Development of self and peer performance assessment on iodometric titration experiment

Nahadi*, W Siswaningsih and H Kusumaningtyas
Departemen Pendidikan Kimia, Universitas Pendidikan Indonesia, Indonesia

*Corresponding author’s e-mail: nahadi@upi.edu

Abstract. This study aims to describe the process in developing of reliable and valid assessment to measure students’ performance on iodometric titration and the effect of the self and peer assessment on students’ performance. The self and peer-instrument provides valuable feedback for the student performance improvement. The developed assessment contains rubric and task for facilitating self and peer assessment. The participants are 24 students at the second-grade student in certain vocational high school in Bandung. The participants divided into two groups. The first 12 students involved in the validity test of the developed assessment, while the remain 12 students participated for the reliability test. The content validity was evaluated based on the judgment experts. Test result of content validity based on judgment expert show that the developed performance assessment instrument categorized as valid on each task with the realibility classified as very good. Analysis of the impact of the self and peer assessment implementation showed that the peer instrument supported the self assessment.

1. Introduction

The scope of the assessment as stated in the educational assessment standard is to include aspects of attitude, knowledge, and skills. Assessment is basically a series of teacher activities in making decisions about the achievement of competencies based on student learning outcomes during the learning process. Assessment in learning can be done through laboratory activities [1]. The number of practicum activities that must be done in the learning of chemistry, then there are many competence psychomotor skills of students who must be assessed by the teacher [2]. Thus, teachers need an assessment to assess it [2]. Performance assessment are important for teachers because all contributions from each individual must be appreciated, especially when the student performs activities outside the classroom or practicum activities within the laboratory [3].

Vocational High School (SMK) is one of the formal education institutions that provide vocational education at secondary education level. SMK is required to produce graduates ready to use in the world of work. One of the competencies that SMK students must achieve, as stated in the decision of the Directorate General of Primary and Secondary Education, in the field of chemistry analysis is performing performance under guidance with quality and quantity that is measured in accordance with work competency standards.

Based on the field survey that the researchers found, the assessment of student performance during practicum activities in some schools, especially in vocational school chemistry analysis, is always done by teachers related to each activity of each student during the practicum. In addition, from the results of interviews of vocational chemistry teachers, it can be seen that in general, in addition to the...
final exam of school, the assessment of student performance during the practicum activities of chemistry in the school more obtained from the results of practical reports that students make after practicum and teacher questioning activities with students about the practicum performed individually. This is a weakness of performance assessment. Student performance assessment methods using written reports during or after students performing practicum only illustrate less information about student performance, even more likely to describe students' knowledge and understanding related to the principles and usefulness of each practicum technique used [4]. There is an alternative form of performance assessment, namely self-assessment, in which students are involved in assessing their own learning outcomes or performance. Students also need to be assessors in practicum activities [5]. Self-assessment can build a tendency to always check every job naturally and make it easier for someone to know for themselves what to learn, so that the person's motivation to learn more will increase [6]. However self-assessment will not increase student achievement to a certain standard if it is not accompanied by an assessment by the teacher [5]. If the teacher allows students to compare their judgment with the teacher's judgment, then the student can know the teacher's thinking and improve the students' skills in the assessment, which, therefore, will tend to be motivated to learn and think better [7].

Another alternative form of assessment is peer assessment, in which students are involved in providing feedback on student learning outcomes or other performance. Peer assessment requires students to give either input or grade, or both, based on the criteria of excellence for the product or work of other students [8]. In a lesson, self-assessment will be better when used in conjunction with peer assessment. The use of self and peer assessment on learning in the laboratory allows for better reciprocity for students and improves teachers' confidence in assessing student performance [3]. Based on the author's analysis of previous research, it was revealed that the use of self and peer assessment can be applied in the assessment of SMK student's performance on acid-base titration practicum, but the performance assessment instrument used cannot be said to be valid and reliable.

Not only acid-base titration, iodometric titration also belongs to a conventional volumetric analysis type. Carry out conventional volumetric analysis included in the Indonesian National Working Standards that serve as a reference of competence of content and basic competencies that SMK students must achieve in chemical analysis field. Based on this, the researcher want to develop performance assessment instrument that can be used by SMK students to assess their own performance or other student performance on iodometric titration practicum. This study aims to produce performance assessment instruments in the form of tasks and rubrics with good quality based on the validity and reliability that can be used then by SMK students to assess their own performance or the performance of other students on iodometric titration practicum.

2. Methods
This study uses a method that refers to the development and validation methods developed by Adams and Wieman [9]. The development performance instrument refers to the step of developing performance assessment according to Harsh [10]. Participants in this study are second grade chemistry analysis field students in one of SMK in Bandung as many as 12 people for reliability test and as many as 12 people for the test stage.

The instrument used in this research is validation sheet and questionnaire. The validation sheet is used to determine the validity of the contents of a performance assessment instrument developed based on expert judgment. Questionnaires are used to determine students' responses to the implementation of self and peer assessment to performance on iodometric titration practicum using performance assessment instruments developed.

This research was conducted through four major stages, namely planning stage, development stage, validation phase, and test stages. In the planning stage, field survey, literature review, and core competencies (KI) and basic competence (KD) are conducted to determine the conformity of the main subject used as the focus of the study with the applicable SMK curriculum. The field survey was conducted to find out the students' performance assessment by SMK teachers on the iodometric
titration practicum through interviews, while the literature review was conducted to find out the concepts or theories which then became the basis for determining the aspects of performance that students need to master when the performance assessment takes place.

At the stage of development, the preparation of performance assessment instruments in the form of tasks and rubrics based on aspects of performance that have been determined. At the validation stage, items in the form of tasks and rubrics are then tested for content validity based on expert judgment and tested for reliability based on Cronbach’s alpha calculations [11]. Validation results are then processed through the calculation of content of validity ratio (CVR) and if the CVR value > 0.736 indicates that items that consist of tasks and rubrics meet valid criteria when validation is performed by 5 experts[12,13]. This is known through the calculation of Pearson's correlation value using SPSS software. Interpretation of the implementation of developed performance assessment instruments based on Pearson's correlation value [14]. A valid and reliable performance assessment instrument is then used in the test stage. The test results are used to determine the implementation of self and peer assessment conducted by students using the developed performance assessment instruments.

3. Results and Discussion
The development of the performance assessment task on the iodometric titration practicum refers to the task development step according to Nitko and Brookhart [15]. The development of the performance assessment rubric on the iodometric titration practicum is based on the Wolf and Stevens rubric development steps that consist of the identification of performance criteria, determining the rating scale (score 1-4), and making a description of each assessment scale [16].

3.1. Content validity
Content validation is performed by expert judgment in the field of assessment instruments related to the suitability of tasks with rubrics, as well as suggestions that may take the form of writing considerations as well as the use of words or sentences in the developed performance assessment instrument. In this study, the experts in this study are three chemistry lecturers and two chemistry teachers. The entire test item that consist of tasks and rubrics is valid after an improvement based on the advice of the experts, so that the validated performance assessment instrument can be tested for its reliability [11].

3.2. Reliability of performance assessment instruments
The reliability of a performance assessment instrument indicates the extent to which the instrument provides a truly reliable picture of a person’s performance. Cronbach’s alpha is obtained from three observations conducted on a small scale, in which observations are made by three observers. The first observations were made to two students who were then assessed by three of the same observers, while the second observation was made to four students, and the third observation was made to six students.

![Figure 1. Recapitulation of Cronbach’s alpha value on performance assessment instrument reliability test of iodometric titration practicum.](image-url)
Figure 1 shows that the reliability of the developed performance assessment instrument against 2 students by 3 observers is very good. This is seen from the value of Cronbach’s alpha on each task at 1.00. Similarly, the reliability of developed assessment instruments against 4 students, overall can be categorized very good, although there is a decrease in the value of Cronbach’s alpha on task number 10 that is equal to 0.789. Based on Cronbach’s alpha value interpretation of instrument reliability, the value is still categorized as good [11]. It shows that the consistency of the observers who do the assessment on the performance of students, especially against 2 students and 4 students, during the practicum is very good. It is also supported by conducive circumstances when observing 2 students so that the observers can concentrate well during the assessment of student performance.

Cronbach’s alpha value decrease in task number 10 on instrument reliability test that was developed to 4 students can be caused by the increasing number of students assessed influence the concentration of the observer to assess each student, so that also influences the consistency of the observer’s judgment become decreased. As for the observation of 6 students, there is a significant decrease in reliability when compared with the observations of 2 and 4 students, especially in tasks 3, 5 and 7 respectively of 0.354; 0.454; and 0.403. All three of Cronbach’s alpha values fall into unacceptable categories. Based on interviews by observers, this may be due to the more students being assessed, the more irregular the laboratory situation. The performance of each student at the same time is increasingly diverse, making it difficult for observers to focus on the different performance aspects at the same time [13].

On the other hand, the Cronbach’s alpha value in task 10 against 6 students is greater than the Cronbach’s alpha value of task 10 against 4 students. This increase in reliability is in contrast to other reliability findings. The observer’s response after performing a performance assessment on 6 students that there were 2 observers who did not have time to observe a particular performance on a student, so that the student is considered not doing a performance that needs to be assessed. But in fact, there is other observer who observe that the student performs the performance as it should be. Therefore, it can be said that the consistency of performance assessment of 6 students by observer using performance developed assessment instrument include as less category [11,13].

3.3. Implementation of self and peer assessment using developed instruments

Implementation of self and peer assessment using the developed instrument can be known by looking at relation between the results of the assessment performed. Measurement of the implementation of the assessment using the developed instrument is done by comparing the results of self-assessment conducted by the students with the results of the observer’s assessment, and compare the results of the student’s assessment of his colleagues with the results of the observer’s assessment.

| Table 1. Correlation between rater (student) and observer on self-assessment. |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Student Code | Observer (O) vs Student (A) | Pearson’s Correlation | Category |
| A1 | O1 vs A1 | 0.628 | Strong |
| A2 | O1 vs A2 | 0.291 | Enough |
| A3 | O2 vs A3 | 0.254 | Enough |
| A4 | O2 vs A4 | 0.106 | Very weak |
| A5 | O3 vs A5 | 0.177 | Very weak |
| A6 | O3 vs A6 | 0.594 | Strong |
| Average | | 0.341 | Enough |

In table 1 it can be seen that the correlation between students with the observer on the results of self-assessment is quite strong. That is, it can be said that the implementation of self-assessment conducted by students on iodometric titration practicum using performance assessment instrument developed quite well. However there are two low Pearson correlation values between the students and
the observer so that the correlation is categorized as very weak. It indicates that there are students who have not been adequately able to carry out self-assessment using the developed assessment instrument. Students still need practice and motivation to develop their ability in self-assessment. In addition, based on student responses, there are students who get difficult when doing self-assessment because they feel not sure related to the quality of their own performance during the practicum whether it is in accordance with what should it be. Therefore, self-assessment using developed assessment instruments would be better if accompanied by teacher assessment results, as has been suggested that self-assessment would not improve student achievement to a certain standard if not accompanied by a teacher’s assessment [7,14].

Table 2. Correlation between rater (student) and observer on peer-assessment.

| Student code | Observer (O) vs Student (A) | Pearson’s Correlation | Category |
|--------------|-----------------------------|------------------------|----------|
| A1           | O1 vs B1                    | 1.000                  | Perfect  |
| A2           | O1 vs B2                    | 0.525                  | Strong   |
| A3           | O2 vs B3                    | 0.960                  | Very strong |
| A4           | O2 vs B4                    | 0.905                  | Very strong |
| A5           | O3 vs B5                    | 0.953                  | Very strong |
| A6           | O3 vs B6                    | 0.896                  | Very strong |
| Average      |                             | 0.873                  | Very strong |

In table 2, it can be seen that the overall correlation between students with observers on peer assessment results is very strong. That is, the implementation of peer assessment that conducted by students on iodometric titration practicum using the developed performance assessment instrument is very good. This suggests that students are able to involve themselves in assessing their peer performance using developed instruments, especially on iodometric titration practicum. With high Pearson correlation value, it can be said that the results of the assessment of students made to their friends almost the same as the results of assessments by teachers or observers. Therefore, it can be said that students are able to conduct peer performance assessment on iodometric titration practicum using developed assessment instruments so well and objective [14].

![Figure 2](image)

Figure 2. Comparison of Pearson correlation of self-assessment and peer assessment.

Based on figure 2, it can be seen that peer assessment is better than the self-assessment done by the students using the developed assessment instrument. Therefore, peer assessment results can be used to support the results of self-assessment performed by students. As the assertion that self-assessment will not improve student achievement to a certain standard if it is not accompanied by an assessment by the teacher [6], a reciprocal of students’ raters for each student judged already representing the reciprocity from teacher’s.
4. Conclusion
The developed performance assessment instrument meets valid and reliable terms against 2 and 4 students with a total of three observers. This is known based on the CVR and Cronbach’s alpha values obtained meet the minimum of specific criteria. In addition, the reliability test of the developed assessment instrument in this study was conducted on a limited scale in which the assessment was performed by three observers. The performance assessment instrument developed can also be used by the vocational school students in performing self and peer assessment, especially on iodometric titration practicum.

5. References
[1] Doran R L 2002 Science educator’s guide to laboratory assessment (Arlington: NSTA press)
[2] Ok triawan T 2015 J. Pendidikan dan Pembelajaran Kimia 4 2 593-604
[3] Wenzel T J 2007 Evaluation tools to guide students’ peer-assessment and self-assessment in group activities for the lab and classroom Journal of chemical education 84 1 182
[4] Hofstein A 2004 The laboratory in chemistry education: Thirty years of experience with developments, implementation, and research Chemistry education research and practice, 5 3 247-264
[5] Boud D and Falchikov N 2006 Assess. and Ev Higher Education 31 4 399-413
[6] Spiller D 2012 Assessment matters: Self-assessment and peer assessment The University of Waikato
[7] Elkoumy A S A K 2010 Student self-assess. in higher ed. alone or plus CPLA Conference 71
[8] Falchikov N 2001 Learning together (London and New York: Routledge Falmer)
[9] Adams W K and Wieman C E 2011 Development and validation of instruments to measure learning of expert- like thinking International Journal of Science Education 33 9 1289-1312
[10] Harsh J A 2016 Designing performance-based measures to assess the scientific thinking skills of chemistry undergraduate researchers Chemistry Education Research and Practice 17 4 808-817
[11] Bhatnagar R, Kim J and Many J E 2014 Candidate surveys on program evaluation: Examining Instrument reliability, validity and program effectiveness American Journal of Educational Research 2 8 683-690
[12] Lawshe C H 1975 A quantitative approach to content validity Personnel psychology 28 4 563-575
[13] Wilson F R, Pan W and Schumsky D A 2012 Recalculation of the critical values for Lawshe’s content validity ratio Measurement and Evaluation in Counseling and Development 45 3 197-210
[14] Sarwono J 2009 Statistik itu mudah: panduan lengkap untuk belajar komputasi statistik menggunakan SPSS 16 (Yogyakarta: Universitas Atma Jaya)
[15] Nitko A J and Brookhart S m 2007 Educational Assessment of Students (New Jersey: Pearson Education Inc)
[16] Wolf K and Stevens E 2007 The role of rubrics in advancing and assessing student learning.