Assessing graduate attributes in large classes without sampling

Said M. Easa
Professor, Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada
Email: seasa@ryerson.ca

Structured abstract

BACKGROUND
Assessing graduate attributes in large classes is a time consuming task. The assessment requires a carefully designed random sampling that ensures the sample is representative of the entire group of students from which the sample is taken. In addition, the assessment becomes more difficult when soft-skill type of attributes, such as ethics and equity, are involved. There is a need for innovative and simple methods that enable the assessment of the entire class.

PURPOSE
The purpose of this paper is to present a new and efficient method for assessing soft-skill attributes in large classes without sampling.

DESIGN/METHOD
The proposed method involves defining the indicator (or learning objective) by knowledge elements (topics) that the student should know or by case studies that involve interactions representing the principles related to the indicator. Multiple-choice questions are then established for the knowledge elements/interactions and processed using scantron sheets. The method involves a weighted-score procedure and performance scales for determining class performance.

RESULTS
The method performed well in assessing four graduate attributes in the common engineering course *Law and Ethics in Engineering Practice*. Application of the method to the lifelong learning indicator is illustrated in this paper and the results show that class performance is sensitive to the weights of the questions.

CONCLUSIONS
The proposed method has shown to be useful in identifying the indicators that need improvements along with specific topics within the indicator. Since class performance is sensitive to the question weights, these weights should be carefully established by the user.

KEYWORDS
Accreditation, assessment, large classes.

Introduction

Outcome-based (OB) assessment focuses on empirically measured outcomes that include a range of skills and knowledge. The OB education is a requirement to join the Washington Accord, established in 1989 (IEA 2013). Graduates of accredited programs in any of the signatory countries of the accord are recognized by the other signatory countries as having met the academic requirements for entry to engineering practice. This recognition has motivated numerous countries to shift to OB accreditation to be able to join the accord and benefit from the global mobility of engineers.

The Canadian Engineering Accreditation Board (CEAB) has recently introduced an OB criterion for the accreditation of engineering programs. The criterion includes 12 graduate attributes. According to this criterion, each engineering program in Canada must have a system in place for continuously assessing these attributes and using the assessment results to improve the program (CEAB 2012). The improvements include curriculum changes, such as modifying course content and changing order of material within a course. An important area of curriculum changes is the linkage of graduate attributes to course design; see for example Barrie (2006) and Jackson (2000). A framework and guidelines for OB assessment in engineering education can be found in Easa (2013).

Four CEAB graduate attributes were assessed at Ryerson University in the fourth-year common engineering course "*Law and Ethics in Engineering Practice.*" The graduate attributes assessed in the course were professionalism, impact of engineering on society and environment, ethics and equity, and lifelong learning. For these graduate attributes, eight indicators were identified and assessed. The indicators were assessed in the midterm exam (law part) and the final exam (ethics part) using multiple choice questions. An innovative method to assess the indicators using the entire class of 446 students was developed. The graduate attributes were assessed using groups of exam questions and associated weights. The following sections present the process of defining the indicators, the proposed method and its application using numerical data.

Defining indicators

The "*Law and Ethics in Engineering Practice*" course addresses the legal and ethical aspects of engineering practice, including Canadian legal system and business organizations, tort liability, business contract law, intellectual and industrial property, principles of arbitration and alternative dispute resolution, the practice of engineering, occupational health and safety, ethical aspects and dilemmas of engineering practice, sustainable development, and international standards for ethical and social responsibility.

Four CEAB graduate attributes were assessed in this course as follows: professionalism (Attribute 8), impact of engineering on society and environment (Attribute 9), ethics and equity (Attribute 10), and lifelong learning (Attribute 12). Indicators for the preceding attributes were identified and assessed in the course. The assessment Attribute 12 is presented here to illustrate the proposed method.

Proposed method

For large classes, it is often necessary to use multiple choice questions for which other methods would not work. To address this issue, a method is proposed here that allows the use of multiple choice questions to define class performance for an indicator. The method involves defining multiple-choice question groups, determining class performance (individual weighted score, performance scales, and aggregate performance), and data processing,

Defining multiple-choice question groups

The idea of this method is to define a group of multiple-choice questions associated with an indicator. The group of questions can be defined based on: (a) the knowledge elements associated with the indicator or (b) the interactions among different players in a case study.

Knowledge element-based groups

In this method, the indicator is represented by a number of knowledge elements (topics) that a student should know to fully understand the indicator (Fig. 1). The number of these elements, *n*, may vary from one indicator to another and among different assessors. A question is then developed for each of these elements. Each element *i* has a weight w_i associated with it, where $i = 1, 2, \ldots, n$. The weight an element represents the importance of that element in engineering practice. For example, if $w_1 = 2$ and $w_2 = 1$, this means that Element 1 is twice as important as Element 2.

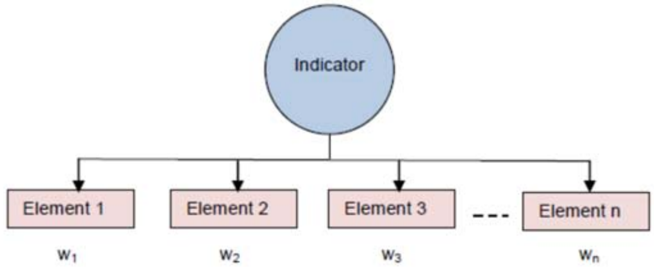

Figure 1: Defining the indicator by knowledge elements and their weights

Interaction-based groups

In this method, a case study that involves a number of players (individuals and/or organizations) is developed. The players in the case study have numerous legal and ethical interactions (Fig. 2). A group of questions related to the interactions of the players is then developed to test student understanding of the underlying principles. Similar to the previous method, there weights associated with the interactions.

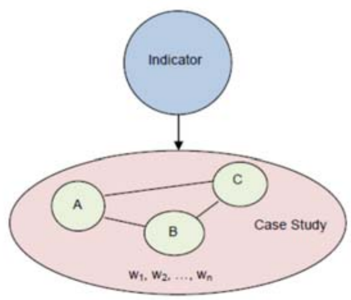

Figure 2: Defining an indicator in a case study by interaction elements and their weights

Determining class performance

Individual weighted score

For each indicator, a group of n questions is designed to address the knowledge elements or the case study related to each indicator. Let the weight of element *i* be defined as w_i. Define Q_{ij} as a binary variable such that $Q_{ij} = 1$ if student *j* answers question *i* correctly and $Q_{ij} = 0$ if student *j* answers the question incorrectly. Let the number of students assessed be defined as *m*. Then, the performance of student *j*, R_j, is given by

$$R_j = \sum_{i=1}^{n} \mathrm{w_i Q_{ij}}, \qquad j = 1, 2, \ldots, m \qquad (1)$$

The maximum performance score by any student equals the total weight, *W*, which is given by

$$W = \sum_{i=1}^{n} \mathrm{w_i} \qquad (2)$$

Performance Scales

In assessing the indicators of the graduate attributes, it is necessary to define scales to assess class performance. The number of scales varies depending on the type of graduate attribute and the indicator. Some methods consider three scales (e.g. poor, average, and excellent), while others consider five scales (e.g. poor, below average, average, above average, and excellent). The definition of the scales may be based on rubrics or some numerical values (Spurlin et al. 2008).

The performance scales in the proposed method are defined using the maximum performance score. For simplicity, a three-scale system is implemented: Excellent (E), Average (A), and Por (P). The limits of the performance scale, S_j, are then defined as follows,

$$S_j = \begin{cases} Excellent & \text{, if } \mathrm{R_j} \geq 0.8\,\mathrm{W} \\ Average & \text{, if } 0.5\,\mathrm{W} \leq \mathrm{R_j} < 0.8\,\mathrm{W}\,, \\ Poor & \text{, if } \mathrm{R_j} < 0.5\,\mathrm{W} \end{cases} \qquad j = 1, 2, \ldots, m \tag{3}$$

Based on Eq. 3, the student performance is considered Excellent if the weighted score is equal to or greater than 80% of the maximum score and is considered Poor if the weighted score is less than 50% of the maximum score. The performance is considered Average if it is between the preceding limits. Table 1 shows the limits of of performance scales for different maximum score for different maximum scores based on Eq. 3. For example, if the indicator is represented by three questions (n = 3) and their weights are w_1 = 2, w_2 = 1, and w_3 = 3, then W = 6 and the performance scales based on Eq. 3 are defined as follows:

S_j = Excellent if R_j is equal to or greater than 0.8x6 = 4.8 or 5.
S_j = Average if R_j is greater than or equal to 0.5x6 = 3 and less than 0.8x6 = 4.8 or 5.
S_j = Poor if R_j is less than 0.5x6 = 3.

Table 1: Limits of performance scales for different maximum scores

Maximum Score, W	Limits of Performance Score, S_j		
	Poor <	Average =, <	Excellent >=
12	6	7, 10	10
11	5	6, 9	9
10	5	6, 8	8
9	4	5, 7	7
8	4	4, 7	7
7	3	3, 6	6
6	3	3, 5	5

Aggregate Performance

Based on the individual student performance, the class performance is represented by the percentage of students in each scale as follows,

$$P_E = 100\frac{m_E}{m} \tag{4}$$

$$P_A = 100\frac{m_A}{m} \tag{5}$$

$$P_P = 100\frac{m_P}{m} \tag{6}$$

where m_E = number of students who has S_j = Excellent, m_A = number of students who has S_j, = Average, and m_P = number of students who has S_j = Poor.

Once the class performance is determined, a decision is made regarding whether improvements are needed for the indicator. In this regard, it is necessary to consider a threshold (TH) and a target (T). The threshold is the minimum acceptable level of performance on a given indicator (Meyer et al. 2010) and the target is the intended level of learning proficiency for that indicator. If the performance of an indicator (excellent plus average) is less than the threshold, this means improvement is needed. Typically, the improvement effort should focus first on such indicators, followed by those with class performances that are below the target.

Although improvements to all elements of the indicator should be made, more attention should be devoted to the particular element (question) in which the students have the most difficulty. This can be determined by defining the deficiency score for each question of the group which is given by

$$D_i = m w_i - \sum_{j=1}^{m} \mathrm{w_i Q_{ij}}, \quad i = 1, 2, \ldots, n \tag{7}$$

where D_i = deficiency in student performance for question i. The first term in the right side of Eq. 7 is the maximum weighted score all students could achieve for question i. The second term is the actual weighted score that has been achieved. More attention for improvement should then be devoted to the question with the largest D_i.

Data Processing

The multiple choice midterm and final exams were conducted using scantron sheets in which students mark the answers. The sheets were then processed through the media printing centre and a report of the results of answering the questions along with summary statistics is emailed to the instructor. The report provides the number of exam questions that each student answers correctly. The report also presents for each student a row whose columns are the exam question numbers. In this row, the questions answered incorrectly are indicated in the respective cell by the choice number that the student selected, and for the questions answered correctly the cells are left blank. This format makes it easy to determine the group questions that are correctly answered.

The row information for each student is then used to determine the questions correctly answered for each group of questions related to the indicator. Clearly, to simplify collecting the data from the results report, the questions of each group are made consecutive in the exams. For each question group, the questions correctly answered by each student are recorded in a spreadsheet. The spreadsheet is then used to analyze the data as previously described.

Application

The eight indicators were assessed in the midterm and final exams using multiple choice questions. The exams involved about 60 and 80 questions, respectively. The question groups for assessing the indicators were subsets of the total exam questions. As previously mentioned, the assessment was performed for the entire class. To illustrate the application of the proposed method, the results of only one indicator (lifelong learning) are presented here.

The indicator related to lifelong learning is “Recognizes the need for ongoing professional development to maintain competence in the field.” This was the simplest indicator assessed as it was defined by three elements as shown in Fig. 3.

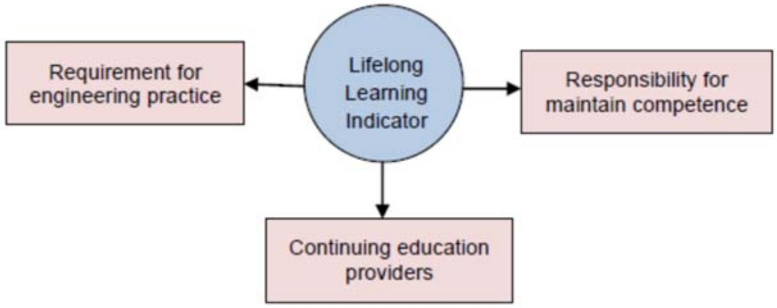

Figure 3: Defining the lifelong learning indicator by three knowledge elements

The elements of this indicator are defined as: (1) requirement for maintaining competence in engineering practice, (2) responsibility for lifelong learning, and (3) available providers. Three questions were developed to address the three elements of the indicator, as follows:

1. A requirement for maintaining competence to practice engineering is that an engineer:
   a) Maintain membership in a technical society.
   b) Take a continuing education course for credit at least once every two years.
   c) Read the monthly magazine for your engineering discipline.
   d) Maintain current knowledge of your discipline.
   e) All of the above

2. Maintaining competence of the professional engineer in practice is the responsibility of:
   a) The professional engineer
   b) The employer
   c) The professional association
   d) All of the above
   e) Both a and b

3. Continuing education providers include:
   a) Technical societies
   a) Universities
   b) Industrial organizations
   c) Private providers
   d) All of the above

The assessment results of only 20 students are presented here to aid the illustration of the method. The question score (Q_{ij}), weighted scores (w_iQ_{ij}), student weighted score (R_j), and student performance scale (S_j) are shown in Table 2. The weights are set to w_1 = 2, w_2 = 1, and w_3 = 3. Based on the student performance scale of Table 2 (last column), the class performance is calculated using Eqs. 4-6 and the results are shown in Table 3.

The sensitivity of class performance to different weighting scenarios is shown in Table 4. Scenario 1 represent the analysis previously presented. The three other scenarios correspond to weights (w_1, w_2, w_3) of (2, 2, 2), (1, 2, 3), and (4, 1, 1). The results show that class performance is sensitive to the weights assigned to the questions. For example, the class performance of Poor for Scenarios 2, 3, and 4 range from 5% to 20%. It is useful to present the results of class performance in terms of the cumulative percentage of the three scales (E, A, and P) as shown in Fig. 4. The threshold and target levels for this indicator were selected as 85% and 95%, respectively. If the cumulative performance of excellent and average is below the threshold, this indicates that attention for improving the indicator is needed. In this example application, if Scenario 4 is the one determined by the assessors, the cumulative percentage of this scenario is 80% which is below the threshold. This means that improvements to this indicator are needed. The deficiencies of the three questions are calculated, based on Eq. 7, as 16, 3, and 3. Thus, more attention should be devoted to the element related to Question 1 (requirement for competence to practice engineering).

Table 2 Original data and total weighted score for each student
(w_1 = 2, w_2 = 1, w_3 = 3)

Student No., j	Question Score			Question Weighted Score			Student Weighted Score, R_j	Student Perfor. Scale, S_j
	Q_{1j}	Q_{2j}	Q_{3j}	$w_1 Q_{1j}$	$w_2 Q_{2j}$	$w_3 Q_{3j}$		
1	1	0	1	2	0	3	5	E
2	1	1	1	2	1	3	6	E
3	0	1	1	0	1	3	4	A
4	1	1	1	2	1	3	6	E
5	1	1	1	2	1	3	6	E
6	1	1	1	2	1	3	6	E
7	1	1	1	2	1	3	6	E
8	1	1	1	2	1	3	6	E
9	0	1	1	0	1	3	4	A
10	1	1	1	2	1	3	6	E
11	1	1	1	2	1	3	6	E
12	1	1	1	2	1	3	6	E
13	0	0	1	0	0	3	3	A
14	1	1	0	2	1	0	3	A
15	0	1	1	0	1	3	4	A
16	1	1	1	2	1	3	6	E
17	1	1	0	2	1	0	3	A
18	1	1	1	2	1	3	6	E
19	1	1	1	2	1	3	6	E
20	1	0	0	2	0	0	2	P

Table 3: Class performance for example application
(w_1 = 2, w_2 = 1, w_3 = 3)

Performance Scale	Number of Students in Scale (m_E, m_A, m_P)	Proportion of Students in Scale (Eqs. 4-6)
Excellent	13	65%
Average	6	30%
Poor	1	5%

Table 4: Class performance for different weighting scenarios

Scenario	Weights (w_1, w_2, w_3)	Class Performance (%)		
		Excellent, P_E	Average, P_A	Poor, P_P
1	2,1,3	65	30	5
2	2,2,2	60	30	10
3	1,2,3	75	20	5
4	4,1,1	75	5	20

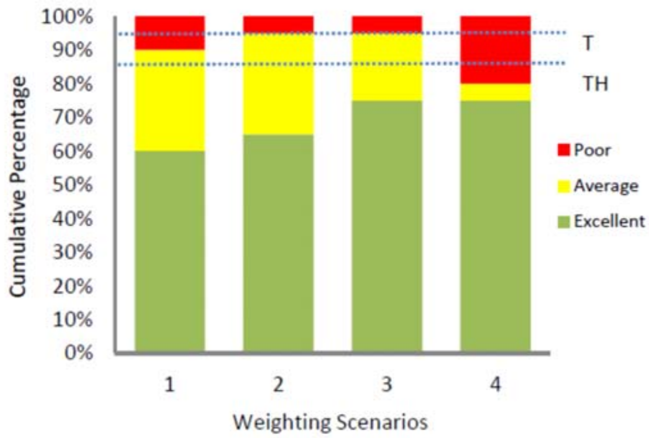

Figure 4: Sensitivity of class performance to different weighting scenarios

Conclusions

An innovative method for assessing graduate attributes in large classes without sampling is presented in this paper. The method involves defining the indicator by knowledge elements that the student should know or by case studies that involve interactions representing the principles related to the indicator. Multiple-choice questions are established for the knowledge elements/interactions and associated weights are specified. Multiple-choice questions are established for the knowledge elements/interactions and processed using scantron sheets or clicker technology. The method involves a weighted-score procedure and performance scales for determining class performance.

The method performed well in assessing four graduate attributes in the common engineering course *Law and Ethics in Engineering Practice*. Application of the method to the lifelong learning indicator is illustrated in this paper. The results show that class performance is sensitive to the question weights and therefore the weights should be carefully established. The proposed method can be used for assessing other soft-skill attributes.

Some limitations of the proposed method should be noted. First, the method cannot be used for assessing all graduate attributes. It is useful only for the graduate attributes that require knowledge of facts. In particular, some ‘soft’ skill attributes (e.g. communication skills and teamwork skills) are the hardest to assess using knowledge of facts and other indicators would be necessary. Second, the method requires setting weights of the knowledge elements and these weights may vary from one instructor to another. Therefore, it is necessary that a consistent procedure for establishing these weights be put in place at the department level. Finally, the results of the method may be sensitive to the number of knowledge areas related to the indicator and this issue should be explored further.

Acknowledgements

The author is grateful to S. Tawfiq for her assistance in the analysis of the assessment data. The author also acknowledges the contributions of the following course instructors: B. Beaumont, A. Wong, and R. Sebastiano (Law part) and M. Rosen and R. White (Ethics part).

References

Barrie, S.C. (2006). Understanding what we mean by generic attributes of graduates. *Higher Education,* Vol. 51, No. 2, 215–241.

Canadian Engineering Accreditation Board (CEAB) (2012). Accreditation Criteria and Procedures. CEAB, Ottawa, Ontario, Canada.

Easa, S.M. (2013). Framework and guidelines for graduate attribute assessment in engineering education. *Canadian Journal of Civil Engineering*, Volume 40, 547–556.

International Engineering Alliance (IEA) (2013). *Washington Accord*. Available from http://www.washingtonaccord.org/Washington-Accord [Accessed on 20 August 2013].

Jackson, N. (2000). Programme specification and its role in promoting an outcomes model of learning. *Learning in Higher Education,* Vol. 1, No. 2, 132–151.

Meyer, J., Land, R., and Baillie, C. (eds). (2010). *Threshold Concepts and Transformational Learning.* Sense Publishers, Rotterdam, The Netherlands.

Copyright statement