Multimedia-Based online Test on Indonesian Language Receptive Skills Development

M Sudaryanto1,2*, D Mardapi2, S Hadi2

1Doctoral Student of Post Graduate Program in Universitas Negeri Yogyakarta, Indonesia
2Evaluation and Research Education of Universitas Negeri Yogyakarta, Indonesia
2Faculty of Teacher Training and Education, of Sebelas Maret University, Indonesia

*memet.sudaryanto2016@student.uny.ac.id, *memetsudaryanto@staff.uns.ac.id

Abstract. The need for computer-based assessment provides an opportunity for test takers to maximize performance in solving questions prepared to their measure language skills. Specifically, this study develops receptive skill tests, i.e. listening, reading, and skills competency responding to the rules. Instruments were tested on the test participants who are elementary level students who join Indonesian language courses. The items were analyzed by Rasch model to determine the level of item difficulty. The results showed that the level of difficulty of the items in the normal distribution in accordance with the ability of test participants. Qualitative analysis shows (1) the weaknesses of using Google-form as a receptive skill test tool, such as connection and unstable internet signal even though tested at college level, test takers still record simultaneous material, the search tool used is not qualified hampering the testing process, (2) the advantages of using Google-form as a receptive skill test tool are there is no need to wait for late test takers, test takers have no chance to cheat and it is paperless, (3) the importance of determining starting point and test rules other. The conclusion of this research is the effectiveness of using Google form as a multimedia-based receptive skill tool capable of minimizing the testing procedure.

1. Introduction

Gaps between the needs of technological developments and supporting conditions in the field of assessment are still not running balanced. The government with all its regulations demands equality between communities in the region and in the city. It is important to analyze the gaps of affordable and easy-to-use modern technologies, especially in the field of assessment, in particular the competency test in the basics. Basic materials such as mathematics and language logic require an overall comprehensive assessment model to map competencies widely.

Assessment is an important aspect to see the mastery of test participants in understanding the materials. A good assessment enables test participants to complete the assessment process with satisfactory results [1]. The ease can be seen from the aspect of the suitability of the level of difficulty with the ability, the openness of the correction process and the answer, as well as the ease of access of the test instrument [2]. On the other hand, the current rating conditions do not match the expectations of test participants (students) and testers (teachers) thus the ratio or difficulty level of the test items seen
from the correct answer proportion is still low. In addition to the level of difficulty that is not reached the competency margin of students, as well as existing facilities in educational facilities, especially in Indonesia, it has not been consummated. Whereas learning facilities such as the use of technology is an important factor that dominant implementation of good education [3].

The test competencies used for the appraisal reference provide analysis results in the form of a capability response which in one logistic parameter is the degree of difficulty. The Rasch model is a direct comparison between the individual and the item. The comparison looks at the participant's ability and item as a parameter of difficulty level [4]. The superiority of the Rasch model compared to other methods, in particular classical test theory, the ability to predict missing data, which is based on a systematic response pattern [5]. The ability of learners and the level of difficulty in the items on the same scale can be compared. The difficulty level of the item has a range that is tailored to the ability of the test participants. It is expected that the level of difficulty of the item becomes a reference for the success of test participants following the learning process.

In particular, item index difficulty is one of the important parameters to determine the feasibility of the test before it is used. On the other hand, these parameters are capable of measuring the competencies of test participants. By looking at the distribution of difficulty level of the item, it will be seen the opportunity to correctly answer the instrument through the characteristic curve test resulted from the measurement of rasch model. The advantages of this measurement model are described (1) more accurately, as the ability to predict lost data is based on a systematic response pattern, (2) the rasch model is able to produce standard error measurement values for the instrument, so as to improve the measurement accuracy, (3) assumptions the rankings generated when the items are prepared, the rasch model can verify whether or not to produce the expected pattern. This calibration is only owned by rasch. A good instrument is certainly capable of working within the upper and lower skill limits of the test participants. On the other hands, there are still language skills tests that test the language knowledge of test participants, rather than testing the use of the language. It is important to collaborate on these two aspects to measure a person's Indonesian language skills to a level equivalent to the difficulty of the item. The right and proper Indonesian language and grammar so far is still needed. Submission of materials using Indonesian language is right and proper to the benchmark understanding of context [6].

Studies in language instruments also need to consider good and proper language rules, which need to be adapted to the needs of communication in good language rules referring to journalistic, scientific, conversational, fictional, written, and spoken language. Language logic is often not realized by the compilers of test instruments. The language used seems to be sufficient. Though judging from the development of language creativity is less meaningful [7]. Especially to measure the receptive skills required measurement media in the form of good video, the media is able to describe the context and facial expressions, gestures, and others. So that the material and media used have ideal hardness, and the thinking of learners is well built [8].

Language as a tool of communication is an important test material for seeing one's cognitive, social, and affective development. Therefore, the constructed test is the listening, reading, and grammar skills used to determine the extent of the linguistic response of the test participants in capturing the linguistic material it receives [9]. On the other hand, the three skills measure the receptive competence of a person by looking at the response given to the item from the low to high difficulty level. Simultaneously the receptive skill tests are also capable of measuring their productive skills. One's writing skills can be seen based on the responses of reading skills, while listening skills reflect the ability to speak. Multimedia-based listening tests carry video in everyday life that demands language foresight and the accuracy of word selection. The reading test tests the ability to generalize paragraphs and analyze the content presented in order to exercise language sensitivity. Both receptive skills are tested by measuring the understanding and acceptance of language stimuli in the form of sounds and writing. The study's study developed a multimedia-based test instrument. This development is still needed to see the flaws and advantages and possibilities of further development. This research develops multimedia-based instruments and is done online.

According to [10] multimedia is combination of communication whereas the form of communication becomes coordinated and clear. [11] Defines multimedia as the mixture of elements such as audio, video, graphics, text, animation, etc as a synergy. Multimedia in the computer context according to [12] is the
use of technology (computer) as the tool of navigation, interaction and communication by combining text, sound, image, animation and video. Based on that understanding, multimedia consists of three factors, namely: (i) there is a tool (computer) in providing audio and video; (ii) there is a tool that help users in searching information and (iii) multimedia as a tool of collecting, processing and communication information.

Assessment of multimedia-based assessment that combines the validation results of an instrument and developed with a diverse media base (multimedia). The use of media is a principle of test requirements that require the achievement of learning objectives, the benefits of assessment, assessment methods, sample coverage, and nirbias. The implemented multimedia on language skills is specified to (1) the video as a listening skill test tool; that is, as a receptive skill requires test takers to more than just capture the meaning of sound, but more broadly about the context and view presented in a video, and (2) drawing as a tool to facilitate the insert clue given to the problem maker as a stimulus and language sensitivity. Explain that each student is unique that’s why its need unique interpretations of what they are studying. In the other side, a teacher should manage the resources needed and help learners to do effectively [13].

Variety of video and images as multimedia tools applied for the achievement of the learning process requires good quality and not bias. The video contains authentic test material, which is based on everyday life, such as cooking activities, television presenter, home dialogue, dialogue at school, and other activities undertaken by native speakers. In addition, the test participants should describe the image media used to facilitate access to the understanding of the test participants, which is containing the hidden clue. The development of multimedia-based tests means combining multiple media that make it easier for test participants to get more information than just everyday skills. Describes (a) assess what is intended to be measured (construct validity); (b) assessed accurately what students know (reliability); (c) be able to provide feedback to students on their learning (support); and (d) easy to construct, administer, and mark the assessment items in a timely manner [14]. Accountability decision-making in a data-driven education. Teacher data collection is focused on authentic assessment.

2. Method
This research is a development research on multimedia-based valuation model and tested online. Instrument developed with google-form application based on multimedia. The developed instruments contain images and videos that support the assessment to maximizing instrument uses. state that research and development in education is a process in developing and proofing the products of education. The processes of reserach and development are tests, evaluation, and refiner to meet the standard. [15].

It is also important to pay attention on the internet security to keep the information remain private [16]. The product of this research is learning materials, which are in the form of test of receptive skills. The second development is done on the development of receptive instruments, such as listening, reading, and responding to the rules. Instrument development tested content (with exploratory factor analysis) and language validity (inter-rater validity). A multimedia-based online rating model is justified with expert judgment with an integrated visual and audiovisual multimedia model.

The data is the response of the test participants, i.e. 450 new students of Sebelas Maret University representing the science and social department, selected by random sampling. Uni-dimensionality test, KMO, and EFA using SPSS 21. Model analysis with rasch model (One Parameter logistic) using BILOG MG 3.0 distributed with Microsoft Excel. There is urgent proof about “student view the assessment as fair, relevant, and useful for improving learning” [17] or what we called as face validity. It means the test seems right in measuring the ability based on subjective judgment, administrative personnel of the examinees [18].

3. Result and Discussion
Test tools in this study using Google-form based multimedia, which consists of 4 parts. The first part is the stuffing of identity and general information. The identity of the test participants is in the form of identity of the name, date of birth, origin of the area, as well as instructions of the test taker. In the second field is a listening session of a number of 10 items, the third field is a reading session of 15 items, and the third field is a session to respond to the rules of 15 items. The concept of multimedia in this
study combines audiovisual media i.e. video recording (1) how to make meat soft, (2) introductory television presenter video when opening the event, and (3) tourism exploration. Third audio quality and image quality on the video is selected at the smallest resolution.

The facilities of online test are tool which insert video and pictures, to insert images there are upload, snap, URL, album, google-drive, and search options. Online test also provide auto-correction procedure on the use of google form. Procedure consists of 5 steps to provide auto-correction. In form creating, it is important to consider the order of questions of the form, the collected data of the respondent as well as the previous data and to make sure that the form in created as it was desired [19]. Step to developing an online test are (a) set google form to "QUIZ" and mark the value and columns of respondents, (b) determine the answer key. The correct answer will have a check mark at the end of the column, (c) completing the test, examinees will see the score directly, (d) wrong and correct answers will be corrected by the system, (e) score presented.

Procedures for developing instrument begin with (1) preparing the material based on the principle of authentic assessment work, i.e. using test materials derived from daily life, (2) uploading videos on a prepared YouTube account, (3) creating linking between Google-form with videos uploaded on YouTube, (4) evaluation of video recording length and reducing unnecessary video part, empty video or introductory tones (5) play trial. Videos are sorted by content and situation difficulty. The video criteria used are (1) derived from daily life, (2) the smallest video resolution of 270x480 making it easier for test takers with internet limit quotas, (3) monologues to focus high concentration, (4) the video's duration ± 60 second to allow test takers to make good use of time. The image media used are (1) the picture/graph of the stem of the dollar exchange rate against the rupiah in one August, (2) the Prameks preliminary train ticket image/table on the first schedule until the last schedule with Solo departure to Yogyakarta, (3) and shoots of plants that are blank images without clue, while the last image (4) plan of the noble university.

Characteristics of images used in reading skill tests are also made based on gradations of difficulty. The first picture, namely the movement of the rupiah against the US Dollar week-1 August has provided complete information such as exchange rates made price range between 9,400-10,200. Besides also the information date of transition, which started from left is August 1 and closed by August 7. Another thing that proves that the image is the picture with the lowest difficulty level is, every intersection between the date and the selling price, always made discounts and price information. The second picture also reads the numbers on the train schedule making it easier for test participants to understand the contents of the data. Third picture demands higher ordered thinking skill, that is finding the problem and solving it in everyday life. While the highest difficulty level is to read the floor plan with the characteristics of the road consisting of two segments and the area that shows the complexity of information in the appendix.

Test maker use general instrument to facilitate students in measuring Indonesian competency. They adapt and adopt the subjects that will be taught from the general textbooks without compile them. In the other hands, the selected test subject also does not suit the environment in which they were used. Test maker use subject which are irrelevant to the context of local environments, in this case are the cultures of the examinees. The Test maker often offers the texts, which have the high level of difficulty in teaching Indonesian language. They also give difficult reading tasks, which should be completed immediately in the classroom whereas it can be delivered to the learners through extensive reading tasks. This burden tasks make the examinees reluctant to complete them because they find it is difficult to complete. Consequently, they cannot develop their language skills.

The development of listening tests is structured on the basis of constructing critical-analytical listening tests. Listening consists of 10 items which are questions based on the video being played. Videos 1 and 2 are used to answer 3 questions, while video 3 answers 4 items, when test participants answer the questions listed, then the test participants are prohibited to return to play video explaining the material. The development of the reading test consists of 15 items consisting of (1) a test tool in the form of an explanatory text for identifying information, (2) reading the floor plan, to see the general information sensitivity and finding supporting information, (3) the table containing the data information, test takers are able to see detailed information well or generalize. The test is developed based on the construct of critical reading. Operational verbs on critical reading skills are applied to the analytical
content and enter the realm of identification, so that later the questions are cornered on the skill of concluding, finding the problem, and developing the concept of the main idea or topic sentence in the paragraph. Using communicative activities while focusing on student-centered learning and working in pairs and groups, through interaction and communication, language learning can occur [20].

The test responds to a rule or grammar developed with the concept of (1) a new understanding of the concept of Indonesian, (2) an analysis of Indonesian errors, (3) correcting language errors, and (4) determining lively and correct language decisions. This test was developed in accordance with the standard Indonesian grammar construct confirmed in Indonesian dictionary to test the language skills, sensitivity, and attitude (like and dislike) The item consists of 15 items. These items are arranged with long sentence tests, there are also true and false tests on raw and non-standard words. Generally, a receptive skill test is a sensitivity test to receive a stimulus of listening, reading, or error analysis prepared by the test taker to develop the knowledge of receiving the response of the test participants.

The principle of the rasch analysis model is the opportunity for the student to answer a single item correctly with the student's ability compared to the difficulty level of the item. Each student has a 50% chance to answer correctly, if the student's ability is equal to the level of difficulty. The opportunity to solve one problem correctly depends on the ability ratio and the level of difficulty. The main goal of the rasch model is to scale measurements at equal intervals and it is also scale the score in one dimension. Raw scores do not have intervals, so these scores can not be used directly to provide interpretation of test takers' abilities [5]. The rasch model uses the score data for each test participant and per item to estimate the pure score (indicates the level of individual ability and difficulty of the item).

There was a time when it was taken for granted, at least in come circles, that it was more difficult to construct reliable and valid test of writing and speaking than of reading and listening. In large part this was because the former seemed to depend on notoriously unreliable subjective scoring while in the case of latter, objective scoring was both possible and appropriate [21]. Based on the triangulation used to explore the effectiveness of test result data. Qualitative analysis shows: (1) The disadvantages of using Google-form as a receptive skill test tool, such as unstable Internet connections and signals even when tested at the university level. Testing in a University is expected to minimize the error in setting test execution. At the time of the test, test takers use mobile to work on Google-form provided. The solution to this problem is the use of Wi-Fi provided by the university, or test maker. (2) Students still record the listening material. In the process of listening, it takes concentration and is not expected to perform other activities. Some test takers use the method of noting to do the test. (3) The search tool (browser) used is not qualified so that it inhibits the testing process; Google form does not work optimally on some browser tools used by the students, resulting in delay work that hampers the individual student process. (4) The advantages of using Google-form as a receptive skill test tool, such as, do not need to wait for students who arrive late and one of the comparable advantages is (a) conventional listening requires loud and audible speakers in the classroom, (b) using paper and pencils test, requires 750 sheets of paper for the questionnaire, and 450 sheets of paper for answer sheets. Both problems are resolved well through tests with online-based multimedia. (5) Students do not have a slit to cheat due to limited time and mobile phone' small screen, students do not directly see the answers of friends A, B, C, or D so they have to finish their own work. (6) The importance of determining starting point and other test rules. “Assessment plays an important role in education because a qualified assessment system would improve the education quality” [4]. In form creating, it is important to consider the order of questions of the form, the collected data of the correspondent as well as the previous data and to make sure that the form in created as it was desired [19].

4. Conclusion
Following the development of the era, the assessment of current language skills requires test participants to master the linguistic aspects holistically without understanding the context used for everyday life. The required assessment model is an effective, efficient, and accurate measure of the dimensions tested. The effectiveness of the assessment can be seen from the autocorrelation of test results and the minimization of the use of test properties (generally paper and pencil). Efficiency in the context of this assessment is the optimization of the use of the test time used. Generally a listening skill test requires the timeliness of the test taker, so the tests can be started simultaneously.
Accuracy to see the error measurement perspective is done to minimize the standard error including the test measure the right dimensions and not biased. Based on the study of the mix method of development of Indonesian multimedia based receptive skill test, it is concluded that online test equipment used in general has significant advantages compared to paper and pencil test. Characteristics of images used in reading skill tests are also made based on gradations of difficulty.

On the other hand, the test tool is capable of directly generating scores on the measurement results of each test participant, so there is a separate satisfaction felt by test participants who know the results and compared with peers. Listening skill tests have a high effectiveness, as test takers are able to easily take the test without considering the delays of others and the readiness of other test takers. The tools needed are not too much and tend to make things easier. Unfortunately, there are two perspectives on reading skill, (1) on the reading skill of graph, diagram, and drawing of the participants with zoom-in and zoom-out capability so that small picture can be enlarged, not with paper and pencils test, (2) on the skill of reading a fairly long text, students tend to get tired easily because the text exceeds one screen mobile phone.

Test tools in this study using Google-form based multimedia, which consists of 4 parts. The results of the measurements can be summarized, (1) the multimedia-based receptive skill assay instruments measure 1 dimension, (2) the level of difficulty of items still has weaknesses in some items that are too difficult and too easy, i.e. items that exceed the difficulty level of +4 and less than -4. Items that tend to exceed +4 are particularly difficult items, namely points 20, 24, and 40. In general, items are too difficult to be in the 2-item reading skill and the skill responds to the rule of 1 item. While the items below the -4 ability, namely point 1, 6, 7, 8, 9, 10, 14, 15, 32, and item 34 are listening skills on 6 items, reading skill on 2 items, and 2 items questions to respond to rules. Roughly, it can be concluded that listening with the help of audiovisual will facilitate the test participants in training their receptive skills, especially listening skills. It would be better to grab the problem with difficulty too easily and the difficulty is too difficult to eliminate from the instrument. The questionnaire can be developed by the teacher to create detail feedbacks of the syllabus and the teaching-learning process. Then, to meet the standard of the university form will administer the questionnaire and the feedback. Hence, the result will improve the future teaching-learning process.

References

[1] William D, 2013 Assessment: The Bridge between Voices from Middle 21, 2 p. 40.
[2] Koh B et al., 2016 Varying levels of difficulty index of skills-test items randomly selected by examinees on the Korean emergency medical technician licensing examination J. Educ. Eval. Health Prof. 13 p. 5.
[3] Lim C P Zhao Y Tondeur J Chai C S and Tsai C-C, 2013 International Forum of Educational Technology & Society Bridging the Gap: Technology Trends and Use of Technology in Schools Source J. Educ. Technol. Soc. 16, 2 p. 59–68.
[4] Djemari M, 2012 Pengukuran Penilaian & Evaluasi Pendidikan Yogyakarta: Nuha Media.
[5] Widhiarso W, 2013 Aplikasi Model Rasch Campuran dalam Mengevaluasi Pengukuran Harga Diri J. Penelit. dan Eval. Pendidik.
[6] Essuman M A and Osei-Poku P, 2015 Evaluation of Selected Textbooks from Ghanaian Primary Schools Int. J. Innov. Res. Dev. 4, 6 p. 310–319.
[7] Winarni R Slamat S Y and Saddhono K, 2018 Development of Indonesian literature textbook with character education through information and communication technology (ICT) learning based Int. J. Eng. Technol.
[8] Cunningsworth A and Horner D, 1985 The role of simulations in the development of communication strategies System.
[9] Farida, Ariani; SLamat Mulyana A, 2009 Pembelajaran Mendengarkan Jakarta: Direktorat Jenderal Peningkatan Mutu Pendidikan dan Tenaga Kependidikan.
[10] Mayer R E, 2017, Using multimedia for e-learning, Journal of Computer Assisted Learning.
[11] Anita Dighe, 2003 Understanding Our Learners March.
[12] Hofstetter I, Aug. 1998 Multimedia applications for local newspapers and local information Comput. Networks ISDN Syst. 30, 13 p. 1223–1232.
[13] Sarjono S Mardapi D and Mundilarto M, 2018 Development of Physics Lab Assessment Instrument for Senior High School Level Int. J. Instr. 11, 4 p. 17–28.

[14] Won M Krabbe H Ley S L Treagust D F and Fischer H E, 2017 Science Teachers’ Use of a Concept Map Marking Guide as a Formative Assessment Tool for the Concept of Energy Educ. Assess. 22, 2 p. 95–110.

[15] Plomp T, 2013, Educational design research: An Introduction, in Handbook of research on educational communications and technology, .

[16] Iskandar A et al., Oct. 2018 Web based testing application security system using semantic comparison method IOP Conf. Ser. Mater. Sci. Eng. 420 p. 012122.

[17] Carmona S and Gronlund A, 1998 Learning from Forgetting Manag. Learn.

[18] Mousavi S F and Rezai V, 1999, Evaluation of scraping treatments to restore initial infiltration capacity of three artificial recharge projects in central Iran, Hydrogeology Journal .

[19] Firth M (Keisen U and Mesureur G (Keisen U, 2010 Innovative uses for Google Docs in a university language program JALT CALL J.

[20] Owen E A and Razali A B, 2018 The Effect of Communicative Activities on Libyan Secondary School Students’ Speaking Performance in Malaysia Int. J. Instr. 11, 4 p. 45–60.

[21] Bunce G et al., 1997 Test results of the g-2 superconducting solenoid magnet system IEEE Trans. Applied Supercond.