Development multiple choice test for chemistry learning in high school

E Nawawi, Y Sartika, F A Rachman
Faculty of Teacher Trainer and Education, Universitas Sriwijaya, Inderalaya, Indonesia

Abstract. The purpose of this research was to develop and produce multiple choice objective tests that are extended to chemistry subjects in class XI of High school that meet the requirements of validity, level of difficulty, distinguishing power, distractor effectiveness, and good reliability. The population in this study were all students in class XI of High school as many as 400 people. The sample of 42 people was determined by the proportionate stratified random sampling technique. This research includes the type of development research according to McIntire. The results showed that: (1) The multiple choice objective test was extended of buffer solution material in class XI High school with basic competencies 3.12 and developed into 5 indicators; (2) test grids were developed into 20 test items to represent all indicators; (3) the assessment rubric of this instrument uses a variety of scores according to the difficulty level of the reasons; (4) the expert material test and the score obtained is 0.890 and the linguist 0.922 belongs to the very high category; (5) the validity test of the items is 18 valid tests, difficulty level test produced 15 medium and 5 grains of the easy category; (6) test the effectiveness of the deceiver (distractor) produced 14 test items passed the test; (7) the reliability test of 0.86 including the very high category

1. Introduction

Education in Indonesia is below international standards. This can be seen from the results of the 2015 Program for International Student Assessment (PISA) Indonesia ranks 57th out of 65 OECD countries, as well as college or school graduates who are not ready to enter the workforce because of the lack of competencies. Provision of skills obtained from educational institutions is not sufficient to be used independently, because what is learned in educational institutions is often just fixated on theory so that students are less innovative and creative [1]. For this reason, the Indonesian government needs to make various reforms in the field of education, especially in improving the quality of education. Quality improvement in the field of education that has been carried out by the government is to improve the quality of the curriculum and the implementation of character learning and life skills, so that college graduates or schools are able to solve various problems of living creatively and become innovative and productive human beings [2].

The measurement researchers have recently expressed interest in classroom assessment intended to support student learning. Growing interest in diagnostic classification models and learning progressions, for example, represents a significant departure from the field's nearly exclusive focus on psychometric models and test development methods needed for large-scale assessment. Measurement science can make meaningful contributions to classroom formative assessment and grading practices, especially
regarding definitions of constructs and valid representation of intended learning goals. But there is also the danger that measurement frameworks can be distorted from high-quality instruction if the emphasis is on quantification rather than the quality of thinking they are encouraged "testing" formats are necessary for the efficacy of large-scale assessments-that are then replicated in computerized testing systems that track students' mastery of standards. Such testing formats are unnecessary in the classroom but often come to dominate classroom assessment [3].

In cases where districts lack the resources to design curricula to help teachers engage with next-generation standards and ambitious teaching practices, states could serve as a resource by providing “replacement units” that model the desired instructional activities with embedded formative and summative assessments [4]. Minister of Education and Culture Regulation of the Republic of Indonesia Number 81A of 2013 in Attachment IV, section VII Number A concerning the Concept of Learning Outcomes Writing states that the Definition of Assessment is the process of collecting information/evidence through measurement, interpreting, describing, and interpreting evidence of measurement results [5]. So that assessment is an integral part of the entire learning process. Assessment can be used as a factor in determining the success of learning processes and results, not just as a way to assess the success of students in mastering the subject matter.

In the assessment, there are four studies, namely the object that is assessed, the criteria as a benchmark, data about the object being assessed, and judgment (judgment) [6]. So that the assessment process consists of determining the object to be assessed, making or determining size criteria, collecting data and making decisions. The tool used as a means in the assessment can be either a test or a non-test. Teachers' conceptions of assessment can be understood in terms of their agreement or disagreement with four purposes to which assessment may be put, specifically, (a) improvement of teaching and learning, (b) school accountability, (c) student accountability, or (d) treating assessment as irrelevant [7].

The test is a set of tools that contain tasks that must be done or a number of questions that must be answered by students to measure the level of understanding and mastery of the required material coverage and in accordance with certain teaching objectives [8]. Scores based on representative samples of test follower behavior are indicators of how far the tested person has the characteristics of the chemistry being measured [9]. Further stated that the test is an instrument or a systematic procedure or procedure consisting of a set of questions or tasks to measure a certain behavior in students with the help of a numerical scale or certain categories. The test is a way to make an assessment in the form of a task or series of tasks that must be done by the child so as to produce a value about the child's behavior or achievement, which can be compared with the values achieved by other children or with the standard values specified [10].

Based on the results of interviews with teachers of High school Negeri 10 Palembang who teach chemistry subjects there are several problems as follows: (1) teachers rarely develop or make new chemical tests, (2) the development of tests conducted by teachers does not follow development procedures correct tests, (3) Tests used by teachers in assessment in the form of ordinary multiple choice objective tests, (4) existing ordinary multiple choice objective tests have not been able to measure higher levels of competency of students, (5) objective tests the usual multiple choices that have not been tested for validity and reliability.

To overcome these problems, it is necessary to have an assessment instrument in chemistry subjects that meet the correct test development procedures. The test to be developed in this study is an expanded multiple choice objective test. Expanded multiple choice test is a test that does not just choose the right answer but requires students to think about the reason for choosing the answer [11]. Assessments in expanded multiple-choice objective tests differ from ordinary multiple-choice objective test assessments that only use a score of 0 (zero) and 1 (one) but use a range of scores for example from a score of 0 (zero) to 5 (five). Because students other than choosing the answer choice option also write opinions regarding the selection of the answer, so the assessment is also as in the assessment test description. From the ideas written by students, the teacher can measure the level of understanding of students towards mastery of deeper learning material.
This study aims to: (1) Identify and determine the dimensions or indicators that must exist in the development of an expanded multiple-choice objective test instrument for the XI grade high school chemistry subjects, (2) compile and determine the lattice of the test instrument multiple choice objectives extended to chemistry class XI for High school, (3) create a rubric of the multiple choice objective test instrument extended to the XI for high school chemistry subjects, (4) test the validity of the expanded multiple choice objective test instruments on subjects chemistry class XI for High school, (5) tested the level of difficulty of the expanded multiple-choice objective test instruments on the XI grade high school chemistry subjects, (6) tested the distinguishing power of the expanded multiple-choice objective test instruments on the high school XI class chemistry subjects, (7) testing the effectiveness of the deceiver (distractor) from the expanded multiple choice objective test instrument on chemistry subjects in class XI for High school, (8) test the reliability of the multiple choice objective test instrument extended to the XI grade high school chemistry subjects

2. Methods
This development research uses Mc. Entire development model with 7 stages. These stages are [12]: Defining the test universe, audience, and purpose. This stage determines the overall test, the subject of this study is the XI for High school 3-grade students in Palembang 10 High school. This research is to test the expanded multiple choice questions which are valid, reliable, have different power, and the level of difficulty that researchers have developed. This study aims to enable students to recognize and train to answer questions in an expanded multiple choice form. The researcher conducted an initial analysis by interviewing one of the 11th-grade chemistry teachers in Palembang's 10th Senior High School. The researcher also analyzes the curriculum in the school by looking at the syllabus and its curriculum to suit the subject matter that will be used as a reference in developing expanded multiple-choice questions on this buffer solution material. The school of Palembang 10 State Senior High School uses the 2013 curriculum so researchers analyze chemical material in high schools that are included in the buffer solution, namely the understanding of buffer solution, the working principle of the buffer solution, the calculation of pH, and the role of buffer solutions in living bodies; Developing a test, At this stage, the completion of the question box, the question card and also the test format is a written test with an expanded multiple choice test type; Composing the test items. The preparation of test items is done by modifying the chemical questions that have already existed in books and the internet. The question chosen was a matter of buffer solution material and then adjusted to the 2013 curriculum. At this stage there were 30 expanded multiple-choice questions; Writing the administration instructions. This stage is writing test instructions. Duration of processing is 90 minutes to work on 20 multiple choice questions expanded buffer solution material. This test is done in writing; Conduct piloting test, The 20 expanded multiple choice questions were carried out by expert tests and limited to 10 students, after which the 20 items were analyzed by Cohan's Kappa Index for expert validation results and Alpha Cronbach for limited testing; Conduct item analysis, After students work on expanded multiple choice questions on the buffer solution material that has been given, then the researcher analyzes the validity of the items, the reliability of the questions, the different power of the questions, and the difficulty level of the answers obtained from students by combining even and odd questions. After analyzing the items using, states description; Revising the test, after analyzing the test items, the test items that are not good are revised. Deciding whether the test items are good or not can use quantitative and qualitative item analysis criteria. One item has the possibility of not fulfilling all the requirements of a good item which is too easy, too difficult or low discriminating power.

3. Results and Discussion
The product that was produced in this study was an expanded multiple choice question of class XI buffer solution material in high school. The results of the analysis on multiple choice questions were expanded to 20 questions tested and found the level of difficulty in 15 questions classified as moderate and 5 questions classified as easy. Based on the distinguishing power there are 18 category questions received and 2 category questions rejected. And for the validity of the question, there are 2 questions classified
as very high, 4 questions classified as high, 12 questions classified as quite high and 2 questions classified as very low. To decide whether these questions should be discarded or revised using the criteria for analyzing the quantitative items by comparing the level of difficulty of the items and the validity of the items and the conclusion that there are 2 questions that need to be removed or replaced. Whereas for the level of reliability, it was found that the reliability of the moderate question was 0.86 which was high.

Table 1. Results of quantitative analysis of multiple choice questions are expanded

| Category                        | Analysis results                           |
|---------------------------------|--------------------------------------------|
| Validity of items               | 2 about very high validity                 |
|                                 | 4 about high validity                      |
|                                 | 12 about enough validity                   |
|                                 | 2 about very low validity                  |
| Test reliability                | 0.86 with high interpretation              |
| The level of difficulty of the  | 15 questions in the medium category        |
| problem                         | 5 easy category questions                  |
| Item distinguishing power       | 18 category questions are accepted          |
|                                 | 2 category questions are rejected           |

Two questions with very high validity are items number 1 and 18. Problem 1 with a distinguishing power of 0.55 categories is very high, the level of difficulty of the item is 0.66, the category is all, the deceiver is functioning (A = 9%, B = 9%, C = 66%, D = 4%, and E = 4%). The problem is as follows: 1. The following is the definition of buffer solution which with answer options: (A). A solution containing strong acid with its conjugate base and strong base with conjugate acid; (B). A solution containing a strong base with conjugate acid and a weak acid with its conjugate base; (C). Solutions containing weak acids with their conjugate bases and weak bases with conjugate acids; (D). Solutions containing weak acids with weak bases and strong acids with strong bases; (E). A solution containing a strong acid with a conjugate base and a weak base with conjugate acid.

Answer Key Question Item No. 1 is C, for reasons: (1). The buffer solution is divided into the acid buffer solution and base buffer solution, score = 1; (2). The acid buffer solution is composed of weak acids and their conjugate bases, and base buffer solutions are composed of weak bases and conjugate acids, score = 1; total score = 3. Students who answered correctly were only 66% even though the level of difficulty in the medium category was due to students not accustomed to answering expanded multiple choice questions [13], even though the problem was only in the realm of Bloom C2 taxonomy. The advantage of question number 1 has a very high grain differentiation power of 0.55, meaning that item number 1 can distinguish high ability groups from low ability groups from a population.

Two items with very low validity are item number 6 and number 11. Problem number 6 with distinguishing factor of 0.00 with the level of difficulty of item 0.66 is medium category, all deceit function (A = 9%, B = 61 %, C = 9%, D = 4%, E = 9%). With the following questions: 6. A buffer system can be made by mixing 50mL of a 0.2 M CH3COOH solution with a solution ... with a choice of answers: (A). 50 mL 0.2 M NaOH; (B). 50mL CH3COONa 0.1 M; (C). 50mL KOH 0.2 M; (D). 50 mL NH4OH 0.2 M; (E). 50mL NH4Cl 0.1 M. The key to answering item number 6 is B, for reasons of the answer: From the data above 50mL CH3COOH (weak acid) 0.2 M can form a buffer solution when mixed with 50 mL CH3COONa 0.1 M by producing the remaining mole of 5 mmol from 10mmol CH3COOH and 5 mmol NaOH. The distinguishing item obtained is 0.00, meaning that item number 6 confuses students in answering, the upper group answers are balanced with the bottom group answers. This means that item number 6 cannot distinguish groups with high ability and groups with the low ability.

Item number 11 with a distinguishing power of 0.00 and the level of difficulty of the item 0.88 category is very easy. Two fools don't work (A = 80%, B = 0%, C = 0%, D = 4%, E = 9%). With the following problem: 11. What is the acid ionization constant value (Ka), if you know the pH of the
mixture of 100 ml 0.2 M KOH and 200 ml 0.2 M HF is 5 - log 6? With answer options: (A). 6 x 10^{-5}, (B). 6 x 105, (C). 9 x 10^{-5}, (D). 9 x 106, (E). 9 x 10^{-6}. The key answer to question number 11 is A, for the reason: the chemical count to find the Ka acid value is weak, problem number 11 is in the realm of applying C4 Bloom Taxonomy. The proportion of the correct answers reaches 80%, there is the possibility of guessing techniques or friend intervention, because the reasons are made a lot wrong. Very low discriminating power really cannot distinguish high ability groups with low ability [14].

4. Conclusion
Based on the results of the research and discussion, some conclusions can be presented as follows. The results showed that: (1) The multiple choice objective test was extended of buffer solution material in class XI High school with basic competencies 3.12 and developed into 5 indicators; (2) test grids were developed into 20 test items to represent all indicators; (3) the assessment rubric of this instrument uses a variety of scores according to the difficulty level of the reasons; (4) the expert material test and the score obtained is 0.890 and the linguist 0.922 belongs to the very high category; (5) the validity test of the items is 18 valid tests, difficulty level test produced 15 medium and 5 grains of the easy category; (6) test the effectiveness of the deceiver (distractor) produced 14 test items passed the test; (7) the reliability test of 0.86 including the very high category

References
[1] Jacobs C L, Martin S N, and Otieno T C. 2008. Science Teacher Education Journal. 192, 1096.
[2] Wagner T. 2010. Overcoming the Global Achievement Gap.
[3] Shepard A L, Penuel W R, & Pellegrino J W 2018 Educational Measuremen: Issues and Practice 37 21
[4] Marion S & Shepard L. 2010. Journal Educational Assessment
[5] Depdikbud, 2013. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 81A Tentang Implementasi Kurikulum. 2013 Jakarta: Kementerian Pendidikan dan Kebudayaan
[6] Alfian A, Aminah N St, and Sarwandi. 2015 Journal Scientific Approach 4, 39.
[7] Brown G T L. 2004. Journal Assessment in Education. 11, 301
[8] Majid A. 2014. Penilaian Autentik Proses dan Hasil Belajar. Bandung: PT Remaja Rosdakarya.
[9] Muhammad N K. 2011. Journal of Educational and Psychological Assessment. 71.
[10] Carlo M. 2011. Journal of Educational and Psychological Assessment. 7, 57.
[11] Jan van den Akker, Koeno Gravemeijer, Susan McKenney and Nienke Nieveen. Introducing Educational Design Research. The Netherlands: University of Twente.
[12] Wulandari, F. 2013. Pengembangan Instrumen soal Pilihan Ganda Untuk mengidentifikasi kesalahan konsep siswa kelas V Materi Pesawat Sederhana di MI Perwanida kota Blitar. Malang: Universitas Islam Negeri Maulana Malik Ibrahim.
[13] Inteni, K. A. S., Candiasa, I. M., Kom, M. I., & Suarni, N. K. 2013. Pengembangan Instrumen Tes Objektif Pilihan Ganda Yang Diperluas Berbasis Web Untuk Mata Pelajaran Tik Kelas Xi Sman Di Kabupaten Karangasem. Jurnal Penelitian dan Evaluasi pendidikan Indonesia, 3.
[14] Hartono, Effendi, Susanti R, Riyanto, Ismet, Eijkelhof H. 2018 Proceedings of the First Indonesian Communication Forum of Teacher Training and Education Faculty Leaders International Conference on Education 2017 174 Atlantis Press 222

5