Rubric for Measuring Psychomotor and Affective Learning Domain

Shelena Soosay Nathan*, Mazniha Berahim and Rosni Ramle

Department of Information Technology, Centre for Diploma Studies, Universiti Tun Hussein Onn, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

ABSTRACT

Designing a reliable measurement of the psychomotor and affective learning domains is a major challenge. One assessment tool, the rubric, provides flexibility in assessing and improves grading consistencies. But students are not being assessed properly as only one rubric is used to evaluate different categories of a project, causing inconsistencies in grading. Thus, an assessment rubric for different categories of a project was created, incorporating the psychomotor and affective learning domains aligned with Bloom’s Taxonomy. To validate the rubrics, intra-class coefficient (ICC) and reliability tests were done using the Statistical Package for the Social Sciences (SPSS) tool. Analysis was done to determine grading consistency and agreement level among two randomly chosen evaluators when using the rubrics and to evaluate whether clearly defined assessment metrics were used in grading projects. The results showed that the psychomotor rubric has strong inter-rater reliability with scores of 0.90 and 0.86; this suggests that variables in the rubric were ‘very good’ at measuring the end product. However, the affective rubric shows slightly weak reliability. This might be due to the different way evaluators assess the same work as some tend to be lenient, while others are strict. The developed rubrics enables evaluators to better assess students so that students obtain justified grades according to the quality of their project.

Keywords: Assessment, affective, ICC, psychomotor, rubric

INTRODUCTION

Designing assessment that covers different learning domains (cognitive, psychomotor, affective) within specific criteria and standards is challenging (University of New South Wales, 2017), especially for the
affective domain. A common assessment tool, the rubric, enables evaluators to assess students’ understanding and creativity, provides flexibility and improves grading consistency (Manson & Olsen, 2010; Meenakshi, 2013; Sharef, Hamdan, & Madzin, 2014; Mustapha, Samsudin, Arbaiy, Mohamed, & Rahmi, 2016). A rubric should be valid and reliable, and to achieve this requires continuous improvement to the tool (Humphry & Heldsinger, 2014; Goldberg & Canty, 2015). Diploma in Information Technology (DAT) offered by the Centre for Diploma Studies (CeDS), Universiti Tun Hussein Onn Malaysia (UTHM), usually have a large number of students, which makes final-year project (FYP) evaluation for the programme challenging, especially when it comes to ensuring fair grading. The FYP is divided into three categories: database management system, multimedia application and hybrid system. Quality of product for each category is evaluated from different aspects. A similar grading scale used for the different categories can result in unreliable evaluation i.e. projects may be underrated or overrated. To address this, a set of rubrics that cater to the different categories is needed to improve grading consistency among the evaluators. The focus of this paper is on developing a validated new set of rubrics as a measurement tool for evaluating FYP end products and presentations for DAT.

Related Work
Curriculum, learning activities, assessment and outcomes must be aligned in order to achieve a meaningful learning experience (Anderson, 2002; Boud & Falchikov, 2006; Martone & Sireci, 2009; Tam, 2014). To see whether a student can demonstrate the outcomes, he or she is assessed using outcome-based assessment (OBA) (Crespo et al., 2010). Assessment of student learning encompasses three learning domains i.e. the cognitive, affective and psychomotor domains (Bloom, 1956). The literature revealed that rubric is a standard assessment tool for evaluating computer science undergraduate FYP (Sánchez et al., 2014; Sharef et al., 2013; Tio, Kong, Lim, & Teo, 2014) and it is used as a scoring tool that lists criteria and level of quality (Andrade, 1997). However, bad rubric design such as being too general (de Sande et al., 2011) or too specific (Fraile et al., 2010; Sánchez et al., 2014) can cause time wastage and an increase in marking load (University of New South Wales, 2017) and could cause the evaluator (Sadler, 2009) or student (Boud, 2010) to lose the overall view of the project. Therefore, rubric designers must create one that is achievable, clarified and suitable for learners’ age and level of education. From the perspective of a computer science project, the psychomotor domain is evaluated based on knowledge in the area while the affective domain is appraised through presentation of the product (Mustapha et al., 2016).
Different rubrics are needed to evaluate each learning domain as each has its own defining characteristics (de Sande et al., 2011; Sánchez et al., 2014; Tio, Kong, Lim, & Teo, 2014).

**METHODOLOGY**

The set of criteria and standards in a rubric covers the learning domains and linked to course learning outcomes (CLOs) and programme learning outcomes (PLOs). In this study, two rubrics were developed to measure the psychomotor (for end product) and affective (for presentation) domains, and they were matched to the CLOs and PLOs of DAT. There are four phases in the rubric development, explained in subsequent sections of this paper.

**Phase 1: Analyse and Identify PLO and CLO for Chosen Type of Assessment**

In this phase, mapping of PLOs to CLOs to type of assessment (log book, project proposal, final report, technical report, end product and presentation) was done. However, this study only focussed on CLO 2, measured by evaluation of end product, and CLO3, measured by project presentation. Table 1 shows that each CLO assessed one learning domain with a specific dominant level of learning.

**Table 1**

<table>
<thead>
<tr>
<th>Course Learning Outcomes</th>
<th>Programme Learning Outcome</th>
<th>Type of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLO 2: To manipulate theoretical and practical knowledge to solve a problem or project</td>
<td>Complex Overt Response (P5)</td>
<td>End product of project</td>
</tr>
<tr>
<td>CLO 3: To demonstrate the project achievement verbally and non-verbally</td>
<td>Organising value (A4)</td>
<td>Project presentation</td>
</tr>
</tbody>
</table>

**Phase 2: Identify Related Level of Learning Domain Covered by CLOs and Criteria of the Domain**

The mapping of CLOs to PLOs was used to brainstorm ideas to design the criteria for the rubrics. Based on the information in Table 1, the dominant level of the learning domain was assigned. All the criteria for each learning domain were listed and the most important were chosen. Other levels related to the dominant level of the learning domain and the keywords for each domain were assigned based on Bloom’s Taxonomy (Bloom, 1956) for each rubric’s criteria.

Two types of criteria (generic and specific) were included in the rubric to cater for the three different FYP categories. Specific criteria were designed based...
on category of project. Table 2 shows the list of general criteria and the level of the psychomotor learning domain for each criterion. The rubric for end product evaluation measured the dominant level of learning, P5, and two supplement level of learning, Guided Response (P3) and Mechanism (P4).

Table 2
Level of learning and its criteria for the Psychomotor learning domain

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td></td>
</tr>
<tr>
<td>i. Follow objectives of project</td>
<td>P3</td>
</tr>
<tr>
<td>ii. Construct a project aligned with current technology that is also marketable</td>
<td>P4</td>
</tr>
<tr>
<td>iii. Calibrate significance and performance of project</td>
<td>P5</td>
</tr>
<tr>
<td>iv. Display innovation, creativity and uniqueness of project</td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td></td>
</tr>
<tr>
<td>Category 1: Construct an efficient database and user-friendly interfaces</td>
<td></td>
</tr>
<tr>
<td>Category 2: Construct an interactive and attractive interface</td>
<td></td>
</tr>
<tr>
<td>Category 3: Construct usable and accurate results</td>
<td></td>
</tr>
</tbody>
</table>

The rubric for project presentation (Table 3) measured the dominant level of the affective learning domain, A4, and two supporting levels of learning, valuing, A3, and internalising values, A5.

Table 3
Level of learning and its criteria for the affective learning domain

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. Follow professional dress code</td>
<td>A3</td>
</tr>
<tr>
<td>ii. Explain end product with good presentation skills</td>
<td>A4</td>
</tr>
<tr>
<td>iii. Organise presentation well in a systematic way</td>
<td></td>
</tr>
<tr>
<td>iv. Prepare attractive and precise poster</td>
<td></td>
</tr>
<tr>
<td>v. Display understanding and knowledge of end product</td>
<td>A5</td>
</tr>
</tbody>
</table>

**Phase 3: Formulate Rubric by Type of Assessment and Align with Its Approaches**

A two-dimensional table was constructed, where the column titles were the scale of performance level and the rows were the learning domains and criteria as listed in Table 2 for rubric of end product and Table 3 for rubric of project presentation. A 5-point Likert scale was used; 1 – very poor, 2 – poor, 3 – fair, 4 – good and 5 – excellent. The rubric for end product contained seven items measuring the psychomotor criteria, while the rubric for the presentation consisted
of five items for the affective criteria. The descriptions of performance were determined by mapping the criteria to scale.

**Phase 4: Validate Reliability of Rubric**

In order to validate the reliability of the rubrics, a reliability test was done using Cronbach’s Alpha and Intra-Class Coefficient (ICC). Cronbach’s Alpha is commonly used to assess the reliability or internal consistency of a scale or test items (Gleam & Gleam, 2003). Inter-Rater Reliability (IRR), also known as inter-rater agreement, is the agreement among raters (Taylor, 2010). It displays how strongly the units in the same group resemble each other in the same set. Scores given by evaluators were analysed using the Statistical Package for the Social Sciences (SPSS).

**RESULTS AND DISCUSSION**

The sample consisted of 47 groups, each having three members. The mean from both evaluators showed a consistent value; evaluator 1 received a score of 39.6 while evaluator 2 scored 40.9. Table 4 shows the descriptive summary for the assessment rubric between two evaluators that was chosen randomly.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Sum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Evaluator 1 (P &amp; A)</td>
<td>28.00</td>
<td>49.00</td>
<td>1861.00</td>
<td>39.5957</td>
<td>4.91961</td>
</tr>
<tr>
<td>Total Evaluator 2 (P &amp; A)</td>
<td>23.00</td>
<td>49.00</td>
<td>1924.50</td>
<td>40.9468</td>
<td>5.84420</td>
</tr>
</tbody>
</table>

**Result of Cronbach’s Alpha for Psychomotor and Affective Rubrics**

A Cronbach’s Alpha value of more than 0.9 is excellent, while 0.7 to 0.8 acceptable, 0.6 to 0.7 questionable and 0.5 to 0.6 poor, while below 0.5 is unacceptable (Gleam & Gleam, 2003). Cronbach’s Alpha for P (Psychomotor) for two evaluators was 0.798, a reasonably acceptable value. This means that items in P exhibited strong face validity and construct validity. Cronbach’s Alpha for A (Affective) for two evaluators was 0.649, a questionable value. However, Loewenthal (2004) stated that an alpha coefficient of 0.6 may be accepted.

**Result of Intra-Class Coefficient Reliability for FYP Evaluators**

Since evaluators were chosen randomly, a one-way random test was used to find the Intra-Class Coefficient (ICC) reliability. The study aimed to determine the reliability of the psychomotor and affective rubrics individually; thus, the results were obtained separately for both domains and evaluators and later compared, as shown in Table 5 to Table 9.
A strong correlation is a nearly perfect prediction for both raters, but actual agreement does not exist. Good agreement is obtained when two values are almost equal and close to 1 (Cicchetti, 1994). The ICC analysis for the total number of 47 groups in terms of the psychomotor domain based on two evaluators were 0.90 and 0.86, while
for the affective domain it was 0.46 and 0.83, respectively. The total ICC score for both evaluators for both learning domains showed that the rubrics were reliable for measuring their assessment. Although the value of the ICC for the affective domain was significant, it was only moderately reliable as the total score obtained was 0.83, which is considered acceptable.

The psychomotor domain showed strong inter-rater reliability with scores of 0.90 and 0.86, respectively, and this suggests that the variables used in the psychomotor rubric were suitable for measuring the end product. However, the affective rubric’s reliability was slightly weak; we believe this was due to the tendency of different evaluators to be strict or lenient when grading student work.

**CONCLUSION**

Reliable rubrics for FYP evaluation that measure the psychomotor and affective domains was established. Usage of the rubrics can be extended to assess students’ performance in conducting projects. Students can also use these rubrics as a guideline when developing an IT project. Evaluators must be briefed before assessing on how to use the rubrics to avoid bias and misunderstanding. Further study is needed to investigate and enhance the rubrics’ validity. One way of doing this is by seeking the opinion of students and evaluators.

**ACKNOWLEDGEMENT**

The authors are grateful to Information Technology Department, CeDS as well as Office for Research, Innovation, Commercialisation and Consultancy Management (ORICC), UTHM for funding this project.

**REFERENCES**


