An Assessment Model and Practical Rubric Design for Thesis Assessment

Ho Sung Kim

Mechanical Engineering, School of Engineering, Faculty of Engineering and Built Environment,
The University of Newcastle, Callaghan, NSW 2308, Australia, Phone +61 2 4921 6211
Fax +61 2 4921 6946, Email: ho-sung.kim@newcastle.edu.au

Abstract - *An assessment method for the final year project thesis is studied. A model integrating assessment structure and functions was refined for practicality. A new concept of 'bound categorization' was developed and as a result two sets of descriptors for quantity and quality respectively were identified within the framework of the model consisting of tasks, criteria, standards, and calculation laws for weighting and marks. The descriptors were translated into practical terms for the assessment rubric design. The descriptors for the quantity was demonstrated to be useful for describing what the student is required to do while those for the quality useful for describing how the student should do in thesis writing.*

Index Term *- Final year project, assessment, thesis, dissertation, rubric, criterion, categorization*

INTRODUCTION

Final year project (FYP) or honours project is a prevalent course globally in university undergraduate programs. The characteristics of engineering FYPs in Australia and Europe have been outlined by Ku [1] and discussed by many stakeholders[2] since 1975 [3]. The FYP has been known as the culmination of learning process and hence important [4] in undergraduate education. It involves multiple skills for implementation and documentation. The documentation is a process to produce a thesis (called 'report' or 'dissertation' in some universities) for assessment after the student chooses their own research project topic and implements the project as part of self-directed learning. 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 skills required. At the completion of a project, the student is required to write a thesis for assessment and then the thesis is marked numerically for grading. The numerical grading requires a high accuracy and hence a high level of sophistication compared to binomial grading (e.g. Fail or Pass for PhD thesis).

Designing a thesis assessment rubric (which is a set of assessment instructions) is a relatively complex task unlike other ones for engineering courses where the course contents are explicitly available (e.g. engineering mechanics, fluid dynamics, etc). The assessment complexity of the thesis involves multiple criteria, various definitions of terminology, and a high level of uncertainty about thesis content, given that the thesis content is not explicitly available until thesis is written. It is, therefore, not uncommon to deal mainly with ambiguous criteria in puzzlement. Webster *et al.* (2000) [5] discussed some common general criteria for FYP thesis and their ambiguity as regards use, meaning and application following Tariq et al (1998) [6] attempted to improve objectivity of criteria for the FYP assessment. Woolf (2004) [7] more specifically stated on the FYP assessment criterion weighting: "The departments are as silent on the weightings that they apply to their criteria as they are on the number of criteria that contribute to a grade.". A more serious concern was raised by Shay (2004) [8] by arguing that the FYP assessment for engineering and social sciences is 'a socially situated interpretive act', implying that the FYP assessment is not reliable. The problems with the FYP thesis assessment have thus received much attention over many years since 1983 [9] but without much shedding light on the puzzlement. Sadler (2005) [10] reviewed broadly on common grading policies in higher education that purport to be criteria based, and found that the concepts of 'criteria' and 'standards' are often confused, and they are interchangeably used 'even though they are not fully equivalent semantically'. This indicates that the meanings of words for assessment components have been interpreted without an adequately conceptualized assessment model. Barrow and Milburn [11], who acknowledges fundamental problems with the educational research in higher education, state that "Defining a word may be a very useful prior stage to analyzing a concept, but it cannot be identified with it. Educational research has often been flawed as a result of failing to appreciate this point…".

As such, the subjective terminology definitions without adequate conceptualization may only add to the confusion. What has really been required for the final year project thesis assessment is a fundamental knowledge based on a model consisting of structure and functions of assessment components including tasks, criteria, standards and principles. The model needs certain rules for consistency. In engineering, the term 'steering wheel' of a car, is often sufficient for a meaning without an exhaustive definition to understand what it is and how it works, etc. It can be described clearly because people know what the car is - the car is a model for structure and functions of components under certain rules. If there was no car model (before the car invention), there could have been many different silly, inconsistent, idiosyncratic definitions. Therefore, an assessment model with a logical consistency and coherence requiring certain rules has been necessary. In response, the author (2009)[12] previously developed a preliminary FYP assessment model capable of identifying various functions of weighted assessment components but it is subject to refinement and further analysis for practical applications. In this paper, the principles of selecting various assessment descriptors based on the model [12] are refined, and the limitation of numerical value adoptability for assessment components is conceptually analyzed for the FYP thesis assessment rubric design.

PRINCIPLES, ANALYSIS, AND REFINED MODEL

Mathematical laws

Thesis marking is a quantification process of meanings of words, phrases, or/and sentences into numerical values. It inevitably employs numerical values in the process, which always need to be calculated. The calculation cannot be conducted without a formula unless arbitrary rules are used. The formulation should be done under certain laws for consistency. Our previous preliminary model [12] incorporates mathematical laws as a back bone given by

$$\text{Mark} = \Sigma f_i X_i = f_1 X_1 + f_2 X_2 + f_3 X_3 + \ldots \qquad (1)$$

where X_i is the weight which is the nominated quantity, and f_i is a student achievement factor which is measurable by the assessor. The subscript i depends on the number of components. Equation (1) is adaptable to any assessment involving a numerical mark (score), including so called 'holistic assessments' if we take a single term in it, and so called 'analytic assessments' if we take multiple terms. The equation is even capable of incorporating conditions (e.g.'if') with conjunctive and disjunctive decision rules for the nominated quantities. It is important to appreciate that the equation is not to impose the inflexibility on assessment but it is to maintain the conceptual consistency. It is also numerical rules for the relationship between 'the parts and the whole', which the marker should obey before the final mark is produced. It is ultimately a bridge for 'traffic control' between meanings and numerals in the quantification process to accommodate the educational values and objectives. When numerals are used, it is essential to know if they are nominated (by the assessor e.g. weighting), or measurable (e.g. marks). When words are nominated as descriptors, it is also essential to know if they are for criteria, tasks, standards, etc. Eventually, we should know how a set of descriptors matches with a set of numerals, requiring analyses to identify the matching properties. Equation (1) has, indeed, been used by many practitioners but implicitly, unknowingly or wrongly for various assessments ranging from a multiple choice exam to a thesis assessment. If the mark calculation is conducted for a complex assessment without a good understanding of the rules in the equation, it would the potential source of puzzlement as evidenced in the literature [7]. It is important to understand the equation for fundamental functions of complex assessment components involving weighting. It expresses the multiplication and addition laws between 'nominated' quantity and 'measurable' quantity. Also, it implies the conservation law of weights should be in operation, given that the total weight nominated should be invariable when a generic descriptor is decomposed into specific ones.

Bound categorization of candidate assessment descriptors

Equation (1) deals with numerical values only such that it is not much meaningful for the FYP thesis assessment unless matching meaningful descriptor properties are analysed and found. Therefore, the way of proper use of Equation (1) is to find appropriate assessment descriptors for tasks for students, assessment criteria and standards. It involves categorization of descriptor properties and decomposition of generic descriptors into specific ones as many as possible as long as the generalizability (Linn 1991) [13] is maintained - the more number of decompositions the more specific description can be made.

When candidate descriptors are selected and examined to consider assigning numerical values, a categorization by descriptor bounds may be conducted as follows. Some descriptors possess indefinite bounds in terms of identity domain within the context of thesis – not allowing us to assign numerical values because the numerical values should be always finite and hence not suitable for an indefinite domain entity. Accordingly, a distinction of one set of descriptors will be made against other set of descriptors, expecting at least two sets of descriptors; one with identifiable bounds (to be referred to as 'bounded descriptors' in this paper) and the other with no bounds (to be referred to as 'unbounded descriptors'). For example, content headings or topics (e.g. 'Chapter 2', 'Introduction', 'Experimental details', or 'Conclusion') may be regarded as bounded descriptors because each one of those is an identifiable domain with bounds within the context of a thesis. It is noted, though, that some of bounded descriptors are not generalizable and hence not usable in some cases. For example, 'Chapter 2' is a small domain compared to that of thesis but not generalizable. Otherwise, appropriate numerical values for the weight (X_i) may be allowed to be assigned to the bounded descriptors.

The other set of descriptors such as 'quality', 'knowledge', 'understanding', 'skills', etc are unbounded because they are adoptable to almost any part of a thesis across the bounds of definable domains within the context of the thesis. Accordingly, a characteristic of unbounded descriptors is found to be its continuous variability depending on which task or domain to be assessed. For example, a student would need to have a particular 'knowledge' to write an introduction of a thesis, which is a different knowledge from that for other parts of thesis. When a judgment is made with reference to standards for tasks, various thoughts and ideas stimulated by the unbounded descriptors come through the assessor's mind in comparing between tasks and standards. Thus, the unbounded descriptors appear to be a set of clues in making decisions. Also, they appear to be suitable for criteria description with tasks and standards, given that the criteria are a means for judging (Etymology of criterion: from Greek *kriterion*: a means for judging [10]).

Further, the unbounded descriptors can later be bounded if necessary when they are combined with bounded descriptors. For example, in the descriptor, 'review quality', the 'review' alone can be a bounded descriptor in a thesis and 'quality' alone can be a unbounded descriptor prior to combining, but the 'quality' in the combined descriptor is bounded by the 'review' so that domain of the 'quality' is reduced to that of the 'review' from the whole thesis domain, allowing f_i to be equal to or smaller than 1 (one) for a 'review' weight value (X_i) in Equation (1).

Now, the bounded descriptors have been found to be numerically finite. Therefore, they are suitable for numerical weighting. When a task is of a finite domain, it allows us to define the task without vagueness (at least being different from criterion) so that the bounded descriptors may be suitable for describing what tasks should be done by the student. For example, it can be said following a definition, "a 'review' (which is bounded) is a task to be done as part of thesis writing" but it is vague to say "'quality' (which is unbounded) is a task to be done as part of thesis writing". It is more adequate to say, "a 'review' should be done with a high level of 'quality' ". The 'quality' is to be achieved rather than to be done, and dependant on how the student performs in writing thesis but the 'review' can exist as a sub-topic in a thesis as a task with description as to what the student is required to do. Hence, the unbounded descriptors are useful for describing how the student should do rather than what they are required to do in writing a thesis. From a practical point of view, it is very important to distinguish between 'what' and 'how' the student is expected to do in a practical assessment rubric design.

On the other hand, the concept of 'bound categorization' here suggests that the criterion quantification in my previous work [12] should not be accepted despite its other merit for clarification of component roles.

The current refined model

The analysis and discussion in conjunction with the 'bound categorization' above may lead to a refined model representing structure and functions for the FYP thesis assessment conceptualization. The structural components of the model include 'tasks' to be conducted by student, 'criteria', 'standards', and 'numerals' for quantification. The functions of the model between components are governed by the mathematical laws expressed by Equation (1). The model requires assessment descriptors to be subjected to the 'bound categorization' to distinguish between different quantitative properties for task and criterion descriptions.

The structural components themselves in the model are not new but what is new about the model is of conceptual consistency allowing us to have a new understanding of the assessment system by knowing what the components are and how they function for quantification of student achievement without confusion created by arbitrariness. A model for complex assessment should be capable of maintaining consistency throughout a range of different assessments including simple assessments such as multiple choice exams, short answer exams, etc. Therefore it should have general features given that a simple assessment is a particular case where only particular features are available or a subset of a complex assessment in a logical sense. The current model possesses such general features, being capable of finding out those particular features of simple assessments. For example, in a multiple choice exam, one of 'tasks' is one of exam questions with a weight, an answer to each question is a 'standard', and a marker makes a decision using the 'right or wrong criterion' with reference to the' standard' (answer). The 'right or wrong criterion' is a single criterion, and knowledge required to answer the question is constant in a mathematical sense for a given exam question rather than variable as for a complex assessment. Therefore, the 'right or wrong criterion' for the multiple-choice exam does not require to be subjected to the bound categorization given that it is intrinsically unbounded already. The bound categorization is useful only when multiple candidate descriptors are dealt with. Thus, the current model appears to be consistent encompassing simple assessments for conceptual consistency.

The model is not to dictate educational objectives and values but to accommodate those for quantification of student achievement. Ultimately, the assessor determines a numerical value in f_i in Equation (1) for marking.

PRACTICAL RUBRIC DESIGN

The practical rubric design may be based on the principles embedded in the refined model. As the first step, a rubric designer for thesis assessment may choose quantity and quality since they are perhaps the most generic descriptors for an assessment. Both two descriptors do not yet have bounds or are just bounded by a thesis domain before they are broken up. The quantity and quality may be related with numerical values in fX (see Equation (1)) where f is for quality to be determined by an assessor and X for quantity represented by weight. The reason is that the quality is to be achieved by the student whereas the quantity is to be nominated. Thus, the quantity and quality are governed by the multiplication law. At this level of specificity, X can be 100 % (weight) and f value depends on student's achievement. As the quantity is further broken into more specific descriptors with weights for tasks to be done by the students, more terms can be introduced in Equation (1).

Task list

One of ways of finding bounded descriptors for quantity is to heuristically list thesis constituents or domains in terms of heading, topic or task. A thesis may be generally divided into three Parts as tasks in terms of sequential stage for thesis writing and hence three weights (X_1, X_2 and X_3) (figure 1) i.e.

$$\text{Mark} = f_1X_1 + f_2X_2 + f_3X_3 \qquad (3)$$

The first Part (associated with X_1) may consist of various bounded descriptors such as 'introduction', 'background', 'literature review', 'hypothesis', and/or 'rationale and objectives', etc. It is made up mainly for collected and processed information from various sources prior to writing about the project execution, and justification of objectives and originality. The second Part (associated with X_2) may consist of bounded descriptors for details about how project was conducted and results were obtained (e.g. experiment, computation, design, manufacturing, results, etc, whichever applicable). The last Part (associated with X_3) may consist of bounded descriptors mostly about processing the results such as analysis, results discussion, inference, reasoning, drawing conclusions, etc. The analysis may be microscopic, mathematical, numerical, etc.

If a rubric designer wants to draw up a matrix, the Parts consisting of descriptors discussed above would be listed in a column while different grades are listed in a row, forming a matrix with ticking boxes. The grades should be related with standards description which might vary from a university to another.

The descriptors in each Part cannot be fully generalized although the general characteristics of the three Parts may be common to most research projects in terms of sequence in thesis writing. For example, ‘background’ or ‘literature review’ in the first Part is not necessarily a separate heading or domain and hence may be written as part of ‘introduction’ depending on a thesis writing style. Therefore, both students and assessors can use them optionally or flexibly, sometimes even individualized descriptors would be used if more accountability of the assessment is required. The assessor would even assign weights to individual descriptors if necessary. Weights may be allocated to each Part as guidelines but they can be altered by the assessor as they see fit. If comments for more specific details are necessary, an adequate space would be allocated as well.

Further, weights and marks for the three Parts may be adjusted after a holistic aspect is taken into account, given that there is some dependence of later Parts on earlier parts. The sequential dependence between different Parts does not mean they are inseparable for weighting in Equation (1) due to the fact that the first Part independently exists for assessment without second or third Part, and so on although it is not true in reverse order.

Criteria description

As part of the rubric, the information as to how the student should write a thesis (criteria description) would be given in a section separately. The information should be shared between the assessors and students because it is a set of criteria from the assessor’s point of view while it is about learning outcomes from the student’s point of view. The student would learn about how the assessor makes judgments so that they would know how they should do in writing a thesis. As such, the rubric designer has to find unbounded descriptors for this section description. One of ways of finding such descriptors may be to break up the quality heuristically into specific descriptors. It is not, though, to find a hierarchical structure governed by the conservation law of the weights - it is rather to form a descriptive vocabulary with a higher level of specificity for clues in making judgment. The ‘quality’ may be broken up into ‘knowledge’, ‘understanding’, ‘skills’, ‘difficulty’, ‘originality’, and ‘outcomes’ as criterion properties. The break-ups are useful for writing detailed information as to how the student should do in writing a thesis. Different sets of break-ups may be possible depending on which point of view the rubric designer takes. At the same time, it is important to realize the limitations on clarity of description due to the limitation of specificity of descriptors. The unbounded descriptors may be in the form of noun at the stage of breaking up but may be modified into appropriately inflected forms, if they are not directly usable without variation, to describe how the student should do the tasks. For example, ‘originality’ can be modified into ‘original’ for a part of sentence. Examples for the first Part are as follows.

> The first Part of task descriptors consists of ‘Introduction/Background’, ‘Literature review’, etc representing documentation of collected information from various sources prior to presenting how project was executed. For assessment, the assessor may examine:
> if information collected is written logically, clearly, critically, analytically, coherently, or/and;
> if information collected is relevant to the topic, or/and;
> if information collected is at an appropriate level of knowledge and understanding, or/and
> if problems are clearly, well defined, or difficult to deal with, or/and;
> if objectives, hypotheses, etc are significant, challenging, or original, etc

The descriptor ‘skills’ under ‘quality’ is highly disciplinary dependant because skills are the ones gained by the student from various formal courses in the curriculum. Examples for skills include mechanical engineering design skill, computer programming skill, presentation skill, etc. Also, the degree of relevance and importance of some descriptors varies. For example, mechanical engineering design skills are applicable highly to the second Part but unlikely applicable to the first Part. However, presentation skills would be common to any discipline so that it is worthwhile describing them as part of information as to how the student should present in writing e.g.

The assessor may examine if thesis is written adequately in terms of structure, grammar, spelling, accuracy, consistency, coherency, logic, clarity, illustration, referencing, etc.

Examples of the second Part is as follows:

The second Part of task descriptors consists of ‘Experiment/Computation/Design/Manufacturing’, etc, which are to be conducted by the student for some sort of results. It could be experiment or/and computational work, etc depending on the project nature. The assessor may examine:
if thesis includes sufficient details about the project execution for reproducibility, or/and;
if adequate methodology is employed, or/and;
if project employs adequately a high level of complexity, difficulty, creativity, skills, knowledge, understanding, etc, in the execution of the project, etc.

The last Part requires a range of student knowledge and skills to write Analysis/Interpretation’ etc for outcomes and finalization. Examples of the last Part are as follows:

The assessor may examine:
if the process involved is adequate, logical, systematic, creative, etc, or/and;
if the process involves a high level of knowledge and understanding, etc, or/and;
if solutions to the problems defined earlier are found, or/and;
if outcomes are significant for practicality or advancement of knowledge/understanding, or/and;
if sufficient evidence with discussion is available for conclusions drawn, etc.

The rubric should be visually efficient for accountability, generalizability, and moderating marks between two different assessors for a thesis.

CONCLUSION

A model (Kim 2009) for FYP thesis assessment structure and functions has been refined by incorporating ‘bound categorization’. Two different sets of descriptors have been analyzed and found as a result of the bound categorization. They are termed as ‘bounded descriptors’ and ‘unbounded descriptors’ applicable respectively to ‘what’ and ‘how’ the student should to do in thesis writing. It is suggested that only bounded descriptors can be practically weighted and unbounded descriptors are useful for criteria description.

REFERENCES

[1] Ku H and Goh S, “Final year engineering projects in Australia and Europe”, *European Journal of Engineering Education*, Vol 00(0), 2010, pp 1-13.

[2] Valderrama E, Rullán M, Sánchez F, Pons J, Mans C, Giné F, Jiménez L, Peig E. “Guidelines For The Final Year Project Assessment In Engineering”, *39th ASEE/IEEE Frontiers in Education Conference*, October 18 - 21, 2009, San Antonio, TX, Session M2J.

[3] Cornwall M G, “Sandwich course training and project work – Do they have the same aims?” *European Journal of Engineering Education*, Vol 1, No 1, 1975, pp 41-47.

[4] Jackson E A, “The importance and assessment of final year projects: case studies from Africa”, *Assessment and Evaluation in Higher Education*, Vol 12, No 2, 1987, pp 83-93.

[5] Webster F, Pepper D and Jenkins A, “Assessing the undergraduate dissertation”, *Assessment & Evaluation in Higher Education*, Vol 25, No 1, 2000, pp 71–80.

[6] Tarq V N, Stefani L A J, Butcher A C, Heylings D J A, “Developing a new approach to the assessment of project work”, *Assessment & Evaluation in Higher Education*, Vol 23, No 3, 1998, pp 221–240.

[7] Woolf H, “Assessment criteria: reflections on current practices”, *Assessment and Evaluation in Higher Education*, Vol 29, No 4, August 2004, pp 439-493.

[8] Shay S B, “The assessment of complex performance: a socially situated interpretive act”, *Harvard Educational Review*, Vol 74, No 3, Fall 2004, pp 307-329.

[9] Allison J and Benson F A, “Undergraduate projects and their assessment”, *IEE Proceedings*, Part A, Vol 130, No 8, November 1983, pp 402-419.

[10] Sadler D R, “Interpretations of criteria-based assessment and grading in higher education”, *Assessment & Evaluation in Higher Education*, Vol 30, No 2, April 2005, pp 175-194.

[11] Barrow R and Milburn G, “A critical dictionary of educational concepts”, 2nd Edition, Harvester Wheatsheaf, New York, London, Toronto, Sydney, Tokyo Singapore, 1990, p62.

[12] Kim H S, “Thesis assessment element weighting and quantitative nature of criterion”, *International Conference on Engineering Education and Research*, Seoul Korea, 23-28 August, 2009.

[13] Linn R L, Baker E L and Dunbar S B, “Complex, performance-based assessment: expectations and validation criteria”, *Educational Researcher*, Vol 20, No 8, November 1991, pp 15-21, .