The development of performance-based practical assessment model at civil engineering workshop in state polytechnic

W S Kristinayanti1*, I.G.A.I. Mas Pertiwi1, Evin Yudhi S.1, W D Lokantara2
1 Civil Engineering Department Politeknik Negeri Bali, Jalan Kampus Bukit Jimbaran, Kuta Selatan, Badung 80364, Bali – Indonesia
2 Faculty of Economics & Business Universitas Gajah Mada, Jalan Sosio Humaniora, Bulaksumur 55281, Yogyakarta, Indonesia
Corresponding email: srikristinayanti@pnb.ac.id

Abstract. Assessment is an important element in education that shall oversees students’ competence not only in terms of cognitive aspect, but alsthe students’ psychomotorin a comprehensive way. Civil Engineering Department at Bali State Polytechnic,a as vocational education institution, emphasizes on not only the theoretical foundation of the study, but also the application throughpracticum in workshop-based learning. We are aware of a need for performance-based assessment for these students, which would be essential for the student’s all-round performance in their studies. We try to develop a performance-based practicum assessment model that is needed to assess student's ability in workshop-based learning. This research was conducted in three stages, 1) learning needs analysis, 2) instruments development, and 3) testing of instruments. The study uses rubrics set-up to test students’ competence in the workshop and test the validity. We obtained 34-point valid statement out of 35, and resulted in value of Cronbach's alpha equal to 0.977. In expert test we obtained a value of CVI = 0.75 which means that the drafted assessment is empirically valid within the triall group.

1. Introduction

The practicum appears in many different forms in especially vocational education — as field placement, cooperative education, sandwich programs, internship, clinical placement and in this case part of class assessment. A well designed practicum assessment can make a significant contribution to learning by putting pressure on lecturers to build on the knowledge available and by asking the students to identify and reflect on what they have studied [1]. Ideally, the practicum provides learners with the opportunity to apply their knowledge, test theory and as a consequence modify their understanding. Insights and understandings from practicum may be hard for students to show and with the complex and allowing for exploration of knowledge to take place in nature, simply demand a better way of assessing it [1,2].

Preliminary studies that have been conducted based on the results of questionnaires from students and lecturers in Bali State Polytechnics showed that 74.8% of students and 62.4% of practicum lecturers stated that the aspects of practicum assessment in the civil engineering workshop were incomplete and not satisfactory. The dissatisfaction with the method of assessment performed is basically caused by the instrument used has not fulfilled the requirements of validity, reliability and practicality. Therefore the urgency of better and more reliable assessment is indispensable for assessing learning outcomes, establishing criteria to ensure that evaluations are not erroneous and determine the range of skills that students gained from learning in such nature and inline with the Minister of Education Regulation No 20 Year 2007 [3].
We proposed an introduction of rubrics as an assessment method for civil engineering workshop, and measure the validity through two methodologies. A rubric is a laid out set of criteria used for assessing a specific work or performance and in return provides more details than a single grade or mark [2]. Rubrics, therefore, will help improve subjective assessment into a more objective one. As an assessment tool, rubrics can reveal and tell the students of what is required in a practicum. It even reduces grading time; they increase objectivity and reduce subjectivity; they convey timely feedback to students and they improve students’ ability to include required elements of an assignment [2, 4, 5].

There are two general criteria for rubrics according to Brookhart, which differs in 1) the number of criteria and judgment set (analytic or holistic), and 2) the type of process to be assessed (general or task-specific) [6]. The rubrics to be developed in this paper is analytical task-specific rubric which suit the nature of civil engineering workshop which requires students to apply timber, steel, concrete and scaffold specific task. Practicum is based on job sheet as explicitly set out in the curriculum by the institution and consists of task such as practice material preparation, painting, cutting, assembling and installation [7]. Based on the circumstances in the field, to date the assessment format used in the workshop at Civil Works refers to the 2007 curriculum in which there is an assessment of the theory, practice and attitude, where the problem is an assessment that contains high subjectivity and performance appraisal aspects, which is incomplete.

In many research, when we look at complex constructs, valid and reliable instruments are needed [8]. Validity, which is the ability of an instrument to measure the properties of the construct under study [9], is an important factor in selecting or applying an instrument. There are three common forms of validity, which include content, construct, and criterion-related validity [10]. Content validity, as the prerequisite for other validity, should receive the biggest priority during instrument development. Validity is not the property of an instrument, but the scores for an instrument used for a specific objective on a group of sample. So, validity evidence should be showed on each study for which an instrument is used [11].

In this study we will use internal and external validity as would be further explained in methodology. The aim of this study is to develop assessment criteria and methodology that is both theoretically valid (Lawshe expert test) and empirically valid (Cronbach test). Therefore in the end it would result in a more objective and structured assessment which is beneficial not only to the students but also the lecturers.

2. Methodology
The study is a research and development study that aims to find out a model of practicum assessment in civil engineering workshop. The problem under investigation was a model of developing students’ competency and a model of assessing students’ vocational competency. We focus on timber, steel, and scaffolding task competencies in which the model is developed in accordance with the empirical findings on the competency assessment model consisting of constructing model of assessing the competency, and validating the model constructed in a rubric modeled assessment by expert test and empirical test, and reliability test.

This study require a population which divided into two subpopulations, they are 1) vocational education experts, language expert, timber steel and scaffolding experts, and construction practitioners, and 2) students and lecturers at the Civil Engineering Department Bali State Polytechnic. The sampling method for experts is the purposive sampling technique, which includes two vocational education experts, one language expert, one timber steel and scaffolding expert, and two construction practitioners who are involved in the rubrics model instrument validation. The sample from students and lecturers was taken by proportional random sampling technique. The validation of the model uses the data from expert judges and is tested through formula and data from students. Each point in the rubrics assessment is also tested with correlation test. In the end external validity test is used on a smaller practicum sample as rated through five lecturers to obtain the Cronbach’s alpha score. The data collection for instrument validation utilized two-period questionnaires administered in two periods.
2.1 Dick-Carey Model of Rubrics Development

The first step of this study is by conducting the rubrics model development by Dick and Carey, which is to formulate conceptual and operational definitions. We construct the variables that will be measured in accordance with the theoretical basis developed thoroughly and operationalize the conceptual definition in accordance with the nature of the instrument complete with indicators to be achieved by the students [12]. Then, the process will proceed to indicators development and instructional writings that will be followed by validation process [12] both theoretical validation and the empirical validation. The first stage of validation is a theoretical validation, i.e. through an expert examination [13] or through a panel that basically examines how far the dimension is appropriate for the construct, how far the indicator is the exact description of the dimension, and how far the points instruction of the instrument are made can precisely measure the indicator. Testing the validity of the content using Content Validity test [14].

After expert test, we will go through internal validity through product moment correlation test between the score of each item and the overall score for the rubrics. Here we see the relationship between each item with the overall score where we can interpret how significance the item in that group of rubrics item [15]. Field trials are done to obtain the scores on each item for this empirical validation process. Through these trials, the instruments were given to a number of respondents as test samples having characteristics similar to or equivalent to the characteristics of the study population. Reliability test then as mentioned in Dick-Carey model would be part of the trial test in which in this study we are using Cronbach’s alpha to determine.

Revision of the instrument is done if after going through the analysis there are items that are not valid or have a low reliability. The revised items are reassembled and recalculated validity and reliability. The research flow is based on Dick Carey's development method as follows [12]:

![Dick-Carey Model Development](image)

**Figure 1. Dick-Carey Model Development**

2.2 Content Validity Test

Content validity or definition validity or logical validity [16] can be defined as the ability of the chosen items to show the variables of the construct in the measure. Content validity revealed the
degree to which items of an instrument, or in this case the rubrics’ items, sufficiently represent the content domain. It gives the answers to the question of to what extent the selected item in an instrument or instrument items is a comprehensive item of the content [8, 17-19]. This type validity provides the basic evidence on construct validity of an instrument [20]. Moreover, it can show the representativeness and clarity of items and help improve the development of the rubrics through achieving recommendations from an expert panel [21]. If an instrument lacks content validity, it is impossible to establish reliability for it [22], and therefore it should be done as a priority. The validity in this study is conducted by six experts (expert judges), i.e. two vocational education experts, 1 language expert, 1 construction expert in timber steel and scaffold workshop, and two construction practitioners, and calculated using Lawshe formula as below [23]:

\[ CVR = \frac{2MP}{M} - 1 \]

where,

- CVR: Content Validity Ratio
- MP: Number of experts declare valid
- M: The number of experts who validate

The content validity range is \(-1 \leq CVR \leq 1\)
- If MP <½ M then CVR < 0 (invalid)
- MP = ½ M then CVR = 0 (received)
- MP> ½ M then CVR> 0 (accepted /valid)

To test the validity as a whole can be calculated using Content Validity Index (CVI):

\[ CVI = \sum CVR \]

Information:

\( CVI = \text{Content Validity Index} \)
\( \sum CVR = \text{Total Content Validity Ratio} \)
\( k = \text{Number of items} \)

2.3 External Validity Test

External validity is related to the generalization of study results. In all forms of research design, the results and conclusions of this study are limited to participants and conditions as defined by the contours of the study and refer to the extent to which generalization of research results for other conditions, participants, time, and place [24].

To test the validity of each item, the scores on the item in question are correlated with the total score. The score of each item is stated as X score and the total score is expressed as score Y, with the index of the validity of each item, it can be seen which items meet the requirements of the validity index [25]. To test the validity of the instrument used product moment correlation formula as below [25]:

\[ r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \]

where,

- \( n = \text{Number of respondents} \)
\[ X = \text{Score item no.1 reached by each respondent} \]
\[ Y = \text{Total Score of each respondent for all items} \]

with \( r_{xy} \) is the correlation coefficient between variables \( X \) and \( Y \), \( N \) is the number of respondents, \( X \) is the scores of each item for each respondent, and \( Y \) is the total score of each respondent.

In addition we will also look at the Cronbach’s alpha value to test for reliability. We are testing whether that the instrument will always elicit consistent and reliable response even if questions were replaced with other similar questions. When the item generated from the rubrics return a stable response, then the item is said to be reliable [26]. Cronbach's alpha is an index of reliability associated with the variation accounted for by the true score of the "underlying construct." Construct is the hypothetical variable that is being measured [27].

Alpha coefficient’s value ranges 0 to 1 and shows the reliability of factors extracted from dichotomous and/or multi-point formatted questionnaires or scales. The higher the score, the more reliable the generated scale is. Nunnaly[28] has indicated 0.7 to be an acceptable reliability coefficient but lower thresholds are also used in other literature. In this study we are using multi-point formatted rubrics scoring from 1= low to 3=high. The overall alpha would be tested against the 35 items in the rubrics.

3. Result and Discussion
3.1. Dick-Carey Model of Rubrics Development
3.1.1 Definition of Concept
The definition of the concept of performance-based assessment is the assessment of the competence of practical work, where the competence is an ability to perform or perform a job or task based on skills and knowledge and support by the work attitude required by the work in the form of preparation stage of the instrument, implementation/work process and stage of presentation of the work [7].

3.1.2. Operational definition
The operational definition of performance-based practicum assessment on timber, steel and scaffolding practices in this study includes competence:
1. Preparatory Practice of Instruments
   a. The ability to understand images consists of accuracy and dexterity / time
   b. The ability to plan materials and tools consists of accuracy and dexterity / time
   c. The ability to calculate the needs of materials and tools consists of accuracy and dexterity / time
   d. The ability to prepare materials and tools consists of accuracy and dexterity / time
   e. The ability to check the tool consists of accuracy and dexterity / time
2. Implementation Practices / Work Process
   a. The ability to paint on materials consists of accuracy, tidiness, dexterity / time
   b. The ability to cut materials consists of skill, accuracy, tidiness, attitude, dexterity / time, work safety
   c. The ability to assemble materials consists of skill, accuracy, attitude, dexterity / time, work safety
3. Practice of Presentation of Work Results
   a. The ability to perform finishing / final settlement consists of skill, dexterity, tidiness
   b. Ability to arrange the work consists of cleanliness, tidiness

3.1.3. Draft Performance Based Practical Assessment Rubric
Based on the grid of the assessment instrument, it is further described as a rubric of performance-based practicum assessment on timber, steel, scaffolding practices at the Civil Engineering Workshop of Bali State Polytechnic, we thus obtain 35 points of assessment items along with descriptions and scores of each item.
3.2. Internal Validity Test (Expert Test)
We calculated the Content Validity Ratio on each score items, and CVR value obtained on each item is in the of between 0.33 to 1.00 so that CVR> 0 (positive) means that all items are declared as valid on the overall score by the experts. Moreover, the category for calculation results for Content Validity Index are as follows [8]:

| Range       | Category                        |
|-------------|---------------------------------|
| 0 – 0.33    | it is not in accordance with    |
| 0.34 – 0.67 | Appropriate                     |
| 0.68 - 1    | Very appropriate                |

The CVI value is under the range of 0.68 - 1 so that it falls into the very appropriate category, which means that the scoring tools in the rubrics modeled assessment, is valid as a whole.

3.3. External Validity Test (Empirical)
External validation test by conducting trials on a small group of 26 students (1 class) doing timber, steel and scaffolding practices using 5 assessors (rater), the results of the assessment show the reliability and validity as follows:

| Reliability Statistics |
|------------------------|
| Cronbach's Alpha       |
| N of Items             |
| .977                   |
| 35                     |

The degree of reliability of the test according to Guilford as follows:

- \( \leq 0.20 \): Very low
- \( 0.40 \leq 0.60 \): Moderate
- \( 0.20 \leq 0.40 \): Low
- \( 0.60 \leq 0.80 \): Height
- \( 0.80 \leq 1.00 \): Very High

Based on the reliability test, the Cronbach's alpha value of 0.977 indicates a very high level of reliability [27]. It is also

| Item-Total Statistics |
|-----------------------|
| Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
| item_1    | 77.2692 | 212.276 | .902        | .976        |
| item_2    | 76.7462 | 216.594 | .851        | .976        |
| Item | Validity | Category | Result |
|------|----------|----------|--------|
| 1    | .902     | Very high| Valid  |
| 2    | .851     | Very high| Valid  |
| 3    | .766     | High     | Valid  |
| 4    | .766     | High     | Valid  |
| 5    | .913     | Very high| Valid  |
| 6    | .902     | Very high| Valid  |
| 7    | .875     | Very high| Valid  |
| 8    | .875     | Very high| Valid  |
| 9    | .902     | Very high| Valid  |
| 10   | .913     | Very high| Valid  |
| 11   | .714     | High     | Valid  |
| 12   | .714     | High     | Valid  |

Table 4. Result of the Correlation Test and Validity
Under the correlation test, item 35, which is for the students’ ability to present the finish product in appropriate manner, has low positive relationship with the overall score. Based on the above validity test, the item 35 is not so significant to the overall score, and thus has low external validity. Otherwise, based on validation test and reliability test above stated there are thirty four items of valid statement on rubrics assessment tool so that the draft rubric of such judgment can be declared valid theoretically [14].

4. Conclusion

Based on the discussion that has been described in the previous section, it is concluded that through the data from analysis of needs obtained information that aspects of performance-based practicum assessment instruments on timber, steel and scaffolding practices in the workshop of Civil Engineering at the Bali State Polytechnic in this study include the competence of the ability to understand drawing ability, ability to plan materials and tools, ability to calculate material and tool needs, ability to prepare materials and tools, tool check ability, material painting ability, material cutting ability, material assembling ability, and finishing / finishing ability. The rubrics modeled assessment instrument consists of 35 items, and after the internal validation test (expert test) all the items are declared valid with CVR value > 0 and the value CVI = 0.75 which means the assessment instrument is very suitable and can be used for the assessment. The results of external consistency reliability test, analyzed using Alpha Cronbach’s formula, obtained value calculated reliability coefficient of 0.977 is classified as very high reliability. The validity test shows that only 34 items are valid theoretically.

5. Acknowledgement

The authors wish to give their gratitude to the fellow college at Bali State Polytechnic for their support and to LP3M PNB for their guidance as to the development of this paper. We would like to thanks the Indonesian Ministry of Research and Technology for their generous funding towards the completion of this project.

|   |    |          |  
|---|---|----------|  
| 13 | .902 | Very high | Valid  
| 14 | .913 | Very high | Valid  
| 15 | .766 | High | Valid  
| 16 | .491 | Medium | Valid  
| 17 | .913 | Very high | Valid  
| 18 | .766 | High | Valid  
| 19 | .832 | Very high | Valid  
| 20 | .406 | Medium | Valid  
| 21 | .445 | Medium | Valid  
| 22 | .445 | Medium | Valid  
| 23 | .832 | Very high | Valid  
| 24 | .445 | Medium | Valid  
| 25 | .766 | High | Valid  
| 26 | .692 | High | Valid  
| 27 | .517 | Medium | Valid  
| 28 | .567 | Medium | Valid  
| 29 | .637 | High | Valid  
| 30 | .851 | Very high | Valid  
| 31 | .851 | Very high | Valid  
| 32 | .445 | Medium | Valid  
| 33 | .902 | Very high | Valid  
| 34 | .913 | Very high | Valid  
| 35 | .113 | Low | Invalid  

6. References

[1] Toohey S., Ryan G., Hughes C. 2006. Assessing the Practicum. Journal Assessment & Evaluation in Higher Education; 1996; 21(3)
[2] Stevens, D. D., & Levi, A. J. 2005. Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning. Sterling, VA: Stylus.
[3] MOE. 2007. Indonesian Minister of Education Regulation Number 20 Year 2007. Jakarta.
[4] Arter, J., McTighe, J. 2001. Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Thousand Oaks, CA: Corwin Press, Inc.
[5] Wilson, M. 2006. Rethinking rubrics in writing assessment. Portsmouth, NH: Heinemann
[6] Brookhart S., Nitko A.2008. Assessment and Grading in Classrooms. Upper Saddle River, NJ: Pearson Education; 2008; (1) 201.
[7] BSP. 2013. Civil Engineering Workshop Teaching Manual. Politeknik Negeri Bali; 2013;
[8] Rubio D., Berg-Weger M., Tebb S., Lee E., Rauch S. 2003. Objectifying content validity: Conducting a content validity study in social work research. Social Work Research. 2003;27(2):94–104.
[9] DeVon H., Block M.E., et al. 2007. A psychometric toolbox for testing validity and reliability. J NursScholarsh. 2007;39(2):155–64.
[10] Lynn M.R. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–5.
[11] Waltz CF, Strickland O, Lenz ER. 2010. Measurement in nursing and health research. 4th ed. New York: Springer Publishing Company; 2010.
[12] Dick W., Cary L. 1990. The Systematic Design of Instruction. Harper Collins: New York.
[13] Zamanzadeh, V., Ghahramanian, A., et al. 2015. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. Journal of Caring Sciences, 4(2), 165–178. http://doi.org/10.15171/jcs.2015.017
[14] Akhyar, M. 2009. A Model Of Vocational Competency Assessment Of Industrial Engineering Students Of Vocational High Schools. Journal of Education 2085-1839.
[15] Kornbrot, D. 2005. Pearson Product Moment Correlation. Encyclopedia of Statistics in Behavioral Science.
[16] Newman I, Lim J, Pineda F. Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. Journal of Mixed Methods Research. 2013;7(3):243–60.
[17] Polit D., Beck C.. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489–97.
[18] Wynd C., Schmidt B., Schaefer M. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25(5):508–18
[19] Yaghmale F. 2003. Content validity and its estimation. Journal of Medical Education.2003;3(1):25–7.
[20] Anastasi A. 1988. Psychological testing. 6th ed. New York: Macmillan; 1988.
[21] Polit D., Beck C., Owen S. 2007. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 2007;30(4):459–67.
[22] Nunnally J., Bernstein I. Psychometric theory. New York: McGraw-Hill. 1994:3.
[23] Lawshe C. H. 1975. A quantitative approach to content validity. Personnel Psychology. 1975; 28 (4):563–75.
[24] Graziano, A.M., Raulin M.L. 2009. Research Methods: A Process of Inquiry 4th Edition. MA: Allyn&Bacon.
[25] Chiang, A. C., Wainwright, K. 2005. Fundamental methods of mathematical economics. New York; McCraw Hill.
[26] Cronbach, L. 1951. Coefficient alpha and the internal structure of tests. Psychometrika.1951; 16: 297-334
[27] Hatcher, L. 1994. *A step-by-step approach to using the SAS(R) system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

[28] Nunnaly, J. 1978. *Psychometric theory*. New York: McGraw-Hill.

[29] Santos J. 1999. *Cronbach's Alpha: A Tool for Assessing the Reliability of Scales*. *Journal of Extension*. 1999; 37(2).