JISE 7 (1) (2018) : 10-24

Journal of Innovative Science Education

http://journal.unnes.ac.id/sju/index.php/jise

Authentic Assessment of Fungi for Vocational School Student: Concept Map, Self Assessment and Performance Test

Ristiana Nugrahani[1,2]✉, Andreas Priyono Budi Prasetyo[2], Retno Sri Iswari[2]

[1] SMK Negeri 1 Bawen
[2] Universitas Negeri Semarang, Indonesia

Article Info

Article History:
Received November 2017
Accepted Januari 2018
Published August 2018

Keywords:
concept map assessment, self assessment, performance assessment

Abstract

The authentic assessment in biology teaching learning at vocational school of Bawen, SMK Negeri 1 Bawern, especially at the department of post harvest technology, has not been fully implemented yet. This study was aimed at developing and analyzing the validity, reliability, effectiveness, and practicality of authentic assessment tools such as concept map assessment (CMA), self assessment (SA), and performance assessment (PA) for vocational students. This research used Research and Development (R & D) approaches. Content and construction validities were verified by experts. Concruent and predictive validity of CMA was verified by score of correlation between CMA score and master map as well as final semester score. Concurrent validity of SA was calculated from corelation scores of scientific attitudes and scientific process skills. Concruent validity of PA was calculated from corelation scores of PA and basic competency 3.5. Concruent and predictive validations were calculated by pearson corelation product moment. Reliability of CMA and PA was analysed by reliability inter-rater reliability techniques. Reliability of SA was calculated by alpha cronbach's methods. Effectiveness and praticability of instruments were calculated from teachers's and students' responses analysed with percentage description techniques. The results showed that the concept map has excellent content and construction validity, excellent concurrency validity, and good predictive validity. Self assessment had excellent content and construction validity and good concurrency validity. PA had excellent content and construction validity and good concurrency validity. CMA is very effective for measuring cognitive aspects, SA is very effective to measure scientific attitudes, and PA of biological process skills. The three types of assessment were claimed to have practicality in terms of time, ease of use, ease of correction, facility supports, and cost required.

✉ Alamat korespondensi:
Jl. Kartini Bawen No.119, Bawen, Semarang, Jawa Tengah 50661
E-mail: ristianaunnes@gmail.com

p-ISSN 2252-6412
e-ISSN 2502-4523

INTRODUCTION

Vocational High School (SMK) has different characteristics compared to Senior High School (SMA), the difference lies in the graduates produced. Learning in Vocational School (SMK) aims to produce graduates who are competent in the world of work in accordance with their respective areas of expertise (Solikha, 2015). In the Department of Agricultural Processing Technology (TPHP), Vocational students are required not only skilled in their field of expertise, but also have a deep understanding of biological concepts and important scientific attitudes to be developed when TPHP students work in the laboratory. The way to measure and strive for the achievements of the goals is through assessment. The goals of biology learning is not only knowledge, skills, but also scientific dispositions. What biology teachers have to do is to assess their knowledge and attitudes by a more appropriate appraisal tool (Prasetyo, 2015). Learning with the 2013 curriculum (K-13) requires the use of authentic assessment as a student achievement tools holistically.

Authentic assessment is a form of assessment that presents an interesting problem in which students must use knowledge to demonstrate effective and creative performance. Assignment of a replica or analogue to the type of problems encountered in everyday life (Wiggins, 1990). Authentic assessments are contrary to traditional judgments, such as multiple choice, matching, and true or false which provide only superficial ideas of what students are studying, and do not show what students can do with the knowledge they have because traditional tests usually only ask students to choose an answer which is true (Wiggins, 1990). Formative terms, alternatives, and authentic judgments are in turn. Authentic assessment types include performance assessment, portfolio, self-evaluation and peer-evaluation, interview-based assessment, play-based assessment, cooperative group assessment, dialogue, journals (Wei et al., 2007), concept mappings, science projects, on diagrams, research articles, research reports, and so forth (McTighe & Ferrera, 1994). An alternative assessment is intended to provide evidence of what students know and can do in the subject matter (Ruiz-Primo & Shavelson, 1996a). However, the manifestation of authentic assessment into the practice of biology teaching in Indonesia is not always easy to do (Prasetyo, 2015; Fatmawati et al., 2013; Sujati, 2014).

The Vocational School of Bawen has implemented K-13 since the beginning of the academic year 2013/2014, but the implementation of authentic assessment has not been optimal. From the preliminary study conducted by the researchers obtained information that the biological assessment applied in Bawen Vocational School has still used a traditional test such as multiple choice, short field, description or matching. The biological assessment is done at the end of the lesson. Cognitive assessment is done after the completion of learning a certain subject. Affective assessment is only done at the end of the semester by Counseling (BK) teachers and mentors, while the psychomotor assessment by assessing the practicum report without the assessment rubric leads to process skills. The same condition is also stated by Hamid (2008) who suggests that the scoring system used in science learning in high school is still dominated by paper and pencil based test. One reason why the teacher never to implement performance and student-self assessment is lack of knowledge and lack of skill to development an authentic assessment instrument and technique (Prasetyo et al., 2016; Sari et al., 2015).

Another problem that has been found is that authentic biological assessments conducted so far have not been linked to the needs of individual skill packages, including TPHP. The lack of integration of substantive vocational content in biology learning resulted in students is not along with the concept, the science process, its application in agribusiness and agrotechnology, and the motivation of low biology learning. Low student motivation is one indicator of low scientific attitude of students that is closely related to curiosity, never give up,

and discipline, and so forth. One solution to some of these problems is the development of authentic biological assessment according to the needs of students of Bawen Vocational School, especially the competence of TPHP department.

The type of assessment to be developed is the concept mapp assessment (CMA), self assessment (SA), and performance assessment (PA). Concept map assessment is able to explain important aspects of organizing cognitive structures and investigating student misconceptions (Ruiz-Primo & Shavelson, 1996b; McClure et al., 1999). PA allows students to make responses, make products, or demonstrate to show what they understand and can do. With PA, teachers can directly observe the application of desired skills and knowledge (McTighe & Ferrera, 1994). SA is a formative assessment in which a student reflects, compares the quality of achievement against predetermined criteria, and makes improvements to his/her own learning (Andrade & Valtcheva, 2009) and can be used for attitude assessment. The developed assessment instrument was used to measure student learning outcomes in KD 3.5, which analyzed the characteristics, reproduction, and role of fungi (yeast and mold) in the TPHP department.

Although the CMA, SA, dan PA have been widely used, testing of the quality of these three assessment forms needs to be done so that the assessment developed can provide reliable measurement results. Developing a quality appraisal tool is important to do because the success of student achievement will be largely determined by the teacher's experience and knowledge in using properly constructed measurements (Rahmawan et al., 2016) and often the assessment results are used to make important decisions in learning. The developed assessment is expected to be an example for other biology teachers to apply authentic assessment in the classroom. The purpose of this study is to: 1) analyze the validity 2) analyze the reliability of the assessment, 2) analyze the effectiveness of the assessment, 3) analyze the practicality of the assessment.

.

METHODS

The type of research conducted is Research and Development (R & D). The research procedure includes identification of potentials and problems, needs analysis, product design, expert validation, 1st product revision, limited-scale trial, 2nd product revision, large-scale trials, 3rd product revisions, and final products. Trial usage is done in Vocational School of Bawen. Research design is pre experimental design with One-shot Case Study method. The population in this study were all students of class X TPHP lesson year of 2016/2017. The subjects of limited-scale trial were 15 students of class X TPHP A, and the subjects of large-scale trial were 34 students of grade X TPHP B. The sample determination was done by purposive sampling technique. The data were collected using interview method, observation, validation sheet, and questionnaire. Interview method was conducted to identify the initial problem. Observation method is used to know the implementation of the assessment. The questionnaire method is used to obtain student and teacher response for data on the developed assessment. Content validity and construction of three types of assessment were obtained with expert review. Concurrent validity of CMA was verified by score of correlation between CMA score and master map referred by Goldsmith dan Davenport theory (Rusilowati, 2011). Predictive validity of CMA was verified by score of correlation between CMA score and final semester score (McClure et al, 1999). Concurrent validity of SA was calculated from corelation scores of scientific attitudes and scientific process skills (Boud, 1992). Concurrent validity of PA was calculated from corelation scores of PA and score in basic competency 3.5 (Wei et al., 2012). Concruent and predictive validations were calculated by pearson corelation product moment. Reliability of CMA and PA was analysed by reliability inter-rater reliability techniques. Reliability of SA was calculated by internal consistency methods and alpha cronbach's formulation. Effectiveness and praticability of instruments were calculated

from teachers's and students' responses analysed with percentage description techniques.

RESULTS AND DISCUSSION

1. Validity

The product developed is CMA to measure the cognitive aspects of fungi material, SA to measure scientific attitudes, and PA to measure the skills of the science process.

Validity of CMA

There are three validity tests performed for CMA instruments, namely content and construction validity, konkuresi validity, and predictive validity. Content validity and construct based on expert review obtained CMA results have excellent validity shown in Table 1.

Table 1. Validity Test Results of CMA

No.	Validity of CMA	Analyze Tecnique	Result	Note
1.	Content and Construction	Descriptive Pertentage	91.98%	very valid
2.	Concurensi on	Pearson korelasion product moment	rxy = 0.932	very valid
3.	Predictive	Pearson korelasion product moment	rxy = 0.580	rather valid

Based on table 1, a CMA instrument has been produced that has excellent content and construction validity criteria, excellent concurrency validity, and good predictive validity. CMA has met the criteria of content validity because CMA has been prepared in accordance with KI 3 and KD 3.5 with the characteristics, health, and role of fungi (yeast and mold) in the TPHP department. Problems are prepared based on their respective indicators of morphological characteristics of fungi, repair yeast and mold, the role of yeast and the role of mold to facilitate students understand the problem with the total map of the resulting concept.

CMA has a validity criterion of construction because it has been equipped with a grid that serves as a guide for the preparation of questions, complemented by a general explanation that serves to provide an explanation of the concept map, especially if there are students who have never known the concept map on previous learning. CMA also comes with a master map and scoring guidelines that make it easier for teachers to correct students' work. Good content and construction validity criteria are the main keys of the valuation instrument and as the main capital prior to the validity test of concurrency. Concurrent validity is performed to determine the reliability of CMA in measuring cognitive students. The validity of concurrency was obtained from the correlation of students' concept map scores with the master map and obtained very valid results. Good concurrency validity indicates that the concept map created by students has approached ideal criteria. The validity of concurrency is also supported by the testing of instruments on a limited scale before being tested on a wide scale. Instruments that have been tested will improve their empirical validity

CMA that has fulfilled the criteria of content validity, construction, and concurrency, then tested the predictive validity to know its reliability in predicting student learning outcomes in the future. Predictive validity can be determined by comparing concept map scores with scores obtained by other valuation techniques because validity with single observations is always suspected (McClure et al., 1999). The predictive validity test results show that CMA has a good enough predictive validity criterion with a validity coefficient of 0.58. Empirically some studies have shown consistent correlations between concept map scores and

student scores from measurements with other techniques, while other studies have shown different results with multiple choice tests (Ruiz-Primo et al., 1997). McClure & Bell (1990) found a correlation of more than 0.50 between concept map scores and final exam scores. However, Novak et al. (1983) reported a different correlation. This correlation is evidence that multiple-choice tests measure different types of learning, ie learning memorization rather than learning as measured by concept maps, meaningful learning (Ruiz-Primo & Shavelson, 1996b).

The CMA fill in the map model is customized with student intake so as to support CMA validity. Preparation of CMA must be appropriate to the student's ability (Taib, 2013) because the overly complex mapping procedure causes students to become focused on mapping procedures thereby degrading the quality of their representation Prior to the assessment, students were given training to do CMA in advance so that students do not experience confusion and affect the assessment results, such as the statement McClure et al. (1999) that teachers should take the time to train students to use CMA before a real assessment.

Validity of SA

There are two validity tests performed on SA, namely content and construction validity and concurrency. The results of the validity test are shown in Table 2.

Tabel 2. Validity Test Results of SA

No.	Validity of SA	Analyze Technique	Result	Note
1.	Content and Construction	Descriptive Percentage	92.22 %	very valid
2.	ConCurrension	Pearson Corelasion product moment	rxy = 0.774	valid

Based on table 2, expert review indicates SA has excellent content and construction validity criteria and good concurrency validity. The expert states that the SA content is in accordance with the scientific attitude indicator based on references from the American Association for Advancement of Science (AAAS), Harlen (1996) and Gegga (1977) in Anwar (2009). SA is stated to meet the criteria of construction validity because it is prepared with language, letters, good appearance and accompanied by manual work so that students easily understand the instructions in the SA assessment sheet. In addition, SA has been equipped with guidance instructions and scoring guidelines to make it easier for teachers to check points that point to indicators of scientific attitudes and make it easier for teachers to score.

However, several other studies found measurements with SA. There is a tendency for inappropriate expectations in students, that is, good students judge themselves lower, while less competent students tend to judge themselves more highly. Some students increase perceptions about their achievements because they are motivated by personal interests (Ross, 2006; Boud, 1992) that allow self-assessment to be invalid. To answer these concerns, the validity of concurrency test is done by correlating the score of scientific attitude with the score of process skill. The correlation test results show that the scientific attitude score has alignment with the skill score of the process with the correlation coefficient of 0.774. This means that students who perform well assess their scientific attitudes are high, and students who have poor performance judge their scientific attitudes. These results indicate that the SA instrument is valid because it describes the actual learning outcomes.

SA validity is obtained due to several techniques that have been done by researchers, namely 1) reviewing the instrument so that the language and sentence used are easily understood by the students; 2) equipping the SA with guidelines for making it easy for students to understand instructions; 3) organize the instrument in accordance with the lesson

learned; 4) working time of 15 minutes, 5) using a simple rubric. Sealaras with the statement Bolivar-Cruz *et al.* (2015), SA validity can be obtained by: 1) improving student training in self-assessment; 2) improving students' experiences in self-assessment; 3) involve students in designing assessment scales (Boud, 1992), given that this process increases their commitment to the system.

Validity of PA

There are two validity tests performed on PA, namely content and construction validity and concurrency. The results of the validity test are shown in the Table 3.

Tabel 3. Validity Test Results of PA

No.	Validity of PA	Analyze Technique	Result	Note
1.	Content and construction	Descriptive Percentage	95.96 %	very valid
2.	Concurention	Pearson corelasion product moment	rxy= 0.770	valid

Based on table 3, the expert review shows the content validity and construction of the PA is excellent and the validity of the concurrency is good. The validity of the contents of the PA is very valid as it has been compiled according to the KI, KD, and process skill indicators ie planning the experiment, experimenting, defining variables, observing, controlling variables, and interpreting data. Practical activities undertaken by students tailored to the department of TPHP, namely innovation making tapai of varied raw materials. Construction validity is obtained because PA has lattice, scribe rubric, and practicum worksheet.

Although Assessment Performance has good content and construction validity, there is still doubt that PA can measure student learning outcomes validly and reliably, especially involving teachers in scoring. The main source of variation is from the appraiser. Therefore, concurrency validity test is done by correlating PA score with skill score from other valuation techniques such as semester final value and cumulative achievement index (Wei *et al.*, 2012). The researcher correlates the score of process skill with psychomotor value on KD 4.5 and obtained valid result with validity coefficient 0,770. It can be concluded that PA can measure process skill, because it has result of measurement parallel with other assessment technique and relative according to students psychomotor learning result.

The validity of PA concurrency is supported by the use of expert judgmentalized rubrics. In line with Jonsson & Svingby's (2007) assertion, the validity of performance appraisals can be enhanced by the use of rubrics, especially if it is analytic, topic specific, and previously performed on rater training. Rubrics have the potential to improve learning and improve teaching, because rubrics create explicit expectations and criteria, which also facilitate feedback and self-assessment.

2. Reliability

Reliability test is done to know the consistency of assessment. Reliability test results are shown in Table 4.

Tabel 4. Reliability of Authentic Assessment

No.	Kinds of Instruments	Coeficient Reliability	Note
1.	CMA	0.975	Very Reliable
2.	PA	0.882	Very Reliable
3.	SA	0.954	Reliable

Based on table 4, CMA, SA, and PA instruments have been produced which have very reliable criteria with reliability coefficients in the range $0.80 < r_{xy} \leq 1{,}00$. This suggests the

assessment instrument provides consistent measurement results.

Reliability of CMA

CMA reliability can be obtained because it has done several things in order to produce a concept map that reflects the students' knowledge: 1) develop a simple CMA, because complex concept maps will degrade the quality of student knowledge representation; 2) use relational valuation method with the help of a simple karate master map, assessment is easy and consistent scores are obtained; 3) involves three rats to improve consistency of the assessment rubric; 4) conduct training to enable students to use CMA.

This result is consistent with McClure *et al.* (1999), that the sources of error in CMA include: 1) variations in student concept mapping, 2) rater variation in assessing CMA, and 3) consistency by which concept maps are evaluated. This third factor relies heavily on the choice of inappropriate assessment methods. The first two factors are assumed to contribute little to the CMA score variation. The CMA fill in the map technique also contributes to the reliability of CMA because the free CMA model has a weakness in terms of scoring (Wibawati & Supahar, 2016). The CMA model fill in the map improves the reliability because the lighter the rater task in scoring.

Reliability of SA

The result of internal consistency reliability test of SA show very high reliability (coefficient raliabilitas 0,954). Reliability SA is not only determined by the characteristics of the assessment, but also by the situation and condition of the students while working (Hartanto *et al.*, 2014). Reliability of SA can be achieved because the researcher gives prior training to the students, using a relatively short time and in accordance with the students' self-assessment needs, using the language easily understood through expert reviews, using examples of events appropriate to the learning, and using simple rubrics. Teachers also provide understanding to students to assess themselves with honest and according to each condition.

In line with Ross (2006), assessment reliability can be gained by easy-to-understand rubrics, discussing familiar student competencies, including perceived performance, and using simple sections. Mistar (2006) found that providing training sessions to students will improve reliability as the quality of self-assessment increases with length of training, however, further research on psychological factors such as motivation, personality, confidence and anxiety is required. Bolivar-Cruz *et al.* (2015) suggests that teachers may warn students about the possibility of applying correction if the results of the self-assessment are significantly different from the reference values. SA reliability can also be enhanced by providing an understanding that students' self-assessment is formative (helping to improve their performance), comparing PA with peer-to-peer assessment and teachers. This is because SA without feedback from tutors can not help students to realize all their mistakes (Taras, 2003).

Reliability of PA

Reliability test of PA performed with inter-rater reliability technique obtained reliability coefficient of 0.882 which means very reliable. Reliability of the assessment instrument is influenced by several factors, namely the workmanship of the task, the atmosphere of the test, and the objectivity of the appraiser. The workmanship test guidance on the PA is easy to understand so as to minimize the score difference between assessors. Conducive atmosphere during the execution of the test will support the smooth implementation of performance tests, so that the assessment results obtained maximized. Reliability PA can be obtained by the use of assessment rubrics that have been validated by experts. Rater is given training in advance on the scoring procedure. Furthermore, three rater independently assessed by observing the students in each study group. Researchers compiled PA with the design of the work steps are relatively the same, but still

accommodate the creativity of students in preparing the experimental design.

The main threat to the reliability of performance assessments is the lack of consistency of raters (Shavelson *et al.*, 1999). Most studies investigating intra-rater reliability using Cronbach alpha to estimate the consistency of raters are considered sufficient. Inter-rater reliability may not be a major concern when assessors are supported by rubrics. Rubric helps assessors to achieve high internal consistency when scoring student performance (Jonsson, 2007). Assisted with the assessment rubric, will increase the objectivity of assessment by rater (Nuraida and Hakim, 2017). Rubric is prepared using Indonesian language is good and true so as not to cause double meaning. The rubric is structured by formulating the best indicator criteria for the maximum scale that can be the number, percentage, accuracy of identification, or work steps achieved by students. If there is 1 indicator criterion that is not met, then the scale scale is reduced, and so on (Khasanah, 2015).

Some factors that may influence an assessor's judgments include experience differences, lack of agreement on the assessment routine, and teacher attitudes about ethnicity and pleasure towards students also influence student judgment. Therefore, in this study rater has been given training before scoring. In addition, the work steps of students are relatively similar, as Wei *et al.* (2012) states, when all students perform the same task or test, and the assessment procedure has been well defined, high reliability will be obtained. However, when students perform different tasks, students choose their own topics or produce unique products, then the reliability can be estimated relatively low.

3. Effectiveness

Effectiveness is defined as something that has the effect, can bring results, and appropriate. The effectiveness test results are shown in Table 5.

Table 5 is a recapitulation of the effectiveness questionnaire showing that CMA, SA, and PA for TPHP students have effectively implemented objectives and assessment functions, as well as providing appropriate learning outcome.

Tabel 5. The Effectiveness of Authentic Assessment

No.	Evaluation Aspects	Presentase (%)	Keterangan
1.	The Objective of Assessment	88.33	Very Effective
2.	The Function of Assessment	89.58	Very Effective
3.	Measurement of Cognitive Aspects	87.50	Very Effective
4.	Measurement of Psychomotor Aspect	95.83	Very Effective
5.	Measurement of Affective Aspects	87.50	Very Effective
Mean		89.75	Very Effective

Effectiveness of CMA

The effectiveness of CMA is shown by 3 aspects. *First*, CMA is able to provide information on student achievement. 58,8% of students have understood the role of fungi reproduction, 88,2% of students have understood the role of yeast, and 67,6% of students have understood the role of fungi. *Second*, the CMA is able to perform the assessment function, which is to provide students complete learning information. A total of 79.4% of students attain the minimum completeness criteria. CMA can be used to identify the weakness of students, that has been identified students' weaknesses are in the concept of reproduction of yeast and mold. Teachers can place students on specific groups based on their weaknesses and strengths. CMA provides information on the success of learning, which is obtained information of students successfully construct the concept of the role of

yeast, while in three other themes, less than 75% of students who have successfully constructed his knowledge. *Third*, CMA can measure students' cognitive with valid and reliable. The effectiveness of CMA measuring cognitive structures and identifying student misconceptions has been demonstrated in other studies indicated by the correlation of CMA scores with scores with other assessment techniques (Rusilowati & Sopyan, 2014).

The effectiveness of assessment tools is not always shown by good learning outcomes, as improvements in learning outcomes are determined not only by assessment instruments, but also the quality of learning and psychological conditions of students. Increased effectiveness of CMA has been done by training students using CMA prior to assessment. According to McClure *et al.* (1999), taking the time to train students working on CMA and designing simple CMA will improve its effectiveness, because complex CMAs tend to take a lot of time to produce concept maps that accurately represent students' knowledge. Sedagkan according to Jennings (2012), the effectiveness of CMA is indicated by the achievement of the main objective of the assessment, which is to motivate students.

Effectiveness of SA

The effectiveness of SA is shown by 3 aspects. *First*, provide information on student achievement in learning. The result of scientific attitude measurement using SA shows 4 students have very good scientific attitude, 27 have good scientific attitude, and 3 have scientific attitude well enough. Attitudes and thinking habits can not be judged directly because they are implicit in students; however, self-assessment is intended to inform students about the emphasis of learning programs and to help students focus on important aspects of learning. If teachers only assess the most easily measured elements, students can only focus on them. *Second*, provide information about student's weaknesses. The scientific attitude that TPHP students must improve is honest and objective. The effectiveness of SA as a tool of self evaluation according to Ross (2000) is to help students who do not have a realistic picture of his strengths and weaknesses. By implementing SA, teachers can determine appropriate learning targets and strategies to address the student's weaknesses. Thirdly, SA is effective in measuring students' scientific attitude with valid and reliable indicated by the correlation of scientific attitude with student process skill. This means that students tend to assess scientific attitudes according to self-achievement and does not occur expectations that are not appropriate in students. The effectiveness of SA as a tool for the assessment of scientific attitudes can be enhanced by training (simulation) before students judge themselves.

Other studies have shown that SA, contributing to higher self-efficacy, greater intrinsic motivation, and stronger attainment. Students will apply known criteria and expectations in science learning and reflect on self-improvement to achieve a defined learning outcome. Student participation in setting self-assessment criteria that refers to expectations that the teacher wants will help students see themselves as scientists and problem solvers (Ross, 2006).

Effectiveness of PA

PA effectively achieving the objectives of the assessment is shown by its ability: 1) providing information on the achievement of competence, ie 32 students from 34 students have good process skills, while 2 students have good enough process skills, 2) teachers can provide feedback directly through experimental activities, 3) teachers can monitor the progress of learning by analyzing student achievement on each indicator of process skills, 4) assessment results can be used as feedback to improve learning methods and activities in accordance with the needs of students.

PA also effectively perform the assessment function is shown by its ability: 1) to provide information of students who achieve learning mastery, ie 94.11% of students have reached the minimum completeness criteria, 2) provide information about student weakness,

that has been identified students' weaknesses are in the ability to control variables , so that the product produced from the experiment does not have the desired quality, 3) the student's weakness becomes the basis for the teacher to improve the work steps undertaken, 4) the teacher can place the students in groups based on weaknesses and advantages, obtained information that all students have successfully planning the experiment as well as identifying and formulating the experimental hypothesis. According to McTighe & Ferrara (1994), the effectiveness of PA is demonstrated by its ability to provide feedback for teachers and students to improve learning. The effectiveness of alternative assessments in Manitoba Education Training is to focus on what students learn and can do, provide opportunities for success for each student, and demonstrate students' knowledge, skills and attitudes, and progress, rather than simply identifying the shortcomings of learning. According to McTighe & Ferrera (1994), with PA teachers can directly observe the application of desired skills and knowledge, because PA is one of the most authentic types of student judgment because it can mimic the type of actual performance.

PA is effective in measuring students' science process skills with valid and reliable measurement results. The results of valid and reliable measurements due to PA have been supplemented with assessment rubric. Rater uses a scoring rubric 1-4 on the basis of student performance levels. Using rubrics improves inter-rater reliability. The effectiveness of the PA measuring the skills of the science process is also supported by the test results which indicate that there is a correlation of the students' PA scores with other assessment techniques. This result is in line with McTighe & Ferrera's (1994) assertion that PA is effective for assessing student products and performance reliably.

The effectiveness of classroom assessment refers to three basic principles; 1) promoting learning, 2) using many sources of information, and 3) providing fair, valid, and reliable information. While the effectiveness of assessment is based on Science Manitoba Education and Training, 1) congruent and integral to instruction, 2) based on authentic task and science learning process, 3) based on criteria known and understood by students, 4) tailored to student's superiority, 5) a collaborative process, 6) focuses on what students can learn and do (McTighe & Ferrera, 1994). Approach to process skills can foster scientific attitudes to develop fundamental skills, so that in the learning of students can understand the concept he learned (Wiyanto, 2009).

4. Practicality of Authentic Assessment

Assessment is practical when it is time efficient, easy to understand, clear assessment criteria, easy to evaluate results, supported by facilities, and economical. The questionnaire results of the practicality of the assessment instruments are shown in Table 6.

Tabel 6. Questionnaire Results of Assessment Practicality

No.	Practicality Criteria	Percentage (%)	Note
1.	Compatibility with time allocation	79.62	Practical
2.	Ease of Use	84.80	Very Practical
3.	Assessment criteria	85.74	Very Practical
4.	Inspection Time of Assessment	77.77	Practical
5.	Compliance with school facilities	81.13	Very Practical
6.	Economic Aspects of Assessment	88.36	Very Practical
Mean		82.90	Very Practical

Based on table 6, authentic assessment has a very high practicality. This means that assessment instruments developed in accordance with time allocation, easily understood by the students, have clear assessment criteria, easy

evaluating results, supported by school facilities, and are economically viable.

The Practicality of CMA

The practicality of CMA is demonstrated by several aspects. First, the average time to work on a CMA theme is 13.5 minutes, so the time to work on 4 CMA themes is approximately 54 minutes, with the estimated time CMA can be used with 2 hours of lesson (2 x 45 minutes). Estimated time for the CMA is comparable to traditional assessment techniques. Second, CMA is easy to use in class, based on students' questionnaire responses, from 34 children, 5 states very easily, 28 states easy, and 1 states less easy to use CMA in class. Three observers said CMA is easy to use as a cognitive assessment tool. Unstructured interviews with students of grade X TPHP obtained information that the results are less satisfactory not because they do not understand how to do, but because students have not prepared for assessment and not yet understand the subject matter. Based on that information, CMA is easy to do, but students need training and habituation using CMA to be trained to construct knowledge and gain meaningful learning that is not just memorizing

Third, based on the questionnaire of the observer response, the concept map assessment with correlational technique is easy to do with the help of master map. Scoring is easier with the help of a simple scoring rubric. Of 34 students, 12 students stated very understanding, 17 students understood, while 5 students expressed less understanding of CMA assessment criteria. This finding is consistent with the research of McClure et al. (1999) that the CMA takes 1-5 minutes, and the easiest to use is the correlational method with the help of the master map. This method can be performed by the assessor with little or no training.

Fourth, the teacher does not take long to assess the concept map. The average time required to assess the 34 sets of student CMA, which comprises 4 CMA themes each set of approximately 3.5 hours. Each 1 CMA theme takes approximately 1.5 minutes to be assessed. According to McClure et al. (1999) the critical consideration of CMA implementation is the time it takes teachers to evaluate and score a concept map. The findings of McClure et al. (1999) indicates the time required to print a concept map ranging from 1 to 5 minutes, depending on the chosen rating method. With such time allocations, assessment with concept maps tends to be comparable to objective assessments such as quizzes or short answer tests.

Fifth, school facilities strongly support the application of CMA due to simple equipment. The availability of facilities at the school will affect the smoothness of the appraisal process, the comprehensive assessment tools will have an impact on the interest of teachers and students in carrying out the assessment. Simple equipment allows this assessment to be followed by all students. Sixth, the CMA is considered very economical, based on the results of the questionnaire analysis, 17 students responded to the assessment with the concept map is very economical, 16 responded economically, and 1 student responded less economically. According to McClure et al. (1999) should optimize the balance between the cost and benefits of the concept map as an assessment. Based on the researcher's experience, the costs used to assess the concept map tend to be comparable to assessments with multiple choice questions, essays, and other traditional assessment techniques.

According to empirical experience, the practicality of CMA relates to three things: 1) the time to train students in concept map thinking techniques to obtain reliable results; 2) the time required for the implementation and evaluation of concept map assessments that are comparable to traditional judgments; 3) cost required implementation of concept map assessment. In the research there are 6 aspects of CMA service, ie time for execution, ease of use, ease of evaluating assessment result, time to evaluate assessment result, school facility support capability, and cost needed (McClure et al., 1999).

Practicality of SA

The practicality of SA in this study was analyzed based on 6 aspects: 1) time for implementation, 2) ease of use, 3) ease of evaluating assessment results, 4) time to evaluate assessment results, 5) cost required. First, based on table 4, SA has practicality in terms of time for assessment. Time spent by all students to complete a self-assessment of approximately 15 minutes. Second, SA is easy to use in the classroom, based on teacher and student questionnaire results in table 4. The SA can be used by students with little prior training and is equipped with simple psychological scoring guidelines. Students work by understanding 15 positive and negative statements, then giving checks on the psychological scale according to the perceived state of learning. This assessment of scientific attitudes was also undertaken by Aydeniz and Kotowski (2014) with 12 total points of statement consisting of six opposite positive and negative statements designed to reflect intellectual attitudes and emotional attitudes. Third, based on the analysis of the questionnaire results, the three observers stated that scientific attitude evaluation using SA instrument is easy to do because the assessment criteria is very clear. The observer gives a score on each item of the statement using the rating guideline. To facilitate the observer in scoring, the researcher displays positive statements on odd and negative numbers on even numbers and provides a different color background on positive statements with negatives.

Fourth, the observer does not take long to give SA score. The total time required for scoring for 34 students is approximately 45 minutes. SA tend to be comparable to assessments with traditional assessment techniques. Fifth, the school facility is very possible to implement SA because the required equipment is easy to prepare, ie paper and complete stationery. Observers and 33 of 34 students said it was very easy to do because the school facilities support the implementation of this SA. The availability of facilities at the school will affect the smoothness of the appraisal process, the comprehensive assessment tools will have an impact on the interest of teachers and students in carrying out the assessment. Simple equipment allows this assessment to be followed by all students. Sixth, based on the analysis of the questionnaire results, SA declared very economical by observers and all students. Costs used for assessing attitudes with SA tend to be comparable to assessments with multiple choice questions, essays, and other traditional assessment techniques.

Practicality of PA

PA developed has 6 criteria of practicality. First, efficient against time. A total of 32 out of 34 students stated the time to implement PA accordingly, meaning that the allocation of 2 times the lesson of the performance appraisal process can be completed. Estimated time for the PA is also proportional to the estimated time required for traditional classroom assessment techniques. Second, easy to use. Based on the questionnaire analysis of student responses, 33 out of 34 students stated that PA is easy to use as a classroom skill assessment tool, while the three observers said it is very easy to use this assessment instrument. Third, easy in scoring. Based on the questionnaire of the observer response, performance appraisal with relational technique is easy to do with the help of the assessment rubric. In harmony with the results of Ruiz-Primo & Shavelson (1996a) research, the PA can be managed efficiently in the assignment of science. Teachers who choose to use activity-based curriculum in the classroom can perform performance appraisals that are combined with scientific inquiry without difficulty.

Fourth, it is easy to evaluate the results of the assessment. The questionnaire analysis shows that teachers do not take long to assess student performance with 6 indicators of process skills. This type of holistic assessment rubric is used as an observer reference to score students. The rubric is prepared by presenting some statements depicting the level of students' ability ranging from the bad to the extraordinary. This type of holistic rubric is usually used for large-scale assessment because it is assumed to be

easy, cheap and accurate (Johnsson & Svingby, 2007), as well as easy and fast to learn and practice. With a good scoring system, score-training sessions can be done for 45 minutes, and after practicing scoring approximately 10 student performance, scoring can be quick and easy (Ruiz-Primo & Shavelson, 1996a). These findings support that PAs have time efficiency criteria.

Fifth, supported by the school facilities. Based on the questionnaire results, students and teachers said it is easy to implement PA because it is supported by school facilities and equipment and materials easily prepared. This convenience is also due to the experiments that students do closely with real examples in everyday life. The availability of facilities at the school will affect the smoothness of the appraisal process, the complex tools will have an impact on the interest of teachers and students in conducting the assessment. Simple equipment allows this assessment to be followed by all students. Sixth, economical, based on the results of questionnaires, students and teachers assess the PA has met the economic criteria. Based on the researcher's experience, the costs used to assess the concept map tend to be comparable to assessments with multiple choice questions, essays, and other traditional assessment techniques.

According to Shavelson & Ruiz Primo (1996a), PA practicality criteria relate to: 1) the effectiveness of time and cost, 2) the time to prepare, manage, and mebereskan assessment? 3) time to assess each student. In this research, there are six things that can describe the nature of PA as a tool for the assessment of process skills, ie time for implementation, ease of use, ease of evaluating assessment results, time to evaluate assessment results, school facility support and cost.

CONCLUSION

Based on the result of the research, the conclusion is obtained that there were authentic biological assessment instruments in the form of CMA, SA, and PA to measure cognitive, scientific attitude, and TPHP students' valid, reliable, effective, and practical TPHP process. For research and development on the same topic, the researcher gives some suggestions for the development and implementation of the assessment in the classroom more effectively, ie 1) the use of CMA needs to be supported by concept-based learning, so that students are accustomed to reading, understanding, and making concept maps 2) students should be included in PA scoring with the objective of knowing the criteria for assessment and motivation to show expected performance; 3) the development of SA to measure scientific attitudes requires comparison of peer assessment so that students are motivated to assess themselves according to actual criteria and avoid expectations which is appropriate to them self.

REFERENCES

Andrade, H. & Valtcheva, A. (2009). Promoting Learning and Achievement Through Self-Assessment. *Educational Theory and Practice Faculty Scholarship*, 12.

Anwar, H. (2009). Penilaian Sikap Ilmiah dalam Pembelajaran Sains. *Jurnal Pelangi Ilmu*, 2 (5): 103-105.

Aydeniz, M. & Kotowski, M.R. (2014). Conceptual and Methodological Issues in the Measurement of Attitudes Towards Science. *Electronic Journal of Science Education*, 18 (3), 1-24.

Bolivar-Cruz, A., Verano-Taroconte, D., & Gonzales-Betancor, S.M. (2015). Is University's Student Self Assessment Acurate? *Springer International Publishing Switzerland,* 21-35.

Boud, D. (1992). The Use of Self-assessment Schedules in Negotiated Learning. *Studies in Higher Education Learning*, 17 (2), 185-200.

Fatmawati, A., Nurhidayati, S., & Gummah, S. (2013). Efektivitas Asesmen Otentik Teknik Saling Silang terhadap Keterampilan Mengajar Mahasiswa. *Jurnal Kependidikan*, 12 (2), 105-113.

Goldsmith, T.E. & Devenport, D.M. (1989). Assessesing Structural Similarity of Graph. Dalam Rusilowati, Pengembangan *Concept Mapping Assessment* untuk Mengukur Kemampuan Mahasiswa Mengkonstruk

Konsep Elektronika. *Jurnal Pendidikan Fisika Indonesia*, (13)16.

Hamid, A. (2008). Pengembangan Sistem Asesmen Otentik Dalam Pembelajaran Fisika Dengan Model Pembelajaran Inovatif di Sekolah Menengah Atas (SMA). *Jurnal Pendidikan Serambi Ilmu*. 35 (42), 1693-4849.

Hartanto, R., Fitrihadijati, H., & Prastiwi, M.S. (2014). Kualitas Instrumen Asesmen Berbasis Kurikulum 2013 untuk Menilai Materi Ekologi SMA. *BioEdu Berkala Ilmiah Pendidikan Biologi*, 3 (3), 528-531.

Jennings, D. (2012). *Concept Maps for Assessment.* UCD Teaching and Learning, UCD, Ireland.

Jonsson, A. & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity, and Educational Concequnce. *Educational Research Review (ELSEVIER)*, 2, 130-144.

Khasanah, A.R.N., Prasetyo, A.P.B., & Susilowati, S.M.E. (2015). Pengembangan Buku Panduan Asesmen Otentik Bagi Guru Biologi SMA/MA. *Unnes Journal of Biology Education,* 4 (2), 179-185.

McClure, J.R. & Bell, P.E. (1990). Effects of an environmental education related STS approach instruction on cognitive structures of pre-service science teachers. University Park, PA: Pensylvania State University (*ERIC Document Reproduction Services No. ED 341 582*)

McClure, J.R., Sonak, B., S., & Suen, H.K. (1999). Concept Map Assessment of Classroom Learning: Reliability, Validity, and Logistical Pacticalitity. *Journal of Reserach in Science Teaching,* (36) 4, 475-492.

McTighe, J. & Ferrera, S. (1994). Performance Based Assessment in the Classroom. *Educational Resource Information Center (ERIC).*

Mistar, J. (2011). A Studi of Validity and Reliability of Self Assessment. *TEFLIN Journal*, 22 (1), 45-58.

Novak, J. D., Gowin, D. B., & Johansen, G. T. (1983). The Use of Concept Mapping and Knowledge Vee Mapping with Junior High School Science Students. *Science Education*, 67 (5), 625-645. [Tersedia di http://onlinelibrary.wiley.com , diakses tanggal 10 Juni 2017 pukul 21.23]

Nuraida, E.F. & Hakim, L. (2017). Pengembangan Penilaian Autentik Aspek Ketermapilan Menggunakan Instrumen Penilaian Kinerja pada Materi Akuntansi Piutang Kelas XI Semester 1 SMKN 2 Tuban. *Jurnal Pendidikan Akuntansi (JPAK) UNESA,* 5 (3), 1-5.

Prasetyo, A.P.B. (2015). Translation of Authentic Aassessment into Biology Teaching Learning Design. *International Conference on Mathematics, Science, and Education 2015* (*ICMSE 2015).* Fakultas MIPA Universitas Negeri Semarang.

Prasetyo, A.P.B., Sukaesih, S., & Hadiyanti, L.N. (2016). Pendampingan Guru Biologi Kota Semarang dalam Mengembangkan Instrumen Penilaian Otentik pada Kurikulum Berbasis Kompetensi. *Unnes Journal of Biology Education*, 5(1), 1167-1175.

Rahmawan, E.F., Sumaryanto, T., & Supriyadi. (2016). Pengembangan Instrumen Penilaian Kinerja Kemampuan Bernyanyi Berbasis Android. *Journal of Education Research and Evaluation,* 5 (1), 81-89.

Ross, J.A. (2006). The Validity, Reliability, and Utility of Self-Assessment. *Practical Assessment Research & Evaluation,* 11,(10).

Ruiz-Primo, M.A., & Shavelson, R.J. (1996a). Rhetoric and Reality in Science Performance Assessments: An Update. *Journal of Research in Science Teaching* , 33(10), 1045-1063.

Ruiz-Primo, M.A. & Shavelson, R.J. (1996b). Problem and Issues in the Use of Concept Maps in Science *Assesment*. *Journal Research in Science Teaching*, (33) 6, 596-600.

Ruiz-Primo, M.A., Shavelson, R.J., & Schultz, S.E. (1997). On the Validity of Concept Map Based-Assessment Interpretation: An Experiment Testing The Asumption of Hierarchical Concept Map in Science. *National Center for Research n Evaluation Satandart, and Student Testing (CRESST).*

Rusilowati, A. & Sopyan, A. (2011). Pengembangan Concept-Mapping Assessment untuk Mengukur Kemampuan Mahasiswa Mengkonstruk Konsep Elektronika. *Jurnal Pendidikan Fisika Indonesia*, 7, 13-16.

Sari, E.N., Rosyidatun, E.S., & Juanengsih, N. (2015). Profil Penilaian Otentik pada Konsep Biologi di SMA Negeri Kota Tangerang Selatan. *Jurnal Penelitian dan Pembelajarn IPA*, 1 (1), 26-41.

Shavelson, R.J., Ruiz-Primo, M.A., & Wiley, E.W. (1999). Notes on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement,* 36 (1), 61-71.

Solikha, D.F. (2015). Bahan Ajar Asam-Basa Menggunakan Konteks Bahan Pengawet Makanan untuk Mengembangkan Literasi Sains SMK Jurusan Teknologi Pengolahan

Hasil Pertanian (TPHP). *Jurnal Penelitian Pendidikan (JPP)*, 15 (2), 59-66.

Sujati, H. (2014). Pengembangan Instrumen Penilaian Otentik dan Validasinya. *Proceeding Seminar Nasional Psikologi UMS: Psykometry Edition*, 47-52.

Taib, E.N. (2013). Respon Siswa terhadap Perangkat Concept Mapping Assessment (CMA) Bervisi SETS. *Jurnal Biotik, 1* (1), 1-66.

Taras, M. (2003). To Feedback or Not to Feedback in Student Self-assessment. *Assessment & Evaluation in Higher Education, School of Education and Lifelong Learnin*g, (28) 5.

Wei, R.C., Cor, K., Arshan, N., & Raymond, P. (2012). *Reliability dan Validity of Perfomance-Based Assessment: Can Performance-Based Assessments be Reliable and Valid? Findings From a State Pilot.* Stanford University.

Wibawati, N. & Supahar. (2016) Pengembangan Instrumen Penilaian Berupa Peta Konsep Fisika untuk Mengukur Struktur Kognitif Peserta Didik Kelas X SMA. *Jurnal Pendidikan Fisika*, 5 (4), 199-208.

Wiggins, G. (1990). The Case for Authentic Assessment. *ERIC Clearinghouse on Test Measurement and Evaluation Washington DC*. American Institute for Research Washington DC.

Wiyanto, S & Marwoto, P. (2009) Pembelajaran dengan Pendekatan Keterampilan Proses Sains untuk Meningkatkan Penguasaan Konsep Suhu dan Pemuaian. *Jurnal Pendidikan Fisika Indonesia*, 5, 42-46.