Using a Delphi Approach to Develop Rubric Criteria

Gayle Lesmond, University of Toronto

Gayle Lesmond is a Research Assistant in the Department of Mechanical and Industrial Engineering at the University of Toronto. Her area of specialization is rubric development and testing.

Nikita Dawe, University of Toronto

Nikita is a M.A.Sc. candidate in the Department of Mechanical and Industrial Engineering at the University of Toronto. She is completing the Collaborative Program in Engineering Education.

Ms. Lisa Romkey, University of Toronto

Lisa Romkey serves as an Associate Professor, Teaching Stream with the Division of Engineering Science at the University of Toronto. In this position, Lisa plays a central role in the evaluation, design and delivery of a dynamic and complex curriculum, while facilitating the development and implementation of various teaching and learning initiatives.

Lisa is cross-appointed with the Department of Curriculum, Teaching and Learning at OISE/UT, and teaches undergraduate courses in engineering & society, and graduate courses in engineering education. Lisa completed an Undergraduate Degree in Environmental Science at the University of Guelph, and a Master’s Degree in Curriculum Studies at the University of Toronto. Research interests include teaching and assessment in engineering education.

Dr. Susan McCahan, University of Toronto

Susan McCahan is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. She currently holds the position of Vice Provost, Innovations in Undergraduate Education. She received her B.S. (Mechanical Engineering) from Cornell University, and M.S. and Ph.D. (Mechanical Engineering) from Rensselaer Polytechnic Institute. She is a Fellow of the American Association for the Advancement of Science in recognition of contributions to engineering education has been the recipient of several major teaching and teaching leadership awards including the 3M National Teaching Fellowship and the Medal of Distinction in Engineering Education from Engineers Canada.

©American Society for Engineering Education, 2016
Using a Delphi Study to Develop Criteria for an Analytic, Competency-Based Rubric

Abstract

Recent developments in post-secondary institutions have motivated a shift towards outcomes-based education. A major impetus for this agenda has been the growing need to provide concrete evidence of student learning and institutional effectiveness to various stakeholders. Given this trend, it is important that research be undertaken to explore valid approaches to learning outcomes assessment.

The research described here involves the development of valid, non-discipline specific, analytic rubrics that assess learning outcomes in five key areas: communication, design, teamwork, problem analysis and investigation. This paper reports on the methodology used to complete the first stage of rubric development; identifying the standards through which student work is evaluated. In particular, a two-stage Delphi study was designed to identify rubric criteria for assessing problem analysis and investigation. The Delphi technique is an iterative research tool used to elicit input from a panel of experts. It typically involves a series of virtual survey rounds in which experts offer their views anonymously and have the opportunity to refine them based on controlled feedback from earlier rounds. Panel members include 11 experts for investigation and 15 experts for problem analysis from faculty and staff. In the first round, participants were asked to propose learning outcome statements or “indicators” that are important for assessing problem analysis or investigation. In the second and final round, these responses were arranged by major outcome areas and sent to participants for their feedback. They were asked to rate how likely they were to use the indicators, and their importance in the curriculum. The focus of this paper is not the results of this study, but the methodological processes involved in designing and administering a Delphi survey to develop tools for learning outcomes assessment. This includes expert selection, survey design, and analysis of expert responses. Special attention is paid to the challenges of conducting a Delphi study.

Introduction

This paper details the process of using a Delphi survey to collaboratively generate criteria for analytic rubrics intended for learning outcomes assessment in multiple contexts. In particular, the research team conducted a two-round Delphi survey to identify the key skills that graduating students should exhibit with respect to two competencies; problem analysis (or “problem solving”) and investigation. The research team used definitions for the competencies provided by the Canadian Engineering Accreditation Board (CEAB), with investigation defined as the “ability to conduct investigations of complex problems by methods that include appropriate experiments, analysis and interpretation of data, and synthesis of information in order to reach valid conclusions” \(^\text{1}\) and is similar to the Accreditation Board for Engineering and Technology (ABET) criteria b. Likewise, problem analysis is defined by the CEAB as the “ability to use appropriate knowledge and skills to identify, formulate, analyze, and solve complex engineering problems in order to reach substantiated conclusions” \(^\text{1}\) and is similar to ABET criteria e. The research team sought to answer the following questions:
1. What are the specific skills/behaviours/attitudes that are important for assessing investigation?

2. What are the specific skills/behaviours/attitudes that are important for assessing problem analysis?

This study is part of a larger research project which seeks to develop non-discipline specific analytic rubrics in problem analysis, investigation, design, communication and teamwork.

The paper is organized into three sections. The first section provides background information on rubrics and the Delphi method. This is followed by a description of the research team’s application of the Delphi process. The final section provides commentary on the Delphi method with a particular focus on the challenges of conducting a Delphi study.

**Background**

Learning outcomes have recently gained considerable attention in institutions of higher education and associated organizations. This is in part a response to the growing demands of stakeholders for more concrete evidence of student learning. This shift towards an outcomes-based approach has also necessitated the development of valid tools for assessment.

One such tool is the rubric, a scoring guide that outlines key performance criteria along a continuum of quality. Rubrics are usually categorized as one of two types; holistic or analytic. Holistic rubrics provide an overall score on the quality of work across multiple criteria (see Table 1) and allow the assessor to enter qualitative comments. Conversely, analytic rubrics provide explicit feedback on the specific individual dimensions of a piece of work (see Table 2) using predetermined descriptors. Rubrics can be created that are a hybrid of these two types, but most are based largely on one or the other of these types.

Rubrics generally include three key components. First, the rubric criteria (or “indicators”) are the specific areas against which student work is evaluated. The second component is the quality scale of a rubric, which articulates the quality of the work. Common examples include:

- Beginner, developing, proficient, strong
- Needs development, emerging, competent, expert
- Fails, below expectations, meets expectations, exceeds expectations

Lastly, in an analytic rubric, quality descriptors articulate the performance expectations at each level of mastery (articulated by the scoring scale).

Although holistic rubrics are arguably easier to develop and may reduce grading time, they lack the additional structure of quality descriptors, which not only provide more detailed feedback to students, but also guide raters in their assessment, thus, facilitating consistency in grading. This is particularly important in the development of rubrics for large courses where multiple raters may be involved in assessment. Their ability to improve inter-rater reliability make analytic rubrics the preferred rubric for this project.
Table 1: Example basic structure of a holistic rubric

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description of Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds Expectations</td>
<td></td>
</tr>
<tr>
<td>Meets Expectations</td>
<td></td>
</tr>
<tr>
<td>Below Expectations</td>
<td></td>
</tr>
<tr>
<td>Fails</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Basic structure of an analytic rubric

<table>
<thead>
<tr>
<th>Rubric Criteria</th>
<th>Scale (Level of Performance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fails</td>
</tr>
<tr>
<td>Indicator 1</td>
<td>Descriptor 1a</td>
</tr>
<tr>
<td>Indicator 2</td>
<td>Descriptor 2a</td>
</tr>
</tbody>
</table>

The Delphi method is an iterative research tool used to elicit expert input. It was developed in the 1950s by Normal Dalkey and Olaf Helmer of the Rand Corporation to assess expert opinions on issues of national defense and technological innovation. The objective was to establish consensus among a panel of experts in order to predict future trends. The Delphi technique is characterized by the following key features:

- Panel of experts: Experts are selected based on pre-determined criteria. Although there is no consensus as to the appropriate number of experts for a Delphi study, Okoli and Pawlowski’s recommendation of 10 to 18 experts was followed for this project.
- Anonymity: Experts are kept anonymous to each other. This allows them to communicate freely without pressure from dominant individuals.
- Iteration and feedback: Delphi involves multiple rounds. Each successive round is based on expert feedback from previous rounds.

According to Linstone and Turoff, Delphi studies are typically characterized by 4 distinct phases. The first phase, subject exploration, corresponds with round one of this study, and often uses an open-ended brainstorming session to identify a thorough list of items or ideas. The second phase, understanding how the subject is viewed by the panel, might involve a review of the responses which is reported back to the experts in the second round, where in this study, they are asked to rate and/or rank each item. This process is repeated until some previously agreed upon level of consensus is reached, fulfilling the needs of the third phase defined by Linstone and Turoff, which involves an exploration of disagreement among experts (if any). The fourth phase is a final evaluation of the results.

The Delphi method has been applied in educational contexts to identify program goals and design curricula. For example, Reeves and Jauch conducted a Delphi study to identify course
subject areas and course hour allocations for an undergraduate business curriculum. They reported that the Delphi resulted in “more rational decision making on curriculum in institutions of higher education”\textsuperscript{10}. More recently, Osborne et al.\textsuperscript{11} used a three-round Delphi study to identify the ideas about the “nature of science” that are agreed upon as essential to the school science curriculum. For this study, the Delphi was important in providing empirical evidence of academic consensus on a long debated topic. The Delphi technique has also been employed in rubric development. For example, Timmerman et al.\textsuperscript{12} conducted a modified Delphi in their development of a universal rubric to measure the scientific reasoning and scientific writing skills of undergraduates. In particular, they conducted multiple one-on-one interviews with pedagogical and content experts. Their feedback was used to refine the rubric criteria and performance levels.

The Delphi is not new to Engineering Education research. Olds et al.\textsuperscript{13} used a three-round Delphi study to identify the key concepts in thermal and transport sciences that are typically misunderstood by students. Similarly, Sadowski and Sorby\textsuperscript{14} conducted a three-round Delphi study to identify fundamental concepts in engineering graphics. Further, Balogh, Criswell and De Miranda\textsuperscript{15} used a three-round modified Delphi to develop a Framework of Knowledge for a Master’s program in structural engineering. Based on these examples, the research team decided that the Delphi technique might offer a useful tool for engaging experts in the process of rubric development.

**Methodology**

*Selection of Experts*

The research team first selected experts from multiple departments within the Engineering school of a large public university (see Table 3 and Table 4). Experts were defined as academic professionals with relevant knowledge in the subject area and/or research experience or interest in the teaching and assessment of the competency.

A total of 11 people agreed to serve as investigation experts and 15 as problem analysis experts. Among the investigation experts, three were recipients of teaching-related awards and one a research-related award. One investigation expert had authored a textbook. Further, four problem analysis experts were recipients of teaching-related awards and seven had received research-related awards. Two problem analysis experts had authored textbooks. All of the experts taught courses in which the competency under study was particularly relevant.

**Table 3: Investigation experts by program**

<table>
<thead>
<tr>
<th>Program</th>
<th>Number of Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomaterials and Biomedical Eng.</td>
<td>4</td>
</tr>
<tr>
<td>Chemical Eng.</td>
<td>1</td>
</tr>
<tr>
<td>Electrical and Computer Eng.</td>
<td>2</td>
</tr>
<tr>
<td>Materials Science and Eng.</td>
<td>1</td>
</tr>
<tr>
<td>Mechanical and Industrial Eng.</td>
<td>2</td>
</tr>
<tr>
<td>Physics</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4: Problem analysis experts by program

<table>
<thead>
<tr>
<th>Program</th>
<th>Number of Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Engineering</td>
<td>5</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>3</td>
</tr>
<tr>
<td>Electrical and Computer Engineering</td>
<td>2</td>
</tr>
<tr>
<td>Mechanical and Industrial Engineering</td>
<td>5</td>
</tr>
</tbody>
</table>

**Round One**

The Delphi survey was administered online in two rounds. Separate surveys were used for each competency. The first round of the survey was created using the web-based survey system Survey Monkey. A hyperlink to the survey was sent to experts via email. They were first given approximately two weeks to complete the survey. The deadline was extended until 10 experts (for each competency) had completed round one.

The survey consisted of a single open-ended question that asked experts to “list all the indicators that are necessary to assess [investigation or problem analysis]”. An indicator was defined as a learning outcome statement used to evaluate a student's work. Experts were provided with examples of indicators to ensure that they were all working from the same definition. There was no limit to the responses that experts could provide. The goal was to generate a comprehensive list of all possible criteria that could be assessed in each competency.

Of the 11 investigation experts who agreed to take part in the project, 10 completed the survey for round one; 12 of the 15 problem analysis experts completed this round. One confirmed problem analysis expert did not complete the survey, but instead provided the research team with a course outline and grading rubric developed for marking case reports in a graduate level course.

**Round Two**

Round one responses were analyzed over multiple sessions by members of the research team. The review process consisted of removing duplicates and indicators deemed irrelevant to the competency through discussion and consensus, and comparing against indicators extracted from an extensive literature review conducted prior to the use of the Delphi technique (the major works that were compiled in this literature review include, but were not limited to, Woods et al.\textsuperscript{16,17}, the Association of American Colleges and Universities’ VALUE rubrics\textsuperscript{18} and the Tuning Educational Structures project\textsuperscript{19}). The responses (indicators) were grouped into major learning outcome domains (“global outcomes”), as follows:

In investigation, the student displays the ability to:

1. Define their research
   1A. Define a research problem, question or gap
   1B. Decompose a complex problem into similar related problems
1. Define the problem
   1A. State the problem in their own words
   1B. Identify primary problem goal(s)
   1C. Characterize the type of problem and the type of solution sought
   1D. Represent the problem visually (e.g., free body diagram, circuit schematic)

In problem analysis, the student displays the ability to:
1E. Identify known information
1F. Recognize unknown information
1G. Identify relevant models, concepts, or theories
1H. Identify relevant assumptions

2. Explore relevant knowledge and solution processes
   2A. Reframe complex problems into interconnected sub problems
   2B. Identify appropriate techniques and methods to solve the problem
   2C. Recognize ‘off-the-shelf’ solutions or routine solution processes
   2D. Incorporate information from other sources into the solution process (e.g. looking up the speed of light)
   2E. Determine methodological constraints
   2F. Predictively compare and contrast alternate solution processes in terms of relevant metrics (e.g., accuracy, precision, efficiency, reliability, feasibility, risk, impact, etc.)
   2G. Combine the information from the problem with the appropriate model, concept or theory to create a solvable system

3. Propose a solution process
   3A. Plan a systematic solution process (i.e., identifies measurable tasks that support sub-goals)
   3B. Modify ‘off-the-shelf” solution process within problem context

4. Implement a solution
   4A. Implement a defined solution process to solve the problem
   4B. Competently use appropriate analysis tools (for e.g., computational, experimental, analytical)
   4C. Express the solution in an appropriate form

5. Evaluate results
   5A. Evaluate whether the solution is suitably accurate for the type of problem
   5B. Determine if the solution is valid and correct, or is clearly not correct
   5C. Interpret and evaluate alternative solutions and select a final solution
   5D. Identify limitations and sources of error or uncertainty in the solution
   5E. Specify conclusions or future work (including improvements or modifications to the solution) that can be supported by the analysis

Round two was administered electronically via Google Forms, a web-based survey tool like Survey Monkey that supported the matrix-type question format required.

Experts were presented with the list of indicators gathered in round 1. They were asked to rate 1) how likely they were to use each indicator in their teaching and 2) the importance of each indicator in the curriculum as a whole, both on a 4-point Likert type scale. A score of 4 indicated the greatest likelihood, and highest level of importance. An even numbered scale was chosen to avoid the tendency of participants to resort to neutral or mid-range responses. The advantage of using such a scale is that experts are required to select one side or the other, thus limiting the possibility of a false consensus.
In the final section of the second round survey, participants were given an opportunity to make any suggestions or additional comments on the consolidated list of indicators. The purpose of this round was to verify that the indicators were appropriately categorized and to obtain further feedback on how they could be improved.

Data collection for round two is ongoing. Of the 11 confirmed investigation experts, 8 have thus far completed the survey for round two. Ten of the 15 problem analysis experts have completed the second survey.

**Discussion**

The Delphi method has served a valuable purpose for expert consultation. In the initial phases of the project, consultations were conducted through face to face meetings with subject matter experts. This process, however, was time consuming, as expert meetings were generally one-on-one to accommodate busy schedules and, therefore, limited the number of experts that could take part in the project. Using this method, for example, resulted in consultations with only five experts in the development of the communication rubric, which was completed in an earlier phase of the project. The Delphi technique, thus, offered a unique opportunity for the research team to systematically consult with a wide range of experts over a reasonably short period of time. Through the anonymity of the process, experts were able to freely share their opinions or express disagreement without pressure from dominant panelists. The technique also encouraged a more systematic way of addressing discrepancies between experts, rather than a less structured, back-and-forth consultation between a few experts as new ideas emerge.

Conducting a Delphi has also helped stimulate new ideas, and improve existing ones. In particular, several indicators suggested by experts had not been previously identified in a review of the literature. For investigation, such indicators included, “maintain a safe working environment and display safe lab practice while performing research”, and “identify necessary resources to carry out the methodology”. Similarly, new problem analysis indicators included “reframe complex problems into interconnected sub problems” and “identify appropriate techniques and methods to solve the problem”.

As expected, experts also proposed indicators identified in earlier studies on the competencies. One such example was “gather existing background information on the topic from literature and other relevant sources” for investigation and “identify primary problem goal(s)” for problem analysis. Although discussed in prior scholarship, these suggestions often resulted in a clearer, more succinct articulation of the indicator. For example, the problem analysis indicator, “outline any required further work” was first generated from a review of the literature conducted in an earlier phase of the project. One similar suggestion from an expert was the “specification of conclusion or decisions that can be supported based on the data and analysis”. The final indicator, “specify conclusions or future work (including improvements or modifications to the solution) that can be supported by the analysis”, a hybrid of the aforementioned suggestion and what was identified in the literature, provided a clearer description of the activity to be assessed.

The Delphi technique, however, is not without its challenges and limitations. The first for this project concerned the selection of panel experts, a critical element in the success of any Delphi
study. In particular, many experts who were approached to take part in the study declined participation. This was particularly true for problem analysis, where only 15 of the 32 experts who were contacted, agreed to take part in the project. The expert pool consists of a very busy group of (primarily) faculty members, with many competing commitments. Another issue related to expert selection concerned the diversity of the panel. Although representativeness is not necessarily important for Delphi samples, some degree of heterogeneity is recommended. This is particularly important in the development of a universal rubric. It is for this reason that the research team sought to include experts from a variety of programs within the Engineering school. However, in the case of problem analysis, the panel was more largely represented by experts from Chemical Engineering and Mechanical and Industrial Engineering. In contrast, Materials Science and Biomaterials and Biomedical Engineering were not represented among problem analysis experts. The generally homogenous nature of the panel may have limited the variety of responses provided in round one of the study.

Also problematic was the low response rate. As mentioned earlier, of the 11 confirmed investigation experts, 10 completed the first survey and 8 have so far completed the second survey. Similarly, while 12 of the 15 confirmed problem analysis completed the first survey, 10 have completed the survey for round two. This diminished willingness and/or interest in the survey continued despite the fact that participants were constantly reminded about the survey. Because of the low response rate, the research team has not been able to conduct any statistical analyses in round two to determine, for example, average ratings and the level of consensus for each indicator. Consideration needs to be given towards incentives for participation in Delphi studies, which require more time and expertise than a standard research survey.

The literature offers recommendations for helping eliminate problems in expert selection. For example, Okoli and Pawlowski recommend the Knowledge Resource Nomination Worksheet (KRNW), wherein researchers begin the Delphi process by identifying the relevant skills, disciplines, organizations etc. that are important for the topic before selecting any potential experts. The objective of this technique is to identify categories of experts without disregarding any important group. For low response rates, Franklin and Hart suggest that experts with a known interest in the area be selected to take part in the study and that communication be consistent throughout the entire Delphi process.

Limitations to the current application of the Delphi technique were also noted in survey design. As previously mentioned, in the first round surveys experts were reminded of the objectives of the study, were given examples of indicators and asked to provide a list of indicators that they believed to be important for assessing the relevant competency. Experts, however, were not provided any more direction on, for example, the scope of indicators or the format in which they should be written. This resulted in a fairly wide range of responses, some of which were more appropriate for “higher level” learning outcome domains, and not the more specific, measurable performance indicators that were sought in this project. Examples of these included, “perseverance” and “attention to detail” in the case of investigation and “demonstrate core technical knowledge in associated fields” in the case of problem analysis. Survey instructions should have thus provided a more adequate, detailed description of the term “indicator”.
Another limitation of the study was that the researchers occasionally encountered difficulty in interpreting experts’ responses to the first round survey. Generally, such responses were removed from further analysis. To help eliminate this problem in future Delphi surveys, the research team recommends that experts be asked to provide explanations and/or examples of each item that they provide to ensure that they are accurately interpreted by researchers.

Finally, an analysis of round one responses revealed that problem analysis experts tended to conflate the process of problem analysis with that of design. Although some overlap exists between these activities, the research team has distinguished between the two based on their respective goals and constraint certainty. In particular, problem analysis is typified by precise goals and constraints. Questions are close-ended, resulting in a single, objective right answer. Conversely, design is open-ended and involves vague goals and constraints. Design activities can result in multiple correct solutions which are assessed in terms of their feasibility and appropriateness for the objectives. This distinction between problem analysis and design, however, was not articulated in the instructions to the first problem analysis survey. As a result, many experts offered problem analysis as well as design oriented indicators. Examples of the latter included the following:

- Identify and/or confirm problem statement and scope
- Identify specific complex engineering problems from broad situational descriptions
- Frame the problem from an engineering perspective
- Determine evaluation criteria for assessing alternative solution plans to open ended problems
- Identify, describe and apply an appropriate model for analyzing unstructured problems in a complex decision making environment
- Identify design or free variables

To address this problem, the researchers not only explained the distinction between problem analysis and design in the second round survey, but also provided an example of a problem that a student may be asked to solve. This provided survey respondents with a more detailed description of “problem analysis” beyond the brief statement provided in round one.

**Conclusion**

Regardless of its methodological limitations, the Delphi technique provided a structured and comprehensive means of obtaining collective expert opinion in areas of uncertainty or disagreement. It is important, however, that researchers recognize the limitations inherent in the process, as described above, and take steps to address them throughout the process. For example, considering the decline in participation through the process, it would be sensible to start with a larger than necessary expert group.

An additional benefit in utilizing the Delphi technique for the purpose of designing competency-based rubrics is that instructors may be more likely to use the final product in the assessment of their students, given their role in co-constructing the key outcomes and indicators. It is hoped that the collaborative nature of this exercise will encourage a common understanding and
measurement of key competencies in our students. The impact of this process on faculty culture change may be one of the most important outcomes of this research methodology.

References


