of National Education and Culture of Indonesia started to implement the computer-based test for the national examination since 2015. The number of students taking Computer-Based National Exams (CBNE) has increased in the last four years. In 2019, it was administered to approximately 7.507.116 (90.9 percent) students from 103.000 Junior and Senior High Schools in Indonesia (Kemdikbud, 2019). The result indicated that the CBNE was a success in minimizing the cheating behavior of students, teachers, and stakeholders (Setiawan, 2017). Even the Minister of Education claims the CBNE able to eliminate up to 99 percent of fraud (Utantoro, 2019). It is hoped that the CBT could improve the quality of test implementation by reducing students' cheating behavior (Bodmann & Robinson, 2004) so that students respond to tests under their respective competence.

The advancement of technology opens the possibility to implement the CBT for the small-scale tests, such as for school-based assessment (John AC Hattie & Brown, 2007) and classroom assessment. Some studies (He & Tymms, 2005; Hwang & Chang, 2011; Nicol, 2007; Scibinetti, Tocci, & Pesce, 2011) show that the use of technology is useful for classroom assessment. Moreover, the Indonesian government has implemented a national standard school exam (NSSE) as deciding student graduation. This policy implicates the importance of enhancing the teacher's role in assessing student learning outcomes, as the teacher made 80% of the NSSE's questions. This policy raises a problem because the school or local government does not have an item bank for the test. During this time, students' graduation at an educational level is determined by a national exam administered by the central government, so that the central government makes all test devices. It would demand the preparation and readiness of stakeholders of local government in planning and implementing an effective and efficient examination. The availability of calibrated item bank could facilitate a high-quality school test. It provides valid and reliable test items as part of high-quality testing. Unfortunately, surveys indicate the item bank has not yet been available at school or local government in Indonesia (Hayati & Mardapi, 2014; Retnawati & Hadi, 2014; Suyata, Mardapi, Kartowagiran, & Retnawati, 2011).

The unavailability of the item bank causes the item test carried out by the teacher in every test, preferably just a few days before the examination. The test items were usually taken from the question in textbooks and student worksheets because the time to construct items is short. Furthermore, the teacher did not conduct an item analysis before being tested to the students so that the quality of the item becomes unclear. Researchers have also shown that the items analysis is rarely conducted by the teachers (Ata & Ani, 2012). If the quality of the item is not clear, then the quality of the test becomes questionable.

A high-quality test should begin from the constructing of good test items. A good item is developed through appropriate procedures and tested first in advanced to obtain the estimation of item parameters. Items that have been proven to have good characteristics can be stored to item bank so that it can be used for other tests. However, if the item bank is not available, then the teacher will spend much time to construct the test items in conducting each test. An item bank provides users with ease in planning tests because it contains a collection of items that are already calibrated and are stored systematically, as stated in (Wood & Skurnik, 1969). Item calibration is an essential activity in developing an item bank because, through the calibration process, high-quality items are obtained. Therefore, the existence of the calibrated item bank is crucial for conducting a high-quality test, even more while performing a computer-based test (Xing & Hambleton, 2004).

Based on recent phenomena, the availability of item banking is essential to facilitate the assessment of student's outcomes. The presence of item banking could accelerate the test construction process because it provides a collection of various items with its relevant characteristic information. Teachers and other users just required to select the questions based on competency that would be tested. It could surely reduce a considerable time for teachers in preparing a test instrument. The availability of the item bank would also increase item security. Though initially, the development of an item bank requires considerable investment, in the long term, it would reduce the high cost related to the development of a new test.

The development of an item bank would provide several advantages. First, the policy of decentralization of the national test program could be introduced without sacrificing test results. It would reduce the cost and the time required for test construction (Rudner, 1998). Second, the quality of the test program could be improved because the number of items being stored in the item bank increases the test security (Millman & Arter, 1984). Moreover, the information of the items characteristics in the item bank could ease the teacher in designing the test instrument. The availability of item banking could indirectly minimize teachers' assessment work because they could then focus on improving the quality of the learning conducted, without having to spend considerable time in test construction (Umar, 1999). Item banking could be developed manually by using item cards, or it could also be computerized. The advancement of technology facilitates the development of a computer-based item banking system so that the user may easily store and manage the items contained in the item bank. The availability of computerized item banks would simplify the construction test instruments process, and it has long-term benefits. This study aimed to develop a computerized item bank for classroom and school-based assessments on economics teaching.

The development of a computerized assessment for classroom assessment has been developed in New Zealand under the name e-asTTle (John A Hattie, Brown, Ward, Irving, & Keegan, 2006). It measures the literacy and numeracy skills of

students in grades 2-10 (Rubie-Davies & Rosenthal, 2016). Some research shows the implementation of e-asTTle has a positive change to the classroom assessment conducted by the teacher (Archer & Brown, 2013) and improves on student achievement (Carnegie-Harding, 2016). The difference between e-asTTle and this study lies in the content and the student grade. The initial stage of developing this item bank is focused on economic teaching, but gradually it is expected to be able to cover all subjects for high school students (grades 10-12).

Methodology

Research Goal

The main focus of the research was to develop a computerized item bank. The goal is tested by 1) validating product quality, 2) calibrating the item parameters, and 3) testing the effectiveness of the product usage. Research and development methods are used to achieve the goal. The item bank was established through 4 stages consisting of planning, items development, system development, and field testing. The development procedure of the item bank illustrated in Figure 1.
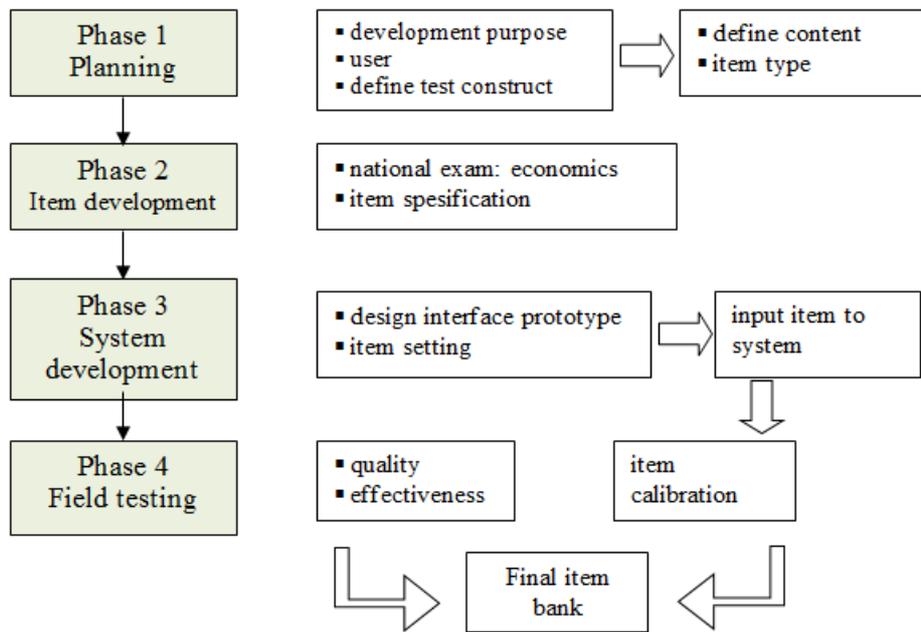

Fig. 1. The Procedure of Item Bank Development

The initial stage was concerned with setting development goals, determining users, and planning test constructs. Based on the test construct, then identified the test content and item type. Both become the basis for establishing tests specification.

Item development is carried out in the next stage. Items are developed by advancing a test specification. Item development for item bank can be done by developing new items or adopting items from standardized tests. In this study, bank items were developed from standardized tests, namely the high school national exam in economics. The high school national exam is a standardized test developed by the Indonesian government under the educational assessment center.

The third stage is the development of the item bank system. In this stage, the database system and prototype system of the item bank are designed. Based on the prototype, the interface of the item bank was developed. The computerized item bank system was developed using PHP and MySQL programming. Furthermore, the items that have been developed in the previous stage are inputted into the item bank system.

The last step is field testing. It was done to test the quality and effectiveness of the item bank and calibrate the items. Field testing results are the basis for improving the item bank system, which then becomes the final item bank.

Sample and Data Collection

The data were collected from the validator, Economics teachers, and students in high school through expert judgment, documentation, and field testing. The expert judgment was conducted to validate the quality of item banking. Five experts on economics, informatics, and educational assessment validated the item bank system by giving a score of 1 to 5 on the validation sheet. The researchers themselves developed it by referring to the ISO 9126 quality model for a computer-based assessment, which consisting of functionality, usability, and reliability aspect. (Valenti, Cucchiarelli, & Panti, 2002). It has been adjusted to make it simple and more relevant to the context and users. Validated aspects

include interface, menu and navigation setting, content, and usability of the computerize item bank. Field testing is carried out by testing the test package to students and the item bank system to the teachers.

Items were adapted from three packages of national high school exam in economics (Puspendik, 2013, 2014, 2015). It is high-stakes testing because the Indonesian education assessment center developed it. Each package contains forty items that are tested to students through tests. So there are 120 items as the initial development of the item bank.

The scope of items includes high school economics contents for the senior high school under the high school curriculum. It deals with the economics introduction, microeconomics, macroeconomics, monetary system, banks, and financial institutions, management, cooperative, and accounting.

The sample of this study was 750 high school students and 30 economics teachers in West Sumatera province who were selected using purposive random sampling techniques. Students are selected from 12th grade because they have mastered all economic content. Student responses are used to calibrate the test by estimating the item parameters. Furthermore, the item's bank system is tested on teachers in practical simulation activities. It was carried out in order to assess the effectiveness of the item bank utilization. Each teacher gives a score of 1 to 4 for the interface, menu and navigation functioning, usability, and usefulness aspects of the item bank on the assessment sheet. It adopts the ISO9126 model for a computer-based assessment evaluation (Valenti et al., 2002).

Analyzing of Data

The quality and effectiveness of item bank system data were analyzed using descriptive statistics based on mean and percentage, while the item parameter data were analyzed using item response theory for two parameters logistic model. The quality of the item bank is determined based on the average rating of the expert judgment. The criteria for determining the quality of the item bank system based on the score in Table 2.

Table 2. The Criteria for Item Bank Quality

	Mean	Criteria
1.	≥ 4.0	Excellent
2.	$\geq 3.0 - < 4$	Very Good
3.	$\geq 2.0 - < 3$	Fair
4.	≤ 2	Poor

Item parameters are estimated with item response theory (IRT) using BILOG MG. It has been used widely because it is not dependent on sample characteristics to estimate the parameters (Carlson & von Davier, 2017). IRT has various models for estimating parameters. The two-parameter logistic (2-PL) is a model for dichotomous items that are able to provide maximum information for 750 samples on a short test (Sahin & Anil, 2017). The 2-PL model contains two-item parameters, i.e., item difficulty and discriminant that can be formulated as follows.

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad i=1,2,3,...n$$

$P_i(\theta)$: probability for a test testee with the ability θ to answer correctly item i

θ : testee ability

a_i : discriminant parameters item i

b_i : item difficulty parameters item i

e   : a transcendental number that is 2.718

D   : scaling factor so that the logistic function approaches normal ogive function (D=1.7)

The b_i parameter is a point on the ability so that the probability of answering correctly an item is 50% (Hambleton & Swaminathan, 2013). The higher b*i* reflects the greater ability needed to answer an item. On the other hand, the higher b*i* indicates more difficult an item. The a_i parameter is an item characteristic related to the item's ability to emphasize differences between participants who can answer correctly and answer incorrectly. Theoretically, the index difficulty and discriminant item lie on the scale $-\infty \leq b \leq \infty$ in item response theory. However, the items are chosen for the item bank should have index discriminant of 0 to 2, and index difficulty of -2 to 2(Baker & Kim, 2017). The goodness of fit items with the 2-PL model is determined based on the chi-square. Item has a good fit if the chi-square probability is greater than 0.01.

Teacher scores on the item bank system were analyzed using mean and percentage. The criteria for determining the effectiveness of the item bank system presented in Table 3.

Tabel 3. The Criteria for Item Bank Effectiveness

	Percentage	Criteria
1.	>85	Very effective/very high
2.	≥ 70 – 85	Quite effective/high
3.	≥ 50 – 70	Less effective/fair
4.	< 50	Poor effective/low

Findings/Results

Item Banking System Development

The first phase concerned with development planning. The main objective of the item bank is to prepare package tests for measuring students' achievement and simultaneously serving as a plan for improvement in the learning process. The item bank could be used by the teachers to implement classroom assessment. The target users are teachers of Economics subject for senior high school. It covered items intended to the competencies in the economics curriculum for the first, second, and third grades at senior high school, called Grades X, XI, and XII, according to the Indonesian national educational system. The item's format was a multiple-choice item. The correcting and scoring of this type of item could be done relatively easier rather than others.

In the second stage, the item was developed by using items from the national examinations of the preceding years. Those items had been prepared based on the standard rules of writing the good items and had been validated by experts. The development of an item bank requires a system to organize the items in such order and manner, so these items are easy to finding and retrieved by users. Rudner (1998) states that an item banking system could be developed based on criteria such as items grouping according to the content and instructional objectives. The specifications of the items in the item bank contained the grouping of items based on grade, competency, and indicator/instructional objective. The answer key of the national examination package chosen was determined so that, with the answer key as a basis, the analysis could be conducted to know the characteristics of the items. Each test item was made complete with related item information concerning its level of difficulty, the power of discrimination, and usability history. Item characteristics were known through item calibration.

The next step was to develop a computerized item bank interface. The item bank was developed based on the web using PHP-MySQL. Nowadays, many web-based assessment systems have been due to technological developments (T.-H. Wang, Wang, Wang, Huang, & Chen, 2004). It had various advantages because it could be accessed both online or offline. The online app allows teachers to add items and manage item banks collectively. The interface of the computerized item bank system is illustrated in Figure 1.
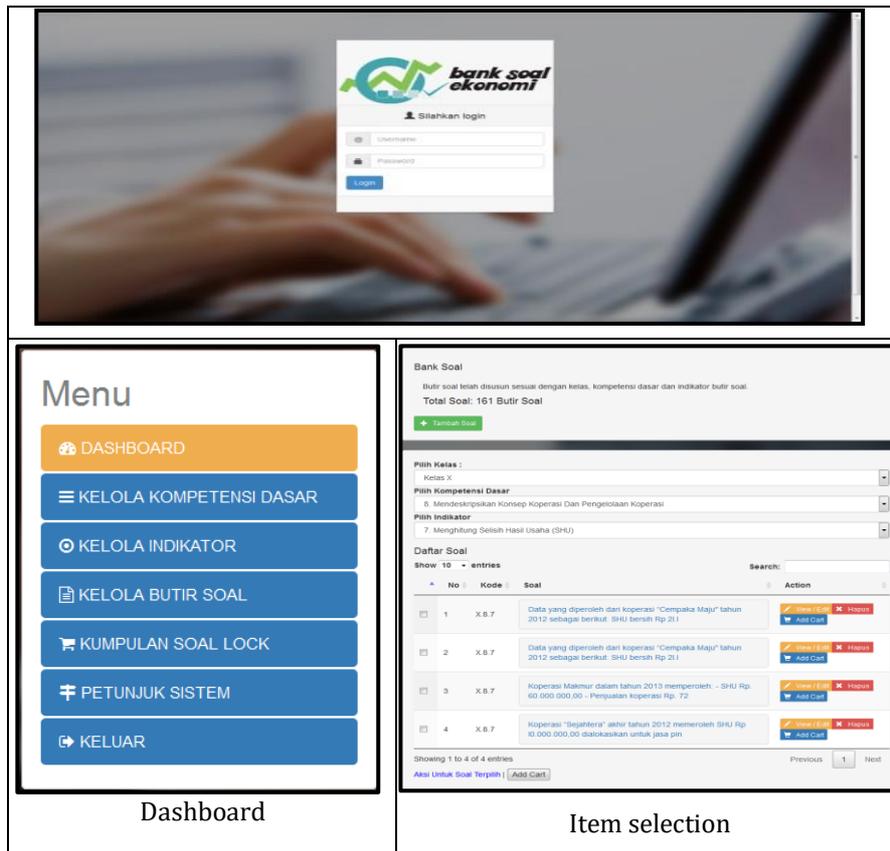
Fig. 2. Main Feature of The Computerized Economics Item Banking

The Quality of The Computerized Economics Item Banking

The quality of the item bank validated by expert judgment. Five experts in economics, informatics, and educational assessment assessed the item bank by providing scores of 1 to 5 for each component of the evaluation. The average score was calculated to determine the quality of the item bank. Table 3 presents the quality of the item bank assessed by the experts.

Tabel 3. The Quality of Computerized Economic of Item Banking

Aspects	I	II	III	IV	V	Mean	Category
Interface						4,20	Excellent
1 Font size accuracy	5	5	5	2	5	4,40	Excellent
2 Legibility of words and sentences	5	4	4	3	5	4,20	Excellent
3 Color composition	5	4	5	3	4	4,20	Excellent
4 Layout	5	4	5	3	4	4,20	Excellent
5 Appearance attractiveness	5	4	4	2	5	4,00	Excellent
Menu dan Navigation setting						4,12	Excellent
1 Menu availability for the operation	5	4	5	4	4	4,40	Excellent
2 Menu functioning	5	5	4	4	4	4,40	Excellent
3 Language accuracy on the menu	5	3	3	4	4	3,80	Good
4 The functioning of navigation for links between components	4	4	3	3	4	3,60	Good
5 Automatic item selection	5	5	4	3	5	4,40	Excellent
Content						4,20	Excellent
1 Number of questions	4	4	5	4	4	4,20	Excellent
2 The accuracy of item grouping	4	4	4	4	4	4,00	Excellent
3 Item information	4	5	2	4	4	3,80	Good
4 Items can be selected, added, edited and deleted	5	4	5	4	5	4,60	Excellent
5 Selected items can be printed	5	4	4	4	5	4,40	Excellent
Usability						4,60	Excellent
1 Ease in selecting items	5	4	5	5	4	4,60	Excellent
2 Ease of understanding menus and buttons	5	5	5	5	4	4,80	Excellent
3 Ease of operation of the item bank system	5	4	4	5	4	4,40	Excellent
Overall Mean						4,28	Excellent

In general, the result shows that the item bank has excellent quality, but some aspects need to be improved. It related to the setting language menu, navigation functioning, and item information. Enhancement to application setting is carried out by revising some terms on the menu and links between documents, especially the connection between competencies and indicators. Then, the program's initial is displayed first followed by the login page. The position of the system instruction is positioned at the top to make it easier for users to understand the usage of the item bank.

The Calibration of Item Parameters

Item parameters were analyzed using the two-parameter logistics item response theory model. It has two parameters, namely item discriminant and difficulty. The discriminant index is the slope of the grain characteristic curve at the difficulty point on a certain ability scale (Brennan, 2006). The higher slope of an item indicates that the greater discriminant index. Theoretically, the discriminant index grains lie on the scale $-\infty \leq a \leq \infty$, but in practice, it differs on a scale of 0 to 2 for a good item (Hambleton, Swaminathan, & Rogers, 1991).

The index difficulty is a function of one's ability (Mardapi, 2017). Someone who has a high ability will have a high probability of answering an item correctly, conversely, those who have low ability will have a low chance of answering an item correctly. The higher difficulty index reflects the more difficult the item (Brennan, 2006). Good items have a difficulty index between -2 to 2. Items that have difficulty index close to or below the -2 indicate the easy items, whereas items that have difficulty index near or located above the +2.00 reveal the difficult item (Hambleton & Swaminathan, 2013). The distribution of item parameters is shown in Table 4.

Table 4. Distribution of Item Parameters

Test Package	Quantity	Items difficulty			Items discriminant	
		Ease	Moderate	Difficult	Good	High
Package A	Frequency	9	26	5	27	13
	Percentage	22.5%	65%	12.5%	68%	33%
Package B	Frequency	7	28	5	29	11
	Percentage	17.5%	70%	12.5%	73%	28%
Package C	Frequency	12	18	10	25	15
	Percentage	30%	45%	25%	63%	38%
	Total items	28	72	20	81	39
	Percentage	23%	60%	17%	67.5%	32.5%

Item calibration results show that most items have a moderate difficulty and a good discriminant index. There are not many items that have high or low index difficulty. This distribution presents the ideal number for achievement tests because it is close to the nature of the normal distribution where most items have a moderate item parameter index, the remaining few items are categorized as easy and difficult. The average score of the parameter items for each test package is shown in Table 5.

Table 5. Mean of Item Parameters

Test package	Mean	
	Item Difficulty	Item Discriminant
Package A	0.32	1.07
Package B	0.47	0.82
Package C	0.93	1.26

Test calibration results show that all test packages have a good mean index difficulty and discriminant. Difficulty index is in the range of values -2 to 2, and the discriminant index is in the range of values 0 to 2. Based on the results, it can be concluded that all three test package has a good quality because it has good parameters attribute an ideal proportion of items. This item parameters information is inputted to each item in the item bank system.

The Effectiveness of the Computerized Economics Item Banking

Field testing was conducted to measure the effectiveness of the item bank. Thirty economics teachers were asked to operate the item bank and use it to plan a classroom test. Then the teachers were being sought to evaluate the utilization of the item bank by giving a score of 1 to 4. The percentage of the score provided by the teacher is computed to determine the level of effectiveness of the item bank. The results are presented in Table 6

Table 6. The Effectiveness of The Computerized Economics Item Bank

Component		Percentage	Effectiveness
A. Usability		90	Very effective
	1 Easy to select items	92	Very effective
	2 Easy to operate	93	Very effective
	3 Easy to understand	92	Very effective
	4 Flexible	85	Very effective
B. Usefulness		94	Very effective
	1 Useful for teachers	98	Very effective
	2 Facilitate teacher assignments	95	Very effective
	3 Improve teacher performance	90	Very effective
C. Feasibility for classroom/school-based assessment		95	Very effective
	Overall	93	Very effective

Overall, the computerized economics item bank is very effective to use, beneficial for teachers, and worthy for classroom and school-based assessment. The results show that the teachers had no difficulty in operating the app because it is presented in Indonesian, so all teachers interested in using it. The recommendation from the teacher lies in the app usage instructions to be more detailed so that teachers who are less familiar with computers can use the application more efficiently.

Discussion and Conclusion

The advancement of information and technology has led to considerable challenges to improve the learning quality and educational assessment. The traditional assessing method began to be abandoned by the users and switched to the utilization of electronic or digital-based assessment. This study developed the computerized economics item bank for classroom and school-based assessment, as well as a tangible form of school readiness to implement a decentralized education that is mandated by the Indonesian government.

One of the factors that affect the success of the implementation of the computerized test is the level of familiarity of users (Al-Amri, 2007). The level of intimacy of the users caused by their experience in using the computer. Then, the design of a computerized item bank should be user-friendly so the user can operate it efficiently (Rice, 2003). The test results show that the teachers have no difficulty in operating the item bank; this indicates that it has a user-friendly attribute. The results also indicate that most teachers are already technology literate because they have no difficulty in operating the item bank. Other research shows that 54% of Indonesian teachers have good computer literacy, 33% have sufficient computer literacy, while the rest only 11% of teachers are less literate. The results show that the Indonesian teachers have sufficient readiness to implement the computerized assessment. It means that implementing a computer-based classroom assessment at the local level in Indonesia is not an impossible thing to realize.

Most teachers require the computerized economics item bank to implement for classroom assessment at school. It suggests that the computerized economics item bank is very beneficial in conducting classroom assessments (T.-H. Wang et al., 2004). The results of this study are in line with (Chien, Wu, & Hsu, 2014) that revealed eighty-five percent of teachers agreed to the use of the technology-based assessment as a useful tool in the classroom. Provide not only enormous benefits but also the computerized item banks will ensure the efficiency of the test and reduce the error (Weiss, 2013). The implementation of computerized assessments has great potential to improve teaching efficiency because it can save time, energy and money. Moreover, it can also increase the validity of the assessment because it contains various measurement tools that are appropriate for the student's skills (Winkley, 2010). Improving the validity can reduce the errors of measurements.

Technological advancements provide significant benefits in assessments to improve learning effectiveness and teaching outcomes. Nowadays, so many learning tools and digital tests have been developed in various forms, such as e-assessment (Morales-Martinez, Lopez-Ramirez, Castro-Campos, Villarreal-Trevino, & Gonzales-Trujillo, 2017), games (Zvarych, Kalaur, Prymachenko, RomashchenkO, & Romanyshyna, 2019), and similar digital assessment tests resulted in the change of schools assessment. Assessment of learning outcomes is designed to be more exciting and enjoyable. Digital tests allow students to evaluate their abilities independently and get feedback directly so that digital tests can be used as a strategy to help students learn effectively (Chen, Ho, & Yen, 2010).

That electronic assessment platforms can develop optimally if supported by the existence of a computerized item bank. The item bank that has been developed in this research has great benefits in developing various digital test in the future. Further development of the item bank urgently needs collaboration between teachers from multiple schools to create more items. The importance of teacher involvement is also shown in the successful development of e-asTTle in New Zealand (Brown, 2014). Most teacher plays an important role in developing the material and the software. Teacher contribution is also needed in this study to participate in enriching the number of items. Intermediate networking is required to increase the number of items continuously so that it can produce a larger item bank. The more items in the item bank will guarantee the security of the items so that the quality of the test program could be improved (Umar,

1999). It could happen due to the quality and quantity of items in the item bank have significant impacts on the accuracy of the test administering (Xing & Hambleton, 2004). Based on the results, the following conclusions can be drawn. The decentralization policy of educational assessment requires the readiness of schools in every region in Indonesia. The development of item banks at the regional level is a supporting tool that can encourage the realization of an effective and efficient decentralization assessment. This study succeeded in establishing a computerized item bank on economics. Furthermore, it is expected to cover all subjects for high school students. The developing of computerized economics item bank has excellent quality based on expert judgment and very high effectiveness based on the teacher's review. Five experts have assessed the quality of the computerized economics item bank based on the interface, menu and navigation setting, content, and usability. The results showed that each of these aspects has excellent quality. Thirty economic teachers evaluated the effectiveness of the computerized economics item bank. They were asked to demonstrate the computerized economics item bank and assessed its effectiveness. Field testing results indicate that the computerized economics item bank has very high effectiveness for usability, usefulness for the teachers, and worthy for classroom and school-based assessment. The item bank contains 120 multiple-choice items calibrated using the 2-PL IRT.Therefore, the computerized economics item bank has a great opportunity to develop various digital tests, so that the cooperation of teachers to maintain the computerized economics item bank is needed.

Suggestions

Item bank development for computerized tests requires many items. Therefore further development of the items bank system and item formats are needed. Format items in this item bank are still limited to multiple-choice, then further research is desired to develop constructed response items. It is relevant to the classroom assessments and makes it possible to assess student skills that cannot be measured by multiple-choice tests. In addition, the item bank contains items for economics only. Further development is directed towards developing items for other subjects to cover all content for high school students. Teacher collaboration is desired to develop more items so that the items bank that has been developed becomes more adequate.

The item bank testing does not yet involve teachers in remote areas. Therefore, it is needed to determine the readiness of teachers and schools for the implementation of computer-based classroom assessment. Testing in rural may show a different profile so that it requires a different approach to the implementation.

Computerized question banks can be used for a variety of computerized test purposes so that these item banks could be the basis in developing various digital assessments for students. Moreover, it would be integrated with computerized tests by involving item randomization systems.

Acknowledgments

References

Abubakar, A. S., & Adebayo, F. O. (2014). Using computer based test method for the conduct of examination in Nigeria: Prospects, challenges and strategies. *Mediterranean Journal of Social Sciences, 5*(2), 47-55. https://doi.org/10.5901/mjss.2014.v5n2p47

Al-Amri, S. (2007). Computer-based vs. Paper-based testing: Does the test administration mode matter? In *Proceedings of the British Association of Applied Linguistics Conference* (pp. 101-110). Edinburgh, Scotland.

Archer, E., & Brown, G. T. (2013). Beyond rhetoric: Leveraging learning from New Zealand's assessment tools for teaching and learning for South Africa. *Education as Change, 17*(1), 131-147. https://doi.org/10.1080/16823206.2013.773932

Ata, N. A., & Ani, W. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota yogyakarta tahun 2012 [Item Analysis of Quality Control Tests for Class XII High School on Economics in Yogyakarta in 2012]. *Indonesian Accounting Education Journal/Jurnal Pendidikan Akuntansi Indonesia, X*(1), 1-26.

Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. New York, NY: Springer.

Bodmann, S. M., & Robinson, D. H. (2004). Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research, 31*(1), 51-60.

Brennan, R. (2006). *Educational measurement*. Westport, CT: Greenwood Publishing Group.

Brown, G. T. (2014). AsTTle–A national testing system for formative assessment: How the national testing policy ended up helping schools and teachers. In M. K. Lai & S. Kushner (Eds.), *A developmental and negotiated approach to school self-evaluation* (pp. 39-56). London, England: Emerald Group Publishing.

Carlson, J. E., & von Davier, M. (2017). Item response theory. In R. E. Bennett & M. v. Davier (Eds.), *Advancing Human Assessment: The Methodological, Psychological and Policy Contributions of ETS* 2017 (pp. 1-69). Princeton, NJ: Springer International Publishing.

Carnegie-Harding, S. (2016). *E-asTTle as a catalyst for change* (Unpublished masters's thesis). Massey University, Palmerston North, New Zealand.

Chen, L.-J., Ho, R.-G., & Yen, Y.-C. (2010). Marking strategies in metacognition-evaluated computer-based testing. *Journal of Educational Technology & Society, 13*(1), 246-259.

Chien, S.-P., Wu, H.-K., & Hsu, Y.-S. (2014). An investigation of teachers' beliefs and their use of technology-based assessments. *Computers in Human Behavior, 31*(2), 198-210.

Cizek, G. J. (2001). An overview of issues concerning cheating on large-scale tests. In *Proceedings of Annual Meeting of the National Conference on Measurement in Education* (pp. 4-30). Seattle, Washington.

Friyatmi, F. (2011). Faktor-faktor penentu perilaku mencontek di kalangan mahasiswa fakultas ekonomi UNP [Determinant factors of cheating behavior among UNP Students in economic faculty]. *TINGKAP, 7*(2), 173-188.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications.* Boston, MA: Springer Science & Business Media.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publication.

Hattie, J. A., & Brown, G. T. (2007). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology Systems, 36*(2), 189-201. https://doi.org/10.2190%2FET.36.2.g

Hattie, J. A., Brown, G. T. L., Ward, L., Irving, S. E., & Keegan, P. J. (2006). Formative evaluation of an educational assessment technology innovation: Developers' insights into Assessment Tools for Teaching and Learning (asTTle). *Journal of Multi-Disciplinary Evaluation, 5*(3), 1-54.

Hayati, N., & Mardapi, D. (2014). Pengembangan butir soal matematika SD di Kabupaten Lombok Timur sebagai upaya dalam pengadaan bank soal[Development of elementary school mathematics items in East Lombok Regency as an effort in the procurement of item banks]. *Journal of Education/Jurnal Kependidikan: Penelitian Inovasi Pembelajaran, 44*(1), 26-38.

He, Q., & Tymms, P. (2005). A computer-assisted test design and diagnosis system for use by classroom teachers. *Journal of Computer Assisted Learning, 21*(6), 419-429. https://doi.org/ 10.1111/j.1365-2729.2005.00148.x

Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023-1031. https://doi.org/10.1016/j.compedu.2010.12.002

Kemdikbud. (2019). *Ujian nasional siap digelar: 8,3 juta peserta dengan 91 persen berbasis komputer* [National examinations ready to be held: 8.3 million participants with 91 percent computer-based]. Jakarta: Kementerian Pendidikan dan Kebudayaan.

Mardapi, D. (2017). *Pengukuran, penilaian dan evaluasi pendidikan* [Measurement, assessment and evaluation of education]. Yogyakarta, Indonesia: Parama Publishing.

Millman, J., & Arter, J. A. (1984). Issues in item banking. *Journal of Educational Measurement, 21*(4), 315-330. https://doi.org/10.1111/j.1745-3984.1984.tb01037.x

Morales-Martinez, G. E., Lopez-Ramirez, E. O., Castro-Campos, C., Villarreal-Trevino, M. G., & Gonzales-Trujillo, C. J. (2017). Cognitive analysis of meaning and acquired mental representations as an alternative measurement method technique to innovate e-assessment. *European Journal of Educational Research, 6*(4), 455-464.

Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education, 31*(1), 53-64. https://doi.org/10.1080/03098770601167922

Novitasari, I. (2012). Sindikasi Jual-beli kunci jawaban ujian nasional 2010 (studi kasus jaringan jual-beli kunci jawaban ujian nasional 2010 di kota X) [Syndication of the answer keys trading for national examination in the 2010 (Case Study of the sindicate in city X)]. *Indonesian Criminology Journal/Jurnal Kriminologi Indonesia, 7*(2), 267-286.

Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal, 93*, 836-847. https://doi.org/10.1111/j.1540-4781.2009.00976.x

Puspendik. (2013). *Ujian Nasional* [National examination]. Jakarta, Indonesia: Puspendik Balitbang Kemdikbud.

Puspendik. (2014). *Ujian Nasional* [National examination]. Jakarta, Indonesia: Puspendik Balitbang Kemdikbud.

Puspendik. (2015). *Ujian Nasional* [National examination]. Jakarta, Indonesia: Puspendik Balitbang Kemdikbud.

Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science, 323*(5910), 75-79. https://doi.org/10.1126/science.1168046

Retnawati, H., & Hadi, S. (2014). Sistem bank soal daerah terkalibrasi untuk menyongsong era desentralisasi [Calibrated regional item bank system to welcome the era of decentralization]. *Educational Science Journal/Jurnal Ilmu Pendidikan, 20*(2), 183-193.

Rice, G. (2003). Designing on line tests: The role of assessment in web-based learning. *Journal of Adventist Education*, (April/May), 31-35.

Rubie-Davies, C. M., & Rosenthal, R. (2016). Intervening in teachers' expectations: A random effects meta-analytic approach to examining the effectiveness of an intervention. *Learning and Individual differences, 50*(6), 83-92.

Rudner, L. (1998). Item banking. *Practical Assessment, Research & Evaluation, 6*(4), 1-3.

Sahin, A., & Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory and Practice, 17*(1n), 321-335. https://doi.org/10.12738/estp.2017.1.0270

Scibinetti, P., Tocci, N., & Pesce, C. (2011). Motor creativity and creative thinking in children: The diverging role of inhibition. *Creativity Research Journal, 23*(3), 262-272. https://doi.org/10.1080/10400419.2011.595993

Setiawan, S. D. (2017). UNBK dinilai kurangi praktik kecurangan saat ujian nasional [UNBK was assessed reducing the practice of cheating during the national examination, *Republika*. Retrieved from https://republika.co.id/berita/pendidikan/eduaction/17/12/24/p1g4wn438-unbk-dinilai-kurangi-praktik-kecurangan-saat-ujian-nasional

Suyata, P., Mardapi, D., Kartowagiran, B., & Retnawati, H. (2011). Model pengembangan bank soal berbasis guru dan mutu pendidikan [Teacher-based item bank development model and education quality]. *Journal of Education/Jurnal Kependidikan: Penelitian Inovasi Pembelajaran, 41*(2), 120-128.

Umar, J. (1999). Item Banking. In G. N. Master & J. P. Keeves (Eds.), *Advanced in Measurement in Educational Research and Assessment* (pp. 207-218). Amsterdam, Netherlands: Pergamon.

Utantoro, A. (2019). UN Berbasis Komputer Tekan Kecurangan [Computer-based national exam press cheating]. *Media Indonesia*. Retrieved from https://mediaindonesia.com/read/detail/225397-un-berbasis-komputer-tekan-kecurangan

Valenti, S., Cucchiarelli, A., & Panti, M. (2002). Computer based assessment systems evaluation via the ISO9126 quality model. *Journal of Information Technology Education: Research, 1*(1), 157-175.

Wang, H., & Shin, C. D. (2009). Computer-based & paper-pencil test comparability studies. *Test, Measurement & Research Services Bulletin, 9*, 1-6.

Wang, T.-H., Wang, K.-H., Wang, W.-L., Huang, S.-C., & Chen, S. Y. (2004). Web-based assessment and test analyses (WATA) system: development and evaluation. *Journal of Computer Assisted Learning, 20*(1), 59-71. https://doi.org/doi: 10.1111/j.1365-2729.2004.00066.x

Weiss, D. J. (2013). Item banking, test development, and test delivery. In K. F. Geisinger (Ed.), *The APA Handbook on Testing and Assessment*. Washington, WA: American Psychological Association.

Winkley, J. (2010). *E-assessment and innovation emerging technologies*. Coventry, England: Becta.

Wood, R., & Skurnik, L. S. (1969). *Item banking*. London, England: King, Thorne & Stace Ltd.

Xing, D., & Hambleton, R. K. (2004). Impact of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. *Educational and Psychological Measurement, 64*(1), 5-21. https://doi.org/10.1177%2F0013164403258393

Zvarych, I., Kalaur, S. M., Prymachenko, N. M., RomashchenkO, I. V., & Romanyshyna, O. I. (2019). Gamification as a tool for stimulating the educational activity of students of higher educational institutions of ukraine and the united states. *European Journal of Educational Research, 8*(3), 875-891.