Abstract Cooperative student performance evaluations are primarily intended to provide formative and summative feedback to students on their performance during work-term employment but also represent a rich source of external data on student performance. These data are collected using the same instrument across the faculty (and, in fact, the institution). Approximately five years ago the evaluation instrument at our institution was revised; a significant objective for the updated instrument was alignment with graduate attributes. In this paper, we examine the data collected through a graduate attributes and continual program improvement lens.

Keywords: graduate attributes; cooperative education; student performance evaluation; employer evaluation; work-integrated learning

1. INTRODUCTION

Employer evaluation is a definitive part of the cooperative education model and is a required component of Co-operative Education and Work-Integrated Learning Canada (CEWIL)-accredited cooperative education programs [1] [2] (note that the Canadian Association for Co-operative Education, CAFCE, was recently renamed CEWIL). At the University of Waterloo, an institution-wide survey instrument is used for this purpose, referred to as the Student Performance Evaluation (SPE) [3]. This instrument was substantially redeveloped several years ago, with alignment to the Canadian Engineering Accreditation Board graduate attributes (GAs) being one of the goals for the new instrument. While the purpose of these evaluations is to provide individual evaluation and feedback to students, they represent a potentially rich source of external evaluation to inform graduate attribute assessment and continual program improvement. This short paper describes the results of an initial exploratory data analysis conducted on the available data.

2. METHODS

2.1. Survey Instrument

The current SPE has been in use since 2013; the complete instrument is available online [3]. The instrument asks employers to respond to 16 Likert-type questions related to different aspects of student performance on a seven-point scale, where 1–2 are described as “Developing”, 3–5 as “Good” and 6–7 as “Superior”; it is also possible to indicate “Not observed”. Employers are also asked to provide an “Overall Performance Rating” on a seven-point scale from “Unsatisfactory” to “Outstanding”, which appears on student transcripts as the grade for that work term. Spaces are then provided for written comments on performance by the supervisor, a response by the student and, optionally, recommendations from the supervisor related future personal and professional development. The guidelines also encourage employers to use the SPE to guide a mid-term review. The same SPE is deployed across all co-operative education programs at the institution.

2.2. Data and Analysis

Data were collated for engineering students for the period 2013–2017. These data represent over 11,000 students in 13 programs, with approximately 5,000 distinct employers. An exploratory data analysis was conducted examining items such as the distributions of the data, correlation of overall performance evaluation with individual ratings, variation across programs, and progression over time.

2.3. Note on Presentation of Results

This work was conducted primarily for internal use to support GA assessment and continual program improvement and is presented in practice (rather than research) stream of the conference. Therefore, the presentation and discussion of results focuses primarily on our learnings about the design, implementation and use of
the instrument in this context and we do not present detailed statistical results.

3. RESULTS

3.1. Overall Performance Rating

The overall performance rating correlated well with the average of the 16 individual ratings. A reliability analysis showed high reliability for the overall performance rating, indicating that it consistently reflects the individual ratings. A positive correlation was found between the overall performance rating and each individual rating; this was strongest for the items interest in work, ability to learn, quality of work, quantity of work, and, problem solving.

Over 90% of students receive ratings in the range 5–7 (7 being outstanding). However, the full range of ratings is used by employers. There is an increase in the average rating with student level (measured by number of work terms completed), reflecting an increased proportion of 7 (Outstanding) evaluations and decreased proportion of 4 (Good) and 5 (Very Good) responses. This observation holds across all programs, although the slopes were not significantly different from zero in approximately half of programs.

3.2. Individual Item Ratings

The averages for each item all fell in a tight band (5.5 to 6.1). “Not observed” was recorded as the response less than 5% of the time, except for the items ethical behavior (19%), appreciation of diversity (22%), entrepreneurial orientation (24%), and written communication (8%). There was a statistically significant increase in average rating with student level for all 16 items. This mostly held when examined by program, except for one program that had above-average ratings in early work terms. Differences in average rating were observed between programs.

3.3. First-Work-Term Ratings

Five of the engineering programs analyzed have two streams, where students have their first work term after either one or two academic terms. Differences in overall performance evaluation for the first work-term were examined based on stream. A statistically significant difference was found for three of the five programs, with an increase in rating for the students having two academic terms.

4. DISCUSSION

Program-level GA assessment is a secondary use of the SPE, which is primarily intended to provide evaluation and feedback from employers to individual students. In terms of the assessment of individual students, the instrument appears to be working well. The overall performance rating appears to be reliable (in the sense that it is correlated with the individual items) and significant variation is seen in the individual item evaluations, suggesting that employers do use the instrument to provide feedback to the students.

From a GA assessment perspective, the SPE data are useful in providing an external assessment. The data generally indicate good student performance across the items (and hence GAs). The slopes are generally weakly positive, suggesting an improvement with academic level; however, this result should be interpreted cautiously. No guidance is provided to employers as to the referent that should be used. Informal discussions with colleagues who have employed coop students suggest widespread disagreement as to whether the referent should be a student at the same level or a senior student. Clearly, this is of fundamental importance to the rating that will be assigned.

A thornier question is whether the data will be able to provide evidence to support continual program improvement. Here, the results are mixed. Variation in ratings is seen across the individual items. This suggests that the instrument may be useful in identifying areas of relative strength and weakness in the programs; presumably it may therefore be useful in detecting changes in those areas, which would provide evidence for or against the effectiveness of program changes. However, it remains unclear whether the instrument will be sensitive enough to measure such changes. The existence of statistically significant differences between some programs suggests that the instrument may be sensitive to changes (though interpretation of this should be made with caution since some differences would be expected given the large number of comparisons possible). However, it is unclear whether those differences correspond to differences in student performance or variation in employer expectation. While the latter might be expected to average out over many evaluations, this may not be the case if they are structural differences that may be associated with the type of employer (for example, a software firm versus an automotive manufacturer). In addition, while the changes are statistically significant it is not clear they are practically significant. Given these challenges, we elect not to place great emphasis on the difference between first-work-term ratings, except to note that it is in contradiction to the finding of other authors [4].

In conclusion, the SPE is a promising instrument for use for GA assessment. Further investigation should be conducted towards the use of the instrument by employers; a particular focus of such investigation should be identification of the appropriate referent, but the investigation should encompass broader aspects such as employer engagement with the instrument, its use and effectiveness in formative assessment through the midterm review, etc. In addition, students should be consulted regarding their experience with and use of the evaluations to determine their usefulness in formative assessment, student opinion of their accuracy in summative assessment and to gain an understanding of how students use the information to guide their personal and professional...
development goals. Finally, we note that the insights we have been able to develop have been facilitated by the existence of five complete years of data; this reinforces the need for careful deliberation before making any changes to the SPE instrument.

Acknowledgements

We are grateful to the Centre for Teaching Excellence at the University of Waterloo for funding this work through a Learning Innovation and Teaching Enhancement (LITE) grant.

References


