Thesis assessment element weighting and quantitative nature of criterion

HO SUNG KIM

Mechanical Engineering, School of Engineering, Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, NSW 2308, Australia.

ho-sung.kim@newcastle.edu.au

Abstract

Assessment elements are analyzed and discussed to find quantitative nature of thesis assessment criteria linking with assessment components. Some of major criterion properties for thesis assessment were found from multiple choice and free response exams, and were further generalized for thesis assessment. A hierarchical structure of assessment elements in terms of compatibility and specificity is proposed for thesis assessment weight calculation. Selection principles of thesis assessment components and criteria were discussed on the basis of the hierarchical structure resulting from the assessment element analysis. An analogy between thesis assessment and a composite material in material science was proposed to represent quantitative nature of assessment elements. Mark allocation based on quantity and quality criteria are governed by the product law whereas weight/mark calculation of individual assessment subordinate elements at the same level in the hierarchical structure is governed by the addition law. It is explained why assessment criterion weights should not set before assessment.

Introduction

In most disciplines, grading is often essential process as part of assessment. Also, it is useful to define assessment elements clearly for both marker and learner in most courses. The assessment elements may include assessment means, content, criteria and weighting. The assessment 'means' or usually called 'items' would include assignments, exams, reports, thesis, seminar, etc. The assessment criteria form the basis not only for a marker to make various judgments in allocating a grade to the learner, but also suggest what the learner is required to do for a grade. Thus, the assessment criteria could be the learning objectives from the learner's point of view. Therefore, clear description of assessment items and criteria is important particularly in a self-learning course to get the learner actively engage with meeting the criteria.

In an engineering course work where the course contents are well defined, assessment items can readily be listed with weights for learners. For example, assignments 30%, exams 50%, etc can be informed to the learner sometimes without showing too much detailed assessment criteria because exhaustive explicit criteria description is not necessarily practical. However, when learners are required choose their own research project topics and, at the end, individuals are to write theses for assessment, the course (project) contents are vague at the beginning and not well defined until each project is completed. At the completion of the research project, individual thesis contents are different from each other. One thesis would have more emphasis on one aspect while others on other aspects. As a result, a thesis is unique in terms of topic, research/knowledge area, or/and analysis skill required. When theses are examined for grading, some inconsistency arises from various sources such as unclear guidelines for assessment items and criteria, examiner expertise, etc. The inconsistency sometime causes serious impacts on students in various ways.

It is important to understand the mechanisms between assessment elements leading to final allocation of marks for adequate development method of assessment. Weighting on assessment items is commonly practiced without much argument. However, weighting on assessment criteria has been a subject for debate. A practical and comprehensive handbook[1] for linking levels, learning outcomes and assessment, points out that the use of effective assessment criteria seems to be a long way behind the learning outcomes and level descriptors. Woolf[2] states that there are no explicit statements about whether some criteria are more important than others or markers may weight the criteria,

either consciously or subconsciously. Oehlers[3] has adopted different criteria for different grades in his proposed sequential assessment for engineering design projects. Warton[4] argued rationale for the assessment criteria for Master's level TESOL/TESP (Teaching English to Speakers of Other Languages/Teaching English for Specific Purposes) assignment which is similar to thesis assessment in terms of the variety of topics. Freeman and Lewis[5] introduced various assessment criteria mainly in terms of presentation for the dissertation. Webster and others[6] discussed some common general criteria and their ambiguity as regards use, meaning and application for the dissertation assessment. As such, none deals with quantitative nature of criteria. Indeed, no fundamental framework for assessment linking assessment items and quantitative criteria has been found in the literature.

In this paper, the relationship between assessment elements is clarified to link between assessment items/contents and criteria for allocating mark on a thesis. Also, the assessment criteria are analyzed to find their fundamental quantitative nature.

Weighting on assessment components for known course content and limitation for multi-criteria

The weighting on assessment items/contents for known course content may be introduced as a starting point to find useful principles applicable to thesis assessment. (Any sub-division for an assessment item, content will be generically referred to as assessment component.) The following equation is a typical form when weighting is applied to assessment items/contents for marking:

$X\_a_i = X(a1 + a2 + a3 + \ldots)$ (1)

where a1, a2, a3, are weights for individual components and X is the highest mark. Let Xi be the highest mark of each assessment component and xi be the mark a student has achieved for component. Thus,

$X_i \geq x_i$ (2)

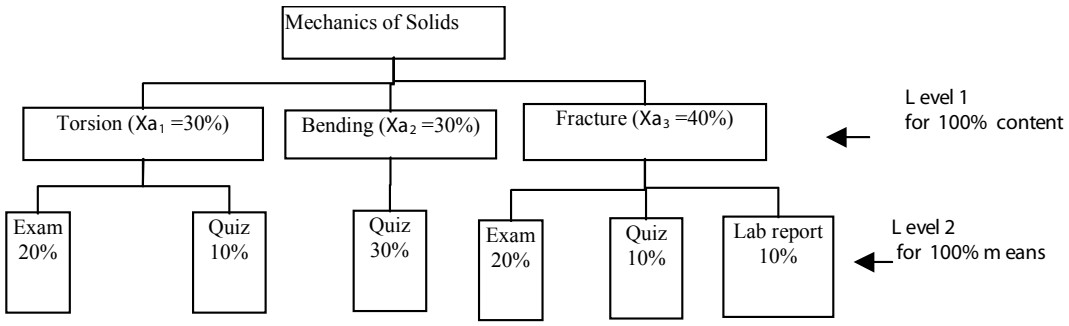
where

$X_i = Xa_i$ (3)

Equation (1) is sometimes used implicitly or explicitly. The weight depends upon the expected learning outcomes and various factors such as difficulty, amount of effort required, etc. Some minor overlap between assessment components may not be a major problem as far as fairness to students is concerned. However, when an overlap exists in the case where entire assessment content is part of other assessment content, the overlap is a manifestation of the result from inseparability of, or unidentified assessment sub-elements in at least one of assessment components. The efficiency of the assessment, consequently, would suffer and students would be confused. Thus, the overlap in assessment components should be minimised in weighting.

A principle of weighting assessment components can be found on the basis of hierarchy of multi-levels with sum of weights without much difficulty. Each level should consist of compatible components. For example, course content is not compatible with assessment item and vice versa so that they should not be on the same level for weighing. An example for a typical course in mechanical engineering, Mechanics of Solids, is shown in *Figure 1*. Each level consists of either course topics/contents or items. The Level 1 in *Figure 1* consists of course contents for $X\_a_i = 100\%$ and Level 2 consists of assessment items for $X\_a_i = 100\%$. In this case, assessment items are subordinate to contents. Alternatively, contents may be subordinate to items. It is important to ensure that the sum of level weights at a level must be 100%.

Figure 1 An example of hierarchical structure for weighting in marking part of a course, Mechanics of Solids.



Quantitative nature of single assessment criterion and achievement factor

A simple assessment task as a subset of thesis assessment may be used to discover part of quantitative nature of thesis assessment criteria. In a multiple choice exam, assessment criteria are usually not explicitly written but used intuitively. The criterion involved is in fact a single criterion, 'right or wrong'. If we adopt '0' and '1' for a student performance in each multiple choice question, we would be able to quantify the performance using the 'right or wrong' criterion so that xi (a mark student has achieved for each question in this case) in Equation (2) is allowed to form an equation for the criterion as

$f_i X_i = x_i \ (\leq X_i)$ (4)

where $f_i = 0$ (wrong) or 1 (right) and X_i is the exam question weight in the case of exam. If the same criterion 'right or wrong' is applied to all the questions, marks would be allocated as

$\_f_i X_i = f1X1 + f2X2 + f3X3 + \ldots$ (5).

The f_i becomes equal to a_i in Equation (1) when the student got it right. Thus, f_i will be referred to as a student achievement factor hereafter. Similarly, we can discover the achievement factor in a different type of examination. In free-response exam questions, right or wrong is not sufficient but requires a range of different answers including partially right answer. In this case, a factor value ranging from 0 to 1 (i.e. 0, 0.1, 0.2…so on, up to 1), depending upon how good the answer is, can be allocated for f_i, and can be used as a superset (which will be referred to as 'how good' criterion) for the 'right or wrong' criterion factor.

Now, the quantitative role of the assessment criterion with the achievement factor (f_i) is clearly found for the 'single criterion versus multiple questions (or items)' assessment. Another finding is that the criterion should be subordinate to each exam question in the hierarchical structure (*Figure 1*) because the criterion follows each question (or item) but not the other way around.

Multi-criteria formula for quantitative thesis assessment

A thesis is one of assessment items produced by a student for assessment after carrying out a project chosen. An assessor then examines the thesis using the various criteria. A thesis is to be examined through assessment criteria. A thesis is a large superset of an exam question and may be subdivided into components, and some components can be assessed using multi-criteria or single criterion depending on how many thesis components and assessment criteria are designed.

If the 'how good' single criterion is adopted for thesis assessment, the quantitative role of the criterion discussed earlier can be extended to a thesis assessment in conjunction with multiple-components. Also, if the single criterion is separable like assessment components and broken up into multiple criteria, an equation for multi-components versus multi-criteria would mathematically be formulated. Thus, $\_a_i$ in Equation (1) becomes

$\_a_i = \_(b_i + c_i + d_i + \ldots)$ (6)

where *b, c, d,* etc are criterion weights for each assessment component. However, success of using Equation (6) for

multiple criteria depends on how the 'how good' criterion can be sensibly separated for weight/achievement factor. The difficulty is that different theses may have different components and different criteria weights should be applied. Further discussion is given in the following section.

Analysis of thesis assessment criterion and its nature

Analysis of criterion

The thesis assessment criterion may be broken into most generally two criteria viz quantity and quality, which can be applicable to any assessment. The quantity in thesis assessment would be to determine if there is sufficient and relevant assessment information consisting of words, figures, data, etc for recognizable thesis coverage of parts/sections such as introduction, objectives, results, analysis, etc for assessment. The quality would then depend upon how words, sentences, paragraphs, sections, etc, are arranged and written, and also how data/information obtained are analyzed to be meaningful for intellectual outcomes. Thus, the quality can be assessed according to the 'how good' criterion. For the 'how good' criterion, attributes such as knowledge/understanding, skill, outcome, originality, etc as being subordinate multi-criteria to the quality the student demonstrated in the thesis would be looked at. Quality in knowledge/understanding would be weighted for its breadth and depth, and for its difficulty to gain in thesis. The quality in skill, which is the learned capability, demonstrated by the student in thesis would be weighted for its difficulty to be gained. Quality in outcome and originality would be weighted for their significance. The skill would further be broken up into thesis presentation, analyses, computer programming, experiment, mathematics, synthesis of principles, interpretation, etc. (Different disciplines would have different skill elements/criteria.) The originality would be largely for project topics, approaches, and outcomes.

In breaking up a single criterion into multi-criteria (such as knowledge/understanding etc) here, it can be found that the level of generality or specificity of descriptors used must be maintained and the sum of weights for each level must be equal to the weight for each assessment component to use Equation (6). For example, knowledge and skill must be at the same level but computation skill must be lower than the level for knowledge/understanding, etc due to different levels of specificity. Thus, it is important to maintain the same level of generality or specificity for setting up multi-criteria.

Also, multi-criteria at a level such as knowledge/understanding, skill, outcome, and originality unlike assessment components are inter-linked with each other and complement each other in thesis writing. They exist usually together in a meaningful assessment component (e.g. introduction). For example, knowledge/understanding in an introduction of a thesis is typically demonstrated together with various skills for originality forming an outcome such that those criteria are blended together. Nonetheless, they have distinctive characteristics allowing us to assign weights in Equation (6) if possible. However, different theses would have different weights on the multi-criteria due to the diversity of thesis topics/styles. Also, different qualities (in terms of weight) are produced depending upon how different attributes (called criteria for assessment) are inter-linked in theses writing. For this reason, it would be wise not to set weights to criteria before assessment begins. However, weights/achievement factors for multi-criteria can be found as a result of assessment to make assessment accountable.

Formulation law for quantity and quality, and its analogy

As discussed, quantity is directly for assessment components covered and the quality is resulted from the quantity (recognizable coverage). It is easily demonstrated that quantity and quality for weight/achievement factor calculation does not obey the addition law. Let the full quantity and quality be quantified as 0.5 each for demonstration purposes so that the highest achievement for the 'how good' single criterion as 1 (one) as the sum. If zero for quantity and 0.2 for quality are given, the how good achievement factor would become 0.2 under the addition law. This violates the principle that zero quantity produces an achievement factor of zero. From this observation and our insight into the nature of quantity and quality may lead us into a discovery of an analogy between assessment quality and density that is one of physical properties in physics. This analogy will help understanding quantity and quality in assessment. The density in physics is defined as Density = Mass / Volume (or Mass = Volume _ Density) so that an achievement factor is equivalent to 'mass', and 'volume' is equivalent to assessment quantity. Thus, the relation between quantity and quality is given by Weight (or Achievement factor) = Quantity _ Quality. Therefore, quantity and quality may be used for the 'how good' criterion weight as well under the product law.

It is would be timely to realize that the quantity (coverage) has already been used as Xi and fi is quality in Equations (5) the for product law calculation. The difference in quantity between an exam consisting questions and a thesis is that quantity in an exam is set by the examiner but the quantity in thesis is produced by the student. Therefore, the quantity is not an issue in an exam and used implicitly but it is one of criteria for thesis assessment.

Conclusion

Assessment elements have been analysed to find the quantitative nature. Hierarchical structures linking assessment components with criteria in terms of compatibility and specificity have been found to be useful for assessment element weights. Some of quantitative nature applicable to thesis assessment criteria has been found from multiple choice and free response exams. Assessment components and criteria are for quantity and quality respectively in weighting. Mark allocation based on quantity and quality criteria are governed by the product law whereas weight/mark calculation of individual assessment subordinate elements at the same level in the hierarchical structure is governed by the addition law. It is explained why assessment criterion weights should not set before assessment.

References

01. Moon J. (2002) The module and programme development handbook - a practical resource for linking levels, learning outcomes & assessment, Selwood Printing Ltd. Great Briton.
02. Woolf H. Assessment criteria: reflections on current practices, Assessment and Evaluation in Higher Education, 29(4), August 2004, 439-493.
03. Oehlers, D. J. (2006) Sequential assessment of engineering design projects at, university level, European Journal of Engineering Education, 31(4), 487–495.
04. Wharton, S. (2003) Defining appropriate criteria for the assessment of master's level TESOL assignments, Assessment & Evaluation in Higher Education, 28(6), 649–663.
05. Freeman, R. and R. Lewis (1998), Planning and implementing assessment, Routledge Falmer, Talylor & Francis group, London, and New York.
06. Webster, F., Pepper, D. and Jenkins, A. (2000), Assessing the undergraduate dissertation, Assessment & Evaluation in Higher