International Conference on Biology Education, Natural Science, and Technology
Vol. 1 No. 1 (2023)

The Feasibility and Effectiveness of High-Level Thinking Tests on Virus to Mapping The Ability of Prospective Science Teacher (Introduction Study on The Development of Integrated Learning Models)

Annur Indra Kusumadani*, Sentot Budi Rahardjo, Sri Yamtinah, Baskoro Adi Prayitno
Faculty of Teacher Training and Education, Universitas Sebelas Maret,
Central Java, Indonesia
*Corresponding Author. E-mail address: aik110390@student.uns.ac.id

KEYWORDS:
Test Research and Development,
Higher Order Thinking Test,
Instrument Measurement

ABSTRACT

This research aims to determine the feasibility and effectiveness of a higher - order thinking test developed on material of virus. The test instrument developed was aimed at mapping the ability of prospective science teacher in the preliminary study of developing an integrated learning model. This research is a *Research and Development* referring to Borg and Gall (1989) with 10 stages of research, namely : (1) Initial research and data collection , (2) Planning , (3) Initial product manufacture , (4) Initial trial (expert validation) , (5) Initial product improvement , (6) Test field trial , (7) Operational product improvement , (8) Operational field test (effectiveness test/external validity test) , (9) Final product revision, and (10) Dissemination. The research involved the following experimental subjects : (1) The initial trial involved virus material experts, research and education evaluation experts ; (2) The small group trial involved 36 student teacher candidates at the University of Muhammadiyah Surakarta for the 2020/2021 academic year; (3) The large group trial involved 60 student teacher candidates at the University of Muhammadiyah Surakarta for the 2020/2021 academic year ; and (4) an operational field test (effectiveness test/external validity test) involving 30 teacher candidates at the Muhammadiyah University of Surakarta for the 2020/2021 academic year. The data analysis techniques used were the validity of AIken at the validation stage, item test analysis in the field trial and Pearson correlation analysis at the operational /effectiveness field test stage. The final results of this research are computer-assisted higher order thinking test with 8 questions of higher order thinking on virus material that has strong internal validity with an Aiken validity index of more than 0.73 (0.84 to 0.92), reliability coefficient 0.814 > 0.213 (r table) which means high reliability, difficulty index of 0.17 (difficult) to 0.455 (medium), and discriminant index (differentiation power) of 0.2 (enough) to 0.745 (good). External validity/effectiveness is 0.626 (strong enough) to 0.831 (strong) which is the correlation of higher order thinking test results developed with other higher-order thinking tests on virus. The conclusion of this research is that the developed test instrument is feasible and effective to measure the higher order thinking ability of virus.

1. INTRODUCTION

Humans who are successful in the era of globalization are humans who are able to think well. Good thinking skills are shown by creative and critical humans so that they are able to produce tools and services needed to meet the needs of world development and solve life's problems. Thinking skills can be empowered through innovative education. Innovative education in the real world is carried out by educators, one of which is by using various learning models and moduls (Setyaningsing, et.al., 2022). The business form of innovative education is the development of

learning models, learning modules, learning media, and evaluation/evaluation instruments. Development of evaluation instruments/assessment is very important because it is strategically valuable to find out the thinking ability map of prospective science teacher as a reflection of learning so that lecturers know the follow-up learning including the development of learning models. Questions as test instruments to measure the success of learning objectives must have good quality. This quality can be measured by the level of cognition expected from students (Mustafidah, 2022).

The low fulfillment of evaluation/ assessment standards is evidenced by the results of the analysis of the question bank used by the lecturers of the Faculty of Teacher Training and Education, Muhammadiyah University of Surakarta for formative and summative evaluations, showing that the percentage of using Bloom's cognitive domain in questions is 10% memorizing (C1), 40% understanding (C2) and 30% apply (C3), 15% analyze (C4), 5% evaluate (C5), while the creative domain (C6) has not been used by lecturers. Therefore, it can be concluded that the test instrument used has not been able to measure higher order thinking skills.

The factor causing the test instrument used has not been able to measure high-level abilities is that the lecturer feels that he has made questions to measure higher-order thinking skills but the questions that are made include low-level thinking questions. In addition, time is also limited because it is felt that it takes longer to make higher-order thinking questions. As a result, the ability of students in higher-order thinking is not mapped so that teachers do not know the extent to which students' abilities in higher-order thinking are. On the other hand, prospective science teacher are required to be able to think at a higher level. Higher order thinking skills are the skills needed in order to remain competitive in the era of globalization.

Moreover, the era of globalization demands the use of technology. The rapid development of technology and communication science has resulted in technology becoming a basic need that must be known and understood by everyone (Sutama, 2021). The world of education is required to adapt to technological developments and there are ongoing demands for 21st century learning skills, the development of the industrial revolution 4.0, and society 5.0. (Ratih, et.al., 2021). This also has a positive impact on the use of technology in tests so that it can simplify and speed up the assessment. The test used by the lecturers of the Faculty of Teacher Training and Education, Muhammadiyah University of Surakarta for formative and summative evaluations has not been computer assisted. The obstacle that was complained by the majority of lecturers was that most of the development of computer-assisted test instruments had a high level of difficulty in manufacturing. On the other hand, one of the advantages of computer-assisted tests is that they are fast and accurate in scoring and profiling abilities.

Based on the problems above, it is necessary to develop a computer-assisted assessment capable of measuring higher order thinking skills that are valid, reliable, objective, and practical. One way that can be done is to use a multiple choice test to analyze the relationship between computer-assisted things with *Microsoft Powerpoint* and *Visual Basic for Applications*. Multiple-choice tests that analyze the relationship between things have reasons in each question statement which is a development of the ordinary multiple-choice question type. The existence of reasons in each question statement can measure students' higher-order thinking skills more deeply and anticipate *ghessing* in answering questions. The test instrument was developed with computer assistance with *Microsoft Powerpoint* and *Visual Basic for Applications* so that it can automatically display student work and display high-level thinking profiles of prospective science teacher. In addition, the test instrument developed is equipped with instructions for use so that it also makes it easier for lecturers to make computer-assisted high-level thinking test instruments with *Microsoft Powerpoint* and *Visual Basic for Applications*. The test instrument developed is limited to the material of the virus. The material on virus must have been taught before students work on the developed test instrument. The test instrument developed can operate on all *operating systems* and all types of laptops/PCs that have been *installed Microsoft Office*.

The purpose of the study was to determine the feasibility and effectiveness of higher order thinking tests on virus materials computer assisted with *Microsoft Powerpoint* and *Visual Basic for Applications* .

2. RESEARCH METHODS

Research model to be carried out is a modification of *Research and Development* (*R & D*) (Borg and Gall, 1989) on the number of trial subjects at each stage of the trial . The research consisted of 10 stages, namely research and initial data collection, planning, initial product creation, initial trial (expert validation), initial product improvement, field trials, operational product improvements, operational trials (effectiveness test/external validity test), final product improvement, and dissemination. The stages in detail are as follows.

Research and initial data collection includes 2 needs analysis activities. An overview of the test instruments used by the lecturers of the Faculty of Teacher Training and Education, University of Muhammadiyah Surakarta for formative and summative evaluations was obtained from the interview answers. The question bank analysis was conducted to determine the percentage of use of the bloom taxonomy level in the questions used by lecturers and to analyze the difficulties of teachers in conducting assessments/ evaluations.

Activities carried out in planning and development refer to the *Delphi Method* (2002). The Delphi method is a technique for planning and developing product development without having to meet directly from all experts and practitioners to make product development. The activities carried out are identification of competency standards and basic competencies as well as the type of material to be developed to determine indicators of learning materials and indicators of higher order thinking questions, making material indicators and questions of higher order thinking skills adapted from Anderson and Krathwohl (2001) then making questions based on the indicators provided. has been established.

The initial trial phase (expert validation) involves 5 material experts, namely lecturers of human anatomy and physiology , 3 an expert on evaluation instruments , namely a lecturer in charge of educational research and evaluation courses. Validation using product validation and assessment sheets. The function of a material expert in research is to validate the content or concept of the material and constructs so that they are able to avoid discrepancies with SK, KD, indicators, and objectives as well as misconceptions on the questions. The function of the evaluation instrument expert is to validate the form of the questions and the suitability of the questions with indicators so that they can ensure that the questions developed actually measure higher-order thinking skills . Content and construct validation by experts using the Aiken Formula (Aiiken, 1985b). After getting validation from the experts in the initial trial (expert validation), revisions were made until the expert declared the test instrument was feasible (good) .

Field trial 1 (small group) involved an microbiology lecturer and 36 student science teacher candidates, Faculty of Teacher Training and Education, University of Muhammadiyah Surakarta for the Academic Year 2020/2021 who took the microbiology course . Lecturers in this case as practitioners. The function of the practitioner is to validate the content of the material, readability, and time sufficiency problem solving , while students comment on the readability of the questions and the adequacy of processing time . The data from discussions and interviews with students are then in percentage. Revision carried out according to the results of discussions and interviews with students . The result of field trial 1 is *draft* I II a computer-assisted higher order thinking test that is ready to be tested in large groups. Field test 2 (large group) was conducted at 60 student science teacher candidates, Faculty of Teacher Training and Education, University of Muhammadiyah Surakarta for the Academic Year 2020/2021 who took the microbiology course . The trial is an internal validation of the questions. Internal validation of the questions includes reliability testing, differentiating power of questions, and the level of difficulty of the questions. The trial has the aim of determining whether the resulting product has the feasibility of being a valid, reliable, and objective question . Field trial data were analyzed, if the results did not meet the internal validity

standard , a revision was made to the product. The results of field trials and operational product revisions are computer-assisted higher order thinking tests that are valid, reliable , objective and practical and ready for operational field tests (effectiveness test/external validity test) .

Operational field test (effectiveness test/external validity test). carried out on 20 prospective science teacher of the Faculty of Teacher Training and Education, University of Muhammadiyah Surakarta in the 2020/2021 academic year who took the microbiology course . External validity refers to *concurrent validity* (Crocker & Algina, 1986). The instrument is said to have *concurrent validity* if the results are in line with the existing criteria. Existing criteria can be in the form of other instruments that measure the same thing. Alignment is indicated by the correlation index (*Pearson Correlation*). The purpose of the trial was to determine the external validity of the test by correlating the results of the developed higher-order thinking test with the results of standardized higher-order thinking tests/other developments on the same material and on the same students so that it was known to what extent the questions developed were effective in measuring thinking skills. high-level college students by looking at the Pearson correlation coefficient. The stronger the correlation, the better the effectiveness/external validity of the test. If the results are not effective, then the product is revised. The result of the operational field test and final product revision is a computer - assisted higher order thinking test which is effective and ready to be disseminated.

The dissemination was carried out through lecturers of the Faculty of Teacher Training and Education, Muhammadiyah University of Surakarta. Dissemination is carried out after the product developed is feasible and effective to measure the higher-order thinking skills of prospective science teacher.

3. RESULTS AND DISCUSSION

The data from the validation results with 10 validators with 4 assessment criteria shows the Aiken validity index of more than 0.73 (0, 84 to 0, 92) for all items. That is, the 8 items developed are items that are valid in content and constructs. In addition, the truth of the concept of the question gets a total percentage of 93.2%, meaning that the truth of the concept is very good. The suitability of the material level for students gets a total score percentage of 98.1%, meaning that the questions are very in line with the students' thinking level. The novelty aspect of the questions also gets a total score percentage of 100%, meaning that the questions used are relatively new and have never been used for tests. The clarity of the meaning of the question gets a total score percentage of 96%, meaning that the clarity of the meaning of the question is very good. Clarity and ease of question terms for students to understand get a total score percentage of 93.7%, meaning that the terms used are clear and easy for students to understand.

Along with content and construct validation, the *software aspect was validated by ICT* experts and the test form used by educational research and evaluation experts. The results of validation with *ICT experts* show that the level of ease of operation of the test gets a percentage of the total score of 100%, meaning that the ease of operation is very good. The specificity of the test function got a total score of 87.5%, meaning that the specificity of the test function was very good. The test communication got a percentage of the total score of 75%, meaning that the test communication was good. Completed instructions get a total score percentage of 75%, meaning that the test instructions are good. The display design and logic design get a total score percentage of 87.5%, meaning that the display and test logic are very good.

The results of the validation with research and education evaluation experts showed that the suitability of SK, KD, and indicators got a total score percentage of 92.7%, meaning that the questions were in accordance with SK, KD, and indicators. The coverage of the questions in the higher-order thinking domain and the knowledge dimension got a total score percentage of 91.9%, meaning that the questions correctly covered the higher-order thinking domain and the knowledge dimension. The clarity of the meaning of the question gets a total score of 94.1%, meaning that the clarity of the meaning of the question is very good. Clarity and ease of question terms for students

to understand get a total score percentage of 93.6%, meaning that the terms used are clear and easy for students to understand.

Field trial 2 (large group) aims to determine the reliability, discriminatory power, and level of difficulty of the higher-order thinking test developed. The reliability, discriminating power, and difficulty level of the higher-order thinking test developed were obtained from the results of tests carried out by students. Student test results were analyzed by items .

The results of the item analysis show that the test reliability coefficient is 0.814> 0.213 (r table), meaning that the higher-order thinking test developed is said to be highly reliable. The formula used to calculate test reliability on *ANSOFT1 software* is KR- 20 (Linn & Groundlund, 2000; Widoyoko, 2010).

The results of the item analysis show that the index of the difficulty level of the test is 0.17 to 0.455, meaning that the higher-order thinking test developed has questions classified as difficult to moderate questions. The smaller the index of difficulty, the more difficult the questions. Measurement of the level of difficulty using the index of difficulty level along with standard limits (Azwar, 2003; Allen & Yen, 1986; and Sudjiono, 2005)

The results of the item analysis showed that the discriminant index (differentiating power) was 0.2 to 0.745, meaning that the higher-order thinking test developed had sufficient to good distinguishing power. The greater the discriminant index, the better the question is in distinguishing students who have good and low-level thinking skills. The measurement of discriminant power uses a discriminant index along with standard limits (Yen, 1992; Crocker & Algina, 1986) .

Operational field test (effectiveness test/external validity test) refers to *concurrent validity* (Crocker & Algina, 1986). The instrument is said to have *concurrent validity* if the results are in line with the existing criteria. Existing criteria can be in the form of other instruments that measure the same thing. Alignment is indicated by the correlation index (*Pearson Correlation*). The results of the effectiveness test (external validity test) between the higher-order thinking test developed with the higher-order thinking test and the development of other tests on the same material, namely the virus in measuring the ability to analyze, show a correlation coefficient of 0.831 with a significance level of 0.000<0.05, meaning that there is a correlation. strong positive. In conclusion, both the developed higher-order thinking test and other developed higher-order thinking tests have good external validity so that they can be used to measure analytical skills.

The results of the effectiveness test (external validity test) between the higher-order thinking test developed and the higher-order thinking test the development of another test on the same material, namely the virus in measuring the ability to evaluate, shows a correlation coefficient of 0.626 with a significance level of 0.000<0.05, meaning that there is a correlation strong positive. In conclusion, both the higher-order thinking test developed and the higher-order thinking test developed by other tests have good external validity so that they can be used to measure the ability to evaluate.

The results of the effectiveness test (external validity test) between the higher-order thinking test developed and the higher-order thinking test the development of another test on the same material, namely the virus in measuring creative ability, shows a correlation coefficient of 0.687 with a significance level of 0.000<0.05, meaning that there is a correlation. strong positive. In conclusion, both the higher-order thinking test developed and the higher-order thinking test developed by other tests have good external validity so that they can be used to measure creative abilities.

Effectiveness test using a comparison of other higher-order thinking test instruments on the same material, namely the virus because it shows the alignment of the measuring instrument, namely measuring higher-order thinking skills on the material of virus. The effectiveness test is more perfect when there is a comparison in the form of a standardized or standardized high-level thinking test. Although there is no standardized comparison of higher-order thinking tests, other higher-order thinking tests can be used for the development of the same material, namely virus.

This means that if the developed higher-order thinking test and other developmental higher-order thinking tests are used to measure higher-order thinking skills of the same students and show a positive correlation, the developed higher-order thinking test and other developmental higher-order thinking tests can is said to be effective for measuring students' higher order thinking skills.

4. CONCLUSIONS and RECOMMENDATIONS

The test instrument developed was a computer- assisted high-order thinking test (*computer based test*). The test consists of 8 items of higher-order thinking on valid virus material with an Aiken validity index of more than 0.73, high reliability with a reliability index of 0.954> 0.257 (r table), objective which is different with a discriminant index of 0.2 (enough) to 0.745 (good) and difficulty index 0.15 (difficult) to 0.467 (medium). The developed test is able to measure higher order thinking skills on the dimensions of factual, conceptual, and procedural knowledge on the material of virus in the microbiology course. The form of the test used is multiple choice analysis of the relationship between things.

The test instrument was declared feasible by 5 material experts, 3 research experts and educational evaluation experts. In expert validation, the Aiken validity index was found to be more than 0.73 (0.83 to 0.93). The test instruments are eligible for field tests 1 and 2 with a reliability index of 0.814> 0.213 (r table) meaning high reliability, discriminant index 0.2 (adequate) to 0.745 (good) and difficulty index 0.17 (difficult) to 0.455 (moderate).

The test instrument was declared effective for measuring higher-order thinking skills in the material of virus after undergoing an effectiveness test by correlating the measurement results of the developed higher-order thinking test instrument with the measurement results of other higher-order thinking test instruments on the same material, namely the virus using *Pearson* correlation formula. There is a strong positive correlation in the measurement results of the ability to analyze with a correlation coefficient of 0.831 and a significance level of 0.000 <0.05. There is a fairly strong positive correlation in the measurement of the ability to evaluate with a correlation coefficient of 0.626 with a significance level of 0.000 <0.05. There is a fairly strong positive correlation in the measurement of creative ability showing a correlation coefficient of 0.687 with a significance level of 0.000 <0.05. The results of the correlation test indicate that the test instrument developed has good effectiveness as well as strong external validity so that it can be used to measure higher order thinking skills.

Other researchers who are interested in continuing to develop higher order thinking tests on virus is expected to pay attention to the limitations of research and research products, so as to make the test instruments that have been developed better.

5. REFERENCES

Aiken, L. R. (1985a) . Content Validity and Realibility of Single Items or Questionnaires. *Educational and Psychological Measurement, 40* , 955-959.

Aiken, L. R. (1985b) . Three Coefficients for Analyzing The Realibility and Validity of Ratings. *Educational and Psychological Measurement, 8* , 131-142.

Allen, MJ & Yen, WM (1979). *Introduction to measurement theory* . Monterey, California: Brookd/Cole Publishing Company.

Anderson, L. W. _ & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing* . New York: Longman.

Azwar, S. (2012). *Preparation of the Psychological Scale. Edition II* . Yogyakarta: Student Library.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1 Cognitive Domain.* New York: David McKay.

Borg, W. R. _ & Gall, M. D. (1983). *Educational research an introduction (4th Ed)* . White Plains: Logman Inc.

Borg, WR and Gall, & D. (1989). *Educational Research.* New York: Longman.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* Forth Worth: Holt, Rinehart, and Winston, INC.

Mustafidah, H., Suwarsito, S., & Pinandita, T. (2022). Natural Language Processing for Mapping Exam Questions to the Cognitive Process Dimension. International Journal of Emerging Technologies in Learning, 17(13)

Ratih, K., Syah, M.F.J., Nurhidayat, & Buckworth, J. (2021). Learning Patterns during the Disruptive Situation in Informal Education: Parents' Efforts and Challenges in the Adjustment of Progressive Learning. *Indonesian Journal on Learning and Advanced Education*, 3(3), 180-193.

Sudjiono, A. (2005). *Introduction to Educational Evaluation* . Jakarta: PT Raja Grafindo Persada.

Sutama & Fajriani, I.N. (2021). WEB-Based E-Learning Media at the Vocational High School Level. *Jurnal Varidika*, 33(2), 129-140

Widoyoko, SEP (2010). *Evaluation of Learning Programs* . Yogyakarta: Student Library.

Yen, WM (1992). "Item Response Theory". in Alkin MC (Eds.), *Encyclopedia of Educational Research* (pp. 657-666). New York: Macmillan Library Reference USA.